Published in Journals: Applied Sciences, Entropy,
Sustainability, Electronics and Energies

Topic Reprint

Artificial Intelligence and Sustainable Energy Systems

Volume III

Edited by
Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

www.mdpi.com/topics

Artificial Intelligence and Sustainable Energy Systems

Artificial Intelligence and Sustainable Energy Systems

Volume III

Editors

Luis Hernández-Callejo
Sergio Nesmachnow
Sara Gallardo Saavedra

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors

Luis Hernández-Callejo
University of Valladolid
Spain

Sergio Nesmachnow
Universidad de la República
Uruguay

Sara Gallardo Saavedra
Universidad de Valladolid
Spain

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of Topic published online in the open access journal *Applied Sciences* (ISSN 2076-3417), *Entropy* (ISSN 1099-4300), *Sustainability* (ISSN 2071-1050), *Electronics* (ISSN 2079-9292), and *Energies* (ISSN 1996-1073) (available at: https://www.mdpi.com/topics/Artificial_Intelligence_Energy_Systems).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

Volume III
ISBN 978-3-0365-7648-0 (Hbk)
ISBN 978-3-0365-7649-7 (PDF)

Volume I-III
ISBN 978-3-0365-7642-8 (Hbk)
ISBN 978-3-0365-7643-5 (PDF)

© 2023 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.
The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editors

Luis Hernández-Callejo

Luis Hernández-Callejo is an electrical engineer at the Universidad Nacional de Educación a Distancia (UNED, Spain), a computer engineer at UNED, and a PhD candidate at the Universidad de Valladolid (Spain). Professor and researcher at the Universidad de Valladolid. His areas of interest are renewable energy, microgrids, photovoltaic energy, wind energy, smart cities, and artificial intelligence. He has participated in numerous research projects, directed many doctoral theses, and is the author of hundreds of scientific articles.

Sergio Nesmachnow

Sergio Nesmachnow is a full professor at the Faculty of Engineering, Universidad de la República, Uruguay. He is a level III researcher (the maximum level) of the National System of Researchers in Uruguay and a visiting professor at renowned universities and research centers in America and Europe. He has more than 400 publications in scientific journals and international conferences and is responsible for more than 50 research projects.

Sara Gallardo Saavedra

Sara Gallardo Saavedra is a professor and researcher at the Campus Duques de Soria of the University of Valladolid, Spain. Her research focuses on the detection, characterization, and classification of defects in photovoltaic (PV) modules through the use of thermography, electroluminescence, I-V curves, and visual analysis. She has participated in numerous national and international R+D+I projects, carrying out active dissemination of the results with high regularity in the scientific production, including scientific publications in high impact factor journals, book chapters, and contributing to congresses on advanced maintenance in PV. She has made a predoctoral and a postdoctoral stay in the Unit of Solar PV Energy in the Energy Department of the Energy Research Center, Environment, and Technology (CIEMAT) in Madrid, and the researcher has collaborated with different institutions such as the University of Gavle in Sweden, the Universidad del Valle in Colombia, the National Polytechnic Institute of Mexico, and the University of Cuenca in Ecuador.

Preface to "Artificial Intelligence and Sustainable Energy Systems"

The problems that affect humanity are numerous and occur in different areas. Energy sustainability, climate change, and the effects derived from pollutants and viruses are some of the most relevant problems. The main objective of researchers is to provide solutions to these and other problems.

In recent years, the use of artificial intelligence has increased considerably. Artificial intelligence is used in different areas: energy, sustainability, medicine, health, mobility, industry, etc. Therefore, it is necessary to continue advancing in the application of artificial intelligence to the aforementioned problems. Energy is a precious commodity, and it is increasingly difficult to dispose of it in a sustainable way. In this sense, renewable energy sources are essential, although the use of conventional energy cannot be forgotten. Therefore, sustainable energy systems, integrating renewable and non-renewable energy sources, smart systems, and new business models, are crucial.

Therefore, in this book, the best accepted and published articles on the topic "Artificial Intelligence and Sustainable Energy Systems" are presented. All articles refer to the themes indicated above.

Luis Hernández-Callejo, Sergio Nesmachnow , and Sara Gallardo Saavedra
Editors

applied sciences

MDPI

Article

Classification of Ground-Based Cloud Images by Improved Combined Convolutional Network

Wen Zhu, Tianliang Chen, Beiping Hou *, Chen Bian, Aihua Yu, Lingchao Chen, Ming Tang and Yuzhen Zhu

School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China; joywenzhu@zust.edu.cn (W.Z.); 222007855005@zust.edu.cn (T.C.); 221901852094@zust.edu.cn (C.B.); yuaihua@zust.edu.cn (A.Y.); 222007855004@zust.edu.cn (L.C.); 222007855035@zust.edu.cn (M.T.); 1200309030@zust.edu.cn (Y.Z.)

* Correspondence: bphou@zust.edu.cn

Abstract: Changes in clouds can affect the outpower of photovoltaics (PVs). Ground-based cloud images classification is an important prerequisite for PV power prediction. Due to the intra-class difference and inter-class similarity of cloud images, the classical convolutional network is obviously insufficient in distinguishing ability. In this paper, a classification method of ground-based cloud images by improved combined convolutional network is proposed. To solve the problem of sub-network overfitting caused by redundancy of pixel information, overlap pooling kernel is used to enhance the elimination effect of information redundancy in the pooling layer. A new channel attention module, ECA-WS (Efficient Channel Attention–Weight Sharing), is introduced to improve the network's ability to express channel information. The decision fusion algorithm is employed to fuse the outputs of sub-networks with multi-scales. According to the number of cloud images in each category, different weights are applied to the fusion results, which solves the problem of network scale limitation and dataset imbalance. Experiments are carried out on the open MGCD dataset and the self-built NRELCD dataset. The results show that the proposed model has significantly improved the classification accuracy compared with the classical network and the latest algorithms.

Keywords: convolutional neural network; classification of ground-based cloud images; combined convolutional network; overlap pooling; attention mechanism

Citation: Zhu, W.; Chen, T.; Hou, B.; Bian, C.; Yu, A.; Chen, L.; Tang, M.; Zhu, Y. Classification of Ground-Based Cloud Images by Improved Combined Convolutional Network. *Appl. Sci.* **2022**, *12*, 1570. https://doi.org/10.3390/app12031570

Academic Editor: Luis Hernández-Callejo

Received: 1 December 2021
Accepted: 29 January 2022
Published: 1 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Affected by short-term weather changes, the output power of PV power generation is easily fluctuated [1,2]. At present, the forecasting of PV power becomes an important method to reduce the impact of power fluctuations. Through power forecasting, the power sector can reasonably dispatch PV resources and reduce the impact of power fluctuations on grid-connected PVs. Dissipation and aggregation of cloud clusters in a short period of time are important factors that cause fluctuations in output power. Besides, solar irradiance is affected directly by the different types of clouds [3]. Different types of clouds have different characteristics, such as thickness, height, and sky coverage, which affect the magnitude of solar radiation received by the ground. Therefore, classification of clouds is crucial for PV power prediction.

There are various forms of clouds belonging to the same category, and different categories of clouds are also a transitional relationship, so they have greater similarity, which brings great challenges to the classification of clouds. In the early days, machine learning based classifiers were often used to classify cloud images. For example, Heinle et al. [4] used the K-nearest neighbors to classify the cloud by extracting the spectral and texture features of the cloud image. Kazantzidis et al. [5] introduced cloud classification by counting the color and texture features of cloud images, and at the same time considered multi-modal information as the input of the improved K-nearest neighbors classifier. Zhao et al. [6]

proposed that the texture, local structure, and statistical feature were used as the input of SVM and achieved an accuracy of 64.1% on the dataset of nine cloud categories.

With the continuous development of deep learning in recent years, the performance of convolutional neural network (CNN) in the field of image classification has been greatly improved. Because of better feature representation capabilities, CNN can mine more deep semantic features from the image. Liu et al. [7] produced a MGCD dataset with 8000 ground-based cloud images and corresponding meteorological data, yielding an accuracy as high as 87.9% with multi-modal fusion algorithm. Ye et al. [8] introduced a CNN to extract the features of cloud images; fisher vector coding and SVM classifier are utilized for cloud images classification. Zhang et al. [9] proposed a CloudNet model and obtained a high accuracy on a self-built CCSN dataset containing 2543 cloud images. Huertas-Tato et al. [10] proposed an ensemble learning algorithm to fuse the output probability vector of CNN and random forest classifier to improve the classification accuracy. In [11], the network named MMFN was proposed, which could learn extended cloud information by fusing heterogeneous features in a unified framework. In [12], the task-based graph convolutional network was introduced to obtain the correlation between cloud images, yielding an accuracy as high as 89.48%.

There are many cloud image datasets available for reference in existing research. It can be divided into satellite cloud images, part-sky ground-based cloud images, and all-sky ground-based cloud images. There are many research methods for the classification of satellite cloud images [13–16], and this classification is extremely effective for macro-level meteorological analysis. However, satellite cloud images cover a large area and have few local details. It is extremely difficult to analyze cloud clusters in a small patch of sky, which makes it not widely used in the field of PV power generation. For the classification of part-sky ground-based cloud images, there have also been many research methods [17–19]. However, part-sky ground-based cloud images have a small field of view and cannot meet the large-scale PV power station requirements. The field of view of the all-sky ground-based cloud images is generally 180°, which can capture most of the sky, as shown in Figure 1. The collected cloud images have clear textures and rich structural features, which are suitable for PV power stations of almost all sizes. However, there are still few studies on this type of cloud images, and the amount of data in public datasets is also small. In some existing studies, the accuracy of classification is low, which cannot well meet the application in PV power generation. In response to the above problems, we made a ground-based cloud images dataset with a larger amount of data and proposed a deep learning-based ground-based cloud image classification method. The main contributions are shown as follows:

(1) By collecting historical cloud images data published by the National Renewable Energy Laboratory (NREL) on the US Measurement and Instrument Data Center (MIDC) website, a ground-based cloud images dataset NRELCD (NREL Cloud Dataset) is constructed. The dataset contains 15,450 cloud images and is divided into 7 categories.

(2) A novel ground-based cloud images classification method by improved combined neural network is proposed; overlap pooling kernels are used in the sub-network to improve the effect of eliminating information redundancy and reduce the risk of overfitting. The improved channel attention module ECA-WS is introduced after the pooling layer, which further enhances the sub-network's ability to express channel characteristics. The synchronization of parameter optimization among sub-networks is realized by improving the sub-networks. The decision fusion algorithm is used to weight the output of the two sub-networks in the combined network to improve the classification accuracy significantly.

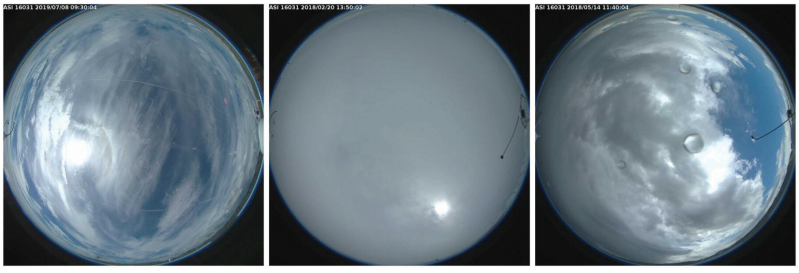


Figure 1. Ground-based cloud images.

The rest of this paper is organized as follows: Section 2 briefly introduces some related work; Section 3 describes a novel ground-based cloud images classification method based on improved combined neural network; and Section 4 presents experimental results and some discussion. At the end of this paper, some remarking points are given in Section 5.

2. Related Work

2.1. Deep Feature Extraction Network

ResNet50 [20] and VGG16 [21] deep convolutional neural networks are introduced to obtain more deep features. ResNet50 is composed of multiple residual blocks, and each residual block added a direct connection channel. The residual learning algorithm can reduce the loss of information when the feature is propagated to the deeper network. VGG16 replaces a larger size convolution kernel by stacking multiple 3 × 3 size convolution kernels, which ensures that the network can learn more complex nonlinear mapping modes while obtaining the same receptive field.

2.2. ECA Attention Mechanism

The attention mechanism can redistribute the originally evenly allocated resources according to the importance of the objects. The contrast between different features is enhanced and useful features are more prominent. Many attention mechanisms such as MAT [22], IHSM&EFRM [23], CBAM [24], SE [25], and ECA [26] have been applied in various visual tasks. MAT module consists of a soft attention unit and an attention transition unit, which allows the transition of attentive motion features to enhance appearance learning at each convolution stage and enrich spatio-temporal object features. This module is of great value in video analysis tasks. IHSM&EFRM are used in the human–object interaction detection task, and they enhance the expression of human and object features, respectively. CBAM module includes channel and space dual attention, which improves the feature extraction ability of the network in multi-dimensions. SE module uses squeeze-and-excitation to learn the relationship among each channel and assign different weights to different channels. Though dimensionality reduction can reduce model complexity, it destroys the direct correspondence between channel and its weight. ECA module proposes a local cross-channel interaction strategy without dimensionality reduction based on the SE. This method can adaptively select the size of the 1D convolution kernel.

The ECA module is shown in Figure 2, where $\boldsymbol{X}$ and $\boldsymbol{X}'$ represent input and output feature maps, and w, h, c represent their width, height, and channel dimensions, respectively. As shown in Equation (1), after using global average pooling (GAP) on the input feature map $\boldsymbol{X}$, a channel coding vector γ_{gap} with a size of $1 \times 1 \times c$ is obtained.

$$\gamma_{gap} = \frac{1}{wh} \sum_{i=1,\, j=1}^{w,h} \boldsymbol{X}_{ij\cdot}, \boldsymbol{X} \in \mathbb{R}^{w \times h \times c}, \tag{1}$$

$$\boldsymbol{\eta}_{gap} = \sigma(\boldsymbol{V}_k^{gap} \gamma_{gap}), \boldsymbol{V}_k^{gap} \in \mathbb{R}^{c \times c}, \tag{2}$$

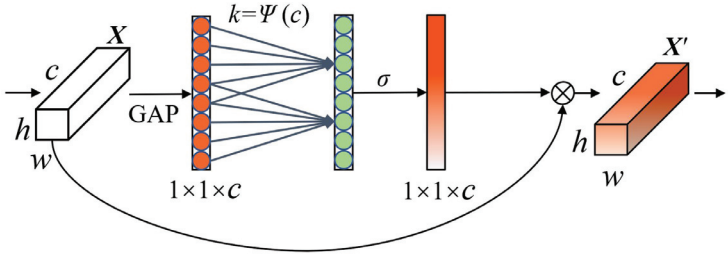


Figure 2. ECA attention module.

As shown in Equation (3), where V_k^{gap} is a band weight matrix. The channel weight vector η_{gap} can be obtained after normalization by the σ (sigmoid) function. The k can be adaptively calculated according to

$$k = \psi(c) = |\frac{\log_2(c)}{y} + \frac{b}{y}|_{odd}, \tag{3}$$

$\psi(\cdot)$ indicates the mapping relationship between c and k. $|\cdot|_{odd}$ indicates the nearest odd number. y and b are custom mapping parameters, here 2 and 1. The feature vector obtained after 1D convolution still maintains its original dimension.

2.3. Decision Fusion

Different networks have different classification probability for the same sample, which is embodied in the output vector of the network. When the output probabilities of each category are close to equal, it can be considered that the corresponding network hardly makes a positive judgment on the sample [27]. If multiple networks are used to make joint decisions, the probability of the sample being correctly classified will greatly increase, as shown in Figure 3.

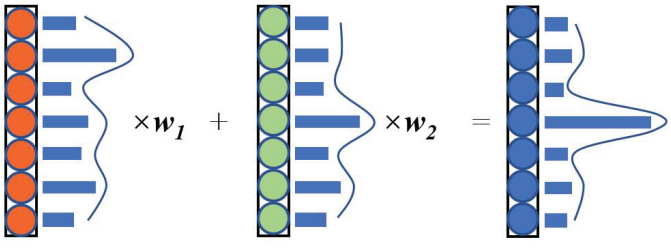


Figure 3. Multiple network fusion decision. When the probabilities of different categories are nearly equal, use a smaller value to multiply the vector. Otherwise, use a larger value to multiply the vector.

3. Our Proposed Methods

The model consists of four parts, which are deep feature combined network, overlap pooling, improved ECA module, and decision fusion. The specific structure of the model is shown in Figure 4.

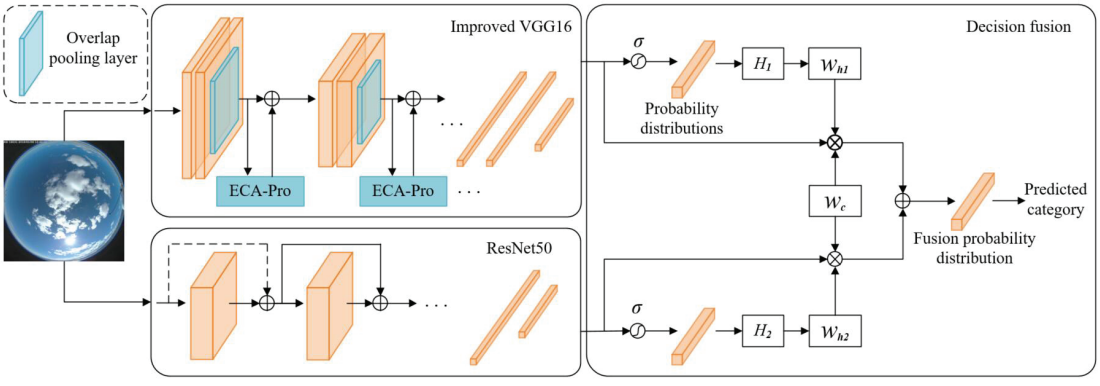


Figure 4. Classification network model of ground-based cloud images.

3.1. Combined Network

The existing ground-based cloud images classification is usually based on a single channel network. However, there are limitations on performance of feature extract. At the same time, there is no such requirement as real-time for the classification of ground-based cloud images in PV power prediction. Therefore, in order to improve the classification accuracy, a combined network is used to improve the ability of the model for feature extraction. Here, ResNet50 and VGG16 networks are used to extract the depth and width feature of images. ResNet50 can extract more deep semantic features in ground-based cloud images because of its depth advantages. However, the network scales four times the width and height dimension of the first convolutional layer, resulting in the loss of some image features during subsequent convolutions. The first-stage convolutional layer of VGG16 can perform feature extraction at the input dimension, so the feature extraction advantage on the width and height dimension is more obvious.

In our experiments, we found that the results of ResNet50 tend to be stable when training on a dataset with less data, while VGG16 undergoes overfitting, which degrades the accuracy. Such results are unfavorable for decision fusion, and the decision fusion algorithm can only perform its best when the performance between sub-networks is close to the same. In this regard, we improve the pooling layer structure and the ECA-Pro module to ensure that the accuracy of VGG16 will not be degraded.

3.2. Overlap Pooling

Most of the clouds in the ground-based cloud images are grayish-white, and the background sky shows a uniform blue color. This results in a high degree of similarity between adjacent pixels of the image. This similarity leads to a higher information redundancy in ground-based cloud images compared to other images and also makes the VGG16 network more prone to overfitting. The network parameter quantity and overfitting caused by the image information redundancy can be reduced through the feature map pooling. However, the 2 × 2 pooling kernels cannot significantly improve the down-sampling quality of high-redundancy images.

As shown in Figure 5, a pooling kernel of size 3 × 3 is used instead of the original pooling kernel. The redundancy of features can be eliminated while the common features of adjacent receptive fields will be extracted. The pooling step is set to 2, which makes the pooling kernels overlap each other and adjacent receptive fields overlap each other as well. At the same time, the feature correlation and the overfitting suppression ability of the network are promoted.

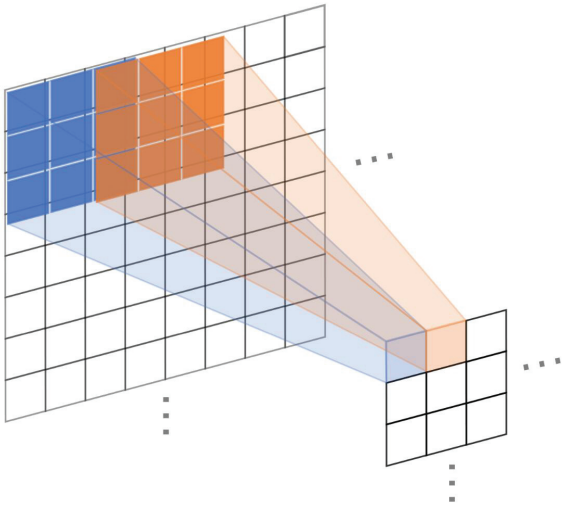


Figure 5. 3 × 3 overlap pooling kernel.

3.3. Improved ECA

Referring to the ground-based cloud image sample in Figure 1, due to the all-sky imager being utilized to acquire the image, only the inscribed circular part of the image is the effective area, and the surrounding four corners are invalid black pixels. Therefore, although GAP has strong noise suppression capabilities, it still does not work well for the special images. To avoid the above-mentioned problems, global max pooling (GMP) is used to prevent the introduction of invalid parts in the feature calculation, which improves the ability of channel feature extraction to a certain extent. The GMP is shown in Equation (5)

$$\gamma_{gmp} = \max_{\substack{i\in[1,w]\\ j\in[1,h]}}(\boldsymbol{X}_{ij\cdot}), \boldsymbol{X} \in \mathbb{R}^{w\times h\times c}, \tag{4}$$

Like global average pooling, the vector γ_{gmp} is multiplied by the band weight matrix $\boldsymbol{V}_k^{gmp}$. The output obtained is the channel weight vector $\boldsymbol{\eta}_{gmp}$

$$\boldsymbol{\eta}_{gmp} = \sigma(\boldsymbol{V}_k^{gmp}\gamma_{gmp}), \boldsymbol{V}_k^{gmp} \in \mathbb{R}^{c\times c}, \tag{5}$$

Then, the ECA attention module is spliced with the pooling layer, GAP and GMP are employed to jointly extract the global features of $\boldsymbol{X}$. It should be noted that GAP and GMP are used here in parallel to aggregate the spatial information of the two feature maps. If GAP is executed before GMP, then after the GMP operation, all the values less than the maximum value in the feature map will be discarded, and this is the useful information processed by GAP.

$$\boldsymbol{X}' = \sigma(\boldsymbol{\eta}_{gap} + \boldsymbol{\eta}_{gmp})\boldsymbol{X}, \boldsymbol{\eta}_{gap} \in \mathbb{R}^{1\times 1\times c}, \boldsymbol{\eta}_{gmp} \in \mathbb{R}^{1\times 1\times c}, \tag{6}$$

where $\boldsymbol{X}'$ represents the output of the entire attention module.

In the ECA module, there are two combinations of GAP and GMP. As illustrated in Figure 6a,b, we named them the weight sharing method (ECA-WS) and the weight independent method (ECA-WI), respectively.

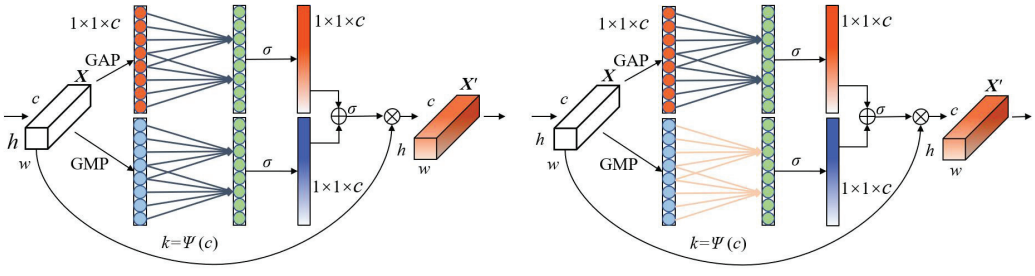


Figure 6. Improved ECA. (**a**) GAP share weights with GMP (ECA-WS); (**b**) GAP and GMP weight independent (ECA-WI).

In ECA-WS, GMP uses the same band weight matrix with GAP, that is, $v_k^{gap} = v_k^{gmp}$. It hardly increases the parameters of the ECA module but improves the feature extraction capability of the module. In ECA-WI, GMP uses the different band weight matrix with GAP, that is, $v_k^{gap} \neq v_k^{gmp}$. However, this method not only increases the network parameters but also reduces the correlation between the two global pooling. This may lead to the instability of the combined network performance, which is also proved by comparative experiments. Therefore, ECA-WS attention mechanism is used to improve the characterization ability of the channel characteristics of the VGG16 sub-network after each stage of pooling.

3.4. Decision Fusion

Considering the significant effects of fusion of multiple networks on the improvement of classification performance, a weighted algorithm is introduced to fuse the output results of the two sub-networks.

Details of the algorithm are shown as follows,

$$\boldsymbol{l}_i = \sigma(\boldsymbol{x}_i) = \frac{1}{1 + e^{-\boldsymbol{x}_i}}, i \in [1, 2], \tag{7}$$

first, the σ function is used to normalize each network output values. Where $\boldsymbol{x}_i$ indicates the output vector of the i-th sub-network.

The normalized probability distribution is expressed as

$$\boldsymbol{P}_i = \frac{\boldsymbol{l}_{im}}{\sum_{j=1}^{n} \boldsymbol{l}_{ij}}, \begin{cases} m \in [1, n] \\ i \in [1, 2] \end{cases}, \tag{8}$$

where $\boldsymbol{P}_i$ is the probability vector output by the network, $\boldsymbol{l}_{im}$ is the value of the m-th category, and n indicates the number of cloud image categories.

The output probability matrix of the transformed combined network is expressed as

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}_1 \\ \boldsymbol{P}_2 \end{bmatrix}, \tag{9}$$

Subsequently, the information entropy H_i of the probability distribution of each network output is

$$H_i = -\sum_{j=1}^{n} \boldsymbol{P}_{ij} \log_2 \boldsymbol{P}_{ij}, i \in [1, 2], \tag{10}$$

The greater information entropy, the greater uncertainty of the network prediction results and the higher probability of being misdiagnosed. Therefore, a lower weight should be assigned. The weight w_{hi} of sub-network is given as follows

$$w_{hi} = \frac{e^{-H_i}}{\sum_{k=1}^{n} e^{-H_k}}, i \in [1,2], \tag{11}$$

The weight w_{cm} of each category according to the number of samples of different categories is defined as follows

$$w_{cm} = \frac{1 - \frac{N_m}{N_{total}}}{\sum_{j=1}^{n}(1 - \frac{N_j}{N_{total}})}, \begin{cases} m \in [1,n] \\ i \in [1,2] \end{cases}, \tag{12}$$

where N_m represents the number of samples in the m-th category and N_{total} represents the total number of samples.

Multiply the sub-network weight w_{ci} and category weight w_{ci} with the original output vector $\boldsymbol{x}$. The weighted matrix $\boldsymbol{W}$ of the combined network is written as

$$\boldsymbol{W} = \begin{bmatrix} w_{h1}w_{c1}\boldsymbol{x}_{11} & w_{h1}w_{c2}\boldsymbol{x}_{12} \cdots w_{h1}w_{cn}\boldsymbol{x}_{1n} \\ w_{h2}w_{c1}\boldsymbol{x}_{21} & w_{h2}w_{c2}\boldsymbol{x}_{22} \cdots w_{h2}w_{cn}\boldsymbol{x}_{2n} \end{bmatrix}, \tag{13}$$

Add $\boldsymbol{W}$ in rows,

$$\boldsymbol{W}' = \left[\sum_{i=1}^{2} w_{hi}w_{c1}\boldsymbol{x}_{i1}, \sum_{i=1}^{2} w_{hi}w_{c2}\boldsymbol{x}_{i2}, \cdots, \sum_{i=1}^{2} w_{hi}w_{cn}\boldsymbol{x}_{in} \right], \tag{14}$$

Take the maximum index of the column to be the final classification decision result,

$$lable = arg\ max(\boldsymbol{W}'). \tag{15}$$

4. Result and Discussion

4.1. Dataset for Classification

The NRELCD and the MGCD dataset are used to prove the effectiveness of the method in this paper. The NRELCD dataset comes from the historical ground-based cloud images published by the National Renewable Energy Laboratory (NREL) in the United States. The image size is 1024 × 1024 pixels and the collection period is from 2018 to 2020. We screened 15,450 images with distinct category characteristics and analyzed the effect of each category on solar radiation attenuation to produce a dataset specific to the PV power sector.

The amount of solar radiation received by the PV panels is a direct factor affecting the power generated by PV. According to the method described in the literature [28], we collected the all-day radiation for days when there was only one cloud genus in the sky and also for the nearest cloudless day to that day. As shown in Figure 7, the two data are compared to see the effect of each cloud category on the solar radiation attenuation. Nearly similar cloud types are combined according to the degree of attenuation of solar radiation by each type of clouds. Cloud categorization is not as clear as other categorization tasks. Clouds are sometimes in a transitional state and their classification is highly controversial. We did not use those transition state cloud images as part of the dataset but instead forced them to be classified in a practical application using a trained classifier. Since the cloud images in the NRELCD have continuity in time, the adjacent images are relatively similar. Random distribution of the dataset to the training set and the test set will cause the accuracy to be high, which is not in line with the actual situation. Therefore, the first 70% of the dataset is used as the training set and the rest as the test set according to the acquisition time.

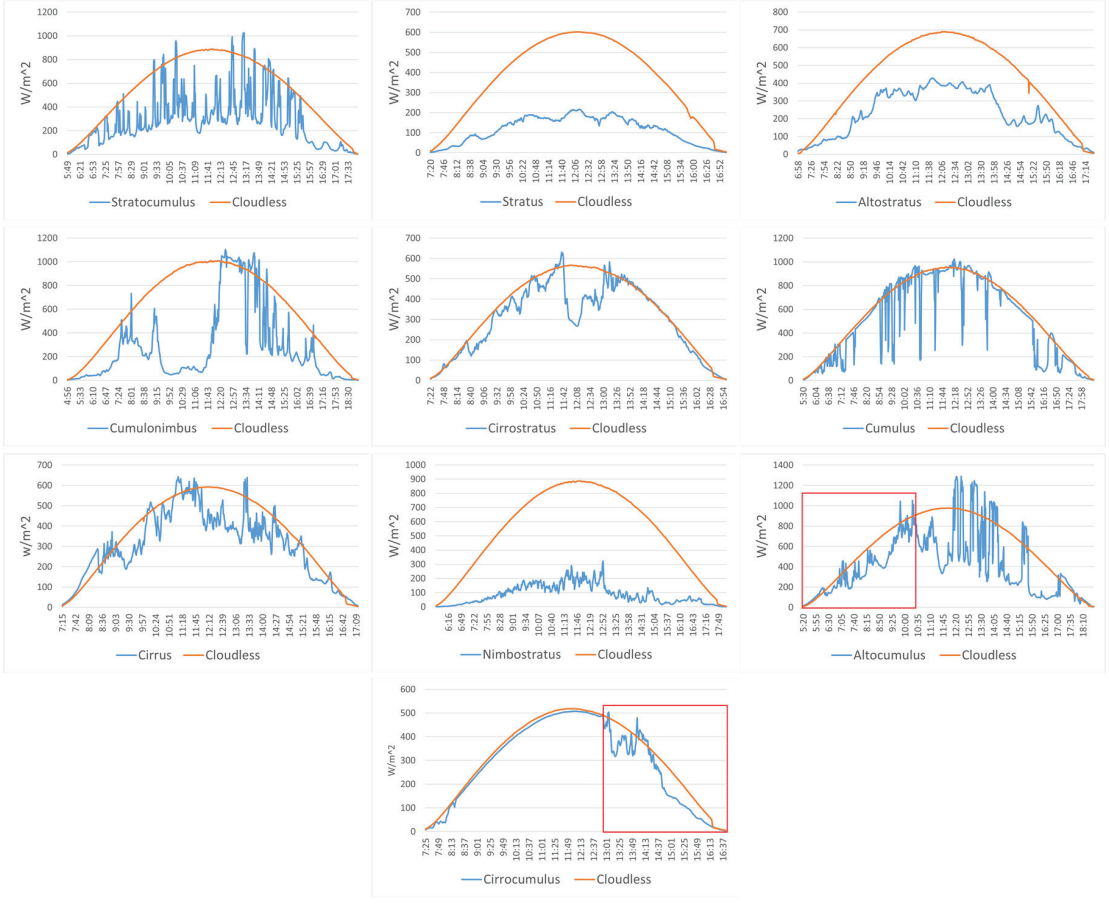
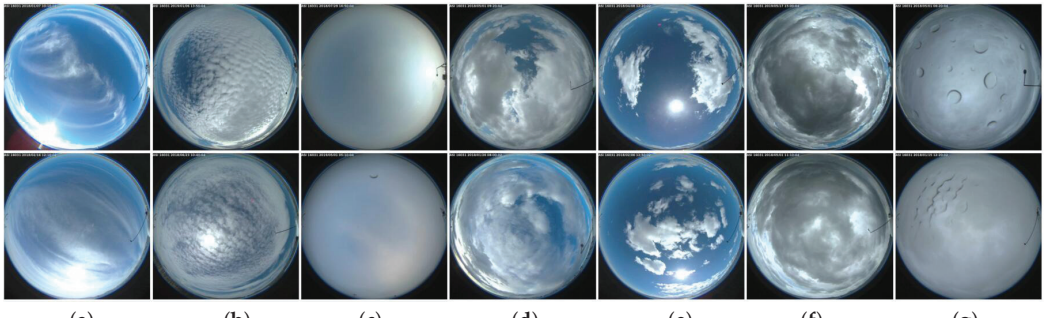


Figure 7. Effect of different cloud genera on solar radiation attenuation. Altocumulus and Cirrocumulus are infrequent, we have boxed the time of occurrence in red.

Different categories of clouds have different probabilities of occurring in different regions. In order to adapt the power prediction to the real environment, there are different cloud classification criteria. We try to test the classification ability of the model under different cloud classification criterias with two datasets with different categories. The MGCD dataset [7] also includes 7 categories, with a total of 8000 ground-based cloud images, each of which corresponds to 4 modals of weather data. The data format of cloud images is JPEG, and the size is 1024 × 1024 pixels. The training set and the test set contain 4000 cloud image samples, respectively. We only use cloud image samples in the dataset for experiments. Mixed clouds in the MGCD dataset are composed of multiple categories of clouds. Objectively speaking, these combinations have different effects on solar radiation, so putting mixed clouds into one category is not reasonable in practice, but this classification criterion is informative for testing model performance. Table 1 shows the number of samples in each dataset, while Figures 8 and 9, respectively, listed the cloud samples of the two datasets.

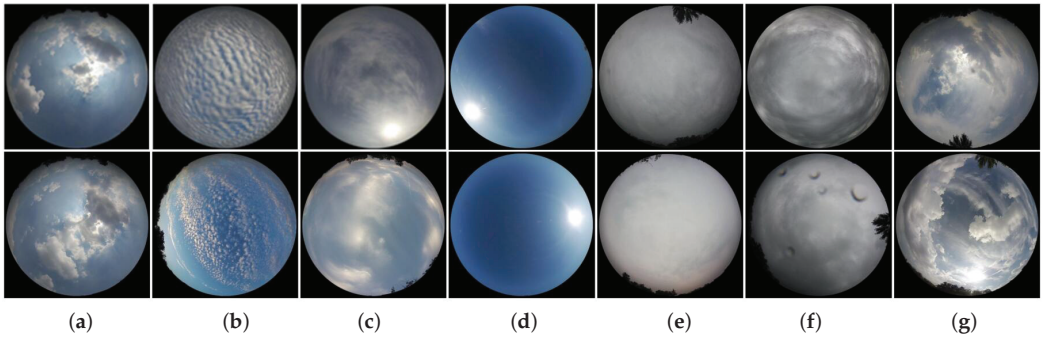
Table 1. Ground-based cloud images dataset.

NRELCD				MGCD			
Type (ABBR.)	**Train**	**Test**	**Quantity**	**Type (ABBR.)**	**Train**	**Test**	**Quantity**
Cirrus and Cirrostratus (Ci and Cs)	1673	716	2389	Cumulus (Cu)	690	748	1438
Altocumulus and Cirrocumulus (Ac and Cc)	876	376	1252	Altocumulus and Cirrocumulus (Ac and Cc)	400	331	731
Altostratus and Stratus (As and St)	1534	658	2192	Cirrus and Cirrostratus (Ci and Cs)	650	673	1323
Stratocumulus (Sc)	1362	585	1947	Clear sky (Clear sky)	650	688	1338
Cumulus (Cu)	1410	605	2015	Stratocumulus, Altostratus and Stratus (Sc, St and As)	500	463	963
Cumulonimbus (Cb)	1293	555	1848	Cumulonimbus and Nimbostratus (Cb and Ns)	600	587	1187
Nimbostratus (Ns)	2652	1155	3807	Mixed (Mixed)	510	510	1020
Total	10,800	4650	15,450	Total	4000	4000	8000



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**)

Figure 8. NRELCD dataset. (**a**) Ci and Cs; (**b**) Ac and Cc; (**c**) As and St; (**d**) Sc; (**e**) Cu; (**f**) Cb; (**g**) Ns. Considering that there are extremely high structural similarities between certain types of clouds, this will make their impact on PV power generation very similar or almost the same. Some changes are made to the International Meteorological Organization's cloud classification standards. The main work is to combine the original cirrus and cirrostratus, altocumulus and cirrocumulus, and altostratus and stratus.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**)

Figure 9. MGCD dataset. (**a**) Cu; (**b**) Ac and Cc; (**c**) Ci and Cs; (**d**) clear sky; (**e**) Sc, St and As; (**f**) Cb and Ns; (**g**) mixed. This dataset combines altocumulus and cirrocumulus, cirrocumulus and cirrostratus, stratocumulus and stratus and altostratus, and cumulonimbus and nimbostratus, respectively.

4.2. Experimental Setup

In the experiment, the Ubuntu 18.04 operating system with 128 G of running memory and an RTX3090 graphics card with 24 G of video memory are implemented. The deep learning framework is Pytorch1.7.1, and the CUDA version is 11.1.

Data augmentation on the cloud images were performed to increase the noise of the dataset and the robustness of the model. The operations included (1) random horizontal flip, (2) vertical flip, (3) grayscale with a probability of 50%, (4) random rotation of 45°, and (5) random change of brightness, contrast, saturation, and hue. The input cloud images' size was adjusted to 224 × 224 pixels, and transfer learning was used to train these cloud images. The learning rate of two sub-networks is fixed to 0.000001, and the Adam optimizer is used to optimize the gradient operation. The batch size of training and testing are both set to 50. The epoch of the NRELCD dataset was set to 200, and the epoch of the MGCD dataset was set to 100.

4.3. Ablation Experiment

Ablation experiments are used to compare several improved methods proposed in this paper. The results are listed in Table 2. Where CN is combined network, OP is overlap pooling, ICN is improved combined network we proposed, CC is the number of correct classifications, P is average accuracy, R is average recall, F1 is average F1-score, Acc is overall classification accuracy, and K (Kappa) is consistency index.

Table 2. Ablation experiment.

Method	MGCD/4000 Test Samples						NRELCD/4650 Test Samples					
	CC	**P**	**R**	**F1**	**Acc**	**K**	**CC**	**P**	**R**	**F1**	**Acc**	**K**
ResNet50	3522	85.67%	86.12%	85.69%	88.05%	0.8593	4373	93.33%	93.12%	93.12%	94.04%	0.9291
VGG16	3488	84.77%	84.86%	84.76%	87.20%	0.8492	4378	93.40%	93.30%	93.27%	94.15%	0.9304
CN	3551	86.23%	87.02%	86.53%	88.78%	0.8679	4411	94.17%	94.23%	94.17%	94.86%	0.9389
CN + OP	3557	86.77%	86.87%	86.69%	88.93%	0.8696	4436	94.88%	94.67%	94.77%	95.40%	0.9453
CN + OP + ECA	3585	87.59%	87.91%	87.55%	89.63%	0.878	4440	94.91%	94.79%	94.84%	95.55%	0.9471
ICN	3603	88.09%	88.15%	87.85%	90.08%	0.8834	4445	95.11%	94.93%	95.01%	95.60%	0.9477

Table 2 illustrates the performance of the model at various stages from the basic structure to the final structure. In the ICN model, accuracy of the MGCD dataset is increased by 2.03% and 2.88%, respectively, compared with the sub-networks ResNet50 and VGG16, while the accuracy of the NRELCD dataset is increased by 1.56% and 1.45%, respectively. Figure 10 is the accuracy curve of the model on different datasets as the epoch increases. As shown in Figure 9a, VGG16 produces overfitting in training on MGCD dataset. At this time, the parameters of the ResNet50 are still being further optimized. The overfitting phenomenon is suppressed effectively on improved VGG16(VGG16 + OP + ECA-WS), and the parameter optimization of the two sub-networks is close to synchronization, which provides a good prerequisite for decision fusion. As illustrated in Figure 10b, improved VGG16 is better than VGG16 in performance on the NRELCD dataset with a larger scale. The optimization process of the two sub-networks is also almost synchronized in time.

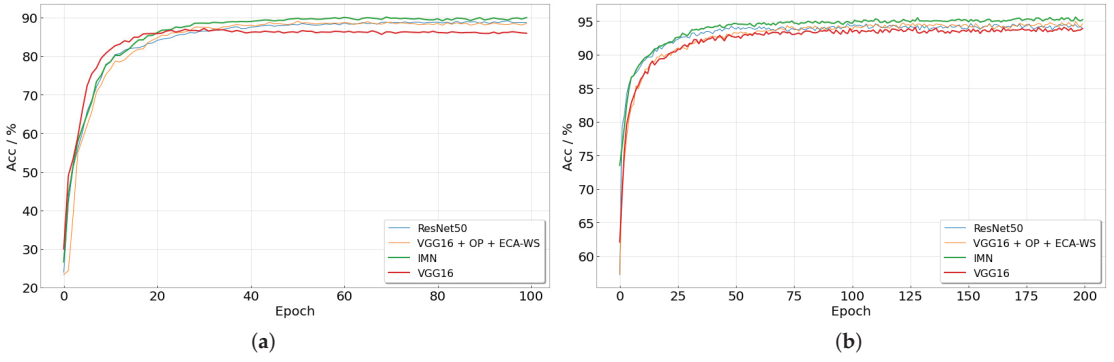


Figure 10. Experimental results of different datasets. (**a**) Accuracy of MGCD, (**b**) accuracy of NRELCD.

4.4. Overlap Pooling Experiment

To verify the effect of overlap pooling, the pooling kernels in the VGG16 were changed to 2×2, 3×3, 4×4, and 5×5.

In Figure 11, it can be seen that the 3×3 overlap pooling kernel has the highest accuracy on the two datasets. The main reason is that the 3×3 overlap pooling kernel has a greater ability to eliminate the redundancy of image feature information than the 2×2 pooling kernel while retaining useful information. At the same time, a larger kernel may reduce information redundancy while causing more useful information to be lost in the pooling process, thereby affecting the overall performance of the network.

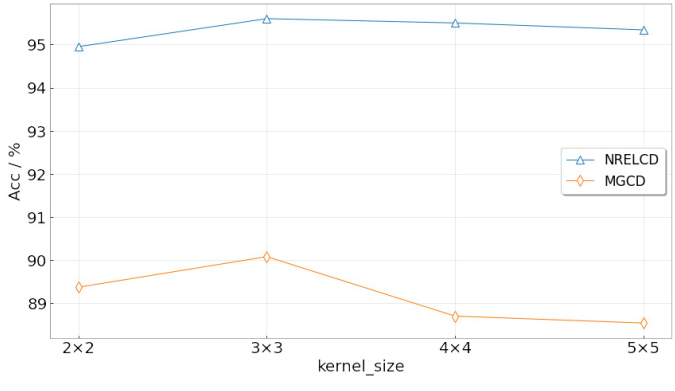


Figure 11. Variation of classification accuracy with respect to the kernel size.

4.5. Attention Mechanism Experiment

To verify the role of the attention mechanism, GRAD-CAM [29] is used to visualize the VGG16, VGG16 and ECA, and VGG16 and ECA-WS. A piece of ground-based cloud image is randomly selected from different categories. The results are shown in Figures 12 and 13. Network's attention to the cloud area can be increased significantly by embedding the ECA attention mechanism into the VGG16 network. However, sometimes it can only focus on part of the cloud or focus part of the attention on the sky background. The ECA-WS attention mechanism we proposed can focus attention on image areas with more inter-class differences, which improves the classification ability of the network.

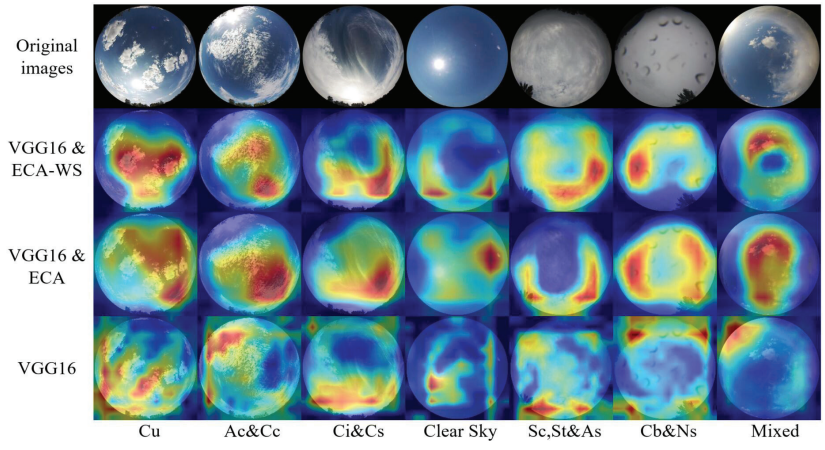


Figure 12. MGCD category heat maps.

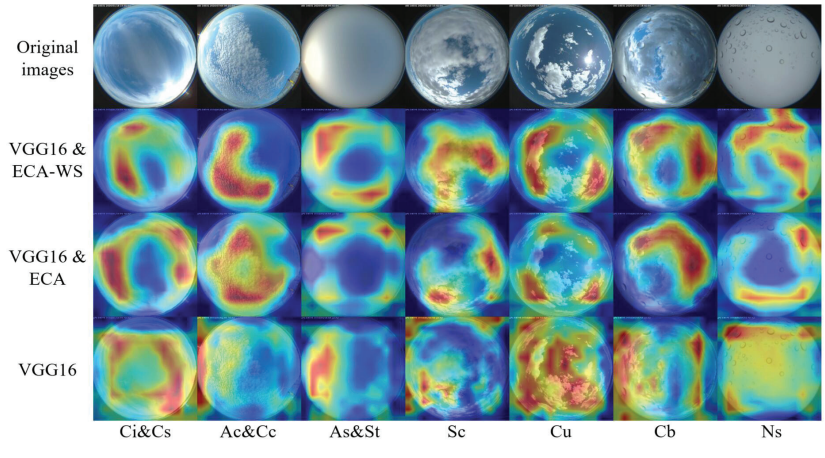


Figure 13. NRELCD category heat maps.

Different attention mechanisms are used for comparative experiments, and the results are listed in Table 3, where backbone is combined network with improved pooling layer. CBAM module, SE module, ECA module, ECA-WI module, and ECA-WS module are introduced, respectively, based on backbone.

Table 3. Comparative result of attention mechanisms.

Attention Mechanisms	MGCD	NRELCD
Backbone + CBAM	88.95%	94.51%
Backbone + SE	89.51%	95.30%
Backbone + ECA	89.63%	95.55%
Backbone + ECA-WI	89.53%	95.55%
Backbone + ECA-WS	90.08%	95.60%

In Table 3, the ECA-WS module has the best performance on improving the accuracy of network classification than other attention modules. However, the ECA-WI cannot improve the performance of the module and may even lead to a decrease.

4.6. Classification Method Experiment

From Table 4, the performance of method [4] based on texture or spectral feature on the two datasets is worse deep learning. Cloud images have more texture features and deep semantic features than other images, and only by acquiring more image features can we satisfy the classification needs of such images. In recent years, CNN has been widely used in ground-based cloud images classification tasks, thanks to its powerful feature extraction capabilities. Method [9] has shown good results on both datasets. Methods [7,11] fuse multimodal meteorological data and CNN, and achieved accuracy rates of 87.90% and 88.63%, respectively.

While good classification accuracy is achieved on other classic CNN models, the one that achieved the highest on our model had accuracy of the MGCD dataset that reached 90.08% and of the NREL dataset, it reached 95.60%. The combined network greatly optimizes the probability distribution of the classification output vector. By comparing with the latest algorithms and single network, the results show that the method in this paper has a greater improvement in classification accuracy, and it also proves the generalization ability of the method. For ground-based cloud images collected in different regions, the model has strong robustness, which will play a positive role in the field of PV power generation forecasting.

Table 4. Comparative experiment of classification methods.

Methods	MGCD	NRELCD
Method [4]	68.90%	75.61%
Method [9]	81.14%	92.17%
Method [7]	87.90%	-
Method [11]	88.63%	-
ResNet50	88.05%	94.04%
VGG16	87.20%	94.15%
GoogleNet	87.53%	93.54%
Inception_v3	88.32%	94.20%
MobileNet_v2	86.92%	93.73%
Ours	90.08%	95.60%

Note. References [7,11] use the meteorological data in MGCD, and NRELCD does not contain meteorological data.

5. Conclusions

In this paper, a combined network-based ground-based cloud images classification method is proposed. Specifically, the ResNet50 and VGG16 networks are combined using decision fusion algorithm, which uses dual weights to weight the output of the sub-network. In addition, to optimize the parameters of the two sub-networks to approach synchronization, overlap pooling is used to replace the original VGG16 pooling layer. At the same time, the ECA-WS module is embedded after the pooling layer to improve the cross-channel interaction capability of the network. We constructed the NRELCD dataset that meets the actual application scenarios and used the MGCD dataset to verify the advanced nature of the network model.

At present, our classification of clouds is only based on image features. In reality, there are many physical characteristics that can provide a basis for cloud classification, such as height, thickness, and speed. In the future, we will consider using these parameters in classification research to improve the performance of the model.

Author Contributions: All authors made significant contributions to the manuscript. W.Z. and T.C. conceived, designed, and performed the experiments, and wrote the paper; B.H. and C.B. collected data and performed the experiments; A.Y. analyzed the data and performed the experiments; L.C., M.T. and Y.Z. revised the paper and provided the background knowledge of cloud classification. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Research and Development Project of Zhejiang Province, Grant Number 2021C04030, and the Public Project of Zhejiang Province, Grant Number LGG21F030004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data collected in this study came from the National Renewable Energy Laboratory, available at https://midcdmz.nrel.gov/apps/sitehome.pl?site=BMS (accessed on 1 May 2021). The MGCD dataset can be obtained from (shuangliu.tjnu@gmail.com).

Acknowledgments: The all-sky cloud images in the NRELCD dataset were obtained by the National Renewable Energy Laboratory in the United States. The MGCD dataset used in the experiment was also allowed by Liu Shuang's team. We would like to express our sincere thanks.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Sol. Energy* **2016**, *136*, 125–144. [CrossRef]
2. Akhter, M.N.; Mekhilef, S.; Mokhlis, H.; Shah, N.M. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renew. Power Gener.* **2019**, *13*, 1009–1023. [CrossRef]
3. Govender, P.; Sivakumar, V. Investigating diffuse irradiance variation under different cloud conditions in Durban, using k-means clustering. *J. Energy South. Afr.* **2019**, *30*, 22–32. [CrossRef]
4. Heinle, A.; Macke, A.; Srivastav, A. Automatic cloud classification of whole sky images. *Atmos. Meas. Tech.* **2010**, *3*, 557–567. [CrossRef]
5. Kazantzidis, A.; Tzoumanikas, P.; Bais, A.F.; Fotopoulos, S.; Economou, G. Cloud detection and classification with the use of whole-sky ground-based images. *Atmos. Res.* **2012**, *113*, 80–88. [CrossRef]
6. Zhuo, W.; Cao, Z.; Xiao, Y. Cloud classification of ground-based images using texture–structure features. *J. Atmos. Ocean. Technol.* **2014**, *31*, 79–92. [CrossRef]
7. Liu, S.; Duan, L.; Zhang, Z.; Cao, X. Hierarchical multimodal fusion for ground-based cloud classification in weather station networks. *IEEE Access* **2019**, *7*, 85688–85695. [CrossRef]
8. Ye, L.; Cao, Z.; Xiao, Y.; Li, W. Ground-based cloud image categorization using deep convolutional visual features. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 4808–4812.
9. Zhang, J.; Liu, P.; Zhang, F.; Song, Q. CloudNet: Ground-based cloud classification with deep convolutional neural network. *Geophys. Res. Lett.* **2018**, *45*, 8665–8672. [CrossRef]
10. Huertas-Tato, J.; Martín, A.; Camacho, D. Cloud type identification using data fusion and ensemble learning. In Proceedings of the Intelligent Data Engineering and Automated Learning (IDEAL), Guimaraes, Portugal, 4–6 November 2020; pp. 137–147.
11. Liu, S.; Li, M.; Zhang, Z.; Xiao, B.; Durrani, T.S. Multi-evidence and multi-modal fusion network for ground-based cloud recognition. *Remote Sens.* **2020**, *12*, 464. [CrossRef]
12. Liu, S.; Li, M.; Zhang, Z.; Xiao, C.; Durrani, T.S. Ground-Based Cloud Classification Using Task-Based Graph Convolutional Network. *Geophys. Res. Lett.* **2020**, *47*, e2020GL087338. [CrossRef]
13. Jin, W.; Gong, F.; Zeng, X.; Fu, R. Classification of clouds in satellite imagery using adaptive fuzzy sparse representation. *Sensors* **2016**, *16*, 2153. [CrossRef] [PubMed]
14. Kostornaya, A.A.; Saprykin, E.I.; Zakhvatov, M.G.; Tokareva, Y.V. A method of cloud detection from satellite data. *Russ. Meteorol. Hydrol.* **2017**, *42*, 753–758. [CrossRef]
15. Christodoulou, C.I.; Michaelides, S.C.; Pattichis, C.S.; Kyriakou, K. Classification of satellite cloud imagery based on multi-feature texture analysis and neural networks. In Proceedings of the 2001 International Conference on Image Processing (ICIP), Thessaloniki, Greece, 7–10 October 2001; IEEE: Piscataway, NJ, USA; pp. 497–500.
16. Chen, X.; Liu, L.; Gao, Y.; Zhang, X.; Xei, S. A Novel Classification Extension-Based Cloud Detection Method for Medium-Resolution Optical Images. *Remote Sens.* **2020**, *12*, 2365. [CrossRef]
17. Luo, Q.; Zhou, Z.; Meng, Y.; Li, Q.; Li, M. Ground-based cloud-type recognition using manifold kernel sparse coding and dictionary learning. *Adv. Meteorol.* **2018**, *2018*, 9684206. [CrossRef]
18. Kliangsuwan, T.; Heednacram, A. Feature extraction techniques for ground-based cloud type classification. *Expert Syst. Appl.* **2015**, *42*, 8294–8303. [CrossRef]
19. Wang, Y.; Shi, C.; Wang, C.; Xiao, B. Ground-based cloud classification by learning stable local binary patterns. *Atmos. Res.* **2018**, *207*, 74–89. [CrossRef]

20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
21. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
22. Zhou, T.; Wang, S.; Zhou, Y.; Yao, Y.; Li, J.; Shao, L. Motion-attentive transition for zero-shot video object segmentation. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), Hilton New York Midtown, NY, USA, 7–12 February 2020; pp. 13066–13073.
23. Zhou, T.; Wang, W.; Qi, S.; Ling, H.; Shen, J. Cascaded human-object interaction recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 4263–4272.
24. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
25. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
26. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 1–12.
27. Chen, W.; Liu, W.; Li, K.; Wang, P.; Zhu, H.; Zhang, Y.; Hang, C. Rail crack recognition based on adaptive weighting multi-classifier fusion decision. *Measurement* **2018**, *123*, 102–114. [CrossRef]
28. Matuszko, D. Influence of the extent and genera of cloud cover on solar radiation intensity. *Int. J. Climatol.* **2012**, *32*, 2403–2414. [CrossRef]
29. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, A.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.

entropy

MDPI

Article

0D Dynamic Modeling and Experimental Characterization of a Biomass Boiler with Mass and Energy Balance

Fateh Mameri [1], Eric Delacourt [1,2], Céline Morin [1,2,*] and Jesse Schiffler [3]

[1] CNRS, UMR 8201–LAMIH, University Polytechnique Hauts-de-France, 59313 Valenciennes, France; fatehmameri@hotmail.fr (F.M.); Eric.Delacourt@uphf.fr (E.D.)
[2] INSA Hauts-de-France, 59313 Valenciennes, France
[3] CNRS, UMR 7357–ICube, University Strasbourg, 67412 Illkirch, France; schiffler@unistra.fr
* Correspondence: Celine.Morin@uphf.fr

Abstract: The paper presents an experimental study and a 0D dynamic modeling of a biomass boiler based on the Bond Graph formalism from mass and energy balance. The biomass boiler investigated in this study is an automatic pellet boiler with a nominal power of 30 kW with a fixed bed. The balances allow to model as time function the flue gas enthalpy flux variation and the thermal transfers between the flue gas and the walls of the boiler subsystems. The main objective is to build a model to represent the dynamic thermal behavior of the boiler. Indeed, small domestic boilers have discontinuous operating phases when the set temperature is reached. The global thermal transfer coefficients for the boiler subsystems are obtained according to an iterative calculation by inverse method. The boiler has an average efficiency of 67.5% under our operating conditions and the radiation is the dominant thermal transfer by reaching 97.6% of the total thermal transfers inside the combustion chamber. The understanding of the dynamic behavior of the boiler during the operating phases allows to evaluate its energy performances. The proposed model is both stimulated and validated using experimental results carried out on the boiler.

Keywords: energy balance; biomass boiler; heat exchanger; 0D modeling; Bond Graph; global thermal transfers; inverse method

Citation: Mameri, F.; Delacourt, E.; Morin, C.; Schiffler, J. 0D Dynamic Modeling and Experimental Characterization of a Biomass Boiler with Mass and Energy Balance. *Entropy* **2022**, *24*, 202. https://doi.org/10.3390/e24020202

Academic Editor: T M Indra Mahlia

Received: 30 November 2021
Accepted: 28 December 2021
Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Biomass plays a significant role in the development of clean and sustainable heat production processes with a large reduction of CO_2 emissions [1]. There are multiple ways to exploit energy potential of biomass, e.g., by pyrolysis [2], gasification or other bio-chemical processes using bacteria to generate gaseous and liquid biofuels or by direct combustion to generate heat and electricity [3–5]. Even if biomass has a lower calorific value than other fuels, such as fossil fuels, this source of energy remains cleaner with some reserves [6]. The biomass can be valued for the simultaneous production of heat and electricity from CHP (Combined Heat and Power) plants [7,8].

In the thermal conversion of biomass, there are multiple physical and chemical processes that have an influence on the performances of industrial and domestic applications, such as furnaces, industrial burners and biomass boilers [9], the exergy analysis must be used in order to find the best way to recover the maximum of mechanical work in a CHP (combined heat and power) unit. Biomass boilers provide a direct conversion of biomass into energy by combustion. They are widely investigated in several configurations according to delivered power: biomass domestic boiler of 24 kW, 27 kW and 32 kW [10–12], industrial biomass boiler of 4 MW [13].

Dynamic modeling of energy systems can be used for the design, the optimization or the control of the studied process. Tognoli and Najafi [14] provided a detailed dynamic model of an industrial fire-tube boiler with five different geometrical configurations. The dynamic model developed consists of two main sections separated on the flue gas side and

the evaporating shell. Both sides are integrated employing an energy balance. Then, a PID tuning was implemented for each boiler to control the vapor pressure, while responding to a demand with variable mass flow rate. The operation of the boilers was simulated to meet four different steam demand profiles. A wood pellet micro-cogeneration system with steam engine was modeled by Bouvenot et al. [15] and implemented in the TRNSYS code. Both theoretical and experimental approaches have been adopted to develop the model. The authors presented the dynamic response of the installation and took into account the steady and transient states. A dynamic model applied to two biomass boilers with nominal power of 6 and 12 kW was presented by Carlon et al. [16]. The model developed with TRNSYS calculates the mass and energy balances of the boilers under time variable inputs. It describes the operation of the boiler under dynamic conditions and provides the chemical composition of the flue gases from the chemical composition of the wood pellets and the value of the excess air and by adopting the assumption of a complete conversion of the mass of fuel. The model has been tested for two modes of boiler operating conditions: full and variable load and steady and transient states. The results of the modeling showed a better agreement with the experimental data during steady operation as well as in dynamic mode.

The modeling of thermofluidic systems related to heat and power generation are also described in terms of mechanical work generation processes. We can note for example, a study on the modeling of ORC (Organic Rankine Cycle) systems investigated by Ziviani et al. [17]. The authors presented an overview of the problems related to ORC modeling and developed an efficient and powerful simulation for an ORC system adapted to the exploitation of low-grade thermal energy. Other physical systems are modeled like ECE (External Combustion Engine), for example Stirling or Ericsson engines [18–20]. Due to their promising future paths for energy cogeneration by coupling them with thermodynamic systems, small biomass boilers were also studied from this type of approach. However, they raised several concerns, ranging from design to dynamic control [21]. Inappropriate power requirement definition and inadequate control can affect the boiler performances and reduce its efficiency. To facilitate the design process and overcome upstream design failing, the modeling represents a very interesting approach.

The dynamic behavior of this kind of system is generally described by non-linear differential equations. A suitable method, as the Bond Graph formalism, is necessary to well understand physical interactions in a such thermofluidic system. Therefore, an appropriate model that represents a system involving energy transfers can be extracted in a structured way [22]. The Bond Graph method is based on a graphic structure representing the power exchanges between different physical entities considered in multidisciplinary dynamic systems. It was initiated by Paynter in 1961 [23] and then developed by Karnopp and Rosenberg [24]. This tool is adapted to the modeling of the physical processes involved in different energy fields (hydraulic, mechanical, electrical, chemical and thermal). Bond Graph formalism allows to develop a parametrized model with an unified language that interprets the power transfers within the system considered explicitly through its graphic structure. Bond Graph investigations have been carried out on energy systems such as hot air engines (Ericsson engine [19]), Heating, Ventilation and Air-Conditioning (HVAC) systems [25], industrial biomass boiler [26], endoreversible heat engine [27], thermo-hydraulic system [28] and in chemical engineering [29]. Ould-Bouamama et al. [30] have developed a dynamic model using Bond Graph methodology for an industrial chemical reactor. The purpose of this application is to design a monitoring and survival platform in case of failure.

The modeling of biomass boilers operation has been the subject of several studies. Mathematical models based on thermodynamic laws have been developed to represent the dynamic behavior of the boilers during operating phases such as start-ups and load changes [31]. Åström and Bell [32] developed a simple non-linear model based on the first law of thermodynamics and configured with the basic design data of the boiler. Sandberg et al. [33] presented a dynamic model based on the mass and energy balances of a biomass boiler to study the effect of fouling on boiler performances. Table 1 summarizes

the experimental and numerical studies of the literature about different systems (boiler, furnace, reactor and engine).

Table 1. Review of energy system modeling.

Reference	Device	Study	Power	Main Objective
Strzalka et al. [8]	Biomass grate furnace	Mathematical modeling	6 kW	Model-based optimization of control strategies of grate furnaces.
Li et al. [9]	Biomass boiler	Thermodynamic modeling		Conventional exergy analysis and advanced exergy analysis of a real biomass boiler.
Kang et al. [10]	Biomass boiler	Experimental investigation	24 kW	Evaluation of the performances of a domestic wood pellet boiler.
Gómez et al. [11]	Biomass domestic boiler	CFD modeling	27 kW	Simulation of the boiler operation under transient conditions. The effect of the parameters influencing the combustion process has been studied.
Ziviani et al. [17]	ORC system	Dynamic modeling (AMESim)		Progress and challenges related to the operation of ORC (Organic Rankine Cycle) systems.
Fénièss et al. [18]	Stirling engine	Theoretical modeling and experimental study	18 W	Establishment of two models, thermal and electrical, and study of the influence of dead volume, the natural frequency of mechanical oscillations and thermal conduction between the hot and cold sides for engine optimization.
Abdulmoneim et al. [22]	Thermal power generation station	Dynamic modeling (Bond Graph)		Modeling of hybrid power plant: pump, boiler, economizer, evaporator, super heater, drum and pipe.
Creyx et al. [19]	Ericsson engine	Dynamic modeling (Bond Graph)		Dynamic model of the expansion cylinder of an open Joule cycle Ericsson engine.
Ould-Bouamama et al. [30]	Chemical reactor	Dynamic modeling (Bond Graph)		Modeling of a chemical reactor for monitoring.
Sandberg et al. [33]	Biomass boiler	Dynamic modeling	157 MW	Biomass boiler dynamic model.
Persson et al. [34]	Biomass boiler and stove	Dynamic modeling (TRNSYS)	10 kW	Development and validation of a dynamic boiler/pellet stove model based on experimental measurements.

Published studies on the dynamic modeling of boilers often refer to black box or grey box models. There are some studies describing white box models but the detail of the modeling is often incomplete (use of components of commercial tool libraries rather opaque or description of physical phenomena modeled without specifying the interactions between them).

In this work, a 0D dynamic model of a domestic biomass boiler is provided using the Bond Graph formalism to simulate its dynamic behavior and to understand all the heat transfers involved in the boiler. The 0D model is based on mass and energy balances. It characterizes all the heat exchanges between the flue gas and the walls of the subsystems constituting the boiler. This dynamic modeling makes sense with domestic boilers whose operation is typically discontinuous unlike larger industrial boilers. The thermal needs of the house are variable which results in intermittent operation of the boiler. The strength of dynamic zonal modeling is to be able to predict the time evolution of different state variables of a complex system by coupling some fields of physics (mechanics, thermodynamics, ...). Moreover, it is possible, during the simulation, to insert time boundary conditions from in-situ measurements. The local evolution of the state variables is much less detailed than with CFD modeling but the dependencies of one zone with another are better taken into account with a 0D dynamic modeling. Moreover, CFD simulations are generally performed in steady state (averaged) because of the high computational cost in unsteady state, contrary to the dynamic 0D model which is able to predict the impacts on coupled systems. CFD and dynamic 0D modeling are therefore to be implemented according to the targeted objectives and the simulations results can hardly be compared. However, they can be efficiently coupled in multi-scale approaches. The objective of this study is to model the dynamic behavior of the boiler during the operating phases in order to take into account the variability of the heat production with regard to the thermal load of the heating network

or of any system which could be connected to it (hot air machine for example in the case of a CHP plant).

Compared to other dynamic modeling, the interest of Bond Graph methodology by its explicit graphic structure is to make clearer the modeling process of coupled multi-physical phenomena with blocks linked together by power links where effort and flow variables as well as causality are explicit. This methodology is very well adapted to model a system with thermal and mass transfers described with linear or non-linear differential equations.

The Bond Graph formalism was not developed in the literature to study the thermal transfers between the different fluids in a low power biomass boiler during the transient operating phases but it is increasingly used for the modeling of thermofluidic systems in general. Thanks to this formalism and its clarity, it is then possible to highlight the boiler components where the thermal transfers must be optimized and to understand the physical interactions. Moreover, there is a lack of experimental data for low power biomass boiler in the literature, these data are essential to develop a dynamic model by considering the real operating cycle of the boiler. An innovative way is used in this study by coupling experimental values and 0D modeling at each time step of the calculation with an analysis of energy performances for a domestic biomass boiler.

In the paper, the biomass boiler is described with all sensors used for the measurement of temperatures and mass flow rates. Then, the methodology is explained and the dynamic model 0D of the boiler is presented. Experimental and numerical results are discussed.

2. Description of the Biomass Boiler

2.1. Experimental Setup

The study is focused on an automatic domestic wood pellet boiler with a power of 30 kW Figure 1), equipped with a water-flue gas heat exchanger whose main role is to recover a part of the heat energy in the flue gas and transfer it in the water. The water circulation in the hydraulic circuit is ensured by a pump. The introduction of the pellets into the burner of the boiler is done by a screw that operates cyclically as long as the temperature of the outlet water is lower than the setpoint temperature. When the setpoint temperature is reached, the pellet supply stops. The primary air arrives through trapdoors located in the lower part of the boiler and its circulation is ensured by an exhaust fan mounted on the top cover of the boiler which is controlled by a lambda probe located in the chimney of the boiler. To dissipate the heat of the working fluid, the hydraulic circuit is connected to two air heaters located to outside of the test cell. In order to carry out an experimental characterization of the boiler, several sensors are installed at different locations in the boiler (Figure 1). An electromagnetic flowmeter with an operating range of 20 to 500 dm^3/h with an uncertainty of 0.5% measures the water mass flow rate ($\dot{m}_w^{exp}$) circulating in the boiler heat exchanger. The flue gas mass flow rate ($\dot{m}_{fg}^{exp}$) is calculated from pressure and temperature measurement in the chimney (Pitot wing system connected to a micromanometer (uncertainty 5% and K-type thermocouple (uncertainty 0.75%) on their measurement ranges respectively). The water temperature at the inlet ($T_{w,in}^{exp}$) and outlet ($T_{w,out}^{exp}$) of the water-flue gas heat exchanger are recorded by two platinum Pt100 probes (uncertainty 0.8%). A K-type thermocouple is inserted at the chimney (uncertainty 0.75%) for the measurement of the flue gas temperature ($T_{fg,exh}^{exp}$). Another type S thermocouple (uncertainty 0.25%) is placed in the central axis of the combustion chamber to measure the instantaneous evolution of the flue gas temperature ($T_{fg,cc}^{exp}$).

K-type thermocouples (uncertainty 0.75%) are placed in the burner ($T_{fg,bur}^{exp}$), on the top and bottom sides of the combustion chamber ($T_{fg,top}^{exp}$ and $T_{fg,bot}^{exp}$). Two other K-type thermocouples (uncertainty 0.75%) are also welded to each side of the combustion chamber wall ($T_{wall,outer}^{exp}$) and ($T_{wall,inner}^{exp}$). A K-type thermocouple (uncertainty 0.75%) is placed at the outlet of the heat exchanger tubes ($T_{fg,exit}^{exp}$) (Figure 1). The flue gas temperature measurements in the burner and the combustion chamber have been corrected from radiative effects. Indeed, with such temperature levels, the radiative dissipation of the thermocou-

ples is significant. Several methods exist to take into account this phenomenon which underestimate the true value of the temperature. The method used is the extrapolation method [35,36] which consists in using two thermocouples with wires of different diameters and therefore with different hot welds diameter (here 0.95 mm and 0.64 mm) placed at the same position. The radiative flux exchanged is assumed to be proportional to the surface of the hot weld, resulting in a zero radiative flux when the surface of the weld is infinitely small. From the two measured temperatures, an extrapolation allows to obtain the temperature value for a zero-weld surface corresponding to an absence of radiation. In our case study (in the flame and its vicinity), as an example, for a 1000 °C temperature measurement, the corrective value to be applied reach 170 °C. In this paper, the superscript "exp" corresponds to experimental measurements. The quantities calculated by the model have no superscript.

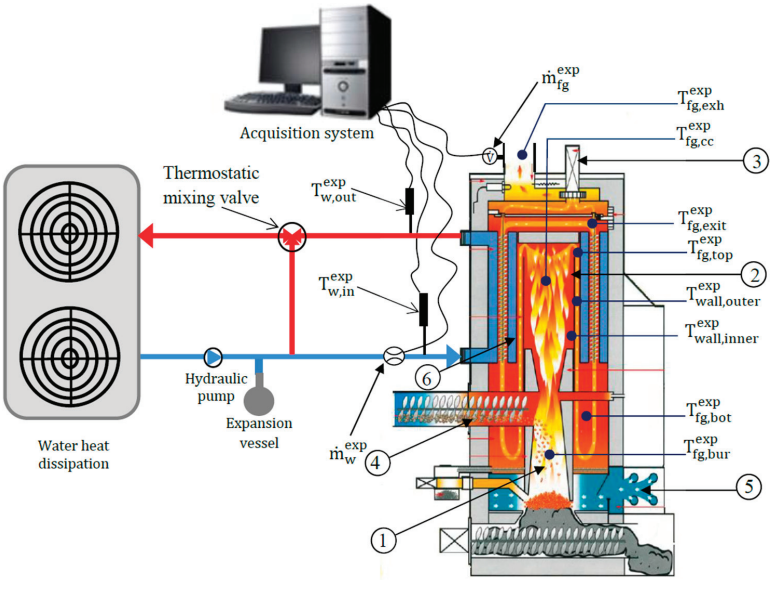


Figure 1. Thermocouples positions and hydraulic circuit. (1) Burner. (2) Combustion chamber. (3) Flue gas extraction. (4) Screw feeder. (5) Air inlet. (6) Water-flue gas heat exchanger.

The flue gas temperature evolution in the combustion chamber measured at radius of 90 mm and height of 330 mm obtained during the boiler operating cycle is plotted versus time and correlated with the pellets mass flow rate (Figure 2). It shows a strong dependence between the quantities of pellet supplied by a feed screw and the temperature increase of the flue gas in the boiler combustion chamber. The burnt gas temperature varies between 600 and 1100 °C. It increases with the arrival of pellets and decreases with their complete consumption. As mentioned above, the pellets are introduced into the boiler burner by a feed screw that rotates with a PWM duty cycle as long as the water temperature at the outlet of the heat exchanger is lower than the set temperature. This operating mode is controlled by a pulse-width modulation control. When the set temperature is reached, the pellet supply stops. Thus, a long stop of the pellet supply (12 min) can be observed in Figure 2. The pellet supply disruptions induce a drop in the burnt gas temperature in the combustion chamber with a temporary delay. The time lapse between the increase and drop of flue gas temperature defines a thermal cycle. The pellets mass flow rate is deduced from the calibration of the feed screw, according to the angular position of the screw. All sensors are connected to a data acquisition system with a dedicated code developed under Labview software.

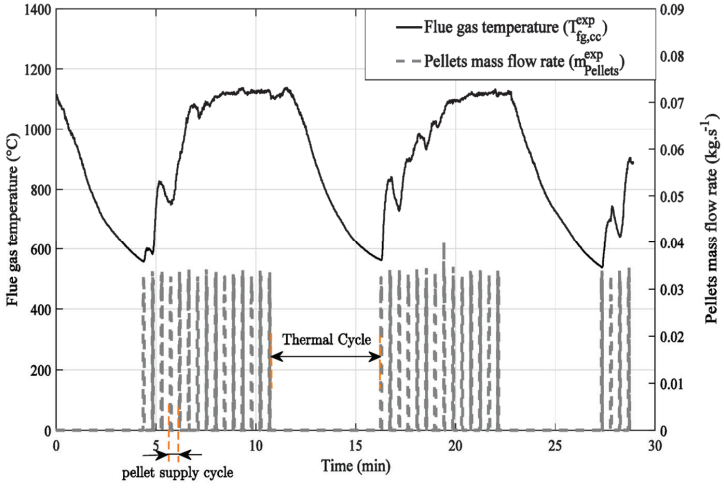


Figure 2. Boiler operating cycle.

2.2. Energy Balance of the Boiler

Although estimated from the rotation speed of the screw during a calibration phase without combustion, the mass flow of pellets is difficult to obtain accurately over short operating times of the screw. In fact, the quantity of pellets introduced by the screw by PWM method is not constant between each cycle because more or less large pellet clusters are detached from the screw. Moreover, the combustion is not instantaneous, it would be necessary to introduce a dynamic combustion model of solid biomass to calculate the heat release as a function of time. The boiler model presented here can be modified in the future by integrating this combustion dynamics. The pellet combustion is therefore not modelled, so the heat generated during the combustion of the pellets has been calculated using the experimental mass flow rate and the experimental temperature of the flue gases in the burner. Then, the heat flux provided by the combustion of pellets is estimated from the following equation:

$$\dot{H}_{fg,bur}(t) = \dot{m}^{exp}_{fg}(t).\left[c_p\left(T^{exp}_{fg,bur}\right).\left(T^{exp}_{fg,bur}(t) - T^{ref}_{fg,bur}\right) + \Delta H^0_{ref}\right] \quad (1)$$

With:

$\dot{H}_{fg,bur}$: heat flux released from pellet combustion (W)

$\dot{m}^{exp}_{fg}$: experimental flue gas mass flow rate (kg.s^{-1})

$T^{exp}_{fg,bur}$: experimental flue gas temperature in the burner (K)

$T^{ref}_{fg,bur}$: reference temperature for flue gas in the burner (298 K)

ΔH^0_{ref}: standard formation enthalpy of gas in the burner (J.kg^{-1}).

Considering the majority presence of N_2 and O_2 (air excess close to 1) in the mixture and for a first approximation, we assume that the mixture is composed as a gas including only pure species. We can therefore assume that $\Delta H^0_{ref} = 0$.

The energy balance of the boiler is established at each time step. It represents the heat exchanges between the flue gas and the boiler structure, the heat flux recovered by the water in the water-flue gas heat exchanger and the losses at the boiler exhaust. The outer wall of the boiler is assumed to be adiabatic because the boiler is very well insulated and the losses with the environment are negligible compared to the other heat flux. The losses

are more significant from the boiler outlet through the exhaust pipe but this part is not modeled here. The heat flux released from the pellets combustion is then given by:

$$\dot{H}_{fg,bur}(t) = \dot{H}_{fg,exh}(t) + \Delta\dot{H}_{w}(t) + \dot{Q}_{wall}(t) \quad (2)$$

With:

$\dot{H}_{fg,exh}$: exhaust heat flux (W).

$\Delta\dot{H}_{w}$: heat flux transferred to the water (W).

$\dot{Q}_{wall}$: heat flux stored in the boiler structure (W).

The flue gases resulting from the combustion of the pellets go through the boiler subsystems (Figure 3 dashed red line) and exchange heat with their walls. Due to the transient phases, the walls store or yield a quantity of heat flux from or to the flue gases: the heat flux stored in the combustion chamber walls $\dot{Q}_{wall,cc}$, in the inner wall of the heat exchanger $\dot{Q}_{wall,HEx}$ and in the walls of the flue gas tubes $\dot{Q}_{wall,tub}$. The walls of the subsystems store some heat flux, consisting of three parts:

$$\dot{Q}_{wall} = \dot{Q}_{wall,cc} + \dot{Q}_{wall,HEx} + \dot{Q}_{wall,tub} \quad (3)$$

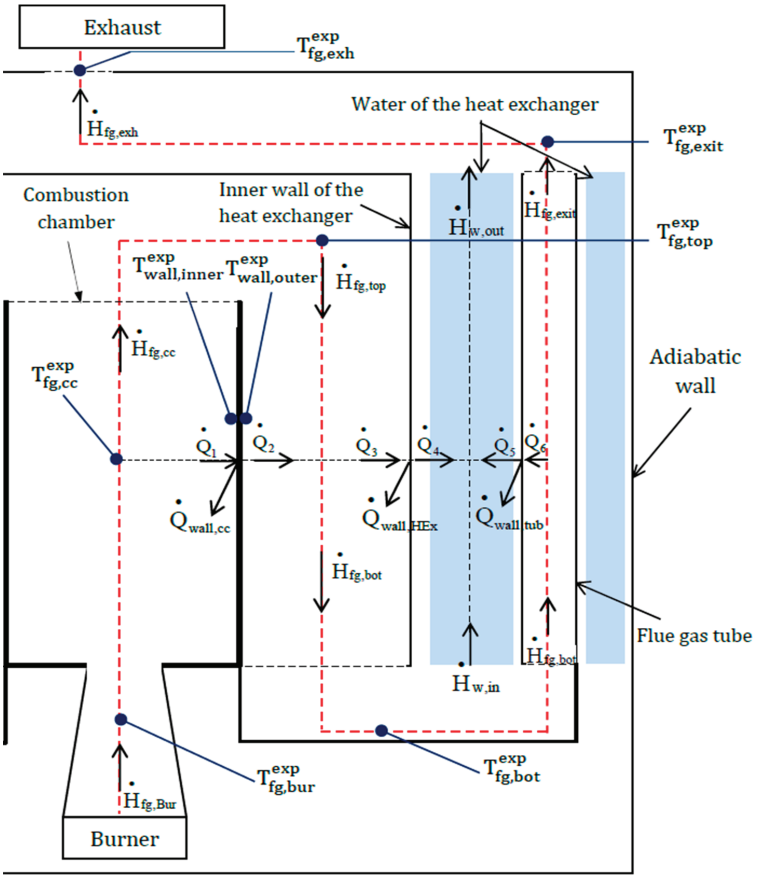


Figure 3. Energy balance of the boiler.

Each of these heat fluxes is calculated in the subsystems from inlet and outlet flux (Figure 3):

$$\dot{Q}_{wall,cc} = \dot{Q}_1 - \dot{Q}_2 \quad (4)$$

$$\dot{Q}_{wall,HEx} = \dot{Q}_3 - \dot{Q}_4 \quad (5)$$

$$\dot{Q}_{wall,tub} = \dot{Q}_6 - \dot{Q}_5 \quad (6)$$

The heat flux transferred to the inner walls of the heat exchanger and the flue gas tubes is partially transferred to the water.

$$\begin{gathered} \Delta\dot{H}_w = \dot{Q}_4 + \dot{Q}_5 - \dot{Q}_{w,st} \\ \Delta\dot{H}_w = \dot{m}_w^{exp} \cdot \left(c_w\left(T_{w,out}^{exp}\right) \cdot T_{w,out}^{exp} - c_w\left(T_{w,in}^{exp}\right) \cdot T_{w,in}^{exp} \right) \\ \Delta\dot{H}_w = \dot{H}_{w,out} - \dot{H}_{w,in} \end{gathered} \quad (7)$$

With $\dot{Q}_{w,st}$: heat flux stored by the water in the heat exchanger (W).

3. 0D Bond Graph Modeling

The modeling of the main components of the boiler system, such as the combustion chamber, the flue gas tubes and the heat exchanger is performed using Bond Graph formalism. The boxes in Figure 4 represent the subsystems of the studied boiler, where the half-arrows characterize the thermal and hydraulic Bond Graph links between the subsystems. The word Bond Graph model describes here the thermal and mass transfers between subsystems. Causalities (I) are also present in order to indicate the variables at the origin of the system dynamics. In Figure 4, the combustion chamber box is not detailed, it includes the flue gas path from burner to the bottom of the heat exchanger. The combustion chamber temperature noted $T_{fg,cc}$ is located inside this box but not appears in the Inlet/Outlet Bond Graph links.

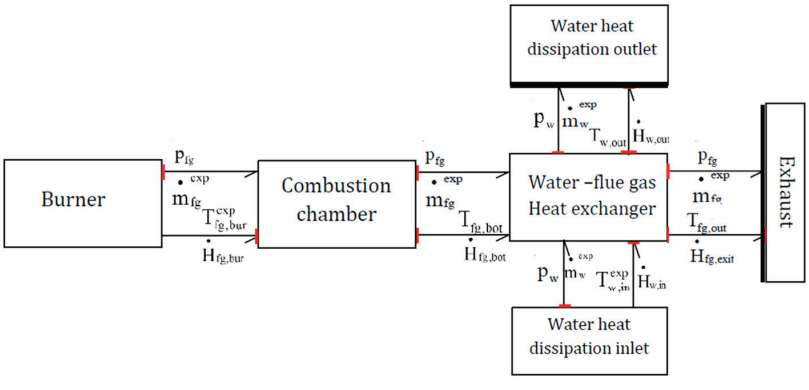


Figure 4. Word Bond Graph model.

3.1. 0D Model of the Boiler

The 0D dynamic model of the boiler is shown in Figure 5 with all the boiler subsystems (burner, combustion chamber, heat exchanger). The time variation of mass flow rate and temperature of both water and flue gas is considered. As input conditions, the experimental flue gas temperature $T_{fg,bur}^{exp}$ in the burner and the experimental mass flow rate of the flue gas $\dot{m}_{fg}^{exp}$ are introduced as time files. The water-flue gas heat exchanger is also modelled by providing the experimental water mass flow rate $\dot{m}_w^{exp}$ and the experimental inlet water temperature $T_{w,in}^{exp}$ as input conditions. The model is therefore stimulated with real limit

conditions and then validated with other experimental measurement obtained at the same time than these limit conditions values. This method improves the validation of the model.

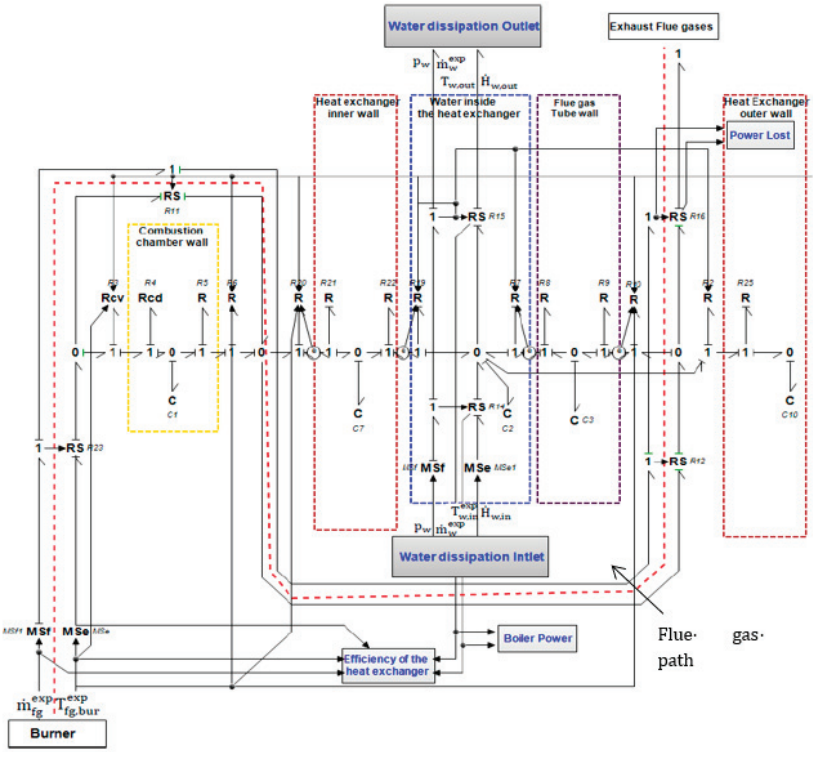


Figure 5. 0D model of the boiler.

The storage and/or removal of thermal energy in the walls in the different zones of the boiler (refractory concrete, walls of the combustion chamber and walls of the heat exchanger) are modeled by the following law:

$$\dot{Q}_{wall,i} = m_{wall,i} \times c_{wall,i} \times \frac{dT_{wall,i}}{dt} = \sum \varnothing_{diss} \quad (8)$$

With:

$\dot{Q}_{wall,i}$: heat flux stored in the wall of the system i (W).

$\varnothing_{diss}$: dissipative fluxes between wall and flue gas (W).

$m_{wall,i}$: wall mass of the system i (kg).

$c_{wall,i}$: wall specific heat of the system i ($J/kg^{-1}.K^{-1}$).

This expression is traduced to a 'C' element (C_1, C_7, C_3, C_{10}) (Figure 5) in the bond graph formalism because the flux is a function of the derivative of the effort:

$$\frac{T_{wall,i}}{\dot{Q}_{wall,i}} \rightarrow C$$

The causality applied to these 'C' elements is always a flux causality because at the beginning of the simulation, it is the temperatures (efforts) that are known and then allow calculation of dissipative heat fluxes (conductive, convective and radiative ones). So, in

these differential equations solving, the value of temperatures of the next time step are obtained from the integration of the flux balance at the current time.

$$T_{wall,i}(t) = \frac{1}{m_{wall,i} \times C_{wall,i}} \int_0^t \dot{Q}_{wall,i} dt \quad (9)$$

where $\dot{Q}_{wall,i}$ is calculated from a heat flow balance between all dissipative fluxes $\varnothing_{diss}$ (Equation (8)) and therefore obtained with a '0' junction centered in the wall.

The dynamic behavior of the boiler depends on the interaction between the both hydraulic and thermal systems. The 'RS' elements (R_{11}, R_{12}, R_{16}, R_{18} and R_{23}) have been used to couple them in order to calculate enthalpy flux (10) from inlet temperature and mass flow rate. The power input of each RS elements is defined with an effort causality which means that the temperature value (effort) of thermal power input is knows at the start of each calculation step. RS elements then calculate the enthalpy flux, which is necessary for each flux balance carried out by the zero junctions on the path of the flue gas labelled by the dashed red line. The enthalpy flux of flue gas at the inlet and outlet of each 'RS' element $\dot{H}_{fg,in}$ and $\dot{H}_{fg,out}$ (Figure 6) is calculated with the following Equations (10) and (11) and the same hypothesis than the Equation (1):

$$\dot{H}_{fg,in}(t) = \dot{m}_{fg}^{exp}(t).\left[c_p\left(T_{fg,in}\right).\left(T_{fg,in}(t) - T_{fg}^{ref}\right) + \Delta H_{ref}^{0}\right] \quad (10)$$

$$\dot{H}_{fg,out}(t) = \dot{H}_{fg,in}(t) \quad (11)$$

With:

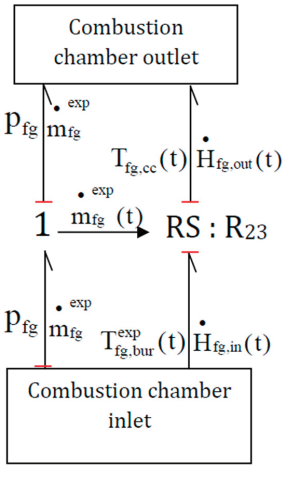


Figure 6. RS element example.

$\dot{H}_{fg,in}(t)$: calculated flue gas enthalpy flux at the RS-element inlet (W), equal to $\dot{H}_{fg,bur}$

$\dot{H}_{fg,out}(t)$: calculated flue gas enthalpy flux at the RS-element outlet (W), equal to $\dot{H}_{fg,bur}$

$T_{fg,bur}^{exp}(t)$: experimental temperature in the burner (K)

$T_{fg,cc}(t)$: calculated flue gas temperature (K) imported from the following '0' junction in the combustion chamber.

$\dot{m}_{fg}^{exp}(t)$: experimental flue gas mass flow rate ($kg.s^{-1}$).

P_{fg}: experimental pressure in the boiler (Pa) supposed constant because pressure losses are low and not easy to model in 0D due to the complexity of the geometry.

The Figure 6 illustrates one of the 'RS' elements. This one is located between the burner outlet and the combustion chamber.

The thermal transfers by conduction and convection are modeled by equations of the following form:

$$\dot{Q}_{diss}(t) = \frac{1}{R_{th}}.\,\Delta T(t) \tag{12}$$

With:

$\dot{Q}_{diss}(t)$: dissipative flux (W).

R_{th}: thermal resistor ($K.W^{-1}$).

$\Delta T(t)$: temperature difference (K).

Here, the flux $\dot{Q}_{diss}(t)$ directly depends on the effort $\Delta T(t)$ (not derivative link) whether it's linear or not. In this case the 'R' element is used in the bond graph formalism:

$\Delta T(t)$ is obtained with a '1' junction which consist in an effort balance and then can calculate the temperature difference.

R_4, R_5, R_{21}, R_{22}, R_8, R_9 and R_{25} quantify the conductive exchanges through the walls. In a cylinder, the thermal conductive resistance R_{cd} is given by:

$$R_{cd,i} = \frac{\ln\left(\frac{r_{2,i}}{r_{1,i}}\right)}{2\pi.\lambda_i.H_i} \tag{13}$$

With:

$r_{2,i}$: outside radius of the system i (m).

$r_{1,i}$: inside radius of the system i (m).

λ_i: thermal conductivity of the system i ($W.m^{-1}.K^{-1}$).

H_i: height for the system i (m).

R_3, R_6, R_{20}, R_{10}, and R_2 deal with the convective transfers between the flue gas and the different walls of the boiler. They are calculated from the convective resistance R_{cv}:

$$R_{cv,i} = \frac{1}{h_{g,i}.S_i} \tag{14}$$

With:

$h_{g,i}$: global thermal transfer coefficient of the system i ($W.m^{-2}.K^{-1}$).

S_i: exchange surface of the system i (m^2).

The global thermal transfer coefficients $h_{g,i}$, including convective and radiative effects, for the different geometrical configurations in the boiler are obtained according to a first stage of simulation by inverse method. Indeed, the radiative effects of the flame or the gases with the walls are complex to model in 0D. This method is based on the energy balances presented in Section 2.2 for each zone. Heat fluxes are calculated by using experimental wall and flue gas temperatures as well as experimental water temperatures (and calculated temperatures by the dynamic 0D model when the experimental measurement is not available). These experimental temperature values are introduced into the model at each calculation time step. They thus allow at each time step to calculate the value of $h_{g,i}$ as illustrated in the following relations (Equation (15)) in order to use it for the calculation of the parietal fluxes in the model. This method allows to obtain temporal evolutions of $h_{g,i}$ coefficients like the one presented in Figure 7 and was implemented only once as a prerequisite to the main simulation. This allowed to determine the global coefficients even in areas where we were unable to place thermocouples probes by using temperatures calculated as close to reality as possible since in places where temperatures were measured, the model took them into account at each time step. This combination of measured and

calculated quantities inside a behavioral model is similar to a HIL (Hardware in the Loop) process.

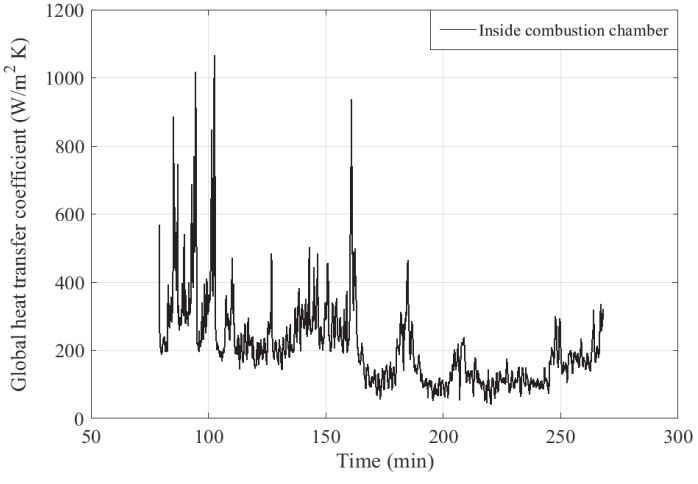


Figure 7. Global thermal transfer coefficient inside the combustion chamber.

By example, the variation of the flue gas enthalpy flux between the inlet and the outlet of the combustion chamber $\Delta\dot{H}_{fg,cc}(t)$ is calculated at each time step. The global heat flux exchanged between the flue gas and the combustion chamber wall $\dot{Q}_{fg,cc}(t)$ is calculated at each time step also. By performing a balance between the two heat fluxes, the value of global thermal transfer coefficient is deduced for each time step (Equation (15)). The time evolution of a global thermal transfer coefficient, including radiative and convective transfers, near the inside combustion chamber is shown in Figure 7. The peaks observed are due to the low temperature difference between the flue gas and the wall of the combustion chamber. With the induced errors on the global thermal transfer coefficients higher than 2000 W.m^{-2}.K^{-1}), they cause discrepancies in the calculation carried out by the dynamic model. This adds complexity to the choice of the computation scheme.

$$\begin{gathered}\dot{\Delta H}_{fg,cc}(t) = \dot{m}^{exp}_{fg}(t)\cdot\left(\underbrace{c_{p,fg}\left(T^{exp}_{fg,cc}(t)\right).T^{exp}_{fg,cc}(t)}_{\text{cc outlet}} - \underbrace{c_{p,fg}\left(T^{exp}_{fg,bur}(t)\right).T^{exp}_{fg,bur}(t)}_{\text{cc intlet}}\right)\\ \dot{Q}_{fg,cc}(t) = h_{g,cc}.S_{cc}\left(T^{exp}_{fg,cc}(t) - T^{exp}_{wall,cc}(t)\right)\\ h_{g,cc}(t) = \frac{\dot{\Delta H}_{fg,cc}(t)}{S_{cc}\left(T^{exp}_{fg,cc}(t) - T^{exp}_{wall,cc}(t)\right)}\end{gathered} \tag{15}$$

With:

$\Delta\dot{H}_{fg,cc}$: variation of the flue gas enthalpy flux between the inlet and the outlet of the combustion chamber (W)

$\dot{Q}_{fg,cc}$: global heat flux exchanged between the flue gas and the combustion chamber wall (W)

c_{pfg}: flue gas specific heat at constant pressure (J.kg^{-1}.K^{-1}).

S_{cc}: combustion chamber exchange surface (m^2).

$h_{g,cc}$: global thermal transfer coefficient for the inner wall of the combustion chamber (W.m^{-2}.K^{-1}).

$T^{exp}_{wall,cc}$: experimental temperature of the inner wall of the combustion chamber (K).

For this example, apart from the peaks mentioned above, an average value of h_g = 200 W.m^{-2}.K^{-1} has been selected. The results presented in Figures 8 and 9 show that this approximation leads to some errors in the calculated water and flue gas temperatures. Indeed, we could identify here two operating regimes:

- h_g = 300 W.m^{-2}.K^{-1} for t = 0–170 min
- h_g = 100 W.m^{-2}.K^{-1} for t = 170–250 min.

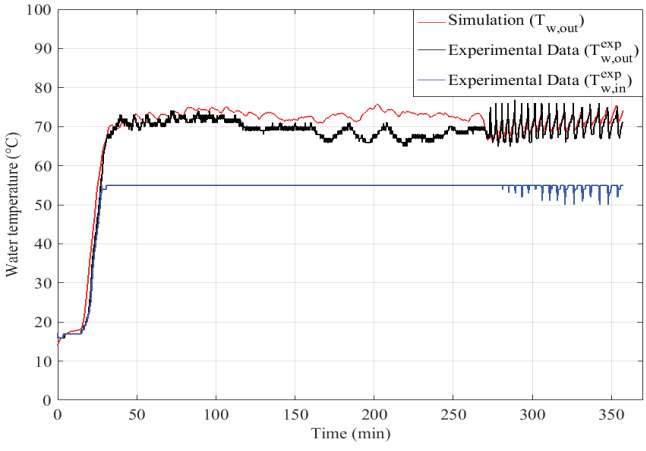


Figure 8. Comparison of experimental and calculated water temperatures at the outlet of the water-flue gas heat exchanger.

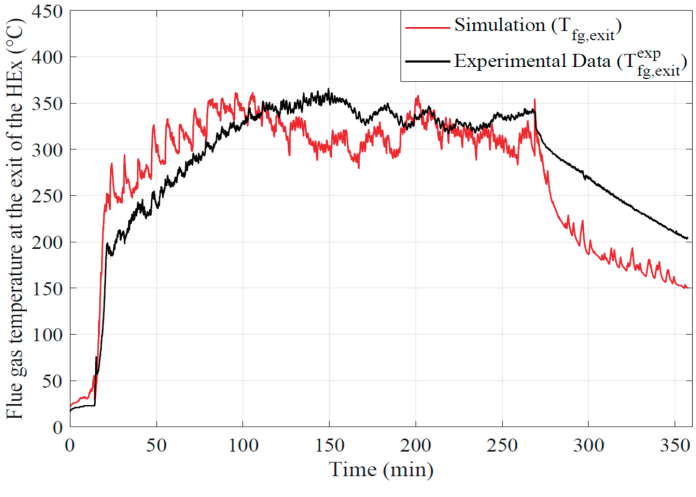


Figure 9. Comparison of experimental and calculated flue gas temperature at the outlet of the flue gas tubes.

These two operating regimes can be identified in Figure 10, the combustion is continuous until t = 170 min and then an operating cycle is set up as presented in Figure 2. The choice of only one value for the global coefficient generates an under estimation of the transfers on the first phase and an over estimation on the second one as it can be noticed in Figures 8 and 9.

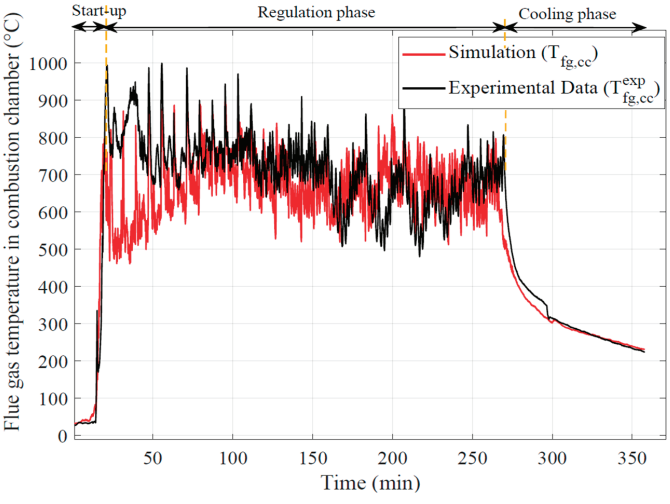


Figure 10. Comparison of experimental and calculated flue gas temperatures in the combustion chamber.

Nevertheless, the change of operating regime is difficult to take into account here in an automated way without modeling the pellet supply mechanism and their subsequent combustion.

In order to differentiate radiative and convective heat exchanges in the boiler, the convective heat fluxes are calculated using the Newton's law and the convective coefficients from Equation (16). Knowing Reynolds number as well as Prandtl, Nusselt numbers was determined from the semi-empirical correlations of Dittus-Boelter [37] and Gnielinski [38], adapted to the studied configurations. Nusselt number then allows to calculate the convective exchange coefficient within the geometric configurations remaining inside the boiler. Table 2 includes the semi-empirical correlations used to calculate Nusselt number.

$$\mathrm{Nu} = \frac{\mathrm{h}.\mathrm{D_h}}{\lambda_{fg}} \tag{16}$$

With:

D_h: hydraulic diameter (m).
λ_{fg}: flue gas thermal conductivity ($W.m^{-1}.K^{-1}$).
h: convective coefficient ($W.m^{-2}.K^{-1}$).

Table 2. Semi-empirical correlations used for the calculation of Nusselt number.

Location	Flow Configuration	Correlations	Valid Range
Combustion chamber and flue gas tubes [37]	Inside a cylinder	$\mathrm{Nu} = 0.023\mathrm{Re}_{D_h}^{0.8}\mathrm{Pr}^{0.4}\left(1+\left(\frac{D_h}{H}\right)^{0.7}\right)$	$0.7 \leq \mathrm{Pr} \leq 120$ $10^4 \leq \mathrm{Re}_{D_h} \leq 1.2\,10^5$ $2 \leq D_h \leq 20$
Passage between the combustion chamber and inner wall of the heat exchanger [38]	Inside an annular duct–fixed walls	$\mathrm{Nu} = 0.023\mathrm{Re}_{D_h}^{0.8}\mathrm{Pr}^{0.4}(r_2/r_1)^{0.14}$	$0.7 < \mathrm{Pr} < 100$ $\mathrm{Re}_{D_h} > 2000$

Finally, the calculation of the water temperature at each instant is obtained by a flux balance represented by the area inside the blue dotted lines in Figure 5.

The flux balance consists of the algebraic sum of the enthalpy input/output fluxes calculated by the 'RS' elements (respectively 'R_{14}' and 'R_{15}') with the convective heat fluxes

calculated with the thermal resistances 'R_2', 'R_7' and 'R_{19}'. The water temperature is then calculated from the integration of the flux balance performed in the 'C_2' element.

3.2. Flue Gas Thermodynamic Properties

To take into account the variation of the thermodynamic properties of the flue gas in the boiler, correlations have been used for each property as a function of the temperature.

In this section, all the correlations used to calculate the thermodynamic properties are detailed. The properties are: density ρ_{fg} from the perfect gas law with R = 8.314 J/mol^{-1}/K^{-1}, thermal conductivity λ_{fg}, dynamic viscosity μ_{fg}, and specific heat c_{pfg} of the flue gas resulting from the combustion of the pellets. The mass fraction used to calculate some properties is evaluated from the molar fraction deduced from Equation (17). The correlation versus temperature of c_{pfg} is introduced in the 0D model to calculate enthalpy flux. The other thermodynamic properties correlation are used to calculate Reynolds and Prandtl numbers in order to deduce the convective transfer coefficients introduced above.

The thermodynamic properties of flue gas are obtained by adding the properties of each species multiplied by the corresponding molar or specific fractions. The mixture of these species is given by the combustion reaction.

From an elementary analysis, the chemical formulation of pellet is $C_{36.725}H_{71.6}O_{30.475}$. Then, the combustion reaction of pellets in air is given as below:

$$\begin{aligned} &C_{36.725}H_{71.6}O_{30.475} + 39.388(1 + e_{air})(O_2 + 3.76N_2) \\ &\qquad \rightarrow 36.725\ CO_2 + 35.8\ H_2O + 39.388.e_{air}.O_2 + 148.097(1 + e_{air})N_2 \end{aligned} \quad (17)$$

With e_{air} the air excess.

As discussed before, it is assumed that the pressure in the boiler remains constant and equal to the atmospheric pressure and the specific fractions of the combustion products also remain constant during the boiler cycles and in the different zones. Knowing that the boiler operates with an air excess of 80% (e_{air} = 0.8), the correlations used are presented in Table 3.

Table 3. Correlations for the calculation of the flue gas thermodynamic properties (i = CO_2, H_2O, O_2, N_2). The constants A, B, C, D and E were fixed for each species and for each property.

Flue Gas Thermodynamic Properties	Correlations	Units	Temperature Range (K)	Min–Max
Density	$\rho_i = \frac{PM_i}{RT}$ $\rho_{fg}(T_{fg}) = \left(\sum_i \frac{y_i}{\rho_i(T_{fg})}\right)^{-1}$; $y_i = \frac{m_i}{m_{tot}}$;	kg.m^{-3}	298–1500	0.23–1.22
Thermal conductivity [39]	$\lambda_i = A + BT_{fg} + CT_{fg}^2 + DT_{fg}^3$ $\lambda_{fg} = \frac{\sum_i x_i \lambda_i M_i^{1/3}}{\sum_i x_i M_i^{1/3}}$, $x_i = \frac{n_i}{n_{tot}}$	W.m^{-1}.K^{-1}	298–1500	2.32 10^{-2}–8.65 10^{-2}
Dynamic viscosity [40]	$\mu_i(T) = A + BT + CT^2 + DT^3$ $\mu_{fg}(T_{fg}) = \frac{\sum_i x_i \mu_i(T_{fg})\sqrt{M_i}}{\sum_i x_i \sqrt{M_i}}$	Pa.s^{-1}	298–1500	1.711 10^{-5}–5.42 10^{-5}
Specific heat [41]	$c_{p,i}(T_{fg}) = A + BT_{fg} + CT_{fg}^2 + DT_{fg}^3 + E/T_{fg}^2$ $c_{p,fg}(T) = \sum_i y_i c_{p,i}(T_{fg})$	J.kg^{-1}.K^{-1}	298–1500	1090–1374

3.3. Solver Scheme

The Bond Graph method uses a system of algebraic-differential equations to describe the dynamic of the modeled system. The accuracy of the dynamic model is based on the choice of the computation scheme used to efficiently solve these differential equations. The resolution scheme used in our model is the Runge-Kutta fourth order formula (RK4) [42,43]

which is a particulate case of Runge-Kutta method. This method is recommended when the required accuracy is very high but it requires more CPU time than simpler methods (for this study about 5 min). This method is based on the iteration principle, i.e., an estimation of the solution is calculated from the previous solution. The principle is to approach the next value y_{n+1} at time t_{n+1} by the current value y_n obtained at time t_n combined with a function taking into account the iteration step (δ) and the estimated slope. The slope is obtained by the weighted average of four slopes (k_1, k_2, k_3 and k_4), where each slope is the product of the iteration step and an estimated slope. The slope is specified by the function F on the right side of the differential equation [44,45].

The following problem is then considered:

$$\dot{y} = F(t, y) \text{ with } y_0 = f(t_0) \rightarrow y = f(t) \tag{18}$$

From a known initial condition, the RK4 method is given by the equation:

$$y_{n+1} = y_n + \frac{\delta}{6}(k_1 + 2k_2 + 2k_3 + k_4) + \left(\delta^5\right) \tag{19}$$

$$\delta = t_{n+1} - t_n,\ 1 < n < N \tag{20}$$

where

$$k_1 = F(t_n, y_n) \tag{21}$$

$$k_2 = F\left(t_n + \frac{\delta}{2}, y_n + \frac{k_1}{2}\right) \tag{22}$$

$$k_3 = F\left(t_n + \frac{\delta}{2}, y_n + \frac{k_2}{2}\right) \tag{23}$$

$$k_4 = F(t_n + \delta, y_n + k_3) \tag{24}$$

The RK4 method is of order 4, this means the error committed at each step is of the order of δ^5, whereas the total accumulated error is of the order of δ^4.

4. Results and Discussion

The experimental results are discussed to characterize the boiler operation. They are then compared to the simulations. In order to validate the model, measurements of the flue gas temperature profiles in the combustion chamber and at the heat exchanger outlet as well as measurements of water temperature at boiler outlet are carried out.

The dynamic model input data are the experimental flue gas temperature in the burner ($T_{fg,bur}^{exp}$), the experimental water temperature at the boiler inlet ($T_{w,in}^{exp}$), the experimental flue gas mass flow rate ($\dot{m}_{fg}^{exp}$) and the experimental water mass flow rate ($\dot{m}_{w}^{exp}$).

The thermal behaviors of the flue gas in the boiler and the water in the heat exchanger are investigated.

From the analysis of the flue gas temperature profiles in the combustion chamber (Figure 10), sudden and fast temperature changes occur during the boiler start-up due to the uncontrolled combustion of a large mass of pellets during this step. Before the combustion start, pellets are heated during several minutes (about 15 min) with an air heater. During the entire control phase (regulation phase between 30–270 min), the flue gas temperatures remain very high and display fluctuations. Then, they decrease progressively during the cooling phase (270–355 min). The fast fluctuations observed are due to the quantities of pellets supplied every 20 s. These fluctuations are also observed on the temperature profiles calculated from the 0D model as a consequence of the limit condition that is an experimental measurement of the flue gas temperature in the burner. We can note that a notable difference exists between the calculated temperature and the one measured during the beginning of the combustion phase (just after the start-up jump). This difference is undoubtedly linked to the fact that the quantity of pellets burning in this phase is very important (accumulation before combustion start), also the gasification is such that the

combustion continues in the upper combustion chamber (above the burner zone). The model only integrates the combustion in the burner and therefore does not integrate this excess of heat release in the combustion chamber area.

This problem would have been the same using a combustion model of the pellets in the burner area. It would be necessary to separate the combustion in the 2 zones (burner and combustion chamber) and thus to find a key of distribution of the combustible gases in each zone. This key is not easy to find because the problem is related to unsteady 3D aero thermochemical phenomena.

The water temperature at the heat exchanger outlet is also examined. As noted for the flue gas temperature, a drastic increase of the water temperature can be observed during the start-up phase. During this phase (0–30 min), the water circulates in closed circuit until to reach a temperature of 325 K (Figure 8). This process is imposed by the mixing valve (3-way valve) resulting in a significant increase of the water temperature. After this step, the hot water is redirected to the cooling circuit. The fluctuations observed during the cooling phase are due to the intermittent operation of the water pump to maintain as long as possible the boiler body closed to the operating conditions if the boiler needs to be restarted.

Here, the differences between the calculated and measured values are not significant. Differences of 5 °C are nevertheless noted in the cyclic operation zone (170–250 min), this is perhaps linked to the overestimation of the global exchange coefficients in this operation mode as mentioned at the end of Section 3.1.

The instantaneous evolutions of the experimental and calculated flue gas temperature at the outlet of the flue gas tube of the water heat exchanger are plotted in Figure 10. The flue gas temperature at the flue gas tube outlet has the same evolution as in the combustion chamber. Nevertheless fluctuations are filtered by the thermal inertia of the different parts of the boiler along the flue gas path. The temperature of the flue gas remains relatively high at outlet of the tubes (~573 K).

As mentioned at the end of Section 3.1, here the under estimation of the wall global exchange coefficients on the stabilized phase and the under estimation on the cyclic phase is notable. In the stabilized phase, the underestimation of the thermal wall fluxes limits the thermal dissipation of the gases and thus also the reduction of their temperature. In the cyclic phase, the overestimation of the fluxes increases abnormally the wall transfers and reduces the flue gas temperature. Significant temperature differences remain during the cooling phase of the boiler.

The time evolution of the water enthalpy flux variation between inlet and outlet, calculated by the dynamic model, is plotted in Figure 11. The heat flux drops to zero when the pump stops.

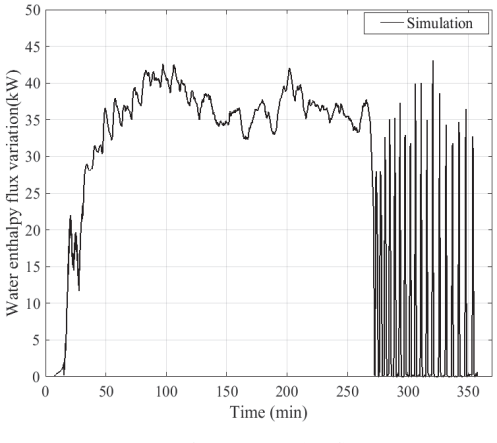


Figure 11. Variation of water enthalpy flux.

The instantaneous thermal power delivered to the water can be calculated as:

$$\Delta\dot{H}_w(t) = \dot{m}_w^{exp}\left(c_w(T_{w,out}).T_{w,out} - c_w\left(T_{w,in}^{exp}\right).T_{w,in}^{exp}\right) \quad (25)$$

With:

$\Delta\dot{H}_w(t)$: water enthalpy flux variation (W).

$T_{w,in}^{exp}$: experimental water temperature at the heat exchanger inlet (K).

$T_{w,out}$: water temperature at the heat exchanger outlet calculated by the dynamic model (K).

c_w: water specific heat at an average temperature of 328 K ($4183/kg^{-1}.K^{-1}$).

The amount of heat flux transmitted to the water remains very low during the start-up phase of the boiler and then increases drastically after the start of the combustion. After the start of combustion, an increase of the heat transmitted to the water can be observed between 30 and 50 min in Figure 11. This progression exists because some heat from combustion is first accumulated by the metal walls inside the boiler before being completely transferred to the water when the walls reach an established thermal regime. Considering only the regulation phase represented by the period from the combustion start time (30 min) to the boiler shutdown (270 min), the heat flux transmitted from the flue gas to the water of the heat exchanger is quite stable and close to 37 kW. The fluctuations observed on the flue gas temperature during the cyclic phase are well absorbed by the inertia of the walls and the water. After the boiler shutdown and during the cooling phase, intermittent operation of the pump is observed. When the pump is shut down, the walls of the heat exchanger transmit heat to the volume of water became motionless in the heat exchanger, which explains the peaks of enthalpy flux as soon as the pump is started up again.

From the dynamic model of the boiler, it is also possible to calculate its efficiency, lost power, heat flux stored by the walls and released by combustion in the burner.

According to the manufacturer, the boiler must have an average efficiency of 85% under nominal operating conditions (thermal power of 30 kW). In this study, the duty cycle (PWM) of the pellet supply screw was modified to increase the power of the boiler in order to saturate the downstream thermal load and thus create thermal control cycles suitable to unsteady operating conditions.

The efficiency can be defined as the ratio of the enthalpy flux variation of the water heat exchanger and the heat flux released from pellet combustion:

$$\eta(t) = \frac{\Delta\dot{H}_w(t)}{\dot{H}_{fg,bur}(t)} \quad (26)$$

The instantaneous evolution of the efficiency calculated by the model at each moment t is presented in Figure 12. Its evolution is drastically affected by the fluctuations of flue gas temperature. The boiler has an average efficiency of 67.5%. This low efficiency value is not surprising because the overpower generated in our test case cannot be fully absorbed by the capacity of the gas-water exchanger of the boiler. The thermal power of the boiler is nevertheless increased (37 kW instead of 30 kW).

Several parameters impacting the response of the 0D model can be highlighted. For example, an influence on the thermal behavior of the flue gas with the mass flow rate can be distinguished at the tube outlet. As the experimental flue gas mass flow rate is used as input condition, during the cooling phase a significant discrepancy between the evolution of the calculated and experimental flue gas temperatures is recorded as shown in Figure 13a.

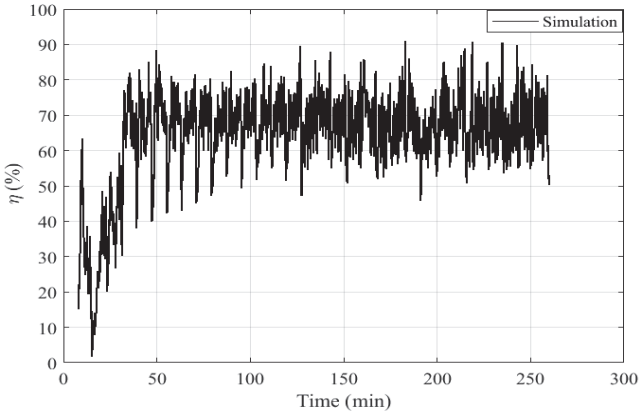
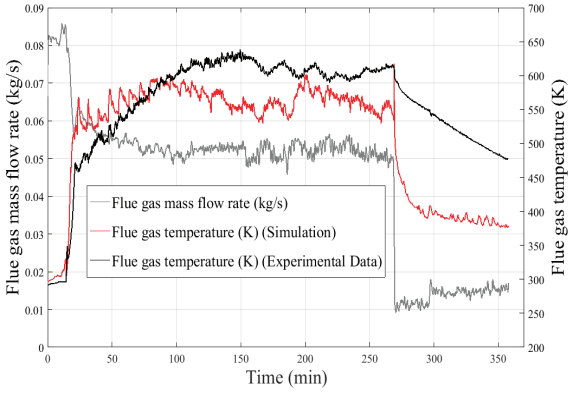
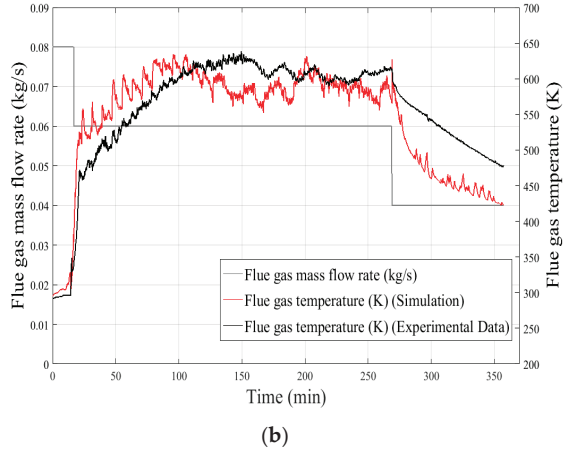


Figure 12. Calculated boiler efficiency.



(**a**)



(**b**)

Figure 13. Influence of the flue gas mass flow rate at the flue gas outlet on the 0D model response. (**a**): With experimental flue gas mass flow rate, (**b**): With a flue gas mass flow corrected for each boiler operating phase.

This discrepancy can be explained by a low value of flue gas flow rates during the cooling phase (Figure 13a) according to the accuracy of the measurement chain (Pitot tube associated with a micromanometer and a thermocouple), which involves a maximum error of 40%. By adjusting the mass flow rate value of the flue gas, staying within the uncertainty range of the flowmeter, a clear improvement of the model response is observed (Figure 13b).

A heat flux balance in four zones of the boiler is performed by using the equations introduced in Section 2.2 and allows to compare the radiative and convective heat flux. The convective heat flux is calculated by using the convective coefficients obtained from the semi-empirical correlations given in Table 2 and the total heat flux ϕ_{tot} is obtained by using the global thermal transfer coefficients calculated from the inverse method by carrying out flux balances inside the boiler using Equations (3)–(6), previously introduced in Section 2.2. The radiative flux can be then deduced from the total heat flux, knowing the convective heat flux.

Due to the presence of large temperature gradients in the boiler, combustion products such as water vapor (H_2O), carbon dioxide (CO_2) and soot particles radiate significantly. Radiation is the dominant thermal transfer in the boiler and must be compared to the total thermal transfers (Table 4).

Table 4. Heat flux balance.

Location	ϕ_{rad}/ϕ_{tot} (%)
Inside the combustion chamber	97.6
Outside the combustion chamber (annular passage)	96.8
Inside the heat exchanger (flue gas side)	96.1
Inside the flue gas pipes	95.6

5. Conclusions

A 0D dynamic modeling of a domestic biomass boiler of low power was developed by using Bond Graph formalism that allows to represent the coupled multi-physical phenomena, to study the thermal transfers between the different fluids during the transient operating phases, to evaluate the energy performances of the boiler and to take into account the variability of the heat production. The local evolution of the state variables is much less detailed than with CFD modeling but the dependencies of one zone with another are better taken into account with a 0D dynamic modeling. A biomass combustion model was not developed in this study but the combustion reaction of pellets in air allowed to calculate the thermodynamic properties of the flue gas in the boiler used in the 0D model. This model based on mass and energy balances was validated with experimental results, in particular the flue gas temperature in several locations of the boiler and the water temperature at the heat exchanger outlet. Some experimental data and 0D modeling at each time step of the calculation were coupled. The thermal transfers between the flue gas and the water circulating inside the heat exchanger and between these two fluids and the boiler structures were simulated. The experimental results showed a dependence of the evolution of the flue gas temperature in the combustion chamber as a function of the quantity of pellets supplied, according to the thermal cycle of the boiler. This directly affects the operating conditions of the boiler and generates important temperature fluctuations in the combustion chamber, which could significantly affect the operation of a hot air machine in the case of a conversion into a micro cogeneration unit. Indeed in this case, the air-gas exchanger of such an installation would be located in the zone where the temperature is the highest and thus closest to the flame. A calculation of the global thermal transfer coefficients by inverse method was done in the subsystems of the boiler. A good agreement between the experimental measurements and the simulation has been found and the origins of the differences have been identified, such as the excess of heat release in the combustion chamber above the burner zone not integrated in the model. It has been shown that the boiler has an average efficiency of 67.5% and the radiation is the dominant thermal transfer in the boiler by reaching 97.6% of the total thermal transfers inside the combustion chamber. The

0D dynamic model of the boiler during the operating phases allows not only to evaluate its energy performances but also to highlight the boiler components where the thermal transfers must be optimized.

The modeling of pellet combustion using a heat release law adapted to solid biomass combustion associated with an efficient identification of the pellet mass flow rate will make it possible to improve this model and make it independent of experimental boundary conditions. The radiative transfers being preponderant but difficult to model in 0D for mutual exchanges between gases, particles and walls, a detection of the different combustion phases and this according to the presence or not of flame in each of the zones would allow to better parameterize the global exchange coefficients which moreover will be able to be identified with the help of the proposed inverse method. The modeling methodology developed will allow the study of a complex unit, such as a CHP plant by coupling the different models for each component.

Author Contributions: Conceptualization, E.D.; Formal analysis, F.M.; Investigation, J.S.; Writing—Original Draft, F.M., C.M. and E.D.; Writing—Review and Editing, C.M. and E.D.; Software, F.M. and E.D.; Methodology, E.D. and C.M.; Validation, all authors; Resources: J.S.; Data curation, J.S.; Visualization, F.M.; Supervision, C.M.; Project administration, C.M.; Funding acquisition, C.M. All authors have read and agreed to the published version of the manuscript.

Funding: F.M. was supported by joint PhD grant from ADEME (French Environment and Energy Management Agency) and Hauts-de-France Region. Grant Number: TEZ15-27.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge CCM (ULCO, Dunkerque France) for the elementary analysis of wood pellets.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

$c_{p,fg}(T)$	Flue gas specific heat ($J.kg^{-1}.K^{-1}$) at constant pressure, function of temperature T
c_w	Water specific heat ($J.kg^{-1}.K^{-1}$)
c_{wall}	Wall specific heat ($J.kg^{-1}.K^{-1}$)
D_h	Hydraulic diameter (m)
H	Combustion chamber height (m)
h_g	Global thermal transfer coefficient ($W.m^{-2}.K^{-1}$)
$\dot{H}_{fg,[area]}$	Enthalpy flux of the flue gas in a specific area (W)
$\dot{H}_{w,[area]}$	Enthalpy flux of the water in a specific area (W)
k_1	First slope of Runge-Kutta fourth order formula
k_2	Second slope of Runge-Kutta fourth order formula
k_3	Third slope of Runge-Kutta fourth order formula
k_4	Fourth slope of Runge-Kutta fourth order formula
m_{wall}	Wall mass (kg)
$\dot{m}_{fg}^{exp}$	Experimental flue gas mass flow rate ($kg.s^{-1}$)
$\dot{m}_{pellets}^{exp}$	Experimental pellets mass flow rate ($kg.s^{-1}$))
$\dot{m}_{w}^{exp}$	Experimental water mass flow rate ($kg.s^{-1}$)
p_w	Water pressure (Pa)
p_{fg}	Flue gas pressure (Pa)
$\dot{Q}_w$	Heat flux transferred to the water (W)
$\dot{Q}_{wall}$	Heat flux stored in the boiler structure (W)
$\dot{Q}_{fg,cc}$	Convective heat flux exchanged between the flue gas and the combustion chamber wall (W)
$\dot{Q}_{wall,HEx}$	Heat flux stored in the heat exchanger wall (W)

$\dot{Q}_{wall,tub}$	Heat flux stored in the walls of the flue gas tubes (W)
r_1	Inside radius (m)
r_2	Outside radius (m)
R_{cd}	Conduction resistance ($K.W^{-1}$)
R_{cv}	Convective resistance ($K.W^{-1}$)
S	Exchange surface (m^2)
t	Time (s)
T_{amb}	Ambient temperature (K)
$T^{exp}_{fg,bot}$	Experimental flue gas temperature at the bottom of the heat exchanger (K)
$T^{exp}_{fg,bur}$	Experimental flue gas temperature in the burner (K)
$T^{exp}_{fg,cc}$	Experimental flue gas temperature in the combustion chamber (K)
$T^{exp}_{fg,exh}$	Experimental flue gas temperature in the chimney (boiler exhaust) (K)
$T^{exp}_{fg,exit}$	Experimental flue gas temperature at the flue gas tubes outlet (K)
$T^{exp}_{w,in}$	Experimental water temperature at the inlet of the heat exchanger (K)
$T^{exp}_{w,out}$	Experimental water temperature at the outlet of the heat exchanger (K)
$T_{fg,in}$	Calculated flue gas temperature at the RS-element inlet (K)
$T_{fg,out}$	Calculated flue gas temperature at the RS-element outlet (K)
$T^{exp}_{fg,top}$	Experimental flue gas temperature at the top of the combustion chamber (K)
$T^{exp}_{wall,inner}$	Experimental temperature of the inner wall of the combustion chamber (K)
$T^{exp}_{wall,outer}$	Experimental temperature of the outer wall of the combustion chamber (K)
T_{wall}	Calculated wall temperature (K)
Subscripts	
air	Air
amb	Ambient
bur	Burner
bot	Bottom
cc	Combustion chamber
cd	Conductive
cv	Convective
exit	Exit
exh	Exhaust
fg	Flue gas
g	Global
HEx	Heat Exchanger
in	Inlet
rad	Radiative
tub	Tube
top	Top
tot	Total
out	Outlet
w	Water
wall	Wall
Superscript	
exp	Experimental Value
Greek symbols	
Δ	Variation of thermodynamic quantity
λ_i	Wall thermal conductivity ($W.m^{-1}.K^{-1}$)
λ_{fg}	Flue gas thermal conductivity ($W.m^{-1}.K^{-1}$)
ρ_{fg}	Flue gas density ($kg.m^{-3}$)
μ_{fg}	Flue gas dynamic viscosity (Pa.s)
η	Boiler efficiency (%)
δ	Iteration step of Runge-Kutta fourth order formula
Dimensionless numbers	
Re	Reynolds number
Pr	Prandtl number
Nu	Nusselt number

References

1. Demirbas, M.F.; Balat, M.; Balat, H. Potential contribution of biomass to the sustainable energy development. *Energy Convers. Manag.* **2009**, *50*, 1746–1760. [CrossRef]
2. Sharma, A.; Pareek, V.; Zhang, D. Biomass pyrolysis—A review of modelling, process parameters and catalytic studies. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1081–1096. [CrossRef]
3. Sonnino, A. Agricultural biomass production is an energy option for the future. *Renew. Energy* **1994**, *5*, 857–865. [CrossRef]
4. Tripathi, M.; Sahu, J.; Ganesan, P. Effect of process parameters on production of biochar from biomass waste through pyrolysis: A review. *Renew. Sustain. Energy Rev.* **2016**, *55*, 467–481. [CrossRef]
5. Soltani, R.; Dincer, I.; Rosen, M.A. Thermodynamic analysis of a novel multigeneration energy system based on heat recovery from a biomass CHP cycle. *Appl. Therm. Eng.* **2015**, *89*, 90–100. [CrossRef]
6. Demirbas, A. Combustion characteristics of different biomass fuels. *Prog. Energy Combust. Sci.* **2004**, *30*, 219–230. [CrossRef]
7. Saidur, R.; Abdelaziz, E.; Demirbas, A.; Hossain, M.; Mekhilef, S. A review on biomass as a fuel for boilers. *Renew. Sustain. Energy Rev.* **2011**, *15*, 2262–2289. [CrossRef]
8. Strzalka, R.; Erhart, T.G.; Eicker, U. Analysis and optimization of a cogeneration system based on biomass combustion. *Appl. Therm. Eng.* **2013**, *50*, 1418–1426. [CrossRef]
9. Li, C.; Gillum, C.; Toupin, K.; Donaldson, B. Biomass boiler energy conversion system analysis with the aid of exergy-based methods. *Energy Convers. Manag.* **2015**, *103*, 665–673. [CrossRef]
10. Kang, S.B.; Kim, J.J.; Choi, K.S.; Sim, B.S.; Oh, H.Y. Development of a test facility to evaluate performance of a domestic wood pellet boiler. *Renew. Energy* **2013**, *54*, 2–7. [CrossRef]
11. Gómez, M.A.; Porteiro, J.; de la Cuesta, D.; Patiño, D.; Míguez, J.L. Dynamic simulation of a biomass domestic boiler under thermally thick considerations. *Energy Convers. Manag.* **2017**, *140*, 260–272. [CrossRef]
12. Zadravec, T.; Rajh, B.; Kokalj, F.; Samec, N. CFD modelling of air staged combustion in a wood pellet boiler using the coupled modelling approach. *Therm. Sci. Eng. Prog.* **2020**, *20*, 100715. [CrossRef]
13. Karim, M.R.; Naser, J. CFD modelling of combustion and associated emission of wet woody biomass in a 4 MW moving grate boiler. *Fuel* **2018**, *222*, 656–674. [CrossRef]
14. Tognoli, M.; Najafi, B. Dynamic modelling and optimal sizing of industrial fire-tube boilers for various demand profiles. *Appl. Therm. Eng.* **2018**, *132*, 341–351. [CrossRef]
15. Bouvenot, J.-B.; Latour, B.; Siroux, M.; Flament, B.; Stabat, P.; Marchio, D. Dynamic model based on experimental investigations of a wood pellet. *Appl. Therm. Eng.* **2014**, *73*, 1039–1052. [CrossRef]
16. Carlon, E.; Verma, V.K.; Schwarz, M.; Golicza, L.; Prada, A.; Baratieri, M.; Haslinger, W.; Schmidl, C. Experimental validation of a thermodynamic boiler model under steady state and dynamic conditions. *Appl. Energy* **2015**, *138*, 505–516. [CrossRef]
17. Ziviani, D.; Beyene, A.; Venturini, M. Advances and challenges in ORC systems modeling for low grade thermal energy recovery. *Appl. Energy* **2014**, *121*, 79–95. [CrossRef]
18. Féniès, G.; Formosa, F.; Ramousse, J.; Badel, A. Double acting Stirling engine: Modeling, experiments and optimization. *Appl. Energy* **2015**, *159*, 350–361. [CrossRef]
19. Creyx, M.; Delacourt, E.; Morin, C.; Desmet, B. Dynamic modelling of the expansion cylinder of an open Joule cycle Ericsson engine: A bond graph approach. *Energy* **2016**, *102*, 31–43. [CrossRef]
20. Lontsi, F.; Hamandjoda, O.; Fozao, K.; Stouffs, P.; Nganhou, J. Dynamic simulation of a small modified Joule cycle reciprocating Ericsson engine for micro-cogeneration systems. *Energy* **2013**, *63*, 309–316. [CrossRef]
21. Gölles, M.; Reiter, S.; Brunner, T.; Dourdoumas, N.; Obernberger, I. Model based control of a small-scale biomass boiler. *Control Eng. Pract.* **2014**, *22*, 94–102. [CrossRef]
22. Abdulmoneim, M.M.; Aboelela, M.A.; Dorrah, H.T. Hybrid modeling using power plant and controlling using fuzzy P+ID with application. *Int. J. Adv. Eng. Technol.* **2012**, *4*, 42–53.
23. Paynter, H.M. *Analysis and Design of Engineering Systems: Class Notes for M.I.T. Course*; University of Michigan: Cambrige, MA, USA, 1961.
24. Karnopp, D.; Rosenberg, R.C. *System Dynamics: A Unified Approach*; John Wiley & Sons Inc: Hoboken, NJ, USA, 1975.
25. Merabtine, A.; Benelmir, R. Modeling of the RHC System with Bond Graphs Approach. *Int. J. Therm. Environ. Eng.* **2013**, *5*, 145–153.
26. Nur Aziz, A.; Nazaruddin, Y.Y.; Siregar, P.; Bindar, Y. Structured Mathematical Modeling of Industrial Boiler. *J. Eng. Technol. Sci.* **2014**, *46*, 102–122. [CrossRef]
27. Dong, Y.; El-Bakkali, A.; Descombes, G.; Feidt, M.; Périlhon, C. Association of Finite-Time Thermodynamics and a Bond-Graph Approach for Modeling an Endoreversible Heat Engine. *Entropy* **2012**, *14*, 642–653. [CrossRef]
28. Aridhi, E.; Abbes, M.; Mami, A. Pseudo bond graph model of a thermo-hydraulic system. In Proceedings of the International Conference on Modeling, Simulation and Applied Optimization, Hammamet, Tunisia, 28–30 April 2013; pp. 1–5.
29. Couenne, F.; Jallut, C.; Maschke, B.; Breedveld, P.C. Bond graph for dynamic modelling in chemical engineering. *Chem. Eng. Process.* **2008**, *47*, 1994–2003. [CrossRef]
30. Ould Bouamama, B.; el Harabi, R.; Abdelkrim, M.N.; Gayed, M.B. Bond Graphs for diagnosis of Chemical Processes. *Comput. Chem. Eng.* **2012**, *36*, 301–324. [CrossRef]

31. Verma, V.K.; Bram, S.; Delattin, F.; de Ruyck, J. Real life performance of domestic pellet boiler technologies as a function of operational loads: A case study of Belgium. *Appl. Energy* **2013**, *101*, 357–362. [CrossRef]
32. Åström, K.J.; Bell, R.D. Simple Drum-Boiler Models. *IFAC Proc. Vol.* **1988**, *21*, 123–127. [CrossRef]
33. Sandberg, J.; Fdhila, R.B.; Dahlquist, E.; Avelin, A. Dynamic simulation of fouling in a circulating fluidized biomass-fired boiler. *Appl. Energy* **2011**, *88*, 1813–1824. [CrossRef]
34. Persson, T.; Fiedler, F.; Nordlander, S.; Bales, C.; Paavilainen, J. Validation of a dynamic model for wood pellet boilers and stoves. *Appl. Energy* **2009**, *2009*, 645–656. [CrossRef]
35. Shannon, K.S.; Butler, B.W. A review of error associated with thermocouple temperature measurement in fire environments. In Proceedings of the 2nd International Wildland Fire Ecology and Fire Management Congress and the 5th Symposium on Fire and Forest Meteorology, Orlando, FL, USA, 16–20 November 2003.
36. Hindasageri, V.; Vedula, R.P.; Prabhu, V. Thermocouple error correction for measuring the flame temperature with determination of emissivity and heat transfer coefficient. *Rev. Sci. Instrum.* **2013**, *84*, 024902-1–024902-11. [CrossRef] [PubMed]
37. Winterton, R.H. Where did the Dittus and Boelter equation come from? *Int. J. Heat Mass Transf.* **1998**, *41*, 809–810. [CrossRef]
38. Gnielinski, V. Heat Transfer Coefficients for Turbulent Flow in Concentric Annular Ducts. *Heat Transf. Eng.* **2009**, *30*, 431–436. [CrossRef]
39. Poling, B.E.; Prausnitz, J.M.; O'Connell, J.P. *The Properties of Gases and Liquids*, 5th ed.; The McGraw-Hill Companies: Henrico, VA, USA, 2001.
40. Krieager, F.J. Calculation of the viscosity gas mixtures. *RAND Corp.* **1951**, *RM-649*, 1–11.
41. Stull, D.R.; Prophet, H. *JANAF Thermochemical Tables*; Defense Technical Information Center: Washington, DC, USA, 1971; pp. 1856–1985.
42. Rajaraman, V. *Computer Oriented Numerical Methods*, 3rd ed.; Prentice-Hall of India: New Delhi, India, 2006.
43. Najafi-Yazdi, A.; Mongeau, L. A low-dispersion and low-dissipation implicit Runge–Kutta scheme. *J. Comput. Phys.* **2013**, *233*, 315–323. [CrossRef]
44. Akanbi, M.A.; Okunuga, S.A.; Okunuga, A.B. Runge-Kutta Schemes for Solving Electrical Network Problems. *J. Sci. Res. Dev.* **2001**, *6*, 31–44.
45. Kim, D.; Stanescu, D. Low-storage Runge–Kutta methods for stochastic differential equations. *Appl. Numer. Math.* **2008**, *58*, 1479–1502. [CrossRef]

electronics

MDPI

Article

An Ultra-Low Power Threshold Voltage Variable Artificial Retina Neuron

Qiguang Wang, Guangchen Pan and Yanfeng Jiang *

Department of Microelectronics, School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; wqg1997@outlook.com (Q.W.); 6171916008@stu.jiangnan.edu.cn (G.P.)
* Correspondence: jiangyf@jiangnan.edu.cn

Abstract: An artificial retina neuron is proposed and implemented by CMOS technology. It can be used as an image sensor in the Artificial Intelligence (AI) field with the benefit of ultra-low power consumption. The artificial neuron can generate signals in spike shape with pre-designed frequencies under different light intensities. The power consumption is reduced by removing the film capacitor. The comparator is adopted to improve the stability of the circuit, and the power consumption of the comparator is optimized. The power consumption of the proposed CMOS neuron circuit is suppressed. The ultra-low-power artificial neuron with variable threshold shows a frequency range of 0.8–80 kHz when the input current is varied from 1 pA to 150 pA. The minimum DC power is 35 pW when the input current is 5 pA. The minimum energy of the neuron is 3 fJ. The proposed ultra-low-power artificial retina neuron has wide potential applications in the field of AI.

Keywords: artificial retina neuron; spike; CMOS; Axon-Hillock circuit; ultra-low power

Citation: Wang, Q.; Pan, G.; Jiang, Y. An Ultra-Low Power Threshold Voltage Variable Artificial Retina Neuron. *Electronics* **2022**, *11*, 365. https://doi.org/10.3390/electronics11030365

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 3 December 2021
Accepted: 17 January 2022
Published: 25 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Compared with the traditional Von Neumann architecture computer, the human brain shows stronger associative memory and thinking in images. It also has a greater potential ability than existing computers in solving complex problems such as function approximation, complex classification and clustering [1]. Moreover, compared with current existing computers, the human brain is not only more powerful, but it is also smaller and consumes less power. Therefore, the realization of the artificial neural network (ANN) to mimic the human brain intelligence has become a hot subject for research recently [2]. The human brain is composed of many complex interconnected neurons, and the information interaction between neurons is what forms the thinking ability. Designing a reasonable and efficient neuron unit is an important point for imitating the thinking ability of the human brain [3,4].

The first-generation ANN consists of threshold gates [5]. Its principle is using the threshold gate to judge the output result by counting the binary sum of the inputs. If the inputs' summation is larger than the threshold value, it is considered to be high level (1); otherwise it is low level (0). It can be seen that the function of the first-generation ANN is very limited and that it can only process binary data. This is still far removed from the real biological neuron. The second-generation ANN is based on the encoding of the frequencies of the neuron pulses [6]. By stacking multiple layers of the neurons and applying a back propagation algorithm, a neural network can be constructed, which is known as deep learning neural network. This network is widely used in machine learning, brain-machine interfaces, image sensors, etc. [7]. Although the second-generation ANN is powerful, its energy consumption and efficiency are still not good enough compared with the biological network. Moreover, there is a big difference in the process of communicating with the spikes of neurons in the human brain in the underlying logic. Faced with these problems, the third-generation ANN has been proposed recently. Its neuron units are much closer to biological neurons, in that they can communicate with each other using spike signals.

For this reason, it is also called spiking neural networks (SNN) [8]. The neuron in the SNN is not activated in every iteration state. It can be activated only when the membrane voltage reaches a certain value. When a spike neuron is activated, it generates a spike signal, which is transmitted to other neurons [9]. After the transmission, its membrane potential is changed accordingly. The spike generated by a biological neuron is used for encoding and processing the biological information. The artificial neural network shows a far superior ability in implementing real-time behavior systems or detailed large-scale simulations of neural systems than other digital tools and simulators [10].

The retina is a key tissue and can obtain visuosensory information efficiently, a subject that has been intensively studied recently [11–13]. By the pre-processing of optical information on the retina, the input light is transferred into the corresponding neural signal, which is encoded into the spike pattern for further transmission into a higher processor. Mimicking the biological retina, the artificial neuron model is designed based on CMOS technology, which is used to convert the optical pixel signals into specific spikes with certain frequencies. Billions of neurons with complex connections could build a large and efficient biological computing system. Figure 1 shows the schematic diagrams of the biological retinal system and the artificial ones. With the very large number of retinal cells in the artificial neuron structure, it is important to optimize the energy efficiency of the artificial neurons by reducing the power consumption. One of the most important issues of the artificial neuron in a neuromorphic system is how to decrease the power consumption.

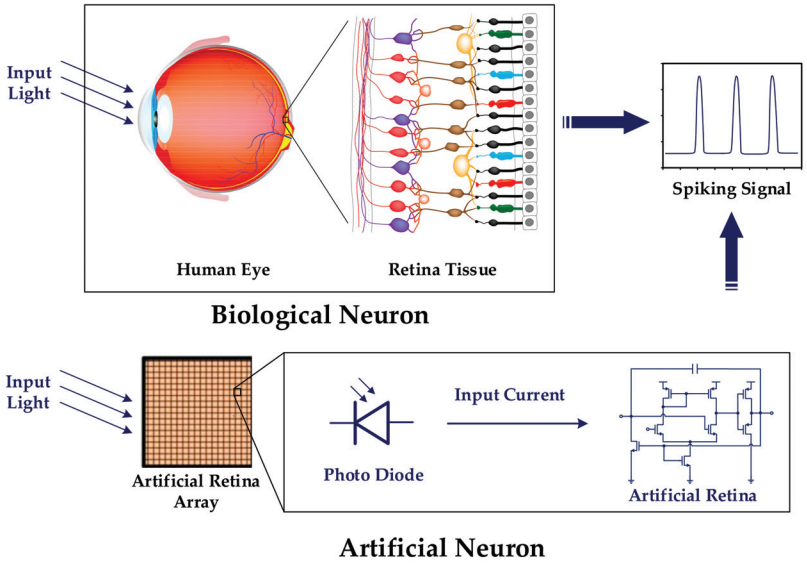


Figure 1. Schematic diagram of the working mechanism of the artificial neurons mimicking the biological retina.

With the continuous investigation of the working mechanism of neurons, some artificial neuron circuits have been proposed. In [14], a neuron circuit based on the leaky integrate-and-fire (LIF) model is proposed. This circuit can realize the spike timing dependent plasticity (STDP) function of the neuron [15]. However, due to the existence of the multiple trans-conductance amplifiers in the circuit, the power is too high to be implemented practically. In [16], a circuit based on the Morris-Lecar (ML) model is proposed, as shown in Figure 2. Because the ML model is similar to the ion transport mechanism of the real neurons, the circuit can be used to mimic the real neurons [17]. However, due to the existence of many conductive paths in the circuit, its static power consumption is

relatively high. Moreover, the adopted large capacitances Cm and CK limit its operating frequency. In [18], the circuit is simplified to reduce its power consumption, making the circuit display excellent merits in terms of its power consumption and area. However, the circuit is unstable and susceptible to the influences of the PVT variables.

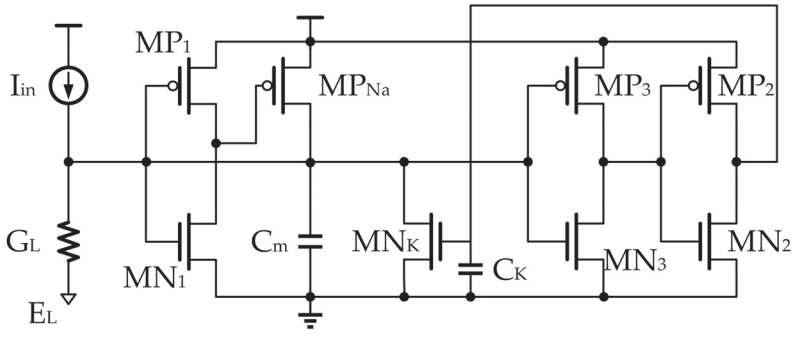


Figure 2. Artificial neuron circuits based on ML model. As drawn in [16].

In this paper, a novel artificial neuron circuit is proposed that has ultra-low power while keeping robust variation tolerance. The circuit shows minimum layout area and can be integrated into large-scale arrays for mimicking the biological systems. The structure of the proposed artificial neuron circuit is described in the paper. The analysis and the results of the artificial neuron retina are reported.

2. The Principle of Axon-Hillock Circuit

Figure 3 shows the Axon-Hillock circuit, which is considered to be the traditional artificial neuron circuit [19]. It was proposed by Mead in 1989 and has been widely used in many works [20,21]. The input current I_{in} is commonly generated by a photodiode, in which different light intensities correspond to different magnitudes of the induced currents. The current I_{in} charges the membrane capacitor C_{mem}. The capacitor C_{mem} is the model of the retinal neuron's membrane with the ionic current across it.

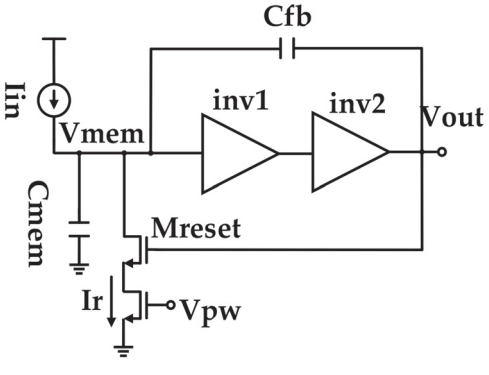


Figure 3. Traditional Axon-Hillock structure. I_{in} is the current generated by the photodiode. C_{mem} is the capacitor of the neural membrane.

The amplifier in Figure 3 is the main part for the generation of the neuron spike, in which two invertors, inv1 and inv2, are included. At the initial state, both the values of V_{out} and V_{mem} are zero. The capacitor C_{mem} is charged by I_{in}, so that the voltage V_{mem} on C_{mem} is pulled up by the charging current. When V_{mem} exceeds the threshold voltage of inv1, the invertor flips and a spike signal is generated by the output port. At this moment, V_{out} is at

a high enough level to turn on the reset transistor, Mreset. The reset current is set by V_{pw}. If it is larger than the input current I_{in}, the membrane capacitor is discharged. Therefore, V_{mem} is decreased continuously until it reaches the amplifier's switching threshold again. In this way, a cycle is finished and the next cycle starts again.

The circuit in Figure 3 can imitate the characteristics of a stimulated retinal neuron, in which the output of the electric spike signal with a certain frequency can be generated and adjusted. It is a kind of classical SNN circuit. Based on the circuit, some research works propose useful solutions on how to reduce the power consumption. In [18], the C_{mem} capacitor is replaced by the parasitic part of the MOSFET of the first-stage inverter. Without the capacitor C_{mem}, the proposed solution can effectively reduce the power consumption of the neuron circuit. In the above design, the spike is related to the threshold voltage of inv1, which is determined by the process characteristics of the MOSFET device. However, the process parameters of MOSFET are generally variable in a certain scale, which easily leads to large deviation and affects the accuracy of the neural network calculation.

To diminish the possible influences of the process variations, a specific reference voltage V_{thr} is introduced, accompanied with a comparator for the implementation of the circuit, as shown in Figure 4 [10]. The adoption of the comparator can increase the process variation tolerance and improve the robustness of the artificial neuron. However, the power is increased according to the additional comparator and the related reference voltage V_{thr}. In this way, further improvement should be addressed to improve its characteristics. The following section shows the detailed information of the improvement.

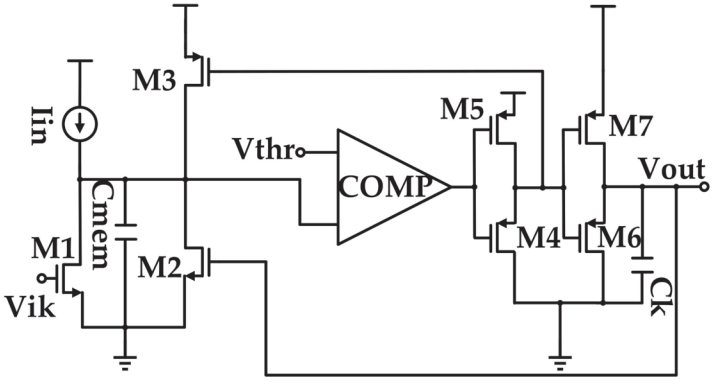


Figure 4. The artificial neuron with the reference voltage V_{thr}. As drawn in [10].

3. Design of Novel Artificial Neuron

Based on the operation principle of the neuron, the new designed artificial neuron is shown in Figure 5. In the design, the adjustable voltage threshold is adopted. With the current charging, I_{in} can inspire the artificial neuron to generate the mimicked spikes with a certain frequency with ultra-low power consumption.

The setting of the voltage threshold is achieved by combining the comparator and the traditional Axon-Hillock circuit. As shown in Figure 5, the inverter with the two devices, M1 and M4, is the main part in the amplifier. At the same time, the inverter composed of M1 and M4 also operates in the comparator. The comparator with a mirror current source includes four transistors, M0, M1, M3 and M4.

With the positive input, the gate of M0 is the input of the reference threshold voltage V_{thr}. With the negative input, the gate of M1 is the input of the membrane voltage V_{mem}. When the voltage V_{mem} is higher than the threshold voltage V_{thr}, the comparator output voltage V_c is zero. Otherwise, when V_{mem} is lower than V_{thr}, V_c is set as V_{dd}. As the first stage inverter of the amplifier, M1 and M4 act as the same function as the inv1 of the Axon-Hillock in Figure 3. Therefore, if V_{mem} is increased to be the threshold voltage V_{thr},

the output of the artificial neuron reaches a high level by the output of the second invertor. For the proposed circuit in Figure 5, M1 and M4 are common-shared by the amplifier and the comparator. The common-shared design can effectively reduce the number of neuron circuits and therefore decrease the power consumption of the artificial neuron.

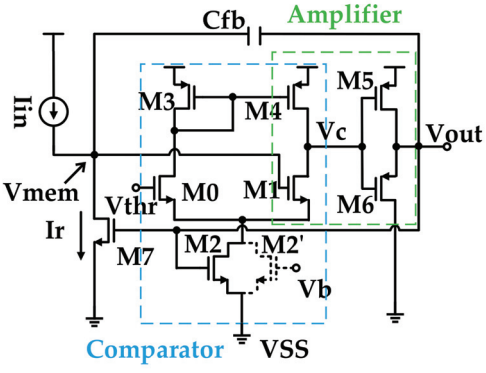


Figure 5. Schematic of the artificial neuron designed in the work. The comparator and the amplifier share the branch of M1 and M4 together. M2 is the tail current source supplied by the V_{out} in this work. M2′ is the common tail current source supplied by the reference voltage V_b.

The power consumption is the key factor to be considered in the artificial neuron design [22]. To reduce the power consumption, one effective solution is to remove the membrane capacitor. As shown in Figure 5, the parasitic capacitance in the negative half cycle of the comparator is used as a part of the capacitance. At the initial state, the output voltage V_{out} is zero. Therefore, the feedback capacitor C_{fb} can also be regarded as the membrane capacitor and charged by the input current I_{in}. The output voltage V_{out} is connected with the gate of M7. Therefore, M7 is not only a switch of the reset current I_r, but also the current source of I_r.

To further reduce the power consumption, the tail current source of the comparator is effectively processed. In general, the offset voltage of the tail current source M2′ is provided by the reference voltage V_b. However, the branch of M2′, M0 and M3 is always at the conduction state because of the existence of V_{thr} and V_b. In this situation, the quiescent current always exists even without the input current. Therefore, the neural circuit still has a large amount of power loss during the sleeping state. To solve the problem, V_b is not used anymore.

As shown in Figure 5, the actual bias of the M2 is provided by the output voltage V_{out} of the artificial neuron. When there is no input current, V_{out} is zero and there i no static current through M2. The reduction of the tail current of the comparator can effectively decrease the power consumption. The operation process of the artificial neuron is shown in Figure 6. At the initial state, there is no light irritation. The output of the photodiode is zero. Therefore, the input current I_{in} is also zero without the light irritation. The membrane voltage V_{mem} is at a low level, which is lower than the threshold voltage V_{thr}. Therefore, the output voltage is low. No reset current is generated because the transistor M7 is at the off-state. As the input increases, Tr also shrinks, and the frequency of the output spike becomes higher.

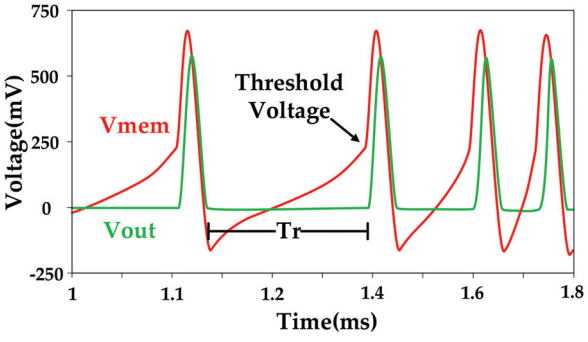


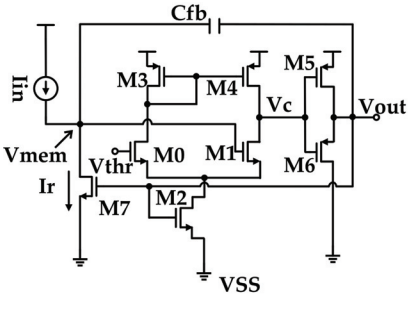
Figure 6. The schematic diagram of the operation process of the artificial neuron. V_{mem} is the voltage of the membrane capacitor. V_{out} is the output of the artificial neuron. Tr is the time of the resting state.

When the light is switched on, the dc current I_{in} is produced by the photodiode. The capacitor C_{fb} is charged by I_{in}. Therefore, the membrane voltage V_{mem} is increased during the charging process. Before V_{mem} reaches the threshold voltage, the output voltage of the neural circuit V_{out} is kept at a low level. When V_{mem} exceeds V_{thr}, the V_c of the first stage inverter is switched to a low level quickly. Meanwhile, V_{out} of the second inverter is quickly changed from 0 to V_{dd}. The membrane voltage V_{mem} is pulled up to the level of V_{out} by the feedback capacitor to maintain the stable state of the comparator and the invertors. As V_{out} rises, the reset current source M7 is turned on and generates the reset current I_r. Because I_r is greater than the input current I_{in}, the membrane voltage V_{mem} decreases back to the threshold voltage V_{thr}. Thus, the output voltage of the comparator and artificial neuron are reset to their initial state. The re-closed reset current source M7 causes the feedback capacitor to be charged by input current again.

The rest time Tr is controlled by the input current, the feedback capacitance and the threshold voltage at the same time. The resting time is inversely proportional to the input current I_{in}, but proportional to the feedback capacitance C_{fb} and the reference voltage V_{thr}.

4. Result and Discussions

The proposed circuit is simulated with SMIC 40 nm CMOS process. The sizes of the transistors in the circuit are shown in Figure 7. In order to reduce the leakage currents of the transistors and suppress the static power consumption of the circuit, the channel length of M7 is set to 120 nm. The feedback capacitor Cfb is set to 5 fF. The power supply voltage is set to 500 mV and the reference voltage Vthr to 50 mV. The default value of the input current Iin is 5 pA.



transistor	W/L (nm)
M0	160/40
M1	160/40
M2	120/40
M3	120/40
M4	120/40
M5	120/41
M6	160/40
M7	120/120

Figure 7. Schematic diagram of the designed circuit, with the information of the sizes of the transistors.

Figure 8 shows the relationship between the output voltage V_{out} and the membrane voltage V_{mem} with different reference voltages V_{thr}. In the figure, the value of V_{thr} is varied from 30 mV to 70 mV. It can be seen that the flip point of the output is changed from 80 mV to 110 mV corresponding to the different V_{thr} values. This means that the reference voltage of the comparator has a proportional effect on the flipping point of the output, while the traditional one depends entirely on the process parameters of the inverter [14–16]. Adoption of the comparator reduces the influence of process parameters on the circuit flipping mechanism and improves the robustness of the circuit [10]. The sweep simulation with Vthr varied from 0 to 100 mV is conducted and the same tendency can be obtained, showing the circuit to have a wide operating range.

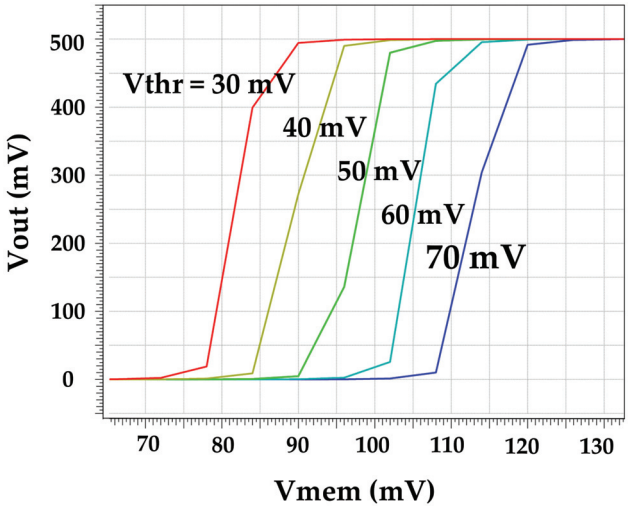
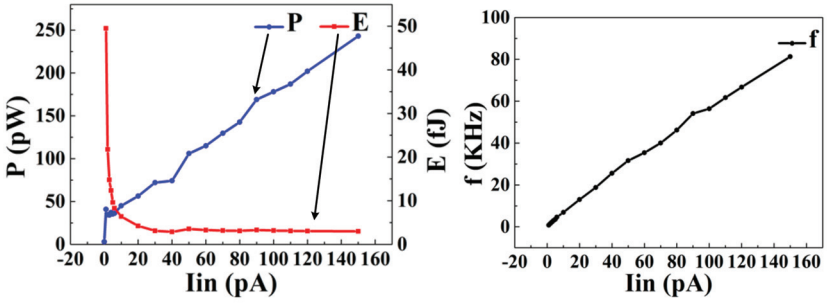


Figure 8. The relationship between the output voltage V_{out} and the membrane voltage V_{mem} under different reference voltages V_{thr} varied from 30 mV to 70 mV.

As illustrated in Figure 5, the tail current of the comparator is cut off by the gate control on M2. The gate of M2 is directly connected with the output voltage. At this point, except for the weak leakage current of M3 and M7, there are no static currents on the other MOS transistors in the neural circuit. The static power consumption can be suppressed effectively. When the input current is not 0, the tail current source M2 is turned on by V_{out}, providing the current for the comparator. When V_{mem} exceeds V_{thr}, the two branches of the comparator are all switched on. All of the transistors in the neural circuit except M6 and M7 have current flowing through.

The total power consumption of the neural circuit and the energy loss by a spike signal in the range of input current from 1 to 150 pA are shown in Figure 9a. The power P is the product of the supply voltage V_{dd} and the DC current. E denotes the energy consumption per spike. With the increment of I_{in}, the charging speed of the feedback capacitor and the frequency of the state are accelerated. The minimum power is 35 pW with the input current 5 pA. The energy consumed by a spike is as low as 3 fJ. The cycle period is shrunk with the increasing of the input current I_{in}. In the tradeoff of the cycle time, the energy consumed by a spike signal is decreased, being opposite to the increment of the circuit power.

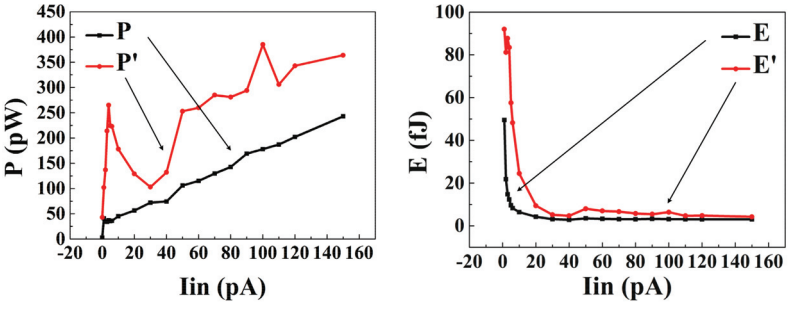


(**a**) (**b**)

Figure 9. Power consumption and the output frequency results of neural circuits. (**a**) The variation of the power P and spike energy E with the input current. (**b**) The variation of the spike frequency with the input current.

As shown in Figure 9b, the frequency of the spike signal is positively correlated with the input current. As the input current increases, the charging speed of the current on the feedback capacitor increases, which can significantly reduce the charging time of the membrane voltage. With the intensity of the input signal triggering the circuit, the output spike signal of the corresponding frequency is generated, which is the artificial neural source that imitates the working mechanism of the biological neuron, and it is also the core of the SNN signal encoding.

As shown in Figure 10, the artificial neurons in different schemes of the tail current sources (with M2 or with M2′) are compared in terms of the power consumption. The voltage offset of the tail current source of M2 is connected directly with the output voltage V_{out}. The tail current source consisting of M2′ is provided with a voltage offset by a separate voltage source. It can be seen, in Figure 10a, that the power consumption in the activated state of the V_{out}-biased tail current source (with M2) is significantly lower than that of the fixed-biased tail current source (with M2′). Similarly in Figure 10b, the power consumed by each spike in the circuit using the V_{out}-bias current source M2 is also reduced.



(**a**) (**b**)

Figure 10. The comparison of power and spike energy between the artificial neurons using the tail current source in M2 and M2′. (**a**) P is the power of the circuit with M2. P′ is the power of the circuit with M2′. (**b**) E is the spike energy of the circuit with M2. E′ is the spike energy of the circuit with M2′.

The neuron unit in SNN is not activated during the iterations, so the power consumption in the standby state accounts for the main part of the total power consumption. When the input current is 0, the DC current is 6.5 pA, which means a standby power consumption of 3.25 pw. However, if a fixed-biased tail current source M2′ is used, the DC current in the

standby state increases to 80 pA due to the presence of the on-state current, which means a static power consumption of 40 pW. This is intolerable in a low-power neuron circuit. The use of a V_{out}-biased tail current source significantly reduces the overall power consumption of the circuit.

In order to verify the influence of the supply voltage V_{dd} and ambient temperature on the working frequency, the circuits are verified under the conditions of the feedback capacitance of 5 fF and the input current of 40 pA [20]. As shown in Figure 11a, with the increase of the power supply voltage V_{dd}, the emission frequency decreases. With the voltage range of 0.44–0.56 V, the variation of the frequency is approximately 1.5%. This means that the circuit is less affected by the power supply ripple and that the circuit is robust to the potential power supply voltage noise. The relationship between the firing frequency and the temperature is shown in Figure 11b. With the temperature increasing, the firing frequency tends to increase. The maximum variation of the emission frequency is approximately 6% in the range of 27–41 °C.

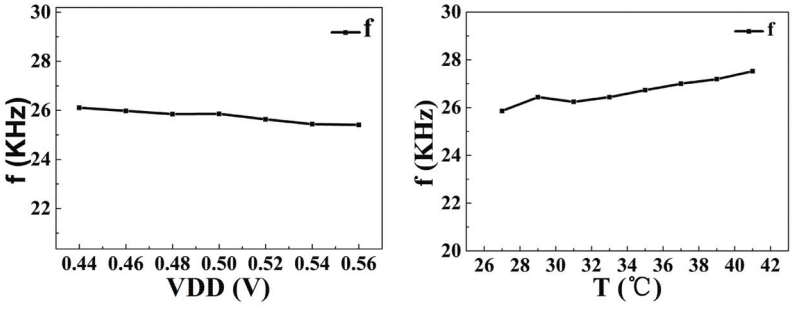


Figure 11. The influence of the supply voltage and the temperature on the transmitting frequency. (**a**) The variation of the firing frequency with the supply voltage. (**b**) The variation of the firing frequency with the temperature.

The transient state simulation results of the neural circuits are shown in Figure 12. When the input current I_{in} is zero, the membrane voltage V_{mem} and the output of the artificial neuron remain at zero. When the input current is 5 pA (I_{in} in Figure 10), V_{mem} is increased with the charging by the input current. Afterwards, by the presence of the reset current source M7, V_{mem} is decreased. In this way, the spikes with certain frequency can be generated, as shown by the Vout result in Figure 12.

Figure 13 shows the layout of the designed neuron circuit. Thanks to the shrinking process size, its area is only 13 μm^2, which makes it easy to implement the integration of the neuron arrays with thousands of the retina cells.

The designed retinal circuit is fabricated based on standard CMOS 40-nm technology. Figure 14 shows the input and output waveforms of the retina with different input currents. Figure 14a shows the input current waveform. After the artificial retina processing, the output voltage is shown in Figure 14b. As the input changes from 6.3 pA to 9.4 pA, the interval between output spikes changes from 0.27 ms to 0.19 ms, that is, the frequency changes from 3.7 kHz to 5.3 kHz. For the performance of the fabricated chip, when the working voltage is 500 mV, the overall power consumption is 23 μW, which is mainly consumed by the reference voltage part. With the input current of 5 pA, and the temperature changing from 25 °C to 40 °C, the output spike frequency is varied within 2.3%. When the supply voltage is varied from 440 mV to 550 mV, the maximum output spike frequency is changed within 7.4%. The artificial neuron circuit can generate the spikes with the frequency ranging from 0.8–80 kHz when the input current is changed from 1 pA to 150 pA. It can be seen that the measured result coincides with the simulated ones.

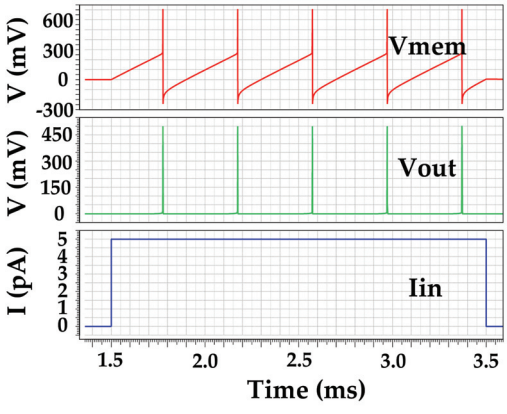


Figure 12. The transient simulation results of neural circuits. I_{in} is the input current generated by the photodiode. V_{out} is the output of the artificial neuron. V_{mem} is the voltage of the membrane capacitor.

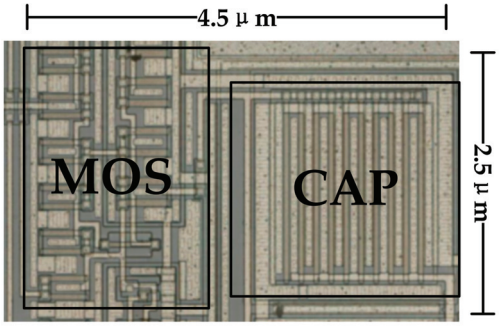


Figure 13. The layout photo of the designed neuron circuit, with area 13 μm^2.

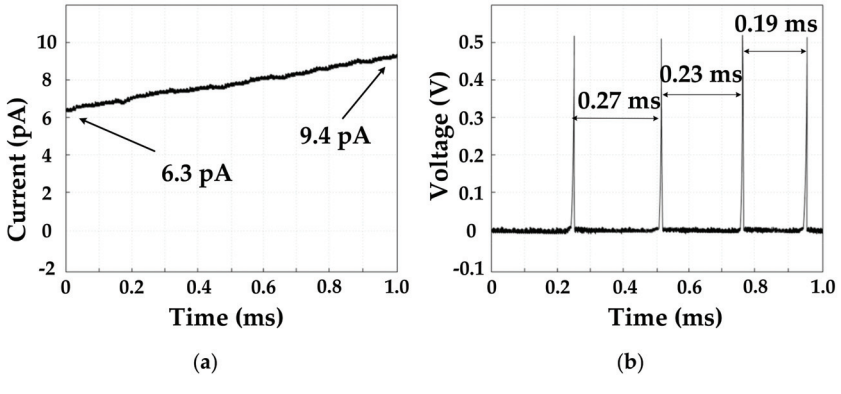


Figure 14. The input and output waveforms of the designed artificial retinal neuron circuit based on the fabricated chip. The different spikes correspond to different input currents. (**a**) The waveform of the input current, in which the amplitude of the input current is increased from 6.3 pA to 9.4 pA. (**b**) The output waveform, in which the output spike voltage is generated by different time intervals.

To get a clear comparison with the other similar published works, Table 1 lists the key results of the designed circuit and the other published works. The results in the paper show better performance, especially in terms of the ultra-low power consumption. The power consumption of the design in the paper is approximately 35 pW, which is compatible with the result in [18]. In [18], the power supply voltage is 200 mV, while the voltage in this paper is 500 mV. The layout area of the design is also smaller than those of the other published results. Besides the low power consumption of the design, the robustness of the circuit is the other advantage. For [18], both the capacitor C_{mem} and the comparator are removed to obtain the low power consumption, with the sacrifice of the robustness of the circuit.

Table 1. Comparison of the design in this paper and the design in others.

Work	Process (nm)	Spiking Frequency (kHz)	Area (μm^2)	Power (W)	Energy Efficiency (pJ/Spike)
[18]	65	15.7	31	30 p	0.002
[23]	65	1900	120	78µ	41
[24]	350	0.1	1887	40 p	17.4
[14]	90	0.1	442	40 p	0.4
[16]	65	26	35	105 p	0.004
This work	40	0.8–80	12	35 p	0.003

The adopted comparator can improve the stability of the circuit. The stability is an important parameter for the neurons used in the network. In a circuit without a comparator, the flip threshold is determined by the threshold of the MOS transistor itself, which is easily affected by the PTV variables. For example, the circuit in [18] is more susceptible to PVT factors without using a comparator. In the simulation results, the spike frequency fluctuates up to 20% by the temperature and up to 25% by the supply voltage. For the circuit in the paper, the fluctuation is controlled successfully within 6% by the temperature and within 1.5% by the voltage. It can be seen that the circuit in the paper can improve the temperature fluctuation by three times and the voltage fluctuation by 16 times when it is compared with [18].

5. Conclusions

The artificial retinal neuron is used to mimic the biological neuron in hardware implementation and is widely used in neuromorphic computing and image sensors. In this paper, a novel artificial retinal neuron with ultra-low power is proposed and demonstrated. With the combination of the comparator and the Axon-Hillock circuit, the artificial neuron not only achieves the setting of the voltage threshold, but also reduces the power loss of the circuit dramatically. In addition, the regulated tail current source by the output of the artificial neuron reduces the leakage current at the static state. The artificial neuron can generate spikes in frequency ranging from 0.8 to 80 kHz when the input current is varied from 1 pA to 150 pA. The minimum DC power is 35 pW at the 5 pA of the input current. The minimum energy consumption of a spike is as low as 3 fJ. It is verified that the proposed artificial neuron circuit can be used to convert the light intensity into the spike signal effectively with ultra-low power consumption.

Author Contributions: Q.W. conceived the idea and wrote most of this paper. G.P. wrote the paper together. Y.J. took part in the discussion, provided expertise and supervised the work. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by a grant from National Natural Science Foundation of China (NSFC)—award number 61774078.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-Art in Artificial Neural Network Applications: A Survey. *Heliyon* **2018**, *4*, e00938. [CrossRef] [PubMed]
2. Carvalho, G.; Pereira, M.; Kiazadeh, A.; Tavares, V.G. A Neural Network Approach Towards Generalized Resistive Switching Modelling. *Micromachines* **2021**, *12*, 1132. [CrossRef]
3. Tacchino, F.; Macchiavello, C.; Gerace, D.; Bajoni, D. An Artificial Neuron Implemented on an Actual Quantum Processor. *Npj Quantum Inf.* **2019**, *5*, 1–8. [CrossRef]
4. Kurenkov, A.; DuttaGupta, S.; Zhang, C.; Fukami, S.; Horio, Y.; Ohno, H. Artificial Neuron and Synapse Realized in an Antiferromagnet/Ferromagnet Heterostructure Using Dynamics of Spin–Orbit Torque Switching. *Adv. Mater.* **2019**, *31*, 1900636. [CrossRef] [PubMed]
5. Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychol. Rev.* **1958**, *65*, 386. [CrossRef]
6. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagnet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
7. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* **2016**, *529*, 484–489. [CrossRef] [PubMed]
8. Joubert, A.; Belhadj, B.; Temam, O.; Héliot, R. Hardware Spiking Neurons Design: Analog or Digital. In Proceedings of the 2012 International Joint Conference on Neural Networks, Brisbane, Australia, 10–15 June 2012; pp. 1–5. [CrossRef]
9. Yang, J.Q.; Wang, R.; Ren, Y.; Mao, J.Y.; Wang, Z.P.; Zhou, Y.; Han, S.T. Neuromorphic Engineering: From Biological to Spike-Based Hardware Nervous Systems. *Adv. Mater.* **2020**, *32*, 2003610. [CrossRef]
10. Indiveri, G.; Linares-Barranco, B.; Hamilton, T. Neuromorphic Silicon Neuron Circuits. *Front. Neurosci.* **2011**, *5*, 73. [CrossRef]
11. Yang, X.; Xiong, Z.; Chen, Y.; Ren, Y.; Zhou, L.; Li, H.; Zhou, Y.; Pan, F.; Han, S. A Self-Powered Artificial Retina Perception System for Image Preprocessing Based on Photovoltaic Devices and Memristive Arrays. *Nano Energy* **2020**, *78*, 105246. [CrossRef]
12. Lee, C.L.; Hsieh, C.C. A 0.5 V/1.8 V High Dynamic Range CMOS Imager for Artificial Retina Applications. *IEEE Sens. J.* **2015**, *15*, 6833–6838. [CrossRef]
13. Yu, Z.; Liu, J.K.; Jia, S.; Zhang, Y.; Zheng, Y.; Tian, Y.; Huang, T. Toward the Next Generation of Retinal Neuroprosthesis: Visual Computation with Spikes. *Engineering* **2020**, *6*, 449–461. [CrossRef]
14. Cruz-Albrecht, J.M.; Yung, M.W.; Srinivasa, N. Energy-Efficient Neuron, Synapse and STDP Integrated Circuits. *IEEE Trans. Biomed. Circuits Syst.* **2012**, *6*, 246–256. [CrossRef] [PubMed]
15. Emelyanov, A.V.; Nikiruy, K.E.; Serenko, A.V.; Sitnikov, A.V.; Presnyakov, M.Y.; Rybka, R.B.; Sboev, A.G.; Rylkov, V.V.; Kashkarov, P.K.; Kovalchuk, M.V.; et al. Self-adaptive STDP-based learning of a spiking neuron with nanocomposite memristive weights. *Nanotechnology* **2019**, *31*, 045201. [CrossRef] [PubMed]
16. Sourikopoulos, I.; Hedayat, S.; Loyez, C.; Danneville, F.; Hoel, V.; Mercier, E.; Cappy, A. A 4-Fj/Spike Artificial Neuron in 65 nm CMOS Technology. *Front. Neurosci.* **2017**, *11*, 123. [CrossRef] [PubMed]
17. Morris, C.; Lecar, H. Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.* **1981**, *35*, 193–213. [CrossRef]
18. Danneville, F.; Loyez, C.; Carpentier, K.; Sourikopoulos, I.; Mercier, E.; Cappy, A. A Sub-35 Pw Axon-Hillock Artificial Neuron Circuit. *Solid State Electron.* **2019**, *153*, 88–92. [CrossRef]
19. Mead, C. *Analog VLSI and Neural Systems*; Addison-Wesley Longman Publishing Co.: Hoboken, NJ, USA, 1989.
20. Olsson, J.A.M.; Hafliger, P. Mismatch Reduction with Relative Reset in Integrate-and-Fire Photo-Pixel Array. In Proceedings of the 2008 IEEE Biomedical Circuits and Systems Conference, Baltimore, MA, USA, 20–22 November 2008; pp. 277–280. [CrossRef]
21. Ganguly, C.; Chakrabarti, S. A Leaky Integrate and Fire Model for Spike Generation in a Neuron with Variable Threshold and Multiple-Input–Single-Output Configuration. *Trans. Emerg. Telecommun. Technol.* **2019**, *30*, e3561. [CrossRef]
22. Bai, Y.; Fan, D.; Lin, M. Stochastic-Based Synapse and Soft-Limiting Neuron with Spintronic Devices for Low Power and Robust Artificial Neural Networks. *IEEE Trans. Multi Scale Comput. Syst.* **2018**, *4*, 463–476. [CrossRef]
23. Schuman, C.D.; Potok, T.E.; Patton, R.M.; Birdwell, J.D.; Dean, M.E.; Rose, G.S.; Plank, J.S. A Survey of Neuromorphic Computing and Neural Networks in Hardware. *arXiv* **2017**, arXiv:1705.06963.
24. Yao, E.; Basu, A. VLSI Extreme Learning Machine: A Design Space Exploration. *IEEE Trans. Very Large Scale Integr. Syst.* **2017**, *25*, 60–74. [CrossRef]

energies

MDPI

Article

Prediction of Gas Concentration Based on LSTM-LightGBM Variable Weight Combination Model

Xiangqian Wang *, Ningke Xu, Xiangrui Meng and Haoqian Chang

School of Computer Science and Technology, Anhui University of Science & Technology, Huainan 232000, China; nkxu999@gmail.com (N.X.); xrmeng@aust.edu.cn (X.M.); hqchang@aust.edu.cn (H.C.)
* Correspondence: xiqwang@aust.edu.cn; Tel.: +86-15309648996

Abstract: Gas accidents threaten the safety of underground coal mining, which are always accompanied by abnormal gas concentration trend. The purpose of this paper is to improve the prediction accuracy of gas concentration so as to prevent gas accidents and improve the level of coal mine safety management. Combining the LSTM model with the LightGBM model, the LSTM-LightGBM model is proposed with variable weight combination method based on residual assignment, which considers not only the time subsequence feature of data, but also the nonlinear characteristics of data. During the data preprocessing, the optimal parameters of gas concentration prediction are determined through the analysis of the Pearson correlation coefficients of different sensor data. The experimental results demonstrate that the mean absolute errors of LSTM-LighGBM, LSTM and LightGBM are 1.94%, 2.19% and 2.77%, respectively. The accuracy of LSTM-LightGBM variable weight combination model is better than that of the two above models, respectively. In this way, this study provides a novel idea and method for gas accident prevention based on gas concentration prediction.

Keywords: coal mine safety; LSTM; LightGBM; LSTM-LightGBM variable weight combination; gas concentration prediction

Citation: Wang, X.; Xu, N.; Meng, X.; Chang, H. Prediction of Gas Concentration Based on LSTM-LightGBM Variable Weight Combination Model. *Energies* **2022**, *15*, 827. https://doi.org/10.3390/en15030827

Academic Editors: Luis Hernández-Callejo, Adam Smoliński, Sara Gallardo Saavedra and Sergio Nesmachnow

Received: 15 November 2021
Accepted: 19 January 2022
Published: 24 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Energy is the engine of economic development and the lifeblood of national economy [1]. Coal is crucial with respect to the energy strategy of China, which is also caused by the feature of resource distribution in China, but also it determines that the solution to energy problems should depend on coal. For a long time, safety has always been one of the important issues during the process of coal mining. Gas accidents are a particularly serious problem. Through the investigation and analysis of coal mine gas accidents, it is found that not accurately grasping the law of gas concentration changes is the main reason for gas accidents [2]. Thus, if the inner rules can be explored and the gas concentration can be predicted relatively accurately [3], it will be of great importance to reduce the occurrence of gas accidents.

So far, many domestic and foreign scholars have conducted a great amount of research on gas concentration prediction [4]. Normally, gas concentration prediction methods can be broadly divided into two categories, one of which is using gas geomathematical modeling methods, and the other of which is based on machine learning methods. However, since the change of gas concentration is not a simple static process, and there are highly complex nonlinear relationship among its the influencing factors, it is still a great challenge for the current gas concentration prediction models to predict gas concentration accurately and efficiently [5].

The prediction of gas concentration using the gas geomathematical model requires detailed measurements of multidimensional attributes of the geological environment surrounding the mine and the underground environment, such as mining depth, permeability of coal seam, stability of coal seam and thickness of the coal seam. Wang et al. [6] constructed the gas concentration prediction equation based on one-dimensional regression

analysis. Zhang et al. [7] established the multivariate prediction model of gas concentration using the actual measured parameters of gas gushing from the mined area. Lu et al. [8] combined the gas gushing characteristics and gas gushing mechanism to construct a mathematical model of gas geology. However, based on the kind of methods for gas concentration prediction, it is not easy to obtain necessary input data, and not possible to achieve real-time prediction. Furthermore, in the process of model building, the prediction equation needs to be adjusted artificially based on experience, and it lacks the consideration of gas concentration time-series correlation.

As machine learning becomes more and more widely used in many fields, machine learning algorithms have been applied to gas concentration prediction. The previous studies focusing on prediction of gas concentration are mainly based on single factor, historical gas data or conventional single machine learning models such as the recurrent neural network (RNN) [9], eXtreme gradient boosting (XGBoost) model [10], the random forest model (RF) [11], backpropagation (BP) neural network [12] and long short-term memory (LSTM) network [13]. These algorithms have been used to predict the gas concentration in the short term. A comparison between the prediction values of gas concentration in several machine learning models demonstrated that LSTM network has a better generalization ability, and it can deal with nonlinear time sequence data on the basis of solving the defect of traditional recurrent neural network [14]. The light gradient boosting machine (LightGBM) [15] operates faster and it is accurate compared with that of XGBoost in the multiple benchmarks and public data set test. To further improve the precision of gas prediction, a few researchers have attempted to predict the gas concentration by combining several single machine learning models. Xun et al. [16] constructed a CNN-LSTM model. Lin et al. [17] combined PSO-BP neural network to predict the gas content of coral beds. Wen et al. [18] developed a BP neural network model based on Gray theory. Xu et al. [19] developed a IGSA-BP combination prediction model that had a better prediction accuracy than that of the single machine learning model. Zhang et al. [20] constructed a prediction model based on a combination of wavelet noise reduction and LSTM. Han et al. [21] constructed a gas concentration residual correction model based on Markov model and Gray neural network. However, majority of the combination models place the first prediction results into another model for the secondary prediction or sum up the prediction results of the two models to utilize the average value. Combination models that adopt this strategy do not "integrate" two single-machine models; this also results in their prediction accuracy still not meeting the needs of underground coal mine safety production.

Considering the drawbacks of the abovementioned studies, in this paper, the historical data of this survey site was selected as the time sequence factor, and the historical data of other survey sites at the working face was selected as a spatial topological factor, and these were combined. An analysis of the correlation between the attribute data and gas concentration is used to define the attribute requirements of the input data. According to the data time sequence and nonlinear characteristics, the variable weight combination model [22] of the LSTM network and the LightGBM model was developed to dynamically predict the gas concentration for the next 10 h. The model conquers the difficulty in obtaining data and inability to predict in real time by traditional gas geomathematical models and improves the accuracy of gas concentration prediction using the improved variable weight combination method of residual weighting. The prediction of gas concentration change trend can be as an important reference for safety management in coal mines to take measures such as gas extraction, water misting, boosting wind speed and other methods in time to ensure a better prevention of gas accidents.

2. Date Source

Since coal is the main source of energy in China, the safety problems related to coal mining have attracted significant attention. A large volume of gas gush is generated in the working face of the gas mine during the process of the production. By referring to the pre-decessor's data collection scale when predicting the gas concentration, [23,24] in this

study, 10,000 sets of data were collected from 11 different survey sites at the working face of a coal mine in Shanxi Province from 19 March 2021 to 24 March 2021. The description of data attribute is shown in Table 1.

Table 1. Data attribute description of each measuring point at a working face.

Measurement Point Name	Measurement Point Description	Index	Max Value	Min Value
MGas	Mixed methane concentration in air entry	% CH_4	0.7	0
EGas	Methane concentration of back air in air inlet drift	% CH_4	0.7	0
Gas1	Methane concentration in the downwind side of the tunnel	% CH_4	0.79	0.16
Gas2	Methane concentration in working face of air entry	% CH_4	0.4	0
YCO1	Concentration of carbon monoxide in the downwind side of tunnel drilling	ppm	6	0
YCO2	Concentration of carbon monoxide at the head of the belt conveyor in the air inlet lane	ppm	6	0
WS	Back air speed in air entry	m/s	1.2	0.2
FC	Dust on working face of air entry	mg/m^3	0	0
ET	Back air temperature in air entry	°C	13.3	10.8
GD	Mixed instantaneous flow in air inlet pipeline	m^3	19.29	0
SM	Smoke on the downwind side of the head of the belt driven into the air entry	mg/m^3	0	0

2.1. Missing Data Processing

Due to various force majeure factors in the data collection, transmission and storage scenarios, some data can be missing. Missing data can cause serious impediments to subsequent data correlation analysis and the construction of gas concentration prediction models. In addition to reducing the validity of the data, it can also lead to inaccuracies in the overall data analysis task and produce incorrect analysis results. Hence, this paper adopts the average method to fill in the missing data. The data filling equation is given as follows:

$$\widetilde{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

In the above formula, $\widetilde{x}$ represents the missing data series, $\sum_{i=1}^{n} x_i$ represents the total of all data in the data set and n represents the number of nonmissing data in the data set.

2.2. Normalization Process

In order to eliminate the impact of the dimensionality between the gas multiparameter time series, it is necessary to perform data normalization. Following data normalization of the raw data, the indicators are in the same order of magnitude and suitable for comprehensive comparative evaluation. Meanwhile, normalization provides a certain degree of numerical comparability of features among different dimensions. The original time series x

is normalized by applying min–max normalization. The normalization formula is given as follows:

$$x* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

where x^* is the normalized value, $x_{\max}$, $x_{\min}$ are the maximum and minimum values of the sample data respectively.

2.3. Feature Selection

After the data have been preprocessed, it is necessary to select meaningful features to input into the machine learning algorithms and models for training. Generally, feature selection is divided into the following two main steps:

2.3.1. Correlation Analysis

In order to fulfill the requirements of gas concentration prediction and to strengthen the situational awareness and extrapolation capability of the prediction model, in this paper, we use the Pearson correlation coefficient to describe the degree of correlation between gas concentration at the working face and its impact factors. The equation is given as follows:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{3}$$

In the above equation, $\rho_{X,Y}$ represent the Pearson correlation coefficient of two continuous variables X, Y, $\operatorname{cov}(X, Y)$ represents the covariance between them, and σ_X and σ_Y represent the standard deviations of the variables X and Y.

2.3.2. Eliminate Redundant Features

Using the Pearson correlation coefficient to obtain the weights of each feature, the features with weights less than a threshold value are eliminated. Afterward, the mutual information is calculated for the features in the remaining data set two by two. Mutual information refers to the extent of information shared between two features. If the value of mutual information is greater than the threshold, the feature with the smaller weight is considered redundant and is removed. The equation for calculating mutual information is given as follows:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{4}$$

In the above formula, $p(x,y)$ is the joint probability distribution function of X and Y, and $p(X)$ and $p(Y)$ are the marginal probability density functions of X and Y.

3. Materials and Methods

3.1. LightGBM

XGBoost should be defined before explaining about LightGBM [25], XGBoost is an improved boosting algorithm of the gradient boosting decision tree (GBDT), which is GBDT in essence, but it strives to maximize the speed and efficiency. Conventional GBDT adopts classification and regression tree (CART) as the base classifier, and XGBoost supports the multiple base classifiers to compensate for the shortcoming in the accuracy of single CART prediction. However, the disadvantages associated to XGBoost are that it stores feature sorting results, which occupy a massive amount of memory, and it severely affects cache optimization.

Compared with that of XGBoost, LightGBM [26] is a relatively new tree-based gradient boosting variant. It adopts the histogram algorithm to ensure that an algorithm utilizes less memory and has a low computational cost. Layer-by-layer growth is a conventional method used for tree based combination (including XGBoost) growth decision trees. LightGBM is different from that of XGBoost, as it does not utilize the conventional decision tree growth strategy and it introduces leaf-by-leaf growth strategy. In contrast to layer-by-layer growth,

leaf-by-leaf growth strategy converges faster and consumes lesser memory. Layer-by-layer growth strategy and leaf-by-leaf growth strategy are shown in Figure 1.

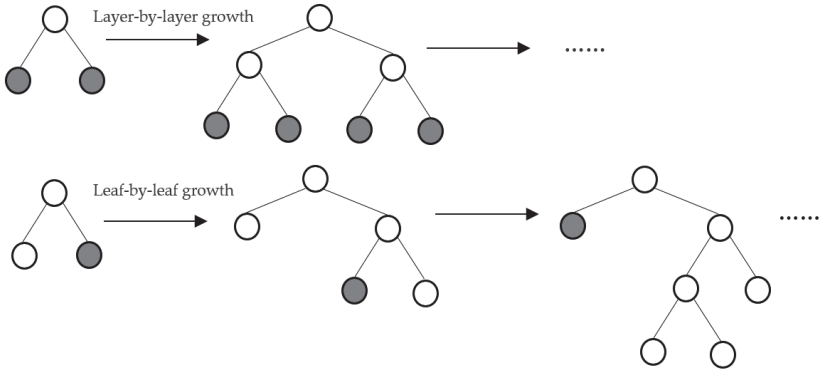


Figure 1. Layer-by-layer growth and leaf-by-leaf growth.

3.2. LSTM

LSTM [27] consists of a set of cyclic subnetworks named according to the memory blocks. Each memory block consists one or multiple self-connected memory cells and three gating units: input gate, output gate, and forget gate. Similar to that of the recurrent neural network (RNN), the hidden unit is horizontally connected back to the hidden unit. However, the hidden unit of RNN is replaced by the memory cell with gating function. The diagram of LSTM structure of a single cell is shown in Figure 2.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\widetilde{C}_t = \tan h(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t \times C_{t-1} + i_t * \widetilde{C}_t \tag{8}$$

$$O_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = O_t \times \tan h(C_t) \tag{10}$$

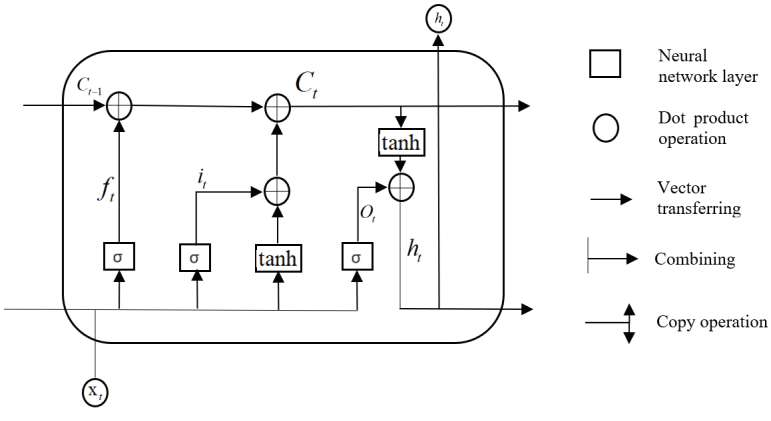


Figure 2. LSTM structure diagram.

In the above formula, f_t represents the forget gate. It is used to control whether or not to filter the hidden cellular state of the upper layer in the LSTM. i_t represents the input gate, $\widetilde{C}_t$ is the cell state at the previous moment, C_t is the cell state at the present moment, O_t represents the output gate, x_t and h_t represent the input and output at the current moment and σ and tanh represent the sigmoid function and hyperbolic tangent function, respectively. The forget gate, input gate, output gate and the weight matrix of the cell state are represented by w_f, w_i, w_o and w_c respectively. b_f, b_i, b_o and b_c represent the offset vector of the forget gate, input gate, output gate and cell state, respectively.

3.2.1. Activation Function

The sigmoid function is used as the activation function for the forgetting, input and output gates in the LSTM. The tanh function is used as the activation function when generating candidate memories. Both are saturated functions. If a nonsaturated activation function is used, the past and present memory blocks will be superimposed all the time, resulting in memory misalignment and making it difficult to achieve the gating effect [28].

Sigmoid is a commonly used activation function in gating structures. It compresses the values to between 0 and 1, which can help update and forget information. In fact, sigmoid activation function is the common choice for almost all modern neural network modules in gating.

Tanh activation function is used to generate candidate memories. This is due to the fact that tanh function has a larger gradient than the sigmoid function, which makes the model converge faster. Likewise, if a nonsaturated activation function is used to generate the candidate memory, it is likely that the output values may explode or the gradient may disappear. Hence, in this paper, we choose tanh activation function as the activation function.

3.2.2. Overfitting

High fit is a key sign of a good model. However, in the process of model fitting, if the pursuit of high *R*-squared is pursued, it is likely that some of the characteristics of the training sample itself will be taken as general properties that all potential samples will have. As a result, this can lead to a reduction in the generalization performance of the model. This phenomenon is called "overfitting" in machine learning and cannot be completely avoided in model training. All we can do is "reduce the risk", and currently, there are several ways to prevent model overfitting:

1. Data enhancement: Employing more data for model training helps to better identify signals and avoid identifying noise as signals.
2. Pretermination: Pretermination prevents overfitting by stopping the iteration of the model before it converges on the training data set.
3. Regularization: Regularization refers to the process of optimizing the objective function or cost function by adding a regular term after the objective function or cost function, typically L1 regular or L2 regular, etc.
4. Dropout: Dropout is implemented by randomly "removing" the hidden units from the neural network after the model training has started.

In order to prevent overfitting of the LSTM model in this paper, pretermination and the addition of a dropout layer are used. First, by recording the best validation accuracy so far during the training process, when after five consecutive iterations, no better validation accuracy is produced, then we can terminate the model early by default. Furthermore, we add a dropout layer to the model to reduce the complex coadaptation between neurons. Once the hidden layer neurons are randomly removed, the fully connected network is sparse, which can effectively reduce the synergistic effect of different features and enhance the generalization ability of the model. Due to the addition of the dropout layer, the model has a certain randomness in prediction, so the 10 predictions of the LSTM model are taken and averaged as the final prediction result.

3.3. Grid Search Algorithm

A reasonable set of model parameters is the basis for building a good model, and the impact of hyperparameters on the effectiveness of the model is crucial. The grid search algorithm refers to an exhaustive list of parameter values. By combining the values determined by the range of values for each parameter and the search step, a "grid" is generated by listing all possible results. Subsequently, the combinations are used to train the model, and an optimal combination of parameters is returned after all combinations have been tried.

3.4. Improved Variable Weight Combination Model

During the gas concentration prediction performed by the conventional combination model, different models are adopted to predict the gas concentration with the same working face. The appropriate weights are assigned to the prediction values, and then combined. The combined prediction model can reduce the effect of random factors of the single forecasting model and effectively improve the prediction precision.

In this study, LSTM-LightGBM equal weight combination model, LSTM-LightGBM residual weight combination model, and improved LSTM-LightGBM variable weight combination model were developed.

3.4.1. Development of Single Machine Learning Model

Ensuring the prediction accuracy and performance of single machine learning model is the basis of determining the combination model—specifically, based on previous research and parameter comparison between LSTM neural network models. Using the grid search algorithm mentioned in Section 3.3 for hyperparameter search optimization of the LSTM model, it is determined that the search range of the first layer cell count is from 20 to 200 with a search step of 20, the search range of the second layer cell count is from 10 to 100 with a search step of 20 and the number of iterations is set to 10 to 40 with a search step of 10. The layer of the network model was set to 2. The activation probability of the dropout layer was set to 0.2, the number of the unit in the first layer was set to 100, the number of units in the second layer was set to 50 and the activation function was set to Tanh. The optimization algorithm adopted the Adam algorithm, and the iteration number was set to 20 times.

Grid search algorithm [29] was used to optimize the superparameter of LightGBM model. The final parameters of the model were set as: max _depth = 6, learning _rate = 0.2, n _estimators = 180, subsample = 0.6, colsample _by tree = 0.85, silent = True.

3.4.2. Weighing of the Residual Combination Model

It is a common method to provide a single model a proper weight to develop the combination model under the condition that the accuracy of the single machine learning model remains the same. This can improve the accuracy of the model [30]. The most extensively used weighting method is equivalent weighting. In general, the method of equivalent weighting is simple, and it has a good universality and participation. However, it does not reflect the importance that the model attaches to the prediction results of different single models, and it is possible that the determined weight is considerably different from that of the actual importance of the prediction results. The residual weighting combination model is expressed as:

$$h(x_t) = \sum_{i=1}^{m} \omega_i(t-1) f_i(x_t) \tag{11}$$

$$\omega_i(t-1) = \frac{\frac{1}{\overline{\varphi_i}(t-1)}}{\sum_{i=1}^{m} \frac{1}{\overline{\varphi_i}(t-1)}} \tag{12}$$

$$\sum_{i=1}^{m} \omega_i(t-1) = 1, \omega_i(t-1) \geq 0 \tag{13}$$

where $w_i(t-1)$ is the weight of the ith model at the moment of $t-1$, $f_i(x_t)$ is the prediction value of the ith model, $h(x_t)$ is the prediction value of combination model, $\overline{\varphi}_i(t-1)$ is the square sum of the predictive errors of ith model at the moment of $t-1$. The central idea of residuals weighting is to assign the weight to describe the importance of the model based on the error between the prediction value and the real value.

3.4.3. Weighting of Improved Variable Weight Combination Model

Compared with that of the conventional prediction method, there are a few improvements in data input dimension in this study. Conventional gas concentration prediction models only adopt the single dimension input model. The improved algorithm proposed in this study adopts multidimension input method based on data correlation analysis. It reveals the constraint of the single dimension input model, and it provides a theoretical basis to explore the relationship between other compounds and gas concentration.

LSTM-LightGBM variable weight combination model was developed using the improved variable weight combination method based on residual weight. The residual weighting model was improved based on weight of the moments obtained in Formula (12), and the optimal m value was calculated. The average of the weights of the first m moment was used for the initial weighting. The expression for the initial weighting is:

$$\omega_j(t) = \frac{1}{m} \sum_{k=1}^{m} \omega_i(t-k) \ (m = 6), \tag{14}$$

After gaining the weight of the models from Formulas (12) and (14), the absolute value of the error between the predicted value and the true value of each combination model at the moment of t is calculated as $\delta_{i,t}$ and $\delta_{j,t}$.

$$\delta_{i,t} = \sum_{i=1}^{m} \omega_i(t) f_i(x_t) - \widehat{f(x_t)} \tag{15}$$

$$\delta_{j,t} = \sum_{i=1}^{m} \omega_j(t) f_j(x_t) - \widehat{f(x_t)} \tag{16}$$

The values of $\delta_{i,t}$ and $\delta_{j,t}$, are compared. If $\delta_{i,t} < \delta_{j,t}$, the new weight $w_j(t)$ of the combination model will replace the previous weight $w_i(t)$. Otherwise, the previous weight will remain unchanged.

3.5. Construction Flow of Prediction Model

The construction flow of the prediction model is shown in Figure 3. The main processes include data preprocessing, prediction of the single machine learning model, construction of the variable weight combination prediction model and the evaluation and analysis of the model prediction [31].

(1) Data preprocessing: Data preprocessing is an important link before data modeling, which fundamentally determines the quality of the data work and the output value. The data in this study was obtained from the working face of a coal mine in Shanxi Province. The data is relatively complete. Therefore, the data are directly normalized. The data attribute and the data correlation are considered and the suitable data from the data set is selected for the model training.

(2) Development of single machine learning model: After the data set is divided according to the scale of the training set:verification set:testing set = 7:2:1, the LSTM model and LightGBM model are trained by the data of the training set, and the data of verification set is used to adjust the parameters and monitor if the model has been fitted. The data

of the test set are placed into two models, respectively, and the prediction results of the single machine learning model are obtained.

(3) Development of improved variable weight combination model. The weight of each single machine learning model is determined by the improved weighting method shown in Section 3.4.3 to ensure that the improved prediction model can be obtained.

(4) Model evaluation analysis: According to the indexes of the model evaluation, the prediction ability of the improved model was compared and the change in the prediction effect of the model is analyzed.

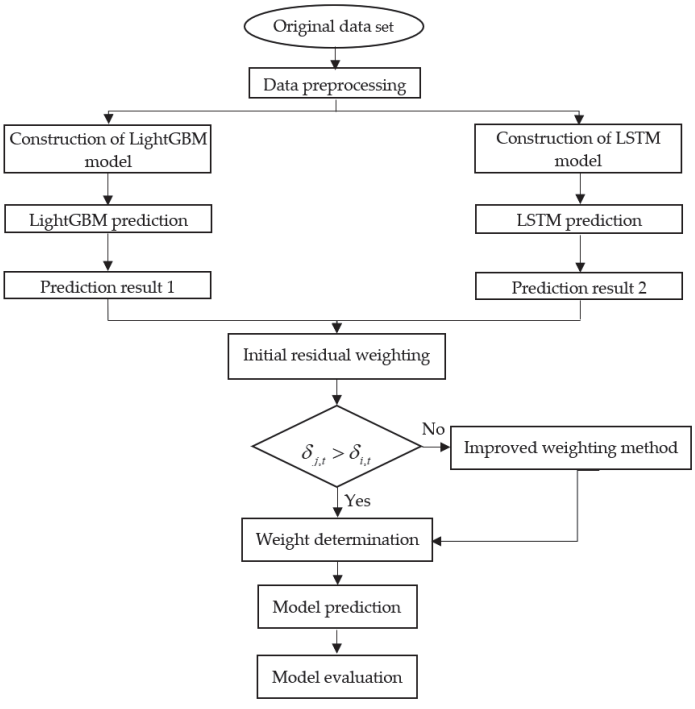


Figure 3. Prediction flow of LSTM-LightGBM variable weight combination model.

3.6. Evaluation Index

The mean absolute percentage error (*MAPE*) is not applicable because the actual value of the data used in this study includes zero. Therefore, the evaluation index used in this study is root mean square error (*RMSE*) and mean absolute error (*MAE*). The formula is as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\widehat{y}_i - y_{pre})^2} \quad (17)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|(\widehat{y}_i - y_{pre})\right| \quad (18)$$

In the formula, m is the number of samples, $\widehat{y}_i$ is the true value, y_{pre} is the forecast results. The actual value will be closer to the predicted value if the value of the loss function is smaller, and this ensures a higher accuracy of the model prediction.

When there is a certain amount of error in the prediction, the value of the root mean square error will also be larger, so the root mean square error is used to characterize the degree of dispersion of the error value. As the error values of the mean absolute error are

absolutized, there is no situation where the positive and negative errors in the mean error cancel each other out. Thus, the mean absolute error can better reflect the actual situation of the prediction errors.

4. Results

4.1. Prediction Factor Analysis

There are multiple transformations and interactions between the gas mixture and other compounds at different measuring points [32]. Therefore, the correlation between the concentration of the gas mixture and other compounds is analyzed.

In statistics, the Pearson product–moment correlation coefficient (PPMMC) [33] is used to measure the correlation between variables. To avoid experimental uncertainties, data from three different coal mines were selected for correlation analysis, and the visualization of the correlation between the mixed gas concentration and the data was determined using heat diagram.

As shown in Figure 4, the "FC" data in this working face are zero, and a correlation with the mixed gas concentration was absent. There is a strong correlation between "EGas", "Gas1", "Gas2" and the mixed gas. However, by calculating the values of mutual information between "EGas", "Gas1" and "Gas2", we found that "EGas" has the largest mutual information value and is greater than the threshold value, so it can be considered that "Gas1" and "Gas2" are redundant features; thus,"Gas1 "and "Gas2" are not used as input data.

Figure 4. Correlation analysis of data.

The four variables "EGas, WS, ET and GD" were selected as the input of the prediction model, and the correlation analysis between the input variables and the mixed gas concentration is shown in Figure 5. According to previous experiments conducted on methane adsorption, an increase in temperature can reduce the gas adsorption capacity and it can effectively promote the rapid desorption and diffusion. Meanwhile, the activity of the methane molecule increases, which promotes the pore expansion of coal bodies, particularly of the small gaps. This significantly improves the methane diffusion of coal bodies. The diffusion coefficient dynamically changes with an increase in the temperature. In this study, the least squares method was used for fitting, as shown in Figure 5a. A positive correlation

between the concentration of mixed gas and the ambient temperature was observed, which revealed the mechanism of the dynamic process of gas diffusion proposed by Liu [34] et al.

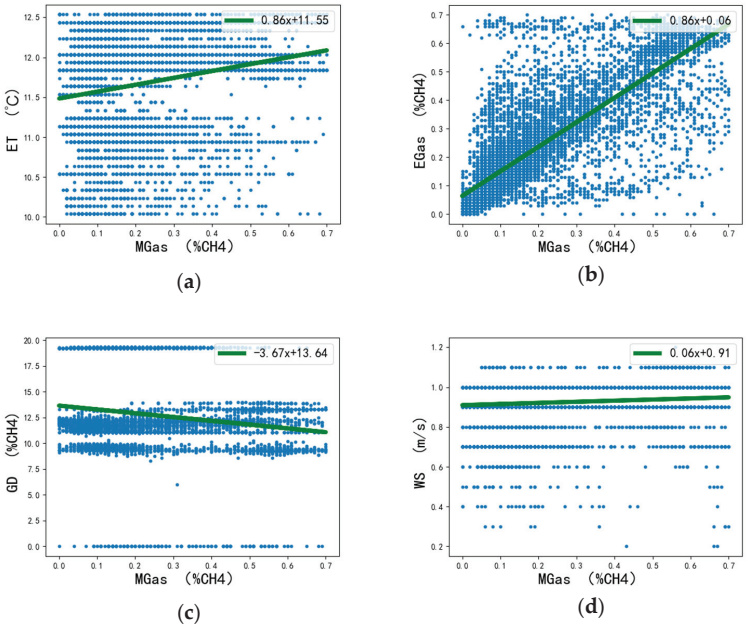


Figure 5. Correlation analysis between gas concentration and input data: (**a**) Scatter plot of correlation between mixed methane concentration and ambient temperature; (**b**) Scatter plot of correlation between mixed methane concentration and back air methane concentration; (**c**) Scatter plot of correlation between the concentration of mixed methane and the instantaneous flow of pipeline; (**d**) Scatter plot of correlation between mixed methane concentration and working velocity and back air.

In this study, the back air methane concentration and mixed methane concentration exhibited a stronger correlation. The back air pipe is mainly used to receive the air flow after cleaning the working face, and a large volume of gas will be produced during the process of production at the mine working face. At the working face, the main gas sources are the falling coal gas emission and coal wall gas. Different gas sources follow different rules of gas emission [35].

4.1.1. Law of Falling Coal Gas Emission

The coal body will crack during the process of mining, causing a change in gas occurrence conditions. A large volume of gas changes into a free state from the adsorption state, and it might enter into the tunnel with the air flow. The volume of falling coal gas emission is closely related to falling coal, the falling coal fragmentation, the content of coal seam gas and residual gas. The intensity of coal falling gas emission is shown in Formulae (19) and (20).

$$q_1 = \frac{q_{10}}{(1+t)^{\alpha}} \tag{19}$$

$$Q_1 = \int_0^T q_1 \theta M dt \tag{20}$$

In the function, q_1 represents the emission intensity per weight of falling coal gas at unit time of $t+1$, unit is $m^3/(\text{min}.t)$. q_{10} represents the intensity of gas emission at initial moment of falling coals with the unit of $m^3/(\text{min}.t)$. t represents the exposure time of

falling coals with the unit of min. α is the attenuation coefficient, Q_1 is the absolute gas emission from falling coals in the process of mining with the unit of m^3/min. M represents the mining weight per unit time with the unit of t/min. θ is the degree of fragmentation.

4.1.2. Law of Coal Gas Emission of in Working Face

The gas released from the coal enters the air stream through the surface of the coal wall according to Duthie's law and the law of diffusion. During the process of continuous mining, fresh coal wall is constantly exposed, mining pressure constantly changes, and the gas pressure balance state near the working face changes. A large volume of gas flow out along the coal cracks and pores gushing lane, the gushing intensity of the coal wall gas is shown in Formulas (21) and (22).

$$q_2 = \frac{q_{20}}{(1+t)^{\beta}} \tag{21}$$

$$Q_2 = \int_0^T q_2 H v dt \tag{22}$$

In this function, q_2 represents gas emission intensity of back coal wall at the time of t + 1 with the unit of $m^3/(\text{min}.m^2)$. The q_{20} is gas emission intensity at the initial moment of coal wall with the unit of $m^3/(\text{min}.m^2)$. t is the exposure time of coal wall, with the unit of min. β is the attenuation coefficient, Q_2 is absolute emissions of coal wall gas in the process of mining with the unit of m^3/min. H is the thickness of coal mining layer with the unit of m. v is the cutting speed of coal mining machine with the unit of m/min.

After entering the lane from the above gas source, methane will form a mixture of gas and air with uneven concentration, and the mixture will migrate by concentration diffusion and convection mixing in the airflow. After fresh air flow passes through the working face of mines, partial methane gas in the mining face is diluted and carried. Therefore, the methane concentration in the back air can accurately reflect the change in the methane concentration in the mining face.

4.2. Model Prediction Analysis and Comparison

To verify the accuracy of the improved LSTM-LightGBM, the LSTM, LightGBM, XGBoost, LSTM-LightGBM (Equivalent weighting) and LSTM-LightGBM (Residual weight) were selected for comparative experiments. The errors of the different models were compared as shown in Figure 6.

From the figure above, it can be observed that the prediction accuracy of the variable weight combination model is higher than that of the single machine learning model and the conventional combination weighting model. The comparison between the values of *MAE* and *RMSE* of the models is shown in Table 2.

Table 2. Comparison between evaluation indexes of each model.

Model	*MAE*	*RMSE*
LSTM	0.0219	0.0306
LightGBM	0.0277	0.0377
XGBoost	0.0253	0.0352
LSTM-LightGBM (Equivalent weighting)	0.0214	0.0276
LSTM-LightGBM (Residual weighting)	0.0201	0.0279
LSTM-LightGBM (Variable weight combination)	0.0194	0.0261

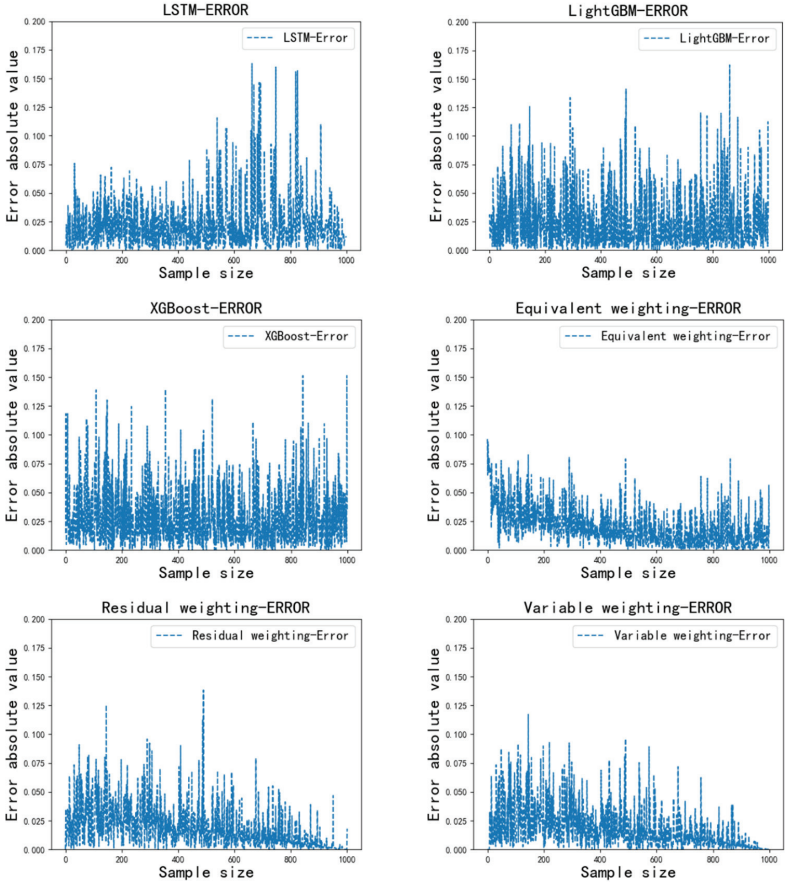


Figure 6. Prediction result and real results of each model.

The *MAE* and *RMSE* values of LSTM model were the average value of the LSTM model which were trained ten times. After the analysis, the *MAE* and *RMSE* values of the improved LSTM-LightGBM variable weight combination model were increased by 3.5% and 6.5%, respectively, compared with that of the LSTM-LightGBM residual weight combination model, and by 11.4% and 14.7%, respectively compared with that of the LSTM-LightGBM single machine learning model. The improved variable weight combination method has a higher prediction accuracy.

4.3. Model Universality Analysis

During the selection of study area, strong local features were observed at different working faces of the coal mine at different locations. To verify the universality of the algorithm, the prediction and analysis of gas concentration were performed in different coal mines. The coal mines selected were Mine A in Shanxi, Mine B in Guizhou, and Mine C in Anhui.

It can be observed from Figure 7 that the prediction error of the modified variable weight combination model is smaller than that of the conventional model, and the increase in Mine A is the most obvious. *MAE* value increased by 18.5% and 29.2%, respectively, compared with that of the LSTM model and the LightGBM model. *RMSE* increased by 22.9% and 30.4%, respectively, compared with the LSTM model and the LightGBM

model. Therefore, the prediction results of the improved variable weight combination model with three different coal mine gas concentrations demonstrated that the prediction accuracy was improved. This demonstrates the universality of the improved variable weight combination model.

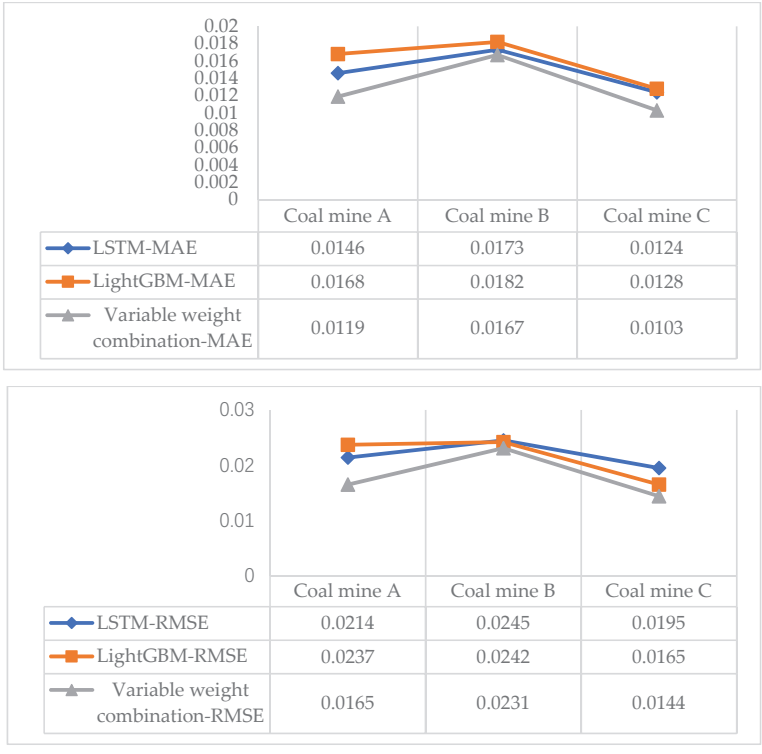


Figure 7. Analysis of evaluation index applied to different coal mines.

5. Discussion

In this study, a variable weight combination model was developed by adopting the methane concentration, wind speed, ambient temperature, gas drainage, and the historical data of mixed gas. Working faces of different mines were selected to predict the gas concentration in the future 10 h. In the improved LSTM-LightGBM variable weight combination model, the *MAE* value and *RMSE* value were 0.0194 and 0.0261, respectively. These values were smaller than that of the prediction values of 0.0224 and 0.0317 obtained in the ARIMA model proposed by Zhang et al. [36] and the 0.0207 and 0.0303 of S-GRU model proposed by Chang et al. [37]. This was because an LSTM neural network with better time sequence prediction and the LightGBM model with better performance in the nonlinear model were predicted in the form of variable weight combination. It considered the time sequence feature of the data and the nonlinear feature of data. For analysis and comparison result of the gas concentration, the improved LSTM-LightGBM variable weight combination model was better than that of the conventional LSTM-LightGBM equivalent weight assignment model and LSTM-LightGBM residual weight assignment model. Considering the difference in prediction error between the LSTM network and the LightGBM model at different moments, the combination model adopted different weights for the prediction values at different moments to combine the advantages of both the models.

In this study, data from coal mine at several locations were selected to explore the performance of the regional model. Additionally, downhole temperature, wind speed and methane gas were selected as prediction factors to determine the effect of factors for gas concentration prediction [38]. To improve the prediction accuracy of gas concentration, suitable factors such as weather and ground surface temperature, depth of coal seam, inclination of coal bed, top and bottom lithology of coal bed should be considered in the future.

6. Conclusions

Based on LSTM and LightGBM model with the variable weight combination model, the prediction method of gas concentration was improved. In this model, the time sequence feature and the nonlinear relationship between the input feature and gas concentration were considered. By the data pre-processing and feature selection, it makes the model converge faster and avoids the degradation of prediction accuracy due to redundant features. Sigmoid function is selected for the activation function of the gate structure of the LSTM model. Tanh activation function is selected to generate candidate memories. These gates increase the convergence speed of the model. Moreover, they guarantee that the model does not suffer from the problem of exploding output values and vanishing gradients. In comparison to traditional single machine learning gas concentration prediction models, LSTM models have a higher prediction accuracy.

Compared with that of single machine learning model and other conventional combination weighting models, the prediction result of the variable weight combination model was closer to that of the real value with a small error. It provides better prediction accuracy, and high reliability. It can give a reference for gas accidents prevention and promote the safety of coal mines.

This study focused on the prediction and analysis of gas concentration using the underground attribute information only including temperature, wind speed, methane gas. Nevertheless, the change of gas concentration is affected by complex factors and conditions [39]. In future research, it is important for us to consider more comprehensive factors of gas concentration, such as roof pressure, minging depth, inclination angle of coal seam and ground weather information.

Author Contributions: Conceptualization, X.W. and N.X.; methodology, N.X.; software, N.X.; validation, X.M. and X.W.; formal analysis, H.C.; investigation, X.W.; resources, X.W.; data curation, N.X.; writing—original draft preparation, N.X.; writing—review and editing, X.W.; visualization, H.C.; supervision, X.W.; project administration, X.M.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Natural Science Foundation of China (51874003, 51474007), Academic Funding Projects for Top Talents in Disciplines and Majors of Anhui(gxbjZD2021051).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository. The partial data presented in this study are openly available in [Gas concentration prediction data set, Mendeley Data], doi:10.17632/p3n7k6hxgw.1.

Acknowledgments: Many people have offered me valuable help in my thesis writing, including my students, my family and the National Natural Science Foundation of China. It is of great help for me to finish this article successfully.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Teng, J.; Qiao, Y. Analysis of coal demand, exploration potential and efficient utilization in China. *Chin. J. Geophys.* **2016**, *59*, 4633–4653.
2. Cheng, J.; Bai, J.Y. Short-term prediction of mine gas concentration based on chaotic time series. *J. China Univ. Mine. Technol.* **2008**, *02*, 231–235.
3. Deng, G. Current status and prospects of coal and gas outburst prediction and prevention technology. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *651*, 32096. [CrossRef]
4. Fu, H.; Liu, Y.Z. Prediction of gas concentration based on multi-sensor-deep long and short time memory network fusion. *J. Sens. Technol.* **2021**, *34*, 784–790.
5. Lai, X.W.; Xia, Y.N. Improved grey gas concentration series prediction based on ensemble learning. *China Work. Saf. Sci. Technol.* **2021**, *17*, 16–21.
6. Wang, Y.S. Mathematical models in the study of gas gush prediction in mines. *Coal Technol.* **2015**, *263*, 185–187.
7. Zhang, Z.X.; Yuan, C.F. Research on prediction of gas gush in mines by gas geological mathematical model method. *J. China Coal Soc.* **1999**, *4*, 34–38.
8. Lu, X.L.; Fu, X.M. Study on the geological pattern of gas in Kongzhuang coal mine. *Coal Sci. Technol.* **1997**, *2*, 73–76.
9. Song, S.; Li, S.G. Research on a multi-parameter fusion prediction model of pressure relief gas concentration based on RNN. *Energies* **2021**, *14*, 1384. [CrossRef]
10. Zhang, D.Y.; Gong, Y. The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. *IEEE Access* **2020**, *8*, 220990–221003. [CrossRef]
11. Wen, T.X.; Zhang, B. A random forest model for coal and gas protrusion prediction. *Comput. Eng. Appl.* **2014**, *50*, 233–237.
12. Yin, G.Z.; Li, M.H. An improved BP neural network-based model for predicting gas permeability in coal bodies. *J. Coal* **2013**, *38*, 1179–1184.
13. Zhang, T.J.; Song, S. Research on gas concentration prediction models based on LSTM multidimensional time series. *Energies* **2019**, *12*, 161. [CrossRef]
14. Li, W.S.; Wang, L. Application and design of LSTM in coal mine gas prediction and warning system. *J. Xi'an Univ. Sci. Technol.* **2018**, *38*, 1027–1035.
15. Sun, Q.M.; Qu, Z.J. Situation-aware multimodal transport recommendation based on particle swarm optimization and LightGBM. *J. Electron.* **2021**, *49*, 894–903.
16. Xun, X.X.; Su, C. CNN-LSTM based coal mine gas concentration prediction. *Mod. Inf. Technol.* **2020**, *4*, 149–150.
17. Lin, H.F.; Gao, F. PSO-BP neural network prediction model for coal seam gas content and its application. *Chin. J. Saf. Sci.* **2020**, *30*, 80–87.
18. Wen, J.Q.; Zhang, Y. Prediction of gas content based on gray theory-BP neural network. *Energy Technol. Manag.* **2020**, *45*, 44–45, 55.
19. Xu, Y.S.; Qi, C.Y. Prediction model of gas gushing based on IGSA-BP network. *J. Electron. Meas. Instrum.* **2019**, *33*, 111–117.
20. Zhang, X.J.; Liu, F. Gas concentration prediction in coal mines based on wavelet noise reduction and recurrent neural networks. *Coal Technol.* **2020**, *321*, 145–148.
21. Han, T.T.; Wu, S.Y. Gas concentration prediction based on Markov residual correction. *Ind. Min. Autom.* **2014**, *216*, 28–31.
22. Kang, J.F.; Tan, J.L. Short-term PM2.5 concentration prediction with the support of XGBoost-LSTM variable weight combination model—Shanghai as an example. *China Environ. Sci.* **2021**, *7*, 1–16.
23. Wang, Y.H.; Wang, S.Y. Research on multi-parameter gas concentration prediction model based on improved locust algorithm optimized long-short-term memory neural network. *J. Sens. Technol.* **2021**, *34*, 1196–1203.
24. Li, D.; Sun, Z.M. Research on AWLSSVM gas prediction based on chaos particle swarm. *Saf. Coal Mines* **2020**, *51*, 193–198.
25. Qiu, Y.G.; Zhou, J. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Eng. Comput.* **2021**, *21*, 1393–1399. [CrossRef]
26. Zhang, X. Ion channel prediction using Lightgbm Model. In Proceedings of the 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Fuzhou, China, 10–12 April 2020.
27. Shu, X.; Zhang, L. Host–parasite: Graph LSTM-in-LSTM for group activity recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 663–674. [CrossRef] [PubMed]
28. Zhang, H.; Zhang, Q. The development and properties of activation function are reviewed. *J. Xihua Univ.* **2021**, *1*, 10.
29. Hossain, M.Z.; Sohel, F.; Shiratuddin, M.F.; Laga, H. A comprehensive survey of deep learning for image captioning. *ACM Comput. Surv.* **2019**, *51*, 118. [CrossRef]
30. Cheng, T.J.; Wang, M. Trend prediction of online public opinion on unexpected events based on variable weight combination. *Comput. Sci.* **2021**, *48*, 190–195.
31. Wang, H.Q.; Liang, W. Scene mover: Automatic move planing for scene arrangement by deep reinforcement learing. *ACM Trans. Graph.* **2020**, *39*, 233. [CrossRef]
32. Peng, S.P.; Gao, Y.F. Theoretical discussion and preliminary practice of AVO detection of gas enrichment in coal seams—A case study of Huainan coalfield. *J. Geophys.* **2005**, *6*, 262–273.
33. Kong, L.; Nian, H. Fault detection and location method for mesh-type DC microgrid using pearson correlation coefficient. *IEEE Trans. Power Deliv.* **2021**, *36*, 1428–1439. [CrossRef]

34. Liu, Y.W.; Wei, J.P. The law and mechanism of temperature influence on the dynamic process of coal particle gas diffusion. *J. Coal* **2013**, *38*, 100–105.
35. Lyu, P.Y.; Chen, N. LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion. *Process Saf. Environ. Prot.* **2020**, *137*, 93–105. [CrossRef]
36. Zhang, Z.; Zhu, Q.J. Construction of ARIMA prediction model for gas concentration based on Python and its application. *J. North China Inst. Sci. Technol.* **2020**, *17*, 23–28.
37. Chang, L.; Zhang, H. An improved GRU gas concentration prediction model. *J. Heilongjiang Univ. Sci. Technol.* **2020**, *30*, 532–535.
38. Yu, G.F.; Fei, W. A compromise-typed variable weight decision method for hybrid multiattribute decision making. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 861–872. [CrossRef]
39. Wang, L.L.; Cao, Q.G. Research on the influencing factors in coal mine production safety based on the combination of DEMATEL and ISM. *Saf. Sci.* **2018**, *103*, 51–61. [CrossRef]

energies

MDPI

Article

Machine Learning Approach for Maximizing Thermoelectric Properties of BiCuSeO and Discovering New Doping Element

Nuttawat Parse [1], Chakrit Pongkitivanichkul [1] and Supree Pinitsoontorn [2,*]

[1] Department of Physics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand; p.nuttawat@kkumail.com (N.P.); chakpo@kku.ac.th (C.P.)
[2] Institute of Nanomaterials Research and Innovation for Energy (IN-RIE), Khon Kaen University, Khon Kaen 40002, Thailand
* Correspondence: psupree@kku.ac.th

Abstract: Machine learning (ML) has increasingly received interest as a new approach to accelerating development in materials science. It has been applied to thermoelectric materials research for discovering new materials and designing experiments. Generally, the amount of data in thermoelectric materials research, especially experimental data, is very small leading to an undesirable ML model. In this work, the ML model for predicting ZT of the doped BiCuSeO was implemented. The method to improve the model was presented step-by-step. This included normalizing the experimental ZT of the doped BiCuSeO with the pristine BiCuSeO, selecting data for the BiCuSeO doped at Bi-site only, and limiting important features for the model construction. The modified model showed significant improvement, with the R^2 of 0.93, compared to the original model (R^2 of 0.57). The model was validated and used to predict the ZT of the unknown doped BiCuSeO compounds. The predicted result was logically justified based on the thermoelectric principle. It means that the ML model can guide the experiments to improve the thermoelectric properties of BiCuSeO and can be extended to other materials.

Keywords: thermoelectric materials; thermoelectric properties; machine learning; BiCuSeO

Citation: Parse, N.; Pongkitivanichkul, C.; Pinitsoontorn, S. Machine Learning Approach for Maximizing Thermoelectric Properties of BiCuSeO and Discovering New Doping Element. *Energies* **2022**, *15*, 779. https://doi.org/10.3390/en15030779

Academic Editor: Luis Hernández-Callejo

Received: 23 December 2021
Accepted: 19 January 2022
Published: 21 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Electricity consumption is increasing continuously as a result of technological progress. Thermoelectric is one of the interesting alternative energy technologies, which can convert heat to electricity and vice versa. This technology provides many benefits, such as environmentally friendly energy sources, scalability, and silent operation. Unfortunately, the generic thermoelectric bulk modules perform with an efficiency of about 3–5% [1], which is lower than other alternative energy sources such as solar cells with an efficiency of up to 30% [2]. In order to develop a better thermoelectric performance, thermoelectric materials, the heart of the technology, need to be better developed. The key performance of thermoelectric materials is determined from the dimensionless Figure-of-Merit (ZT), defined as $ZT = \frac{S^2\sigma}{k}T$ [3] where T, σ S, and k are the absolute temperature, electrical conductivity, Seebeck coefficient, and thermal conductivity, respectively. Various methods have been investigated to enhance ZT, and thus, the performance of the material.

Traditional approaches to investigate thermoelectric materials are by experiments and computational methods based on density functional theory (DFT). In general, experimenting requires expertise, instrument, and advanced technology, which consume considerable resources. Furthermore, it is difficult to control overall variables and may require a long acquisition period. Alternatively, the computational simulation needs less time and is profitable in complete control over the essential variables. Nonetheless, there are also many challenges for the DFT simulation related to microstructures of material. It needs high-performance computing apparatus, usually in large computing clusters, which is difficult

to be accessed by individuals. Additionally, the simulation was merely employed to some specific systems and required approximations to minimize runtime on complex systems. To accelerate the development and discovery of novel thermoelectric materials, machine learning (ML) becomes an attractive approach. ML is a data-driven method that utilizes statistical mathematics to analyze the data. It can predict micro and macro properties and the correlation between the parameters of the materials [4].

To accelerate the material research, advances and applications of ML have been developed continuously [4–7]. The ML was currently supported by several online databases, algorithms, and frameworks [8,9]. The ML model for predicting materials properties was usually implemented via a classical algorithm, such as regression, determined by $y_i = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n$, $i = 1, 2, \ldots n$, where y_i is the target or predicting value, a_i is the regression coefficient automatically calculated by an ML algorithm, and x_i is the feature or descriptor for representing the character of materials. Even though there are many ways to generate the features, Magpie is the software that originates features for material science by using physical properties. They are operated with mathematics requiring only chemical formula [10]. Furthermore, the features have the potential to build an ML model with advantages in a comfortable and quick method for searching new candidate materials [11,12]. With many advantages, ML has the potential to be a new approach to accelerate the discovery of thermoelectric material with high performance.

Related Work

Recently, ML applications in thermoelectric materials have been increasingly investigated due to high accuracy and less time-consuming. For example, Iwasaki et al. reported the ML model that accelerated the discovery of new candidate materials by generating features from the chemical formula confirmed with the experiment [12]. In their investigation for the spin-driven thermoelectric effect (STE) device, the descriptors for training the ML model were generated automatically from the composition with a composition-based feature vector (CBFV) [13]. The results showed that some features, such as atomic weight, spin, and orbital angular momenta, play an important role in thermopower. In addition, Wang et al. studied the $Cu_xBi_2Te_{2.85+y}Se_{0.15}$ system with ML [14]. The correlation between microstructure and thermoelectric properties was investigated with the principal component analysis (PCA) and the regression algorithm. Furthermore, apart from predicting the properties of new materials, ML could design the experimental conditions to obtain a high *ZT* value. Hou et al. presented an effective way to find the optimal chemical composition of the $Al_2Fe_3Si_3$ thermoelectric compound [15]. With the Bayesian Optimization (BO) algorithm, ML can be applied to the experiment effectively. The power factor can be improved by about 40% compared to the sample with the initial Al/Si ratio of 0.9. Moreover, the author claimed that the framework of this study could be extended to the extrinsic doping of $Al_2Fe_3Si_3$. These related works can be summarized in Table 1.

Table 1. Summary of the research investigating thermoelectric properties with ML.

Datasets	Input	Output	R^2	Remark	Ref.
112	temperature, chemical potential, atomic radius, etc.	Thermopower	-	Thermopower improved an order of magnitude	[13]
17	chemical composition	ZT	0.99	design experiment condition providing high *ZT*	[14]
5	Al/Si ratio	Power factor	-	increase 40% of power factor	[15]

The previous related research generally exploited the data from the first principle calculation or from one laboratory. Our present work made a contribution over the previous related research by exploring the experimental datasets available in literature to

construct the ML model. We then used the model to predict the thermoelectric properties of BiCuSeO. BiCuSeO is a class of thermoelectric oxides considered a new candidate for high-performance p-type thermoelectric materials [16]. Even though the material was only discovered in 2010, thermoelectric researchers have paid much attention to this compound, and continuous publications have been reported since then [17–24]. This compound has a complex ZrSiCuAs layered structure, as shown in Figure 1. It consists of the conducting $(Cu_2Se_2)^{2-}$ layers alternatively stacked by the insulating $(Bi_2O_3)^{2+}$ layers. Due to distinct functionalities and the weak bonding between these two layers, BiCuSeO showed outstanding thermoelectric properties and outperformed most thermoelectric oxides [25]. Therefore, intense research interest is focusing on BiCuSeO to lift the thermoelectric performance and *ZT* even higher. The most common approach to enhance *ZT* is by extrinsic doping some elements into the BiCuSeO structure to lower thermal conductivity, increase carrier concentration, and optimize electrical transport properties [25–27]. Nevertheless, since there are numerous available dopants, tedious experiments are required. Therefore, ML could be a wise choice to address the issue by providing guidance for appropriate effective doping of BiCuSeO.

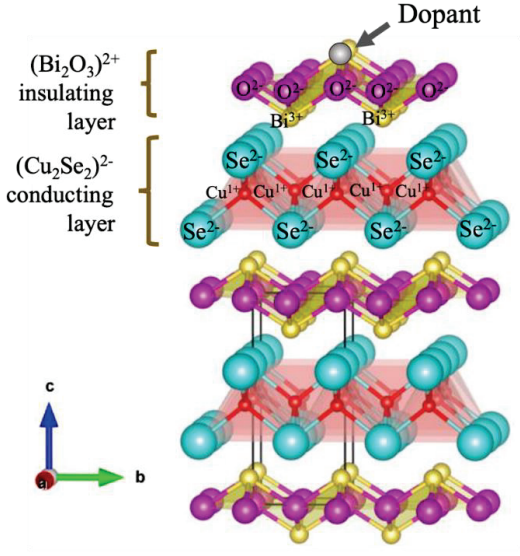


Figure 1. The crystal structure of BiCuSeO consists of conducting $(Cu_2Se_2)^{2-}$ layer and insulating $(Bi_2O_3)^{2+}$ layer. It also shows the dopant substituted at the Bi site.

In this work, the ML model was constructed to provide the guidelines for effective doping of the BiCuSeO system. The ML model was built and tested by collecting data from available published articles (2010–present). Step-by-step, we improved the accuracy of our model so that the predicted *ZT* value from the model closely matched with the experiment. We then extracted the features/descriptors representing the characteristics of materials and discussed their correlation to the physical parameters of the materials. Finally, we used the ML model to predict the suitable dopants in the BiCuSeO system, which can improve thermoelectric properties and lift the *ZT* of the doped compound with respect to the pristine BiCuSeO. We truly believe that our work and technique would be very useful for experimental researchers working to improve the thermoelectric properties of the BiCuSeO compounds.

2. Materials and Methods

Thermoelectric databases for the BiCuSeO compounds were collected from published articles from 2010 to the present (available in the supplementary information, Table S1). They were then tabulated in Excel for the convenience to import into the Jupyter Notebook software. The descriptors or features for building classical ML models were generated from the collected chemical formula via Magpie. The physical and chemical properties of the element were manipulated by mathematical operators, such as average, summation, min, mode, max, and median, and a total of 154 features were obtained. Then, the total datasets were split into a training set (85%) and a test set (15%). The training set was used to teach ML to find the pattern of the data, whereas the test set was used to test the accuracy of the model. Due to the small size of datasets compared to other ML research in materials science [11], the models were built by using different regression algorithms [28], namely, forest regression (RF), Gradient Boosting Regressor (GBR), kneighbor regressor (KN), extraboost tree (ET) and xgboost (XGB). These regression algorithms were determined from a simple linear relationship according to:

$$y_i = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n,\ i = 1,\ 2,\ \ldots\ n, \tag{1}$$

where y_i is the target or predicting value, a_i is the regression coefficient automatically calculated by an ML algorithm, and x_i is the features or descriptors for representing the character of materials. The algorithm which showed the best performance was selected.

Two metrics were used to evaluate the model's accuracy, i.e., (1) the coefficient of determination (R^2) and (2) the root mean squared error (*RMSE*). The R^2 was determined by:

$$R^2 = 1 - \frac{SSE}{SST}, \tag{2}$$

where

$$SSE = \sum_{i=1}^{n}\{y_i - \hat{y}_i\}^2,$$

and

$$\mathrm{SST} = \sum_{i=1}^{n}\{y_i - \overline{y},\}^2$$

and the *RMSE* was determined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

where y_i, $\hat{y}_i$, and $\overline{y}$ are defined as experimental, predicted, and average target value or *ZT*.

The features or descriptors that are important to the model were exposed automatically via the function method from the regression model. Additionally, before bringing the model to use, a final step was to validate the model by Leave One Out Cross Validation (LOOCV). Finally, we used our ML model to predict the *ZT* value of the BiCuSeO compounds doped at the Bi site ($Bi_{1-x}A_xCuSeO$, where A is the dopant and x was set to 0.02). To discover a candidate to maximize the *ZT* value, the dopant (element A) was not in the original datasets and could possibly be done by experiments. Converting the materials into the numerical feature vectors benefits thermoelectric material researchers to build the ML model and discover new candidate material with the only chemical formula.

In the next section, we presented the results for improving the ML model step-by-step until obtaining the desirable ML model. The processes along with the thermoelectric principle of BiCuSeO material were discussed.

3. Results and Discussions

Firstly, the data of BiCuSeO research reporting *ZT* values were extracted from literature (a total of 264 datasets). Then, the ML model was constructed using CBFV to generate 154 features from the chemical formula. Due to relatively small datasets compared to other

ML research in materials science [11], several regression algorithms were employed. The algorithm which showed the best performance was selected.

The results from the ML model are plotted in Figure 2. The *x*-axis is the experimental *ZT*, referring to the reported *ZT* values extracted from the literature. The *y*-axis is called the 'predicted *ZT*', the *ZT* values predicted from the ML model based on the exact chemical formula of BiCuSeO compounds. All related features (a total of 154 features) were included in the model. The orange circles represent data from the training set, whereas the blue squares refer to data from the test set. The dotted line plotted as a guide-to-eyes is an ideal line when the predicted value perfectly matches the experiment. We evaluated the accuracy of the model using two metrics: (1) the coefficient of determination (R^2), and (2) the root mean squared error (*RMSE*). R^2 accounts for how well the model can capture the correlation between the features and the *ZT* value, whereas *RMSE* is used to evaluate the model accuracy regarding the error from prediction. The perfect fit would result in the R^2 of 1 and *RMSE* of 0.

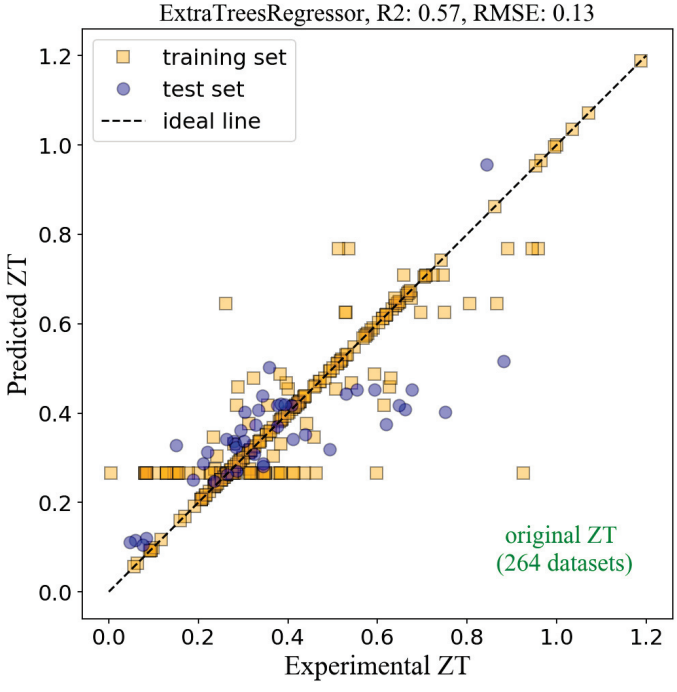


Figure 2. The plot of the Predicted *ZT* versus the Experimental *ZT* from the ML model using ET regressor. The total datasets of 264 datasets were used, resulting in the R^2 of 0.57 and the *RMSE* of 0.13.

Figure 2 shows the R^2 of 0.57 and the *RMSE* of 0.13 from the test set. The R^2 value is relatively low, implying that the model is not very accurate. The model inaccuracy lies in the original data from the experiment database. The reported *ZT* values of the pristine BiCuSeO from several research groups varied significantly. For example, Farooq et al. reported the *ZT* of 0.25 [29], but Yang et al. reported the *ZT* of 0.42 [30] for the same compound (BiCuSeO). These points are explicitly shown in Figure 2, where the orange squares line up horizontally at the 'predicted *ZT*' around 0.3. The discrepancy was due to the experimental details, such as processing parameters, microstructures, etc., which strongly affect the ML performance because the ML models were trained with the

features that were extracted from chemical formulas only. The variations from experimental parameters were not included in the ML model, resulting in the model's inaccuracy.

To improve the model's accuracy, we had to eliminate the experimental dependent variables. To do that, we normalized the experimental ZT by the ZT of the pristine BiCuSeO from each publication. For instance, Farooq et a. reported the ZT of BiCuSeO and $Bi_{0.99}Cd_{0.01}CuSeO$ of 0.25 and 0.43 [29], while Yang reported the ZT of BiCuSeO and $Bi_{0.98}Pb_{0.02}CuSeO$ of 0.42 and 0.66 [30]. By normalizing, the 'experimental $ZT_{\text{normalized}}$' of Farooq's BiCuSeO and $Bi_{0.8}Cd_{0.2}CuSeO$ became 1.0 and 1.72, whereas 'experimental $ZT_{\text{normalized}}$' of Yang's BiCuSeO and $BiCu_{0.8}Zn_{0.2}SeO$ were 1.0 and 1.57. The normalization can be determined as $ZT_{normalized} = \frac{ZT_{doped}}{ZT_{undoped}}$. In other words, by using this process, the 'experimental $ZT_{\text{normalized}}$' of the pristine BiCuSeO from any publication was turned into unity. The 'experimental $ZT_{\text{normalized}}$' of the doped BiCuSeO thus indicated the ratio of improvement between the doped BiCuSeO and the pristine BiCuSeO. The ML model was then reconstructed such that the ZT was only related to the chemical formulas, and other experimental dependent variables were eliminated.

The results from the ML model after normalizing all 264 datasets are presented in Figure 3, with the R^2 of 0.78 and *RMSE* of 1.48 for the test set. The R^2 of 0.78 in Figure 3 is larger than the R^2 of 0.57 in Figure 2, indicating the improvement of the model's accuracy. However, the higher *RMSE* (1.48) in Figure 3 compared to Figure 2 (*RMSE* = 0.13) does not mean that its prediction's error is worse. In fact, it is incorrect to compare the *RMSE* between the two figures because the data ranges are not the same. The scales in both axes in Figure 2 range between 0 and 1.2, whereas Figure 3 ranges from 0 to 20.0. Hence, it is expected that the *RMSE* in Figure 3 tends to be higher.

Although the R^2 for the ML model in Figure 3 is relatively high, there are still outliers that deviated from the ideal line, for instance, the orange square and the blue circle on the right of the figure, leading to the reduction of R^2. This situation occurred even when the selected features in the model were already optimized. Therefore, we tried improving our ML model further by analyzing the original datasets. We found that the outliers and inaccuracy of the model could be from the different doping sites in the BiCuSeO compound. In general, doping elements in BiCuSeO is done by substituting atoms at different sites, written in a chemical formula $Bi_{1-x}A_xCu_{1-y}B_ySe_{1-z}C_zO_{1-w}D_w$, where A, B, C, and D are dopants. Sometimes, dual dopings were done at one or more sites. The purpose of doping in each site is different, such as lowering thermal conductivity, bandgap engineering, and tuning electrical transport properties [17]. We assumed that our ML model could not capture the pattern from the data including all variations. Therefore, we analyzed the data and grouped the datasets into a few sub-groups. The major sub-group (145datasets) was the BiCuSeO compound doped at the Bi site (Figure 1), for instance, $Bi_{0.98}K_{0.02}CuSeO$ [31]. This group is vital from the thermoelectric perspective. The BiCuSeO structure consists of two layers: the conducting $(Cu_2Se_2)^{2-}$ layers and the insulating $(Bi_2O_3)^{2+}$ layers. The electrical transport pathway is mainly limited to the Cu_2Se_2 layers, whereas the Bi_2O_2 layers behave as a charge reservoir [32]. Thus, doping at the Bi site provides extra charge carriers for thermoelectric power factor tuning without interrupting the carrier transport. Therefore, the ML was reconstructed based on these datasets.

Figure 4 shows the results from the ML model based on 144 datasets for the Bi-doped BiCuSeO. The R^2 was considerably increased to 0.89, with the *RMSE* of 0.40, indicating the improvement of the model's accuracy. However, decreasing the amount of data and using many features (154 features) could lead to overfitting, which means the model shows high performance on the training dataset but low performance on the test set [33]. To address the issue, we exported the features or descriptors representing the material characteristics from our ML model and ranked them according to their importance to the model. There were a total of 154 generated features, but the first 30 important features are shown in Figure 5. We then optimized the ML model by including only the important features. We have tried including the first 3, the first 6, the first 9 . . . and so on important features in the model. The best-performance model was obtained when the first 12 important features (as

highlighted in Figure 5) were used. Figure 6 shows the results from such a model, with the R^2 of 0.93 and the *RMSE* of 0.33 for the test set, an improvement in accuracy from the model in Figure 4. If one compared the model in Figure 6 to the primitive model in Figure 2, the accuracy performance increased >63%. However, before bringing the model to use, the generalization of the model was carried out via Leave One Out Cross Validation (LOOCV). This method is appropriate, particularly for small-size datasets [5]. The validation resulted in the *RMSE* of 0.71 for the training dataset, which means that the predicted $ZT_{\text{normalized}}$ values from the model have an error of ± 0.71.

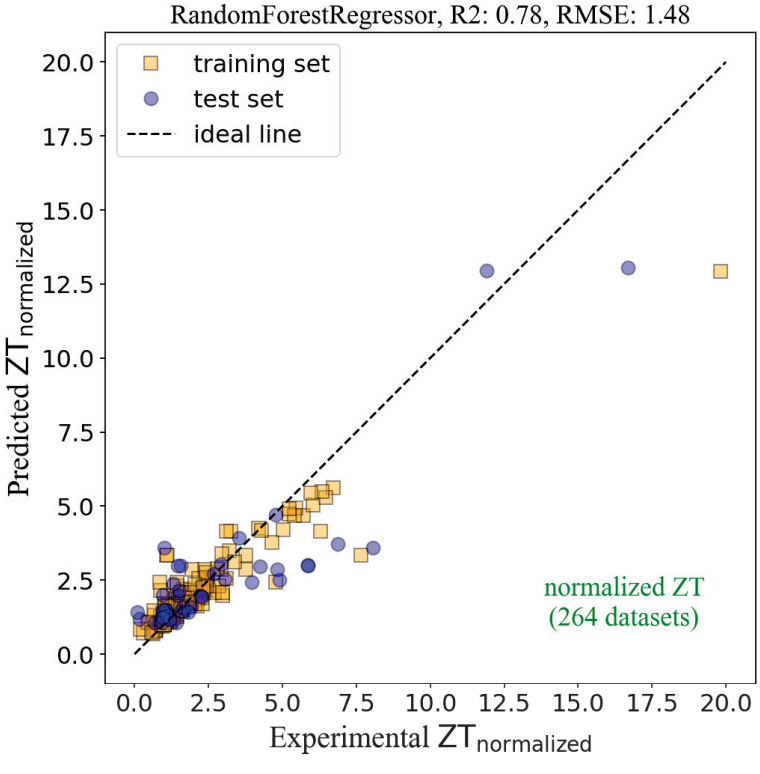


Figure 3. The plot of the Predicted $ZT_{\text{normalized}}$ versus the Experimental $ZT_{\text{normalized}}$ from the ML model using RF regressor. The total datasets of 264 datasets were used, resulting in the R^2 of 0.78 and the *RMSE* of 1.48.

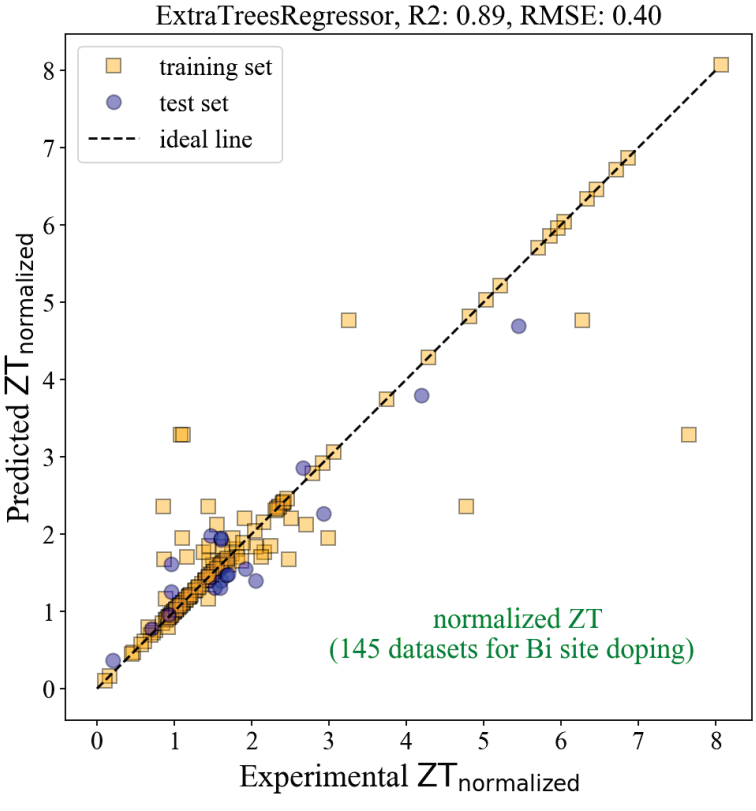


Figure 4. The plot of the Predicted $ZT_{\text{normalized}}$ versus the Experimental $ZT_{\text{normalized}}$ from the ML model using ET regressor. The total dataset of 145 datasets was used, resulting in the R^2 of 0.89 and the *RMSE* of 0.40.

The physical meaning of the important features in Figure 5 is worth discussing. The most important feature is the min_NUnfilled. The prefix min refers to the minimum number of the element's properties obtained from Magpie software, whereas the NUnfilled accounts for the total number of unfilled electrons in electronic shells (s, p, d, f). For example, the NUnfilled of He is 0 from its electronic configuration ($1s^2$), whereas the electronic configuration of Na is $1s^2\ 2s^2\ 2p^6\ 3s^1$ resulting in the NUnfilled of 1. In the case of the BiCuSeO compound, the NUnfilled of Bi, Cu, Se, and O is 3, 1, 2, and 2, respectively, and hence, the min_NUnfilled of BiCuSeO is 1, according to the minimum NUnfilled of Cu. For the doped compound, such as $Bi_{0.94}Mg_{0.03}Pb_{0.03}CuSeO$, the min_NUnfilled of this compound is 0 because the NUnfilled of Mg equals 0. By using Pearson correlation analysis, it was found that the lower the min_NUnfilled, the higher the $ZT_{\text{normalized}}$. The lowest min_NUnfilled (0) was found in the BiCuSeO doped with, for example, Mg, Ca, Sr, Ba. These elements are divalent ions (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}). When they were substituted for Bi^{3+}, an extra +1 charge was generated for charge neutralization. This extra charge increased the carrier concentration of the BiCuSeO system, leading to optimization of power factors [17,34,35]. Therefore, it is reasonable for min_NUnfilled to be the most important feature for our ML model.

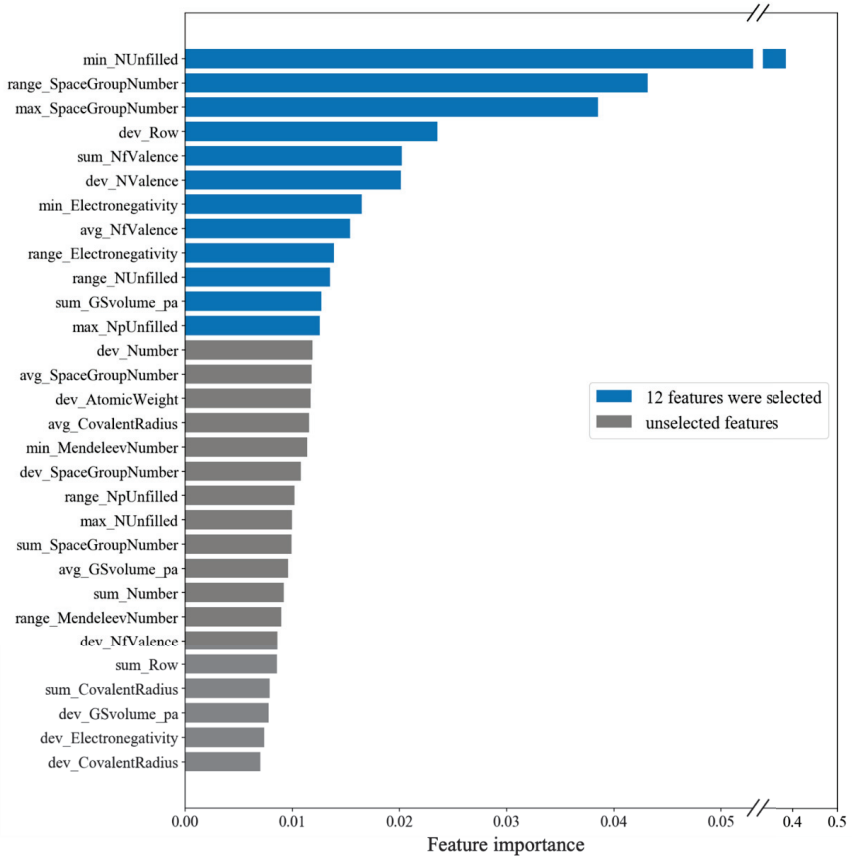


Figure 5. Exported features from the ML model, ranked according to their importance. The first 12 features are: 1. min_NUnfilled = minimum of total number of unfilled valence orbitals of the elements in the material ($Bi_{1-x}A_xCuSeO$), 2. range_SpaceGroupNumber = range of space group of T= 0 K ground state structure of the elements, 3. max_ SpaceGroupNumber = maximum of space group of T = 0 K ground state structure of the elements, 4. dev_Row = deviation of row on periodic table of the elements, 5. sum_NfValence = summation of number of filled f valence orbitals of the elements, 6. dev_NValence = deviation of total number of valence electrons of the elements, 7. min_Electronegativity = minimum of Pauling electronegativity of the elements, 8. avg_NfValence = average of number of filled f valence orbitals of the elements, 9. range_Electronegativity = range of Pauling electronegativity of the elements, 10. range_NUnfilled = range of total number of unfilled valence orbitals of the elements, 11. sum_GSvolume_pa = DFT volume per atom of T = 0 K ground state, 12. max_NpUnfilled = maximum of number of unfilled p valence orbitals of the elements.

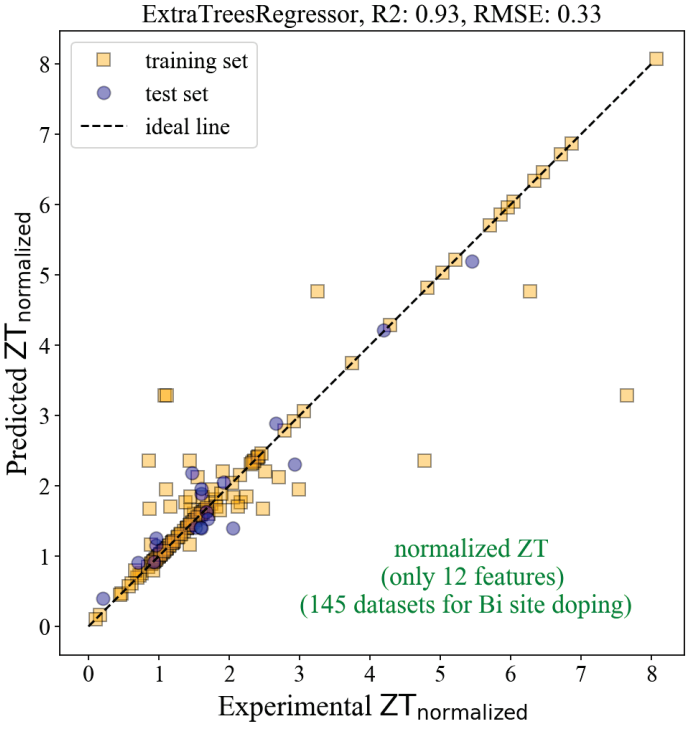


Figure 6. The plot of the Predicted $ZT_{normalized}$ versus the Experimental $ZT_{normalized}$ from the ML model using ET regressor. The total dataset of 145 datasets was used with the first 12 important features, resulting in the R^2 of 0.93 and the *RMSE* of 0.33.

Finally, we used the optimized ML model to predict $ZT_{normalized}$ of the doped BiCuSeO at Bi-site ($Bi_{1-x}A_xCuSeO$, where A is the dopant and x = 0.02). We selected some elements that were not already in the model datasets, and such elements could be synthesized experimentally. Figure 7 shows the predicted $ZT_{normalized}$ value for some candidate materials. The highest $ZT_{normalized}$ belongs to the Si-doped compound, which is reasonably justified. It was reported that doping light elements at the Bi-site in BiCuSeO could promote carrier mobility from the decreased carrier scattering [36]. Since Si can be considered as a light element, doping Si for Bi is likely to promote carrier mobility and increase *ZT*. Moreover, the DFT simulation of the Si doping at Bi-site showed the increased electrical conductivity, with a slight decrease in the Seebeck coefficient, from the modified electronic band near the Femi level, resulting in a large power factor. On the other hand, the Cl-doped compound exhibited the lowest $ZT_{normalized}$ value from the model. This result is understandable. The previous experiment reported that doping Cl at Se-site negatively affected the *ZT* value, by increasing both electrical resistivity and thermal conductivity [37]. Thus, Cl is unlikely to be a good candidate for doping in BiCuSeO.

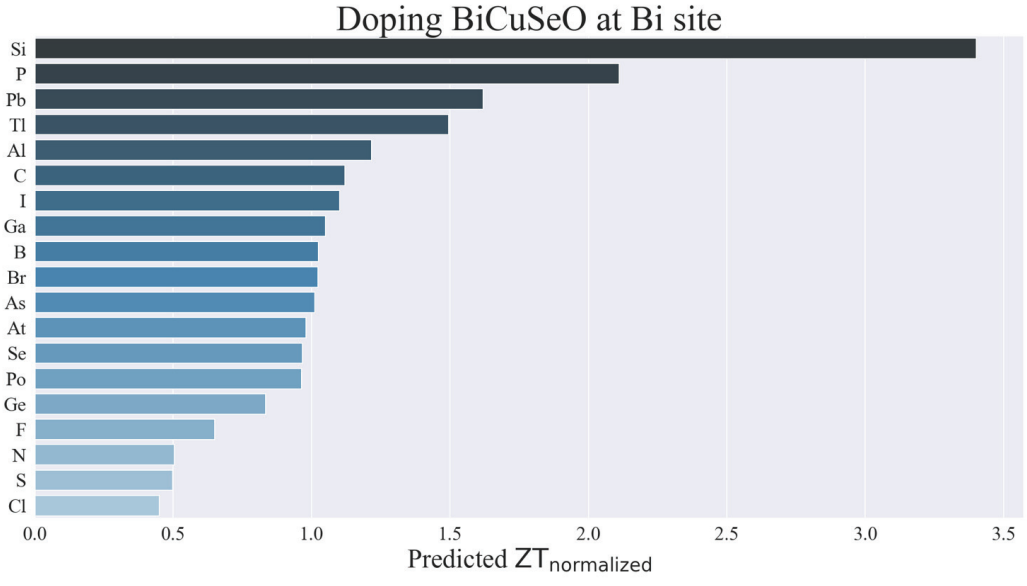


Figure 7. Predicted $ZT_{\text{normalized}}$ values for the selected $Bi_{0.98}A_{0.02}CuSeO$ compounds, where A is the dopant shown in the *y*-axis.

The step-by-step development of the ML model with improving performance was presented. It was used to guide a new candidate material for enhancing *ZT* value. However, the limited data from experiments was an obstacle to constructing the accurate ML model. Apart from that, it was also found that training the ML model requires both good and bad results. Generally, most published articles reported only good results (large *ZT*), but in fact, various data (positive or negative results) are necessary to improve the ML model.

4. Conclusions

We have developed the ML model for predicting the thermoelectric Figure-of-Merit (ZT) of the BiCuSeO compounds. The model was improved step-by-step to achieve relatively high accuracy. The ML initially showed a relatively low R^2 of 0.57. We then improved the model's accuracy by normalizing the experimental ZT of the doped BiCuSeO with the pristine BiCuSeO. The modified ML model showed improved accuracy with an R^2 of 0.78. Furthermore, we selected the data for the BiCuSeO doped at Bi-site only and reconstructed the model. The R^2 increased to 0.89, indicating the enhanced model's accuracy. Last but not least, only 12 important features were used in the model, which resulted in the increased R^2 to 0.93 and the RMSE of 0.33. Furthermore, the most important feature, min_NUnfilled, was discussed and correlated to the physical parameters of materials. The model predicted the substantial ZT improvement for the Si-doped BiCuSeO material, which is scientifically sound from the thermoelectric principle. Therefore, the ML model of this work can provide a guideline for experimental researchers for improving the thermoelectric properties of BiCuSeO and can be extended to other thermoelectric materials.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/en15030779/s1, Table S1: The sources of literature data, showing the dopants, the substitution sites, and the references.

Author Contributions: N.P.: writing—original draft; N.P.: data collection; N.P. and C.P.: software; N.P., C.P. and S.P.: methodology, and validation; N.P., C.P. and S.P.: formal analysis; C.P. and

S.P.: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Thailand Research Fund (TRF) in cooperation with Synchrotron Light Research Institute (public organization) and Khon Kaen University (RSA6280020), the Research and Graduate Studies of Khon Kaen University, and the Development and Promotion of Science and Technology program, Thailand.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and the code that support the results within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zoui, M.A.; Bentouba, S.; Stocholm, J.G.; Bourouis, M. A Review on Thermoelectric Generators: Progress and Applications. *Energies* **2020**, *13*, 3606. [CrossRef]
2. Shockley, W.; Queisser, H.J. Detailed Balance Limit of Efficiency of p-n Junction Solar Cells. *Int. J. Appl. Phys.* **1961**, *32*, 510–519. [CrossRef]
3. Rowe, D.M. *CRC Handbook of Thermoelectrics*; CRC Press: Boca Raton, FL, USA, 1995.
4. Schmidt, J.; Marques, R.G.; Botti, S.; Marques, M.A.L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **2019**, *5*, 83. [CrossRef]
5. Liu, Y.; Zhao, T.; Ju, W.; Shi, S. Materials discovery and design using machine learning. *J. Mater.* **2017**, *3*, 159–177. [CrossRef]
6. Wei, J.; Chu, X.; Sun, X.Y.; Xu, K.; Deng, H.X.; Chen, J.; Wei, Z.; Lei, M. Machine learning in materials science. *InfoMat* **2019**, *1*, 338–358. [CrossRef]
7. Chen, A.; Zhang, X.; Zhou, Z. Machine learning: Accelerating materials development for energy storage and conversion. *InfoMat* **2020**, *2*, 553–576. [CrossRef]
8. Wang, T.; Zhang, C.; Hichem, S.; Zhang, G. Machine Learning Approaches for Thermoelectric Materials Research. *Adv. Funct. Mater* **2020**, *30*, 1906041. [CrossRef]
9. Recatala-Gomez, J.; Suwardi, A.; Nandhakumar, I.; Abutaha, A.; Hippalgaonkar, K. Toward Accelerated Thermoelectric Materials and Process Discovery. *ACS Appl. Energy Mater.* **2020**, *3*, 2240–2257. [CrossRef]
10. Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2016**, *2*, 16028. [CrossRef]
11. Na, G.S.; Jang, S.; Chang, H. Predicting thermoelectric properties from chemical formula with explicitly identifying dopant effects. *npj Comput. Mater.* **2021**, *7*, 106. [CrossRef]
12. Iwasaki, Y.; Takeuchi, I.; Stanev, V.; Kusne, A.G.; Ishida, M.; Kirihara, A.; Ihara, K.; Sawada, R.; Terashima, K.; Someya, H.; et al. Machine-learning guided discovery of a new thermoelectric material. *Sci. Rep.* **2019**, *9*, 2751. [CrossRef] [PubMed]
13. Murdock, R.J.; Kauwe, S.K.; Wang, A.Y.-T.; Sparks, T.D. Is domain knowledge necessary for machine learning materials properties? *Integr. Mater* **2020**, *9*, 221–227. [CrossRef]
14. Wang, Z.-L.; Adachi, Y.; Chen, Z.-C. Processing Optimization and Property Predictions of Hot-Extruded Bi–Te–Se Thermoelectric Materials via Machine Learning. *Adv. Theory Simul.* **2019**, *3*, 1900197. [CrossRef]
15. Hou, Z.; Takagiwa, Y.; Shinohara, Y.; Xu, Y.; Tsuda, K. Machine-Learning-Assisted Development and Theoretical Consideration for the Al2Fe3Si3 Thermoelectric Material. *ACS Appl. Mater. Interfaces* **2019**, *11*, 11545–11554. [CrossRef] [PubMed]
16. Barreteau, C.; Berardan, D.; Dragoe, N. Studies on the thermal stability of BiCuSeO. *J. Solid State Chem.* **2015**, *222*, 53–59. [CrossRef]
17. Zhao, L.; Bérardan, D.; Pei, Y.; Roux- Byl, C.; Pinsard-Gaudart, L.; Dragoe, N. Bi1−xSrxCuSeO Oxyselenides as Promising Thermoelectric Materials. *Appl. Phys. Lett* **2010**, *97*, 092118. [CrossRef]
18. Li, J.; Sui, J.; Pei, Y.; Barreteau, C.; Bérardan, D.; Dragoe, N.; Cai, W.; He, J.; Zhao, L. A High Thermoelectric Figure of Merit ZT > 1 in Ba Heavily Doped BiCuSeO Oxyselenides. *Energy Environ. Sci.* **2012**, *5*, 8543–8547. [CrossRef]
19. Li, F.; Wei, T.-R.; Kang, F.; Li, J. Enhanced Thermoelectric Performance of Ca-Doped BiCuSeO in a Wide Temperature Range. *J. Mater. Chem. A* **2013**, *1*, 11942. [CrossRef]
20. Li, J.; Sui, J.; Barreteau, C.; Berardan, D.; Dragoe, N.; Cai, W.; Pei, Y.; Zhao, L.-D. Thermoelectric properties of Mg doped p-type BiCuSeO oxyselenides. *J. Alloys Compd.* **2013**, *551*, 649–653. [CrossRef]
21. Liu, Y.-c.; Zheng, Y.-h.; Zhan, B.; Chen, K.; Butt, S.; Zhang, B.; Lin, Y.-h. Influence of Ag doping on thermoelectric properties of BiCuSeO. *J. Eur. Ceram. Soc.* **2015**, *35*, 845–849. [CrossRef]
22. Liu, Y.; Ding, J.; Xu, B.; Lan, J.; Zheng, Y.; Zhan, B.; Zhang, Z.; Lin, Y.; Nan, C.W. Enhanced Thermoelectric Performance of La-Doped BiCuSeO by Tuning Band Structure. *Appl. Phys.* **2015**, *106*, 233903. [CrossRef]
23. Ren, G.; Butt, S.; Zeng, C.; Liu, Y.; Zhan, B.; Lan, J.; Lin, Y.; Nan, C. Electrical and Thermal Transport Behavior in Zn-Doped BiCuSeO Oxyselenides. *J. Electron. Mater.* **2015**, *44*, 1627–1631. [CrossRef]

24. Zhang, X.; Chang, C.; Zhou, Y.; Zhao, L.-D. BiCuSeO Thermoelectrics: An Update on Recent Progress and Perspective. *Materials* **2017**, *10*, 198. [CrossRef] [PubMed]
25. Li, F.; Ruan, M.; Chen, Y.; Wang, W.; Luo, J.; Zheng, Z.; Fan, P. Enhanced thermoelectric properties of polycrystalline BiCuSeO via dual-doping in Bi sites. *Inorg. Chem. Front* **2019**, *6*, 799–807. [CrossRef]
26. Das, S.; Valiyaveettil, S.; Chen, K.-H.; Suwas, S.; Mallik, R. Thermoelectric properties of Pb and Na dual doped BiCuSeO. *AIP Adv.* **2019**, *9*, 015025. [CrossRef]
27. Feng, B.; Li, G.; Pan, Z.; Hu, X.; Liu, P.; Li, Y.; He, Z.; Fan, X.a. Enhanced thermoelectric performances in BiCuSeO Oxyselenides via Er and 3D modulation doping. *Ceram. Int.* **2019**, *45*, 4493–4498. [CrossRef]
28. Han, G.; Sun, Y.; Feng, Y.; Lin, G.; Lu, N. Machine Learning Regression Guided Thermoelectric Materials Discovery—A Review. *ES Mater. Manuf.* **2021**, *14*, 20–35. [CrossRef]
29. Umer Farooq, M.; Butt, M.; Gao, K.; Zhu, Y.; Sun, X.; Pang, X.; Khan, S.; Mohmed, F.; Mahmood, A.; Xu, W. Cd-doping a Facile Approach for Better Thermoelectric Transport Properties of BiCuSeO Oxyselenides. *RSC Adv.* **2016**, *6*, 33789–33797. [CrossRef]
30. Yang, D.; Su, X.; Yan, Y.; Hu, T.; Xie, H.; He, J.; Uher, C.; Kanatzidis, M.G.; Tang, X. Manipulating the Combustion Wave during Self-Propagating Synthesis for High Thermoelectric Performance of Layered Oxychalcogenide Bi1–xPbxCuSeO. *Chem. Mater* **2016**, *28*, 4628–4640. [CrossRef]
31. Lan, J.; Ma, W.; Deng, C.; Ren, G.-K.; Lin, Y.-H.; Yang, X. High thermoelectric performance of Bi1−x K x CuSeO prepared by combustion synthesis. *J. Mater. Sci.* **2017**, *52*, 11569–11579. [CrossRef]
32. Barreteau, C.; Pan, L.; Pei, Y.-l.; Zhao, L.; Bérardan, D.; Dragoe, N. Oxychalcogenides as new efficient p-type t hermoelectric materials. *Funct. Mater. Lett.* **2013**, *6*, 1340007. [CrossRef]
33. Ying, X. An Overview of Overfitting and its Solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [CrossRef]
34. Wen, Q.; Zhang, H.; Xu, F.; Liu, L.; Wang, Z.; Tang, G. Enhanced thermoelectric performance of BiCuSeO via dual-doping in both Bi and Cu sites. *J. Alloys Compd* **2017**, *711*, 434–439. [CrossRef]
35. Kang, H.; Li, J.; Liu, Y.; Guo, E.; Chen, Z.; Liu, D.; Fan, G.; Zhang, Y.; Jiang, X.; Wang, T. Optimizing the thermoelectric transport properties of BiCuSeO via doping with the rare-earth variable-valence element Yb. *J. Mater. Chem. C* **2018**, *6*, 8479–8487. [CrossRef]
36. He, T.; Li, X.; Tang, J.; Lou, X.n.; Zuo, X.; Zheng, Y.; Zhang, D.; Tang, G. Boosting thermoelectric performance of BiCuSeO by improving carrier mobility through light element doping and introducing nanostructures. *J. Alloys Compd* **2020**, *831*, 154755. [CrossRef]
37. Zhou, Z.; Tan, X.; Ren, G.; Lin, Y.; Nan, C. Thermoelectric Properties of Cl-Doped BiCuSeO Oxyselenides. *J. Electron. Mater.* **2017**, *46*, 2593–2598. [CrossRef]

energies

MDPI

Article

Sources and Sectoral Trend Analysis of CO_2 Emissions Data in Nigeria Using a Modified Mann-Kendall and Change Point Detection Approaches

Ogundele Lasun Tunde [1], Okunlola Oluyemi Adewole [2,*], Mohannad Alobid [3], István Szűcs [3] and Yacouba Kassouri [4]

[1] Department of Physics, University of Medical Sciences, Ondo 351104, Nigeria; logundele@unimed.edu.ng
[2] Department of Mathematical and Computer Science, University of Medical Sciences, Ondo 351104, Nigeria
[3] Faculty of Economics and Business, Institute of Applied Economic Sciences, University of Debrecen, H-4032 Debrecen, Hungary; mohannad.alobid@econ.unideb.hu (M.A.); istvan.szucs@econ.unideb.hu (I.S.)
[4] Department of Economics and Finance, Nisantasi University, Istanbul 25370, Turkey; yacouba.kassouri@nisantasi.edu.tr
* Correspondence: ookunlola@unimed.edu.ng

Abstract: In Nigeria, the high dependence on fossil fuels for energy generation and utilization in various sectors of the economy has resulted in the emission of a large quantity of carbon dioxide (CO_2), which is one of the criteria gaseous pollutants that is frequently encountered in the environment. The high quantity of CO_2 has adverse implications on human health and serious damaging effects on the environment. In this study, multi-decade (1971–2014) CO_2-emissions data for Nigeria were obtained from the World Development Indicator (WDI). The data were disaggregated into various emission sources: gaseous fuel consumption (GFC), liquid fuel consumption (LFC), solid fuel consumption (SFC), transport (TRA), electricity and heat production (EHP), residential buildings and commercial and public services (RSCPS), manufacturing industries and construction (MINC), and other sectors excluding residential buildings and commercial and public services (OSEC). The analysis was conducted for a sectorial trend using a rank-based non-parametric modified Mann–Kendall (MK) statistical approach and a change point detection method. The results showed that the CO_2 emissions from TRA were significantly high, followed by LFC. The GFC, LFC, EHP, and OSEC had a positive Sen's slope, while SFC, TRA, and MINC had a negative Sen's slope. The trend analysis indicated multiple changes for TRA and OSEC, while other sources had a change point at a particular year. These results are useful for knowledge of CO_2-emission sources in Nigeria and for future understanding of the trend of its emission for proper environmental planning. The severe effects of CO_2 on the atmospheric environment of Nigeria may be worsened in the future due to some major sources such as transportation services and electricity generation that are inevitable for enviable standard of living in an urban setting.

Keywords: CO_2; emission sources; WDI data; trend analysis; Mann-Kendall

Citation: Tunde, O.L.; Adewole, O.O.; Alobid, M.; Szűcs, I.; Kassouri, Y. Sources and Sectoral Trend Analysis of CO_2 Emissions Data in Nigeria Using a Modified Mann-Kendall and Change Point Detection Approaches. *Energies* **2022**, *15*, 766. https://doi.org/10.3390/en15030766

Academic Editors: Luis Hernández-Callejo and João Fernando Pereira Gomes

Received: 22 October 2021
Accepted: 10 December 2021
Published: 21 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Globally, carbon dioxide (CO_2) is one of the common environmental pollutants, and it accounts for more than 70% of the greenhouse effects. It also has varieties of adverse effects on human health and environmental implications [1]. It affects the amount of solar radiation that penetrates through the atmosphere and reaches the surface of the earth as well as outgoing radiation from the surface of Earth. The potential atmospheric and ecological implications of CO_2, among others, consist in global warming, irregular rainfall pattern, over-flooding, extreme weather phenomena, sea-level rises, alterations in crop growth, and disruptions to aquatic water systems. The excessive accumulation of CO_2 in the atmosphere alters the atmospheric radiation budget either by the absorption or emission

of heat [2]. The human activities that contribute to the atmospheric CO_2 are agriculture, urbanization, deforestation, mining, transportation, fuel combustion, waste disposal, and burning. Apart from anthropogenic activities, the atmospheric concentration of CO_2 could also be increased from natural emission sources such as organic decomposition, ocean release, and volcanic eruption in the regions of the world that are prone to tectonic activity, although in a small quantity compared to that emitted from human activities.

The increasing economic activities and technological advancement in the developed and developing countries has greatly boosted economic growth and development, leading to the setting of industries and diversification in industrial activities. In Nigeria, the economic activities cover industrial, manufacturing, agricultural, financial, educational, and tourist sectors. These sectors heavily and solely depend on petroleum and petrochemical resources as the primary sources of energy for effective operations and existence of the sectors due to an epileptic power supply from the natural grid. Despite the contribution of economic activities to gross domestic product (GDP), each sector remains a potential source of CO_2 emissions into the atmosphere. Apart from these, the Nigerian population had increased tremendously in the past few decades. The United Nations reported Nigeria as Africa's most-populous country with about 60% growth from 1990 to 2008 and 170 million people as of 2012. It is projected to reach 0.5–1 billion people by the early 21 century [3]. Other anthropogenic activities such as deforestation, bush burning, and transportation could also increase CO_2) in the atmosphere. The oil and gas sector also emits a large quantity of CO_2 due to gas flaring, illegal refinery, illegal oil refining, frequent pipeline explosions, gas leakage, and venting. Industrial processes such as petroleum-processing refinery, smelting, cement industry, and mining contribute significantly to the atmospheric load of CO_2. In urban centres, road transportation has been reported to be a major source that contributes about one-fifth of the total CO_2 emissions [4,5]. The International Energy Agency [6] found road transportation as a whole to be responsible for 20% of CO_2 emission. In the residential area, most houses rely on kerosene, fuel wood, generators for domestic purposes, cooking, lighting of rural and urban household bulbs, and other electrical appliances where power supplies and distribution are often inconsistent and unreliable. The industries also depend on diesel-powered backup generators for production activities as an alternative to the epileptic power supply [7]. Open burning is a common practice in the urban areas where the waste disposal and management are poorly organised. This contributes to a significant amount of CO_2 emissions into the atmosphere.

A large number of studies have been conducted on CO_2 emissions globally by employing different statistical tools in developed and developing countries. Most studies ascertained both local and global atmospheric implications of CO_2 and reported energy consumption due to economic growth and development as the main reason for increased CO_2 concentration in the atmosphere. This might be due to the fact that trend analysis requires long-time-period data of several decades, which were not available by the real-time measurement in most studies. Short monitoring periods and a small sample size could not be used for trend analysis due to difficulty in identification of rates of change as well as their interpretations. Moreover, the historical to present-day CO_2 emissions sources and distribution among different sectors of the economy are scarce. The analysis of multidecade and historical CO_2 emissions data are of paramount scientific and practical relevance in the prediction of a country's economic development and as well in developing a framework for its emission-reduction strategies to safeguard human health and to sustain environmental quality [8].

The monitoring of CO_2 is highly imperative as it forms the basis of data collection for decision making, regulatory purposes, and future forecasting. However, very few studies have been reported on continuous measurements of CO_2 concentrations. Despite the fact that several studies have reported various emission sources of CO_2 at both local and regional levels, the monitoring periods in most studies are short, yielding a small data size. The inconsistency in data-collection procedures also posed a difficulty in obtaining CO_2 emissions data. These drawbacks are overcome in this study by employing secondary

emissions data collected by an international monitoring organisation. Therefore, this study models CO_2 emissions from various sectors in Nigeria using data sourced from the World Bank database.

Previous studies on the trend analysis of CO_2 emissions were conducted in advanced and emerging countries [9–14]. For instance, [10] found evidence for statistically significant trends in CO_2 emissions of the following countries, namely, India, South Korea, the Islamic Republic of Iran, Mexico, Australia, Indonesia, Saudi Arabia, Brazil, South Africa, Taiwan, and Turkey, including the world total. The authors concluded that projections for CO_2 emissions are influenced by several factors, including fuel consumption types, economic growth rates, and political initiatives. In the case of CO_2 emissions from car travel in Great Britain, [11] found that, although CO_2 emissions continued to increase, the growth rate became substantially lower in the beginning of the 2000s. [12] demonstrated that based on a medium GDP growth rate for 2015–2030 under a business-as-usual scenario, the CO_2 emission trends of China's primary aluminium industry could increase by 60%. By looking at sectoral CO_2 emissions in 41 world regions, [2] found that CO_2 emissions will continue to increase in the construction sectors in all countries. Based on the survey of the available studies, little or no consideration has been given to a natural-resource-rich country like Nigeria. To the best of our knowledge, this is the first study exploring the possible trends and tipping point in CO_2 emissions across different sectors in an oil-dependent developing country. This study departs from the former studies and attempts to fill several gaps in the existing literature in at least two points. First, our study contributes to the growing literature on the determinants of CO_2 emissions by examining trends and the tipping point in CO_2 emissions in order to find out the sector contributing the most to environmental pollution in the study area. Secondly, it is well established that CO_2 emissions are not globally uniform across different sectors [15]. The consideration of CO_2 emissions by sectors could provide new insights on future development and sustainable management of CO_2 emissions across economic sectors.

Parametric and non-parametric statistical tests are some of the approaches used to detect trends of environmental data such as hydrological and hydrometerological data. Parametric tests had been considered to be more powerful, but their main drawback is that the data must be independent, identical, and normally distributed [16]. These are rarely true in environmental data. The non-parametric test overcomes this limitation in its approach via Mann-Kendall trend analysis [8,16–20]. Therefore, this study employed World Development Indicator (WDI) data to conduct the trend analysis of CO_2 emissions from different sources and sectors in Nigeria using a non-parametric Mann-Kendall test. The research questions driving the study were: is there a presence of monotonic trend in the disaggregated CO_2 emission data? what is the magnitude of the trend change? and what is the change point of the trend in each of the CO_2 emissions data? The remainder of the article is structured as follows: Section 2 discusses the methodology; Section 3 presents statistical tools employed to answer the research questions put forward in the study; results, discussion, and policy implications are presented in Sections 4 and 5 has the conclusion.

2. Methodology

2.1. Source of Data and Coverage

This study was conducted in Nigeria, a country in sub-Saharan Africa with per capita CO_2 emissions of around 2.8 tonnes. CO_2 emissions data in the country were sourced from the World Development Indicators of 2018 with respect to sources and sectors. Detailed information on the selected sources and sectors is presented in Table 1. Based on the available information, the study used data from 1971 to 2014 indicating a total of 44 data points for each of the variables. Recent period could not be incorporated due to availability of the data.

Table 1. Variable Description.

Variable ID	Description. CO_2 Emissions from:
GFC	Gaseous fuel consumption (% of total)
LFC	Liquid fuel consumption (% of total)
SFC	Solid fuel consumption (% of total)
TRA	Transport (% of total fuel combustion)
EHP	Electricity and heat production, total (% of total fuel combustion)
RBCPS	Residential buildings and commercial and public services (% of total fuel combustion)
MINC	Manufacturing industries and construction (% of total fuel combustion)
OSEC	Other sectors, excluding residential buildings and commercial and public services (% of total fuel combustion)

Source: World Development Indicator (WDI), 2018.

2.2. The Mann-Kendall Test

The Mann-Kendall (MK) test is rank-based non-parametric statistical method that is used to determine the monotonic upward or downward trend of a time series and long-term data in predicting the future outcome by using historical records [8,21]. The basic assumptions of the MK test are that a value can always be declared less than, greater than, or equal to another value; that data are independent; and that the distribution of data remain constant in either the original or transformed units [17]. The MK test had been used widely by several researchers most especially in climatological, hydrological, visibility, and air-pollution studies [12,13,19,20,22–24]. In this study, the MK test was applied to CO_2 emissions data based on its invariance to data transformation and wide application in environmental studies. Briefly, the fundamental equations for calculating Mann-Kendall statistics S, V(s), and standardized test statistics Z are as follows [17,25,26]:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sig(X_j - X_i) \tag{1}$$

here $i = 2, 3, \ldots n$; $j = 1, 2, \ldots i-1$ and

$$sig(X_j - X_i) = \begin{cases} +1 if (X_j - X_i) > 0 \\ 0 if (X_j - X_i) = 0 \\ -1 if (X_j - X_i) < 0 \end{cases} \tag{2}$$

A normal approximation to the Mann-Kendall test with variance V(s) is given as:

$$V(s) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=i}^{q} t_p(t_p-1)(2t_p+5)\right] \tag{3}$$

here $p = 2, 3, \ldots q$, t_p is the number of ties for the p^{th} value, and q is the number of tied values.

$$Z = \begin{cases} \frac{S-1}{\sqrt{VAR(S)}} if S > 0 \\ 0 if S = 0 \\ \frac{S+1}{\sqrt{VAR(S)}} if S < 0 \end{cases} \tag{4}$$

where X_i and X_j are the time series data in the chronological order, n is the length of the time series, t_p is the number of ties for p^{th} values, and q is the number of tied values. A positive Z value implies an upward trend in the data series, while a negative value indicates a downward trend. Additionally, if $|Z| > Z_{1-\alpha/2}$, (H_o) is rejected. This shows the presence

of a statistically significant upward or downward monotonic trend in the data series. The critical value of $Z_{1-\alpha/2}$ for an alpha level of 0.05 from the standard normal table is 1.96. To reach this objective, the first step is to identify the major sectoral pattern of emission contributions to the total atmospheric CO_2 load. The magnitude of the trend is estimated by Sen's slope approach, which is the slope interpreted as a change in measurement per change in time.

$$Q_{\prime} = \frac{x_{t'} - x_j}{t' - t} \tag{5}$$

$Q^{'}$ is the slope between data points $x_t^{'}$ and x_t, which are the data measurement at time $t^{'}$ and x_i. The Sen's slope estimator is simply given by the median slope:

$$\beta = \begin{cases} Q_{\frac{N+1}{2}}{}', Nisodd \\ \frac{1}{2}\left(Q'_{\frac{N}{2}} + Q'_{\frac{N+2}{2}}\right) \end{cases}, Niseven \tag{6}$$

N is the number of calculated slopes. A positive value of β indicates an increasing trend, and a negative value indicates a decreasing trend in the time series data.

Peculiar features of time series data, which can affect the results of the MK test, are the presence of seasonality patterns and serial autocorrelation [25]. However, significant serial correlation present in time series data can be accounted for by using the modified MK test via a non-parametric block bootstrap technique, which incorporates the Mann-Kendall trend test [21,27]. The block bootstrap is a powerful approach in the presence of series that are serially autocorrelated [28,29]. This technique uses the predetermined block lengths in resampling the original time series, thus retaining the memory structure of the data. If the value of the test statistic falls in the tails of the empirical bootstrapped distribution, there is likely a trend in the data.

The existence of seasonality patterns and serial autocorrelation in the time series data was checked by plotting the autocorrelation coefficients against lags (correlogram). The autocorrelation coefficients for lag were calculated as follows [24,25].

$$r_k = \frac{\sum_{i=1}^{n-k}\left[\left(x_i - \bar{x_-}\right)\left(x_{i+k} - \bar{x_+}\right)\right]}{\left[\sum_{i=1}^{n=k}\left(x_i - \bar{x_-}\right)^2\right]^{1/2}\left[\sum_{i=k+1}^{n=k}\left(x_i - \bar{x_-}\right)^2\right]^{1/2}} \tag{7}$$

3. Statistical Analysis

The variables were summarised using descriptive statistics and one-way analysis of variance, while preliminary time series analysis was done using a time plot and the autocorrelation function (ACF) to detect patterns, seasonality, and serial autocorrelation. Each of the research questions were answered using the block bootstrap Mann-Kendall test, Sen's slope, a sequential plot, and the signed Wilcoxon test. All analysis was carried out in the R statistical software package with a concentration on the modifiedmk, trend change, and Wilcoxon functions

4. Results, Discussion, and Policy Implications

The descriptive statistics of the variables are presented in Table 2. It showed that CO_2 emissions from the transport (TRA) sector were significantly higher than those from other channels, while emissions from liquid fuel consumption (LFC)and electricity with heat production (EHP)were next to it in the rank with the former greater than the latter. In addition, CO_2 from gaseous fuel (GFC), manufacturing industries and construction (MINC), as well as residential buildings and commercial and public services (RBCPS) were statistically different from each other. However, no statistical significance was found in the CO_2 emissions from solid fuel consumption (SFC) and other sectors, excluding residential buildings and commercial and public services (OSEC). The time plot of the series presented

in Figure 1 showed that the series have a sinusoidal pattern with a significant peak and a trough at different periods indicating likelihood of the trend. The autocorrelation function (ACF) of each variable is depicted in the Figures 2 and 3. The vertical line (a "spike") on the graphs corresponds to each lag, while the height of each spike showed the value of the autocorrelation function for the lag. The spike that rises above or below the dashed lines indicated statistically significant autocorrelation at that lag. The occurrence of spikes above and below the dashed line of the ACF plot in the variables considered in the study clearly revealed the presence of serial autocorrelation. As earlier noted, the presence of serial autocorrelation posed a serious setback for the traditional Mann-Kendall test, and this can result in the detection of a false trend. One of the robust methods that account for serial autocorrelation in Mann-Kendall test is the block-bootstrapped Mann-Kendall Test (BBMKT). Hence, this resampling and modified version of MKT was used to calculate the test statistic for each of the variables. The number of bootstrapped simulations and the block length used for the test were 2000 and 5, respectively. The result of the test is presented in Table 3. The table shows the Z-value, the S-value, the Sen's slope, and the change period, which indicated the BBMKT test-statistic value, the direction of the trend, the magnitude of the trend, and the period in the series where the direction of the trend was reversed. Consequently, the null hypothesis that there is no existence of a monotonic trend in GFC, SFC, EHP, RBCPS, and MINC data cannot be rejected at 5% level of significance in that the absolute values of the test statistic in these variables were greater than 1.960. However, the presence of a monotonic trend in the CO_2 emissions from LFC, TRA, and OSEC cannot be ascertained at a 5% level. In terms of the direction of the trend, the positive value of the S-statistic for GFC, LFC, EHP, and OSEC indicated that these variables had an upward trend during the period under consideration, while emissions from SFC, TRA, RBCPS, and MINC had a downward trend. The rate of change in the trend as depicted by Sen's slope revealed that GFC, LFC, EHP, and OSEC increased annually by 63.2%, 39.6%, 47.0%, and 6.5%, respectively, while the annual rate of decline in emissions form SFC, TRA, RBCPS, and MINC were estimated to be 2.3%, 11.9%, 24.1%, and 21.8%, respectively. Figure 4 presented the sequential trend-change-detection plot. The point of intersection of the prograde and the retrograde on each plot marked the change point. The probable change point for GFC, LFC, and EHP were detected to be 1986, 1975, and 1987, respectively while that of SFC, RBCPS, and MINC were 1982, 2004, and 1984, respectively. However, the plots of TRA and OSEC intersected at several locations because there were no clear trends in them. In summary, a statistically significant monotonic trend was found in the GFC, SFC, EHP, RBCPS, and MINC data. The rates of increase in emissions from GFC and EHP were found to be 63.2% and 47.0%, respectively. This indicated that emissions from GFC accounted for about 57.4% of the increment due to it and EHP. Additionally, emissions from SFC, RBCPS, and MINC had reduced over the period under consideration with 2.3%, 24.1%, and 21.8%, respectively. This indicated that emissions from RBCPS and MINC were 10 and 9 times more likely to reduce when compared with the rate of reduction in SFC, respectively.

Table 2. Descriptive statistics of the disaggregated emissions data.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
GFC	44	16.540 ^{d}	9.721	1.091	8.390	23.443	34.411
LFC	44	43.371 ^{b}	19.251	15.518	32.507	57.404	76.819
SFC	44	0.436 ^{g}	0.537	0.009	0.088	0.471	2.071
TRA	44	47.935 ^{a}	5.060	35.389	43.818	52.239	56.305
EHP	44	27.540 ^{c}	6.718	13.587	22.983	32.363	39.062
RBCPS	44	9.853 ^{f}	4.088	2.465	6.577	12.171	17.292
MINC	44	11.493 ^{e}	3.410	4.250	9.611	13.972	18.430
OSEC	44	3.191 ^{g}	2.979	0.028	1.453	4.065	11.406

The supercrpts indicates significant difference at 5% level with a > b > c > d > e > f > g. Mean separation was done by Duncan Multiple Range Test (DMRT).

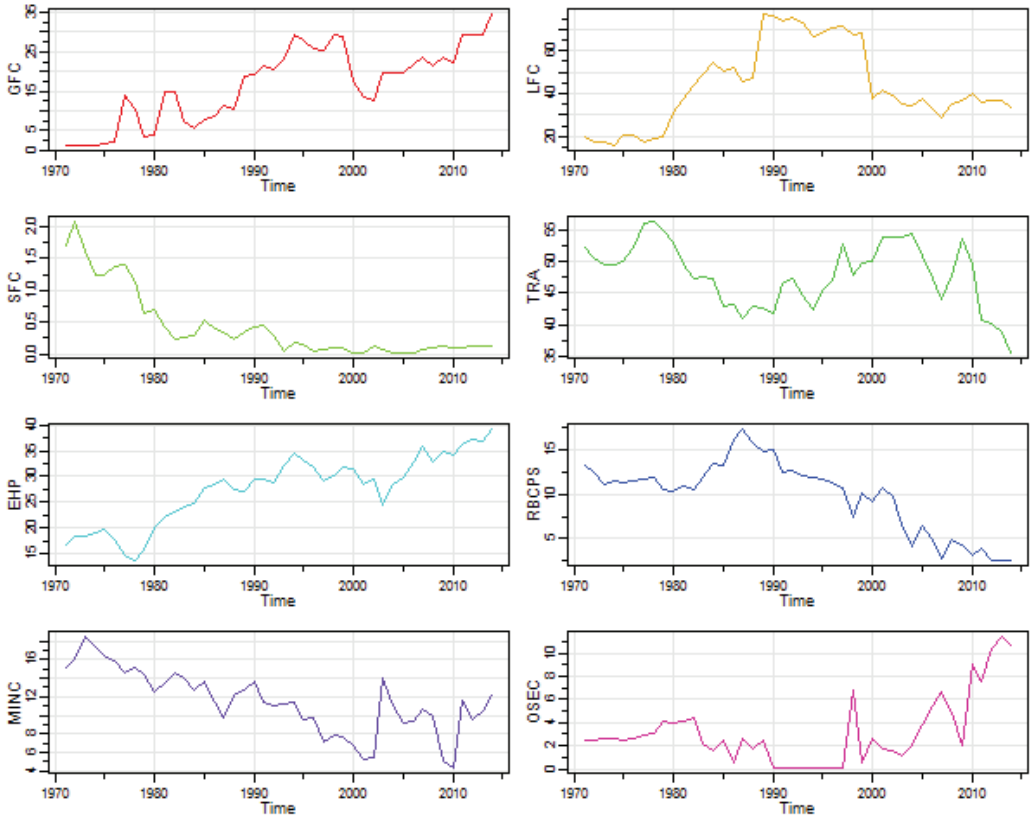


Figure 1. Time plot of CO_2 emissions data.

Table 3. Modified Mann-Kendall analysis results.

Variable	Z-Value	S-Value	Sen's Slope	Change Period
GFC	6.423	636	0.632	1986
LFC	1.224	122	0.396	1975
SFC	−5.795	−574	−0.023	1982
TRA	−1.912	−190	−0.119	Multiple
EHP	6.908	684	0.470	1987
RBCPS	−5.229	−518	−0.241	2004
MINC	−5.836	−578	−0.218	1984
OSEC	1.588	158	0.065	Multiple

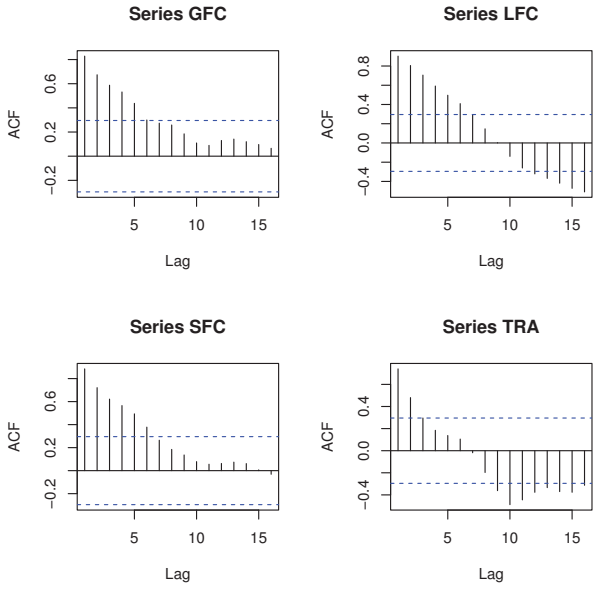


Figure 2. Autocorrelation function of CO_2 emissions data.

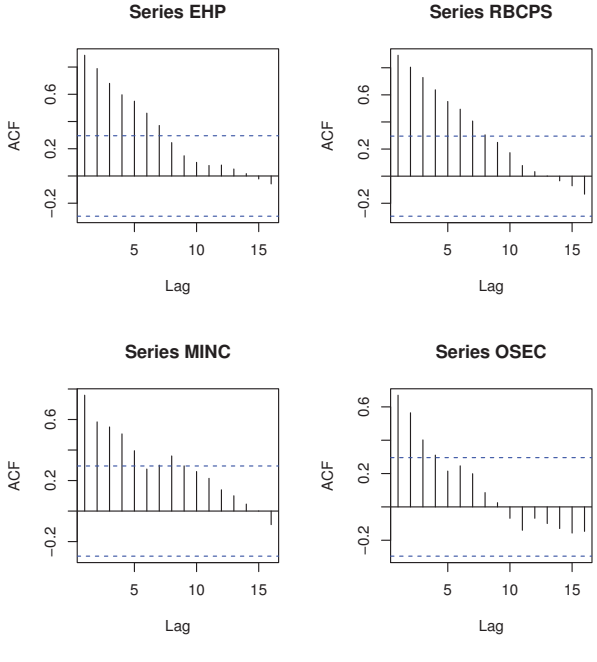


Figure 3. Autocorrelation function of CO_2 emissions data.

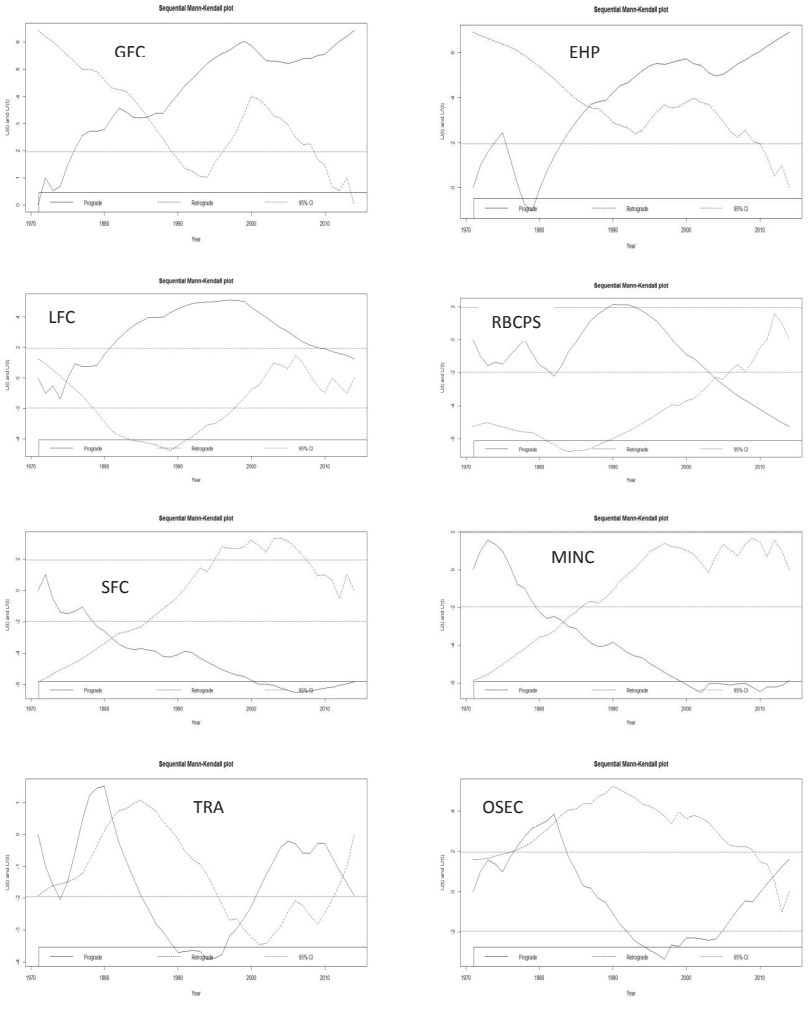


Figure 4. Sequential trend-change-detection plot.

The subject of trend detection in CO_2 emissions data has received a great deal of attention lately, especially in connection with the anticipated changes in global climate [30]. The Mann-Kendall approach used in our study allowed us to detect different trend types not only monotonically but also non-monotically. The simulation results indicate the significance of monotonic trends in CO_2 emissions from gaseous fuel consumption, solid fuel consumption, electricity and heat production, residential buildings, and manufacturing industries and construction sectors. Based on these findings, one may claim that controlling carbon emissions from the aforementioned sectors mainly depends on monitoring the inherent monotonic trends. As a result, the forecasting exercise of CO_2 emissions trends in these sectors can be improved based on information about the monotonic trends of CO_2 emissions. However, the presence of non-monotonic trends in CO_2 emissions stemming from liquid fuel consumption, the transport sector, and other sectors excluding residential buildings make it difficult to predict the future trends of CO_2 emissions based on the dynamics of CO_2 emissions in these sectors. This is partly due to the complex structure of CO_2 emissions in these sectors [30–34]. By looking at the direction of the trends, we

observed that GFC, LFC, EHP, and OSEC displayed an upward trend with annual rates of 63.2%, 39.6%, 47.0%, and 6.5%, respectively. Governments should improve energy efficiency in the gas fuel consumption sector, the liquid fuel consumption sector, fuel combustion from other sectors, and electricity and heat-production sectors and should save energy by converting the current raw fuel sources from heavy to light oil, promoting research and development activities of low-carbon fuels, particularly those with fuel consumption. This will reverse the current positive trend observed in fuel consumption. Another key finding is that the annual rate of decline in emissions from SFC, TRA, RBCPS, and MINC were estimated to be 2.3%, 11.9%, 24.1%, and 21.8%, respectively. Despite evidence of decreasing trends, it is important to indicate that the rate of decrease is relatively lower compared to the annual increasing rate of GFC, LFC, EHP, and OSEC. This indicates that further efforts are required to formulate appropriate incentives to increase energy conservation and reduce emissions from SFC, TRA, RBCPS, and MINC—for example, by creating a fuel standard for public transport and particular vehicles.

5. Conclusions

The study focused on the trend analysis of carbon dioxide emissions, which was disaggregated into various components based on the sources and sectors of the economy generating them. The modified Mann-Kendall test, which corrects for the presence of a seasonal effect via block bootstrap was used to study the likelihood of a monotonic trend as well as its direction, magnitude, and change point in the dataset. Based on the results, gaseous fuel, solid fuel, liquid fuel, etc. all showed an upward trend and breaks at various periods covered in the study. However, no specific break can be established in emissions from transportation. The conclusion from the study suggests that CO_2 emissions from various sectors have maintained an upward trend and this portends serious health implications and environmental hazards. This most-dangerous and prevalent greenhouse gas is the major cause of climate change, which results in food-supply disruptions, extreme water, and increased wildfires, as well as respiratory diseases through smog and air pollution. An industrialized country like Nigeria needs an enhanced Carbon-Dioxide Emissions Reduction Strategy (CDERS) to reverse the growth rate. This underscores the fact that all human activities that trigger this leading greenhouse gas have to be discouraged. It is also paramount to note that there are few or no sectors of the economy globally that do not contribute to greenhouse gas production. Ranging from manufacturing to agriculture to transportation to power production and so on, all release greenhouse gases to the atmosphere, and reduction in all emissions can be achieved through advanced practices that deviate from fossils fuels. Another potent dimension of CDERS is the use of technologies that decrease greenhouse gas emissions, and this includes swapping fossil fuels for renewable sources, boosting energy efficiency, and discouraging carbon emissions by putting a price on them. Bearing in mind sustainable development, the improvement in energy efficiency and a workable energy supply system with low or no CO_2 is imperative. The results of this study create a scenario for a better understanding of CO_2 emissions in Nigeria as well as the knowledge of emission trends. As part of CO_2-reduction measures, there is a need for development of a framework for emissions control, proper ecosystem balance, utilization of greenbelts, and the development of control measures for the mitigation of anthropogenic CO_2 emissions. The energy-efficiency strategies and conservation practices should also be considered for future CO_2 emission reduction.

Author Contributions: Conceptualization: O.O.A. and O.L.T.; methodology: O.O.A. and O.L.T.; software: O.O.A. and Y.K.; validation: O.O.A. and O.L.T.; formal analysis: O.O.A.; resources: M.A. and I.S.; data curation: O.O.A. and Y.K.; writing—original draft preparation: all authors; writing—review and editing: O.L.T., I.S., M.A., and Y.K.; visualization: O.O.A.; supervision: O.L.T. and M.A.; funding acquisition: I.S. and M.A. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used for this study can be accessed publicly at https://databank.worldbank.org/source/world-development-indicators.

Acknowledgments: This publication was supported by the construction EFOP-3.6.3-VEKOP-16-2017-00007 ("Young researchers from talented students—Supporting scientifc career in research activities in higher education"). The project was supported by the European Union and co-financed by the European Social Fund.

Conflicts of Interest: The authors declared no competing interest.

References

1. Akpan, G.E. Electricity Consumption, Carbon Emissions and Economic Growth in Nigeria. *Int. J. Energy Econ. Policy* **2012**, *2*, 292–306.
2. Zhang, L.; Liu, B.; Du, J.; Liu, C.; Wang, S. CO_2 emission linkage analysis in global construction sectors: Alarming trends from 1995 to 2009 and possible repercussions. *J. Clean. Prod.* **2019**, *221*, 863–877. [CrossRef]
3. UN (2012). UN (United Nations) Economic and Social Affairs, 2013. World Population Prospects: The 2012 Revision. New York. Available online: http://esa.un.org/unpd/wpp/ (accessed on 13 July 2012).
4. Santos, G. Road transport and CO_2 emissions: What are the challenges? *Transp. Policy* **2017**, *59*, 71–74. [CrossRef]
5. Fontaras, G.; Zacharof, N.G.; Ciuffo, B. Fuel consumption and CO_2 emissions from passenger cars in Europe Laboratory versus real-world emissions. *Prog. Energy Combust. Sci.* **2017**, *60*, 97–131. [CrossRef]
6. IEA. CO_2 Emissions from Fuel Combustion by Sector in 2014, in CO_2 Emissions from Fuel Combustion, IEA, 2016. In CO_2 Highlights 2016. Excel Tables. 2016. Available online: http://www.iea.org/publications/freepublications/publication/CO$_{2}$-emissions-from-fuelcombustion-highlights-2016.html (accessed on 5 November 2020).
7. Gottesfeld, P.; Pokhrel, A.K. Review: Lead exposure in battery manufacturing and recycling in developing countries and among children in nearby communities. *J. Occup. Environ. Hyg.* **2011**, *8*, 520–532. [CrossRef]
8. Bhuyan, M.D.; Islam, M.M.; Bhuiyan, M.E.K. A Trend Analysis of the Temperature and Rainfal to predict Climate Change for Northwestern Region of Bangladesh. *Am. J. Clim. Chang.* **2018**, *7*, 115–134. [CrossRef]
9. Aydin, G. The Modeling of Coal-related CO_2 Emissions and Projections into Future Planning. *Energy Sources Part Recover. Util. Environ. Eff.* **2013**, *36*, 191–201. [CrossRef]
10. Köne, A.I.; Büke, T. Forecasting of CO_2 emissions from fuel combustion using trend analysis. *Renew. Sustain. Energy Rev.* **2010**, *14*, 2906–2915. [CrossRef]
11. Kwon, T.H. Decomposition of factors determining the trend of CO_2 emissions from car travel in Great Britain (1970–2000). *Ecol. Econ.* **2005**, *53*, 261–275. [CrossRef]
12. Li, Q.; Zhang, W.; Li, H.; He, P. CO_2 emission trends of China's primary aluminum industry: A scenario analysis using system dynamics model. *Energy Policy* **2017**, *105*, 225–235. [CrossRef]
13. Wu, R.; Wang, J.; Wang, S.; Feng, K. The drivers of declining CO_2 emissions trends in developed nations using an extended STIRPAT model: A historical and prospective analysis. *Renew. Sustain. Energy Rev.* **2021**, *149*, 111328. [CrossRef]
14. Zhang, X.; Zhang, H.; Yuan, J. Economic growth, energy consumption, and carbon emission nexus: Fresh evidence from developing countries. *Environ. Sci. Polution Res.* **2019**, *26*, 1090–1094. [CrossRef]
15. Miura, T.; Tamaki, T.; Kii, M.; Kajitani, Y. Efficiency by sectors in areas considering CO_2 emissions: The case of Japan. *Econ. Anal. Policy* **2021**, *70*, 514–528. [CrossRef]
16. Salmi, T.; Maatta, A.; Anttila, P.; Ruoho-Airola, T.; Amnell, T. *Detecting Trends of Annual Values of Atmospheric Pollutants by the Mann-Kendall Test and Sens Slope Estimates. The Excel Template Application Makesens*; Air Quality; Quality No. 31; Finnish Meteorological Institute: Helsinki, Finland, 2002.
17. Hirsch, R.M.; Alexander, R.B.; Smith, R.A. Selection methods for the detection and estimation of trends in water quality. *Water Resour. Res.* **1991**, *27*, 803–813. [CrossRef]
18. Helsel, D.R.; Hirsch, R.M. Statistical Methods in Water Resources. In *Techniques of Water Resources Investigations, Book 4, Chapter A3*; Geological Survey: Reston, VA, USA, 2002.
19. Nalley, D.; Adamowski, J.; Khalil, B.; Ozga-Zielinski, B. Trend detection in surface air temperature in Ontario and Quebec, Canada during 1967–2006 using the discrete wavelet transform. *Atmos. Res.* **2013**, *132–133*, 375–398. [CrossRef]
20. Araghi, A.; Mousavi-Baygi, M.; Adamowski, J. Detection of trends in days with extreme temperatures in Iran from 1961 to 2010. *Theor. Appl. Climatol.* **2016**, *125*, 213–225. [CrossRef]
21. Kendall, M.G. *Rank Correlation Methods*; Charles Griffin: London, UK, 1995.
22. Martinez, C.J.; Maleski, J.J.; Miller, M.F. Trends in precipitation and temperature in Florida, USA. *J. Hydrol.* **2012**, *452–453*, 259–281. [CrossRef]
23. Araghi, A.; Mousavi-Baygi, M.; Adamowski, J.; Malard, J.; Nalley, D.; Hasheminia, S.M. Using wavelet transforms to estimate surface temperature trends and dominant periodicities in Iran based on gridded reanalysis data. *Atmos. Res.* **2015**, *155*, 52–72. [CrossRef]

24. Araghi, A.; Mousavi-Baygi, M.; Adamowski, J.; Martinez, C.J. Analyzing trends of days with low atmopsheric visibility in Iran during 1968–2013. *Environ. Monit. Assess.* **2019**, *191*, 249–263. [CrossRef] [PubMed]
25. Wilks, D.S. Statistical methods in the atmospheric science. In *International Geophysics*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2014
26. Safari, B. Trend Analysis of the Mean Annual Temperature in Rwanda during the Last Fifty Two Years. *J. Environ.* **2012**, *3*, 538–551. [CrossRef]
27. Kundzewicz, Z.W.; Robson, A.J. *Detecting Trend and other Changes in Hydrological Data, World Climate Program-Data and Monitoring*; WMO/TD-No. 1013; World Meteorological Organization: Geneva, Switzerland, 2000; Volume 45, pp. 1–158.
28. Khaliq, M.N.; Quarda, T.B.M.; Gachon, P.; Sushama, L.; St-Hilaire, A. Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. *J. Hydrol.* **2009**, *368*, 117–130. [CrossRef]
29. Onoz, B.; Bayazit, M. Block bootstrap for Mann-Kendall trend test of serially dependent data. *Hydrol. Process.* **2012**, *26*, 3552–3560. [CrossRef]
30. Paraschiv, S.; Paraschiv, L.S. Trends of carbon dioxide (CO_2) emissions from fossil fuels combustion (coal, gas and oil) in the EU member states from 1960 to 2018. *Energy Rep.* **2020**, *6*, 237–242. [CrossRef]
31. Andreoni, V.; Galmarini, S. European CO_2 emission trends: A decomposition analysis for water and aviation transport sectors. *Energy* **2012**, *45*, 595–602. [CrossRef]
32. Bilgili, F.; Kuşkaya, S.; Gençoğlu, P.; Kassouri, Y.; Garang, A.P.M. The co-movements between geothermal energy usage and CO_2 emissions through high and low frequency cycles. *Environ. Sci. Pollut. Res.* **2020**, *28*, 63723–63738. [CrossRef] [PubMed]
33. Kassouri, Y.; Kacou, K.Y.T.; Alola, A.A. Are oil-clean energy and high technology stock prices in the same straits? Bubbles speculation and time-varying perspectives. *Energy* **2021**, *232*, 121021. [CrossRef]
34. Song, Y.; Zhang, M.; Shan, C. Research on the decoupling trend and mitigation potential of CO_2 emissions from China's transport sector. *Energy* **2019**, *183*, 837–843. [CrossRef]

energies

MDPI

Article

Wind Speed Prediction for Offshore Sites Using a Clockwork Recurrent Network

Yuxuan Shi *, Yanyu Wang and Haoran Zheng

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China; wangyanyu@shu.edu.cn (Y.W.); zhrzhr@shu.edu.cn (H.Z.)
* Correspondence: shiyuxuan@shu.edu.cn

Abstract: Offshore sites show greater potential for wind energy utilization than most onshore sites. When planning an offshore wind power farm, the speed of offshore wind is used to estimate various operation parameters, such as the power output, extreme wind load, and fatigue load. Accurate speed prediction is crucial to the running of wind power farms and the security of smart grids. Unlike onshore wind, offshore wind has the characteristics of random, intermittent, and chaotic, which will cause the time series of wind speeds to have strong nonlinearity. It will bring greater difficulties to offshore wind speed predictions, which traditional recurrent neural networks cannot deal with for lacking in long-term dependency. An offshore wind speed prediction method is proposed by using a clockwork recurrent network (CWRNN). In a CWRNN model, the hidden layer is subdivided into several parts and each part is allocated a different clock speed. Under the mechanism, the long-term dependency of the recurrent neural network can be easily addressed, which can furthermore effectively solve the problem of strong nonlinearity in offshore speed winds. The experiments are performed by using the actual data of two different offshore sites located in the Caribbean Sea and one onshore site located in the interior of the United States, to verify the performance of the model. The results show that the prediction model achieves significant accuracy improvement.

Keywords: clockwork recurrent network; offshore site; strong nonlinearity; wind speed prediction

Citation: Shi, Y.; Wang, Y.; Zheng, H. Wind Speed Prediction for Offshore Sites Using a Clockwork Recurrent Network. *Energies* **2022**, *15*, 751. https://doi.org/10.3390/en15030751

Academic Editors: Luis Hernández Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 2 January 2022
Accepted: 15 January 2022
Published: 20 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the increasingly severe global climate problem, the sustainability of traditional fossil fuels is facing huge challenges, and the development of renewable energy (RE) is becoming inevitable [1]. RE, including wind energy, geothermal energy, and solar energy, cannot only reduce carbon emissions, but also achieve sustainable development [2,3]. As one form of RE, wind energy is widely used around the world on account of its wide distribution, huge reserves, and environmental friendliness [4]. At the same time, wind power is also one of the most commercially viable and dynamic RE sources due to its low cost and permanent nature. On account of its relatively mature technology and commercial conditions for large-scale development, wind energy has been the fastest growing energy source in recent years. [5]. According to the data from the Global Wind Energy Council, global wind power is accelerating its deployment, driven by the carbon-neutral trend. The latest data show that the total global wind power bidding volume in the first quarter of 2021 is 6970 MW, 1.6 times that of the same period last year [6].

However, wind energy resources are susceptible to environmental changes, such as geography, climate, and seasons. It brings great difficulties to wind power utilization. In addition, the ecological problem with wind power is that it may disturb birds. Therefore, accurate offshore wind speed prediction is of great help to the development of wind power. However, there are still some factors that affect the prediction accuracy, among which the major challenge is historical data. Regrettably, potential offshore sites have not had enough records of wind speed for various reasons in the past. Consequently, it is a major

technical challenge for risk assessment using only short-term records of historical wind speed data. Nevertheless, unlike onshore wind, offshore wind has the characteristics of random, intermittent, and chaotic, which will cause the time series of wind speeds to have strong nonlinearity [7], inevitably bringing greater difficulties to offshore wind speed predictions.

Within past studies, scholars have proposed various wind speed prediction methods. There are three main categories, including physical models, statistical models, and machine learning models. Physical models make predictions by monitoring the terrain, climate, and other factors. Among the physical models, numerical weather prediction (NWP) is a commonly used model that simulates physical interactions in the atmosphere based on conservation equations (kinetic energy, potential energy, and mass) [8,9]. However, different locations and fields bring about variability in the NWP models and their model resolutions. The resolution of the model data seriously affects the prediction accuracy and the datasets are hard to obtain [10]. Statistical models mainly use historical data to make predictions. The commonly used statistical models are Gaussian process regression (GPR) [11,12], autoregressive (AR) [13], autoregressive moving average (ARMA) [14], autoregressive integral moving average (ARIMA) [15], and seasonal ARIMA [16]. However, when the nonlinear characteristics are prominent, the prediction performance of these models decreases significantly [17]. Comparatively, machine learning is often performed to predict wind speed because of its ability to fit stronger nonlinearity, which includes the multi-layer perceptron (MLP) [18], back propagation neural network (BPNN) [19], radial basis function neural network (RBFNN) [20], support vector machine (SVM)/support vector regression (SVR) [21–26], echo state network [27], deep belief networks [28], and convolutional neural network (CNN) [29]. However, these models still have various problems in their application, such as getting stuck in local optimum solutions, overfitting, and low convergence rates.

Recently, the recurrent neural network (RNN) is proposed to model sequential data or time series data [30]. RNN, as a type of artificial neural network that uses a simple but elegant mechanism, addresses the drawback of vanilla neural networks and keeps the characteristic of the autoregressive model. It brings to RNN the ability to solve the nonlinear problem of time series data. Therefore, RNNs achieve great performances when modeling sequential data and have become one of the most valuable breakthroughs in deep learning model preparation in recent decades. Meanwhile, many studies on wind speed prediction have emerged in recent years, which use RNN models [30,31] or hybrid RNN models [32–36]. At the same time, researchers constantly optimized the network structure of the RNN to improve its performance. Several new models based on RNNs, such as long and short term memory networks (LSTMs) [37–48], bidirectional LSTM (BiLSTM) [49], gated recurrent units (GRUs) [50], clockwork recurrent neural networks (CWRNNs) [51], and dilated recurrent neural networks (DRNNs) [52], have been proposed to solve problems of RNN, including vanishing gradients and the long-term dependency, and improve the performance of RNNs.

CWRNN, which adopts a special mechanism to solve problems of simple RNNs and contains an even smaller number of parameters than simple RNNs, was proposed in 2014 [53]. CWRNN breaks up neurons in the hidden layer into different parts, and neurons in the same part work at a given clock speed. At the same time, only a few parts are activated. It makes CWRNN have a certain memory mechanism that can solve the long-term dependency problem. Additionally, it has shown better performances than common RNNs and even LSTM in various tasks. Xie et al. applied CWRNN to muscle perimysium segmentation. They utilized CWRNN to handle biomedical data, and experiment results show that CWRNN outperforms the other machine learning models [54]. Feng et al. used CWRNN to estimate the state-of-charge of lithium batteries and showed that this method achieves impressive results [51]. Lin et al. proposed a trajectory generation method for unmanned vehicles based on CWRNN. The performance of the CWRNN method is verified by experiments. The study also compared CWRNN with LSTM in several

metrics [55]. Achanta et al. investigated CWRNN for statistical parametric speech synthesis. The experimental results show that the architecture of the CWRNN is equivalent to the RNN with LI units, and outperforms the RNN with dense initialization and LI units [56]. Presently, the methods based on CWRNN have been used in various fields, such as speech recognition and stock prediction [57]. As far as we know, it has not been used in wind speed prediction.

To solve the strong nonlinear problem and achieve a higher prediction accuracy, an offshore wind speed prediction method is proposed, which is based on the CWRNN. In the proposed method, the hidden layer is subdivided into several parts and each part is allocated a different clock speed. Under the mechanism, the long-term dependency of RNNs can be easily addressed. The trained CWRNN model can output an instantaneous prediction for data from the previous sampling step. The experiments are performed to validate the performance of the model by the actual wind speed data of two different offshore sites and one onshore site.

The main contributions of this study are as follows:

- An offshore wind speed prediction method is proposed based on the CWRNN. Compared with the other RNNs, the CWRNN adopts a special mechanism to solve long-term dependency. The experiments prove that the method can effectively solve the problem of strong nonlinearity in offshore wind speed, and improve the prediction accuracy by over 38% in terms of the different kinds of evaluation criteria compared with simple RNNs.
- The method fully exploits the ability of RNNs to solve nonlinear problems with time series data. Compared with the traditional machine learning models, the proposed method keeps the characteristics of the autoregressive model, which improves the performance in prediction accuracy.
- Hyperparameters, such as the number of network parts that are the key influencing factors of the model, and the different part periods are thoroughly analyzed, which seriously affect the performance of predicting the offshore wind speed.

The rest of the paper is organized as follows: Section 2 introduces the related theory; Section 3 describes the overall implementation process of this method; Section 4 presents the experiment results; the results are discussed in Section 5; and Section 6 summarizes the whole paper.

2. Theoretical Background

There is an inherent concept of sequential data that incrementally progresses over time. As we all know, traditional neural networks (NNs) are good at solving nonlinear problems and perform well in most cases. However, they lack the inherent trend for the persistence of sequential data. For example, a simple feedforward NN cannot really understand the meaning of a sentence according to the order of input data in the context. The RNNs settle the shortcomings of the original NNs with an ingenious mechanism, which gives them the advantage in time modeling. This section provides a brief overview of the RNN, LSTM, and CWRNN.

2.1. RNN

RNN is a specific NN that is designed to model sequential data or time- series data. The principle of RNN is to feed the output of the previous layer back to the input of the next layer, which gives RNN the ability to predict the output of the layer. In the RNN, the neurons in different layers of the NN are compressed into a single layer, as shown in Figure 1.

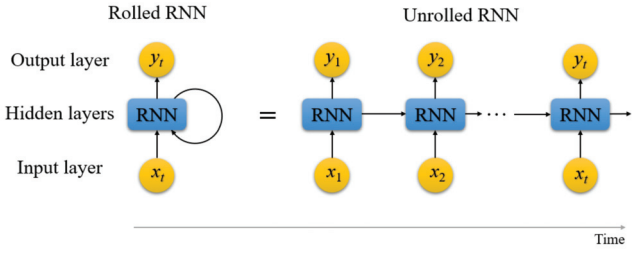


Figure 1. The structure of a simple RNN.

As seen in Figure 1, at time t, the input is a combination of the input at t and the output at a previous time, $t-1$. This feedback mechanism improves the output of the time step t. The calculation formula for output y_t^O at time step t is:

$$y_t^H = f_H\left(W_H \cdot y_{t-1}^H + U_I \cdot x_t\right) \tag{1}$$

$$y_t^O = f_O\left(W_O \cdot y_{t-1}^H\right) \tag{2}$$

where W_H, U_I,W_O are the weight matrices of the hidden layers, input layer, and output layer; x_t is defined as the input vector at t; and y_t^H and y_{t-1}^H are defined as the hidden neurons at different times. $f_H(\cdot)$ and $f_O(\cdot)$ are defined as different activation functions. Here, the biases of the neurons are omitted.

RNNs must use a context when making predictions and, in this case, must also learn the required context. The shortcoming of the RNN is that, when training the model, the gradient can easily vanish or explode, which is mainly because of the lack of long-term dependency. Researchers proposed some techniques to solve the problems, such as LSTM, which uses a gate mechanism.

2.2. LSTM

LSTM, as a special type of RNN, can keep long-term information from the input sequence, which makes up for the difficulties of RNN in learning long-term information, and solves the problems of RNN gradient disappearance and gradient explosion. The framework of the LSTM unit is shown in Figure 2. LSTM and RNN have the same chain structures, but their repeating modules are different. Unlike the repeating module in a standard RNN that contains a single layer, LSTM has multiple layers of neurons. These neurons constitute the forgetting gate, the input gate, and the output gate of LSTM. The status updates and output updates for the three gates are described below.

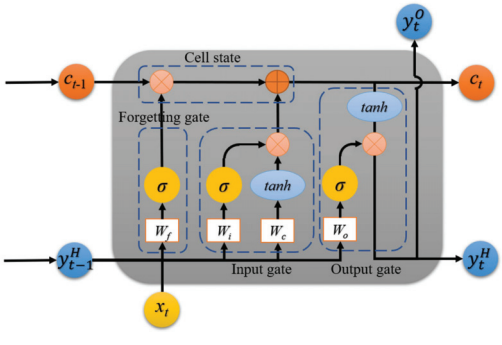


Figure 2. The structure of LSTM.

Forgetting gate: this gate control unit determines how much information the cell state discards. The status update, f_t, of the forgetting gate at the time, t, is as follows:

$$f_t = f_O\left(W_f \cdot y_{t-1}^H + U_I^f \cdot x_t\right) \tag{3}$$

where W_f is defined as the weight matrix of the forgetting gate, and U_I^f is defined as the weight matrix between the hidden layer of the forgetting gate and the input layer.

Input gate: this gate control unit determines to what extent the input information, x_t, at the current moment is added to the memory cell stream. The status update, i_t, of the input gate is as follows:

$$i_t = f_O\left(W_i \cdot y_{t-1}^H + U_I^i \cdot x_t\right) \tag{4}$$

where W_i is defined as the weight matrix of the input gate, and U_I^i is the weight matrix between the hidden layer of the input gate and the input layer.

After the work of the input gate and the forgetting gate is completed, the state of the memory cells, c_t, is updated as follows:

$$\widetilde{c_t} = f_H\left(W_c \cdot y_{t-1}^H + U_I^c \cdot x_t\right) \tag{5}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \widetilde{c_t} \tag{6}$$

where W_c represents the weight matrix of the memory cells, and U_I^c is the weight matrix between the hidden layer of the memory cells and the input layer.

Output gate: after the internal memory cell state is updated, the output gate controls how much memory can be used in the network update at the next moment. The state update, o_t, of the output gate at the time, t, is as follows:

$$o_t = f_O\left(W_o \cdot y_{t-1}^H + U_I^o \cdot x_t\right) \tag{7}$$

where W_o is defined as the weight matrix of the output gate; U_I^o is the weight matrix between the hidden layer of the output gate and the input layer; and b_o represents the offset.

Finally, the network output at moment t is:

$$y_t^H = o_t \cdot f_H(c_t) \tag{8}$$

$$y_t^O = f_O\left(W_O \cdot y_t^H\right) \tag{9}$$

To alleviate the gradient exploding and vanishing problems, an LSTM block that embeds three gates into the hidden neurons of the RNN is generally applied to process the time series data, and achieves a good result in most cases. It is easier to understand that the complex network structure increases the stability and ability of the model. However, it also makes the network computationally more expensive. Meanwhile, the performance of the complex deep learning neural network models, especially LSTMs, depends on the quantity and diversity of the data.

2.3. CWRNN

The structure of the CWRNN is close to that of a simple RNN with three layers. The difference between these two models is that the CWRNN divides the neurons of the hidden layers into n parts; each part has a clock speed, T_i, where $T_i \in \{T_1, T_2, \cdots, T_n\}$. Therefore, each part handles the input data at a different frequency, as shown in Figure 3. The parts with a long clock speed can handle long-term information, and the parts with a short clock speed are used to handle the continuous information.

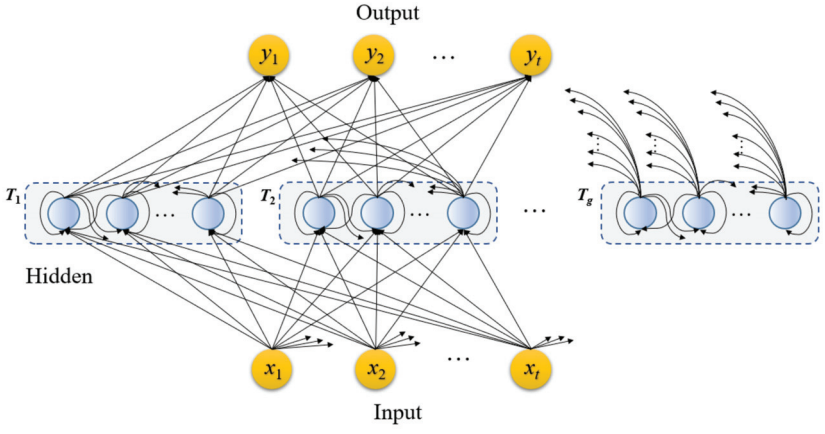


Figure 3. The framework of the CWRNN.

W_H and W_i are defined as the weight matrices of the hidden and input layers, respectively, which are divided into n blocks. At the same time, W_H is also an upper triangular matrix, as shown in Figure 4. At any time step, t, only the related rows of the work parts W_H and W_i are activated. Then, the output vector, y_H, was updated in the same way. The other parts keep the output values unchanged. The update mechanism is shown in Figure 4.

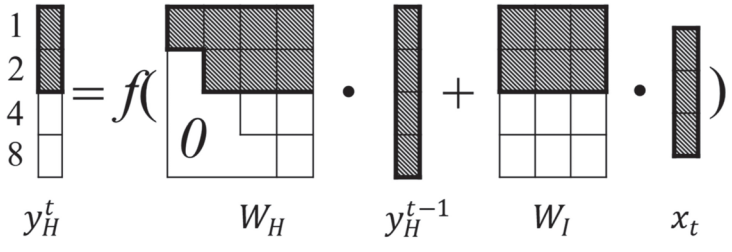


Figure 4. Update process of the hidden units at $t = 6$.

$$W_H = \begin{pmatrix} W_{H_1} \\ \vdots \\ W_{H_n} \end{pmatrix} \quad W_i = \begin{pmatrix} W_{I_1} \\ \vdots \\ W_{I_n} \end{pmatrix} \tag{10}$$

$$W_{H_i} = \begin{cases} W_{H_i} & \text{for } (t \text{ MOD } T_i) = 0 \\ 0 & otherwise \end{cases} \tag{11}$$

Therefore, the parts with a long clock-speed handle the long-term information, and the parts with a short clock-speed handle the continuous information. The two parts are independent of each other and work well.

Having the same number of hidden neurons, the CWRNN processes much faster than a simple RNN, because only the corresponding parts are updated at each step. In the case of this exponential clock setting, when $n > 4$, the CWRNN can run faster than the RNN, which has the same neurons [53].

3. Framework of the Prediction Method

3.1. The Procedure

The framework of the proposed method is described in Figure 5. The procedure is divided into four steps.

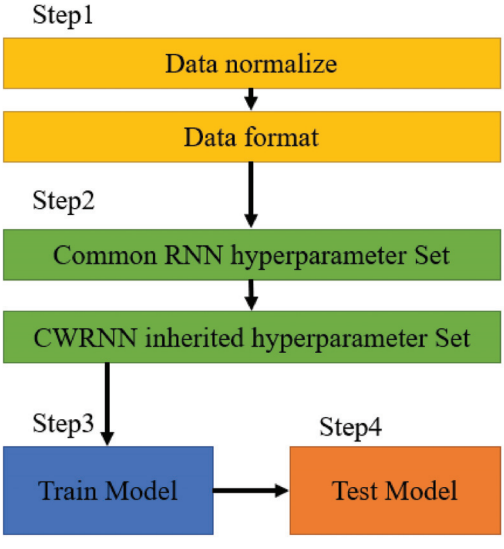


Figure 5. The procedure of the proposed method.

Step 1: data processing. Wind speed raw data are normalized to [0, 1] at first, then preprocessed to the format required for the CWRNN model.

Step 2: model setting. The hyperparameters are set to fit the model, including the hidden layer parts, length of series input, and number of neurons. The influence of these hyperparameters will be discussed later, in detail.

Step 3: train model. For model training, we used a mini-batch stochastic gradient descent and Adam optimizer to minimize the mean square error (MSE) for the prediction vectors. The parameters can be trained through the back propagation of standard error.

Step 4: model test. Some prediction and evaluation indexes of the training model, such as the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2), are performed to verify the prediction performance.

3.2. Dataset

The experimental datasets are from three wind speed measure sites, among which two are located offshore in the Virgin Islands, between the Atlantic Ocean and the Caribbean Sea, and the other onshore site is located in Humeston, Iowa, U.S.A. [58,59]. This study first conducts experiments on two offshore wind speed datasets to verify the proposed model, and then conducts experiments on the onshore wind speeds to verify the generalization of the model. Three data sets and their division in the model are described in Figure 6. The data are collected from 2012–2014. The sampling period in the data set is 10 min and each dataset has 3000 points. Table 1 shows the data of the wind speed at three different locations. It depicts the minimum, average, maximum, and standard deviation values (Stdev).

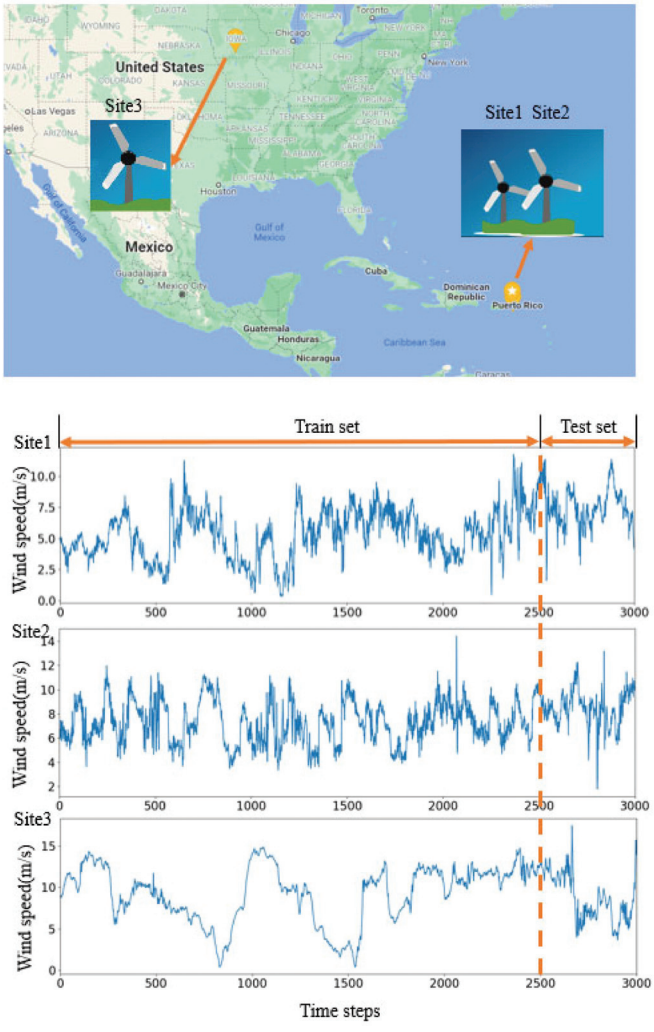


Figure 6. Datasets of Site1, Site2, Site3, and the data segmentation method.

Table 1. Data statistics on the wind speed at the three locations.

Site	Metrics			
	Average(m/s)	Maximum(m/s)	Minimum(m/s)	Stdev (m/s)
Site1	5.6655	11.7630	0.3600	2.0553
Site2	7.4647	14.4030	1.8014	1.7486
Site3	9.1397	17.4560	0.3870	3.3416

3.3. Evaluation Metrics

To quantitatively describe the performance of all the methods, four different indicators, MAE, MAPE, RMSE, and R^2, are used to analyze the results. The calculation formula of each indicator is shown in Table 2. For all the formulas, y_i is the true value, $\hat{y}_i$ is the predicted value, $\overline{y}_i$ is the average of the samples, and N is the length of the samples.

Table 2. Calculation formulas for the four evaluation indicators of the experiment.

Evaluation Metrics	Equations
MAE	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
MAPE	$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert$
RMSE	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2}$

4. Results

The proposed method was programmed with Python using Tensorflow and Keras. The following results and discussions were accomplished on a laptop computer with a system of Windows 10, an Intel Core i5-1135G7 @2.40 GHz, and 16 GB of memory. The source codes of the baseline models will be publicly available on the website [60].

4.1. Comparison with the RNNs

In reference [53], the CWRNN demonstrates that it outperforms both the RNN and LSTM networks in the experiments. In this study, to verify the advantages of CWRNNs, three other RNN models, including simple RNNs, LSTMs, and BiLSTMs, were used to make offshore wind speed predictions. The same dataset was used to train and evaluate the models. All the models have the same hyperparameters, which are shown in Table 3. The prediction results are shown and described in Figure 7 and Table 4.

Table 3. The numerical metrics of the prediction results by CWRNNs and RNNs of Site1.

Hyperparameters	Settings (All Models, including the RNN, LSTM, BiLSTM, and CWRNN, have the Same Hyperparameters)
Input numbers	60
Hidden layers	1
Hidden neurons	200
Dense layers	1
Optimizer	RMSprop
Learning rate	10^{-3}
Epoch	200
Batch size	100

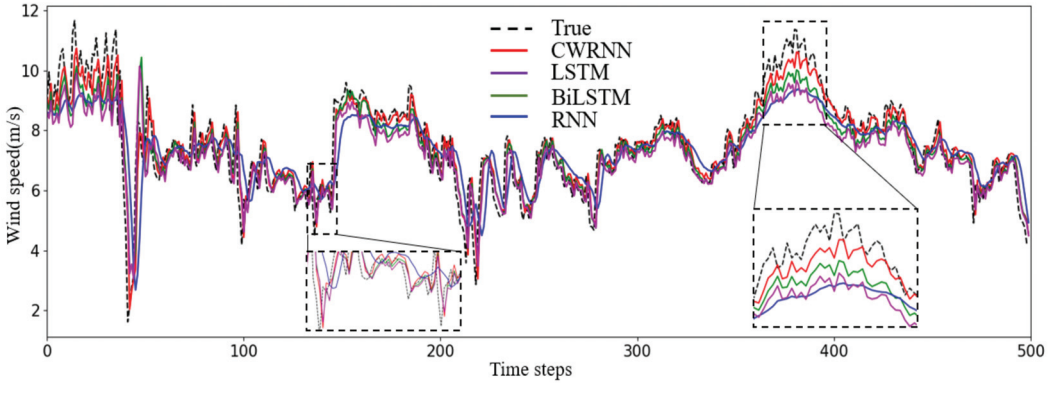


Figure 7. Comparison results of the proposed model with RNNs of Site1.

Table 4. The numerical metrics of the prediction results by CWRNNs and RNNs of Site1.

Model	Parameters	Run Times (Mean Value of 10 Times)	Evaluation Metrics (Mean Value of 10 Times)			
			MAE	MAPE	RMSE	R^2
Simple RNN	40,601	128.3919 s	0.7207	10.8733	1.0116	0.5988
LSTM	161,801	743.7911 s	0.6222	8.5401	0.8304	0.7296
BiLSTM	323,601	1666.8021 s	0.5443	7.8204	0.7551	0.7764
CWRNN	40,801	77.7866 s	0.4572	6.7873	0.6566	0.8310

As shown in Figure 7, compared with the true data for Site1, the prediction curves of all the RNNs are close to the real curve of the true wind speed data, which means they have all captured the tendency of true wind speed. It relies on the powerful ability of RNNs in a modeling time series. In contrast to other RNNs, the prediction curve of the CWRNN appears to be closer to the real curve, which verifies that the CWRNN has a better performance in solving strong nonlinear problems.

Table 4 lists the corresponding MAE, MAPE, RMSE, and R^2 values. The indexes of the RNN are the worst because the RNN cannot remember long-term dependency due to the vanishing gradient. In comparison to the other RNNs, CWRNNs achieves great accuracy, with lower MAE, MAPE, RMSE and higher R^2. Furthermore, it can be observed from Table 4 that the CWRNN almost has the same parameters as the simple RNN, but the LSTM and BiLSTM have large parameters, which are computationally expensive; hence, the LSTMs are slow, which is also shown in Table 5. In comparison to all the RNNs, the CWRNN resulted in fewer runtimes because only parts were updated at every step.

Table 5. Average and standard deviation of prediction results by the CWRNNs and RNNs of Site1.

Model	Evaluation Metrics							
	MAE		MAPE		RMSE		R^2	
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
Simple RNN	0.7207	0.0974	10.8733	1.5279	1.0116	0.0896	0.5988	0.0749
LSTM	0.6222	0.0539	8.5401	0.6652	0.8304	0.0715	0.7296	0.0492
BiLSTM	0.5443	0.0334	7.8204	0.5151	0.7551	0.0362	0.7764	0.0217
CWRNN	0.4572	0.0044	6.7873	0.0311	0.6566	0.0044	0.8310	0.0023

Table 5 shows the mean and standard deviation values of the metrics of the prediction results. All the metrics data in the following figures are the average of 10 times.

As shown in Figure 8, compared with the true data of site2, the same conclusion as Site1 can be obtained. Compared with the other RNNs, the prediction curve of the CWRNN still appears to be closer to the real curve, by which the performance of the CWRNN has been verified again. These numerical results can also be obtained from Table 6. Compared with the other RNNs, the CWRNN also achieves better accuracy, with a lower MAE, MAPE, and RMSE, and a higher R^2, which shows that the CWRNN can deal with strong nonlinear problems.

To verify the generalization of the proposed model, Site3, which is an onshore wind power station, was selected for verification. Compared with the offshore sites, the wind speed of Site3 changes more slowly, as is shown in Figure 9. From the figure, it can be observed that the RNN is still the worst model among all the RNNs. The reason may be that we set the same hyperparameters in the experiments, which included the input length. The RNN has a poor ability in its long-term dependency. The numerical result in Table 7 also verifies the conclusion. The CWRNN continues to show the best prediction results in both the onshore and offshore wind speed data, which verified that the CWRNN has a better performance in wind speed predictions.

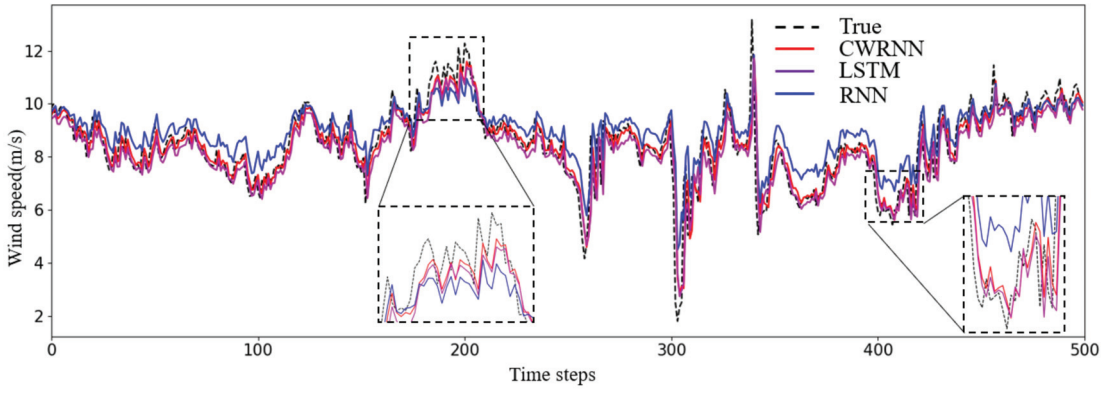


Figure 8. Comparison results of the proposed model with the RNNs of Site2.

Table 6. The numerical metrics of the prediction results by the CWRNNs and RNNs of Site2.

Model	Evaluation Metrics (Mean Value of 10 Times)			
	MAE	MAPE	RMSE	R^2
Simple RNN	0.6719	9.5407	0.8794	0.6362
LSTM	0.4952	6.3373	0.7256	0.7523
CWRNN	0.4430	5.8871	0.6799	0.7825

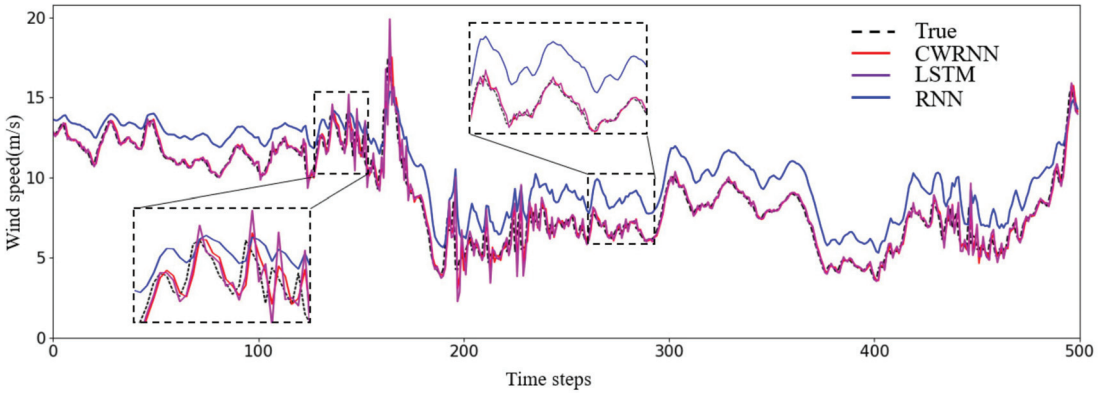


Figure 9. Comparison results of the proposed model with the RNNs of Site3.

Table 7. The numerical metrics of the prediction results by the CWRNNs and RNNs of Site3.

Model	Evaluation Metrics (Mean Value of 10 Times)			
	MAE	MAPE	RMSE	R^2
Simple RNN	1.6685	22.5562	1.7986	0.5955
LSTM	0.4315	5.7073	0.7715	0.9288
CWRNN	0.3843	5.0672	0.6446	0.9480

The evaluation metrics of all three sites are recorded together, as shown in Figure 10. It can be seen that the model achieves a better performance at all three sites, which means

the proposed method has good generalization. Furthermore, Site3, which was an onshore site, achieved the best performance out of all of the sites; its wind speed could be more easily predicted in comparison to the other offshore sites.

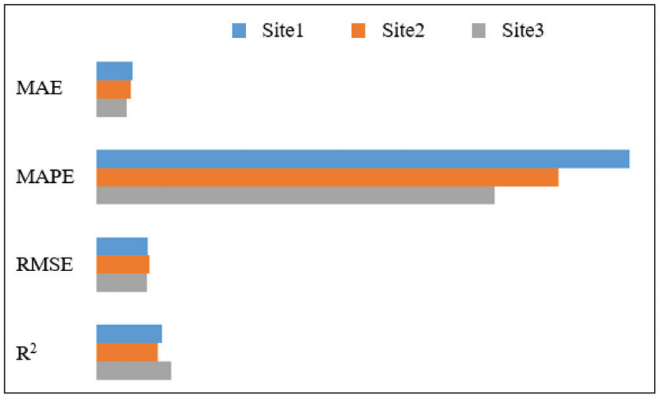


Figure 10. Evaluation metrics of Site1, Site2, and Site3.

4.2. Comparison with the Traditional Neural Networks

In order to verify the powerful ability of CWRNNs for time series prediction, the proposed method was compared with the traditional neural networks. In this experiment, the MLP, BPNN, and CNN, as traditional neural networks that are powerful machine learning models often used in different fields, were tested to perform the time series prediction task. The results are shown and described in Figure 11 and Table 8.

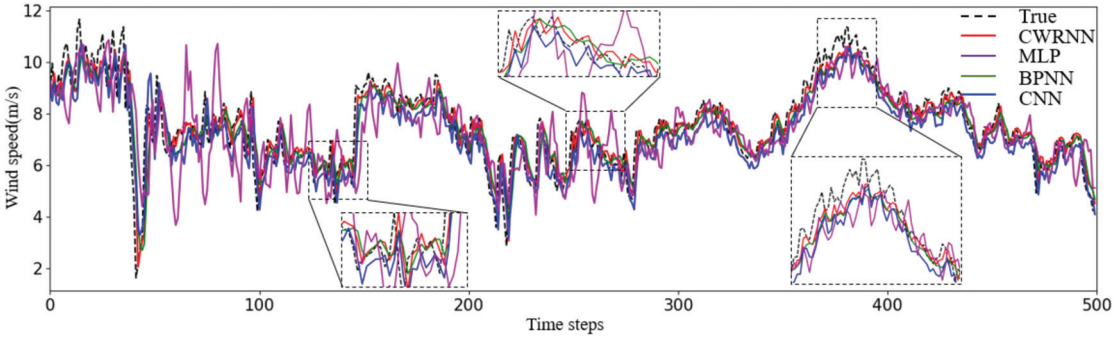


Figure 11. Comparison results of the proposed model with traditional NNs.

Table 8. The numerical metrics of the prediction results by CWRNNs and traditional NNs.

Model	Evaluation Metrics (Mean Value of 10 Times)			
	MAE	MAPE	RMSE	R^2
MLP	0.86	24.9	1.18	0.45
BPNN	0.53	7.99	0.76	0.78
CNN	0.61	8.58	0.79	0.76
CWRNN	0.46	6.79	0.66	0.83

It is obvious from the figure that MLP achieves the worst result. MLP, as a typical simple NN, has shortcomings, such as a slow learning speed, easily falling into local

extremum, and learning may not be sufficient. The result shows that MLP fails to learn from the wind speed data. The results also show that BPNN and CNN have worse performances in wind speed prediction. In most cases, BPNN and CNN have the powerful ability to solve nonlinear problems. However, they are not good at dealing with time series. Compared with the traditional neural networks, CWRNN appears to be more powerful in time series processing. Table 8 shows the numerical metrics of the prediction results, which further illustrates the above conclusion.

4.3. Comparison with Different Hyperparameters

There are many hyperparameters to set up a CWRNN model. Some hyperparameters are shared by RNN models, such as hidden layer parts, hidden layer neurons, the number of hidden layers, and the length of time series inputs. In essence, the CWRNN is a type of RNN that has the same network framework and mechanism of the backward pass of the error propagation. Therefore, the influence of the shared hyperparameters on the network is roughly the same. However, the CWRNN has some unique hyperparameters. The following experiments will focus on the specific parameters of CWRNNs.

4.3.1. Comparison with Different Part Numbers

The number of hidden layer parts is an important hyperparameter of the CWRNN, which has a great impact on the performance of the model. In the experiment, by changing the value of the hyperparameter, the influence on the accuracy of the model is evaluated. By setting different numbers for the hidden layer parts and training the model, we then used the evaluation metrics to evaluate the model's accuracy. The number of parts was set as (2, 4, 5), with all other parameters being the same.

The results are shown and described in Figure 12 and Table 9. From the results, we find that the least number of parts has the worst accuracy. When the number of parts increase to 4, we achieved the highest prediction accuracy. When the number raised to 5, the accuracy was lower than 4 parts, and higher than 2 parts. However, at the same time, the cost time of training the model significantly increased. Therefore, the value of four parts was the best choice in this study.

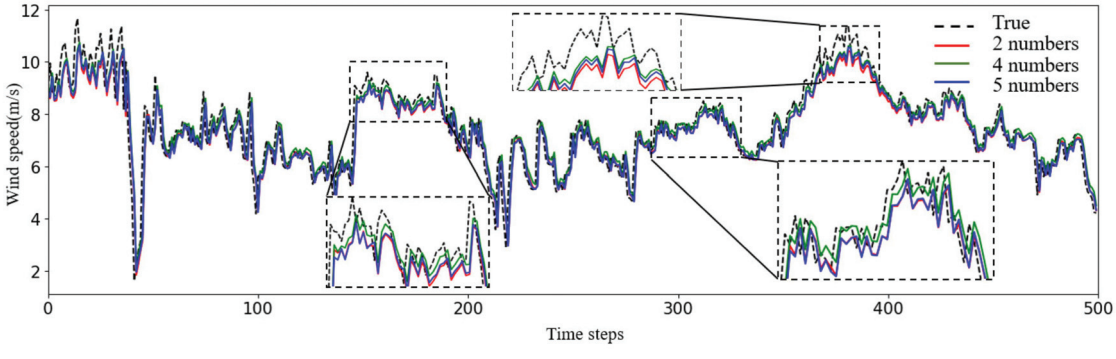


Figure 12. Comparison results of the proposed model with different part numbers.

Table 9. The numerical metrics of the prediction results with different part numbers.

Part Number	Values	Evaluation Metrics (Mean Value of 10 Times)			
		MAE	MAPE	RMSE	R^2
2	[1,2]	0.4835	6.9156	0.686	0.8155
4	[1,2,4,8]	0.4572	6.7873	0.6566	0.8310
5	[1,2,4,8,16]	0.4719	6.8052	0.6725	0.8227

4.3.2. Comparison with Different Part Periods

The part period is another hyperparameter that is unique to CWRNNs. The exponential series is often used as the part period. However, some other functions can be used for the part period, such as the linear function, Fibonacci function, logarithmic functions, or even fixed random periods. Different part periods will cause the different performances of the model. In this experiment, four different part periods were used to test the performance of the CWRNN. All the hidden layer parts were set to 4 and the other parameters were the same.

The results are shown in Figure 13 and Table 10. The four part periods were the linear series, odd series, triple series, and exponential series. Compared with the other series, the part period using the exponential series resulted in the model achieving the best performance. The result of the triple series shows great competitiveness, which means that the series gap increases with the increase in the number of periods and is thus a better choice.

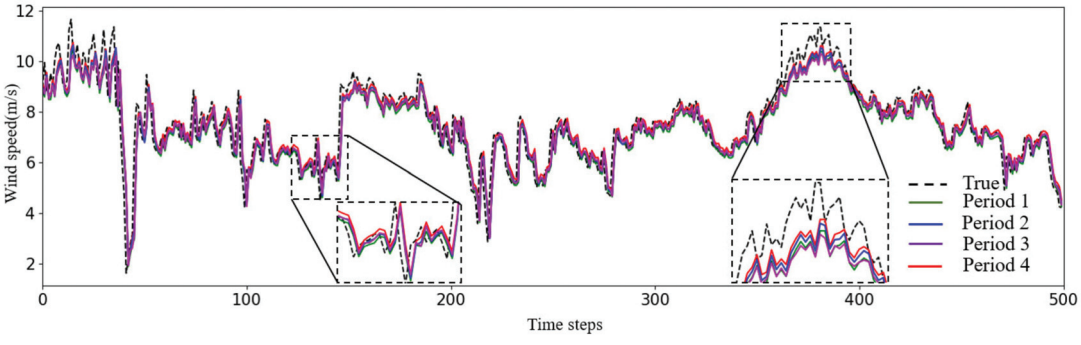


Figure 13. The results comparison of the proposed model with different part periods.

Table 10. The numerical metrics of the prediction results with different part periods.

Part Periods	Values	Evaluation Metrics (Mean Value of 10 Times)			
		MAE	MAPE	RMSE	R^2
1	[1,2,3,4]	0.4948	7.0096	0.6973	0.8094
2	[1,3,9,27]	0.4655	6.7413	0.6673	0.8254
3	[1,3,5,7]	0.4806	6.9184	0.6822	0.8175
4	[1,2,4,8]	0.4572	6.7873	0.6566	0.8310

5. Discussion

An offshore wind speed prediction method using CWRNNs is proposed and is verified by the wind speed dataset of offshore and onshore sites. The results are further discussed and analyzed in the following contexts:

(1) As is commonly known, RNN is excellent at modeling sequential data with a simple mechanism. However, with the increase in the dependency length, which means more context is needed, the RNN cannot learn from the input data. There are some techniques to improve the RNN. LSTM, which uses the gating mechanism, is proposed to solve problems, including vanishing gradients and long dependency. It is easier to understand that the complex network structure increases the model stability. However, the performance of most machine learning models, especially complex deep learning neural network models, depends on the quantity and diversity of the data. Naturally, if a machine learning model has a lot of parameters, it needs a proportional number of samples to perform well.

The CWRNN is another type of RNN, which breaks up the neurons in the hidden layer into different parts, and the neurons in the same part work at a given clock speed

to address long term dependency. The parameters of the CWRNN are close to the simple RNN. This indicates that the CWRNN is more suitable for the case of a small sample size than LSTM. Meanwhile, the CWRNN employs an ingenious mechanism for activating neurons parts at different clock speeds, which can efficiently learn the long-term time series information, thus solving strongly nonlinear problems. At the same time, the CWRNN only updates neuron parts at a specific clock rate, which reduces the computation cost.

(2) There is an inherent concept of sequential data or time series data that incrementally progresses over time. As we know, traditional NNs are good at solving the nonlinear problem and perform well in most cases. However, they lack the inherent trend of persistence for obtaining sequential data. A simple feedforward NN cannot really understand the meaning of a sentence according to the order of input data in the context. CNNs have been extremely successful in the computer vision field. However, they have difficulties in dealing with time series data. The RNN, as a type of neural network, keeps the characteristics of the autoregressive model, and also has the ability to model sequential data. Furthermore, for the human neural system, the vision channel and the memory channel are different channels that have different mechanisms.

Recently, the attention mechanism is one of the most valuable breakthroughs in deep learning model preparation in the last few decades. Unlike the vanilla RNN approach, it proposes to help monitor all the hidden states in the encoder sequence for making predictions. It can assign the weight values to the extracted information to highlight the important information that the attention mechanism seems to break the barriers between the vision channel and memory channel. However, it still has a great number of parameters, which also need a large number of sample data. For now, the CWRNN is a good choice to solve strong nonlinear problems with limited samples.

(3) Hyperparameters can directly impact the performance of machine learning models. Therefore, to achieve the best performance, the optimization of the hyperparameters plays a crucial role. In addition to the common parameters of the RNNs, the CWRNN has some unique parameters. The setting of these parameters requires a complex parameter tuning process and the appropriate parameters will result in a great improvement to its performance.

In this study, some unique parameters were discussed, which were based on the experiment results. However, the common parameters of the RNNs still affect the model performance. Considering the shared RNN parameters together with the intrinsic parameters of the CWRNN will be a big project. Tuning these parameters requires further research.

6. Conclusions

This study proposes an offshore wind speed prediction method based on CWRNNs. The CWRNN breaks up neurons in the hidden layer into different parts, and neurons in the same part work at a given clock speed to address long term dependency, which can effectively solve the problem of strong nonlinearity in offshore wind speed. The performance of the proposed method is verified by three datasets from two different offshore sites and one onshore site. The experimental results show that the proposed model achieves a significant improvement in its prediction accuracy.

Author Contributions: Conceptualization, Y.S.; methodology, Y.S.; software, Y.S.; validation, Y.S., Y.W. and H.Z.; data curation, Y.W.; writing—original draft preparation, Y.S.; writing—review and editing, Y.W. and H.Z.; visualization, Y.S.; supervision, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China under Grant No. 2018YFB1307400.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the support from the DER AI Lab of Shanghai University and the State Grid Intelligence Technology Corporation of China for the development of the machine learning model and the dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jung, C.; Taubert, D.; Schindler, D. The temporal variability of global wind energy—Long-term trends and inter-annual variability. *Energy Convers. Manag.* **2019**, *188*, 462–472. [CrossRef]
2. Wang, J.; Song, Y.; Liu, F.; Hou, R. Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models. *Renew. Sust. Energ. Rev.* **2016**, *60*, 960–981. [CrossRef]
3. Yang, K.; Tang, Y.; Zhang, Z. Parameter Identification and State-of-Charge Estimation for Lithium-Ion Batteries Using Separated Time Scales and Extended Kalman Filter. *Energies* **2021**, *14*, 1054. [CrossRef]
4. Qian, Z.; Pei, Y.; Zareipour, H.; Chen, N. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Appl. Energy* **2019**, *235*, 939–953. [CrossRef]
5. Zhang, J.; Draxl, C.; Hopson, T.; Monache, L.D.; Vanvyve, E.; Hodge, B.M. Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. *Appl. Energy* **2015**, *156*, 528–541. [CrossRef]
6. Wang, J.; Qin, S.; Jin, S.; Wu, J. Estimation methods review and analysis of offshore extreme wind speeds and wind energy resources. *Renew. Sust. Energ. Rev.* **2015**, *42*, 26–42. [CrossRef]
7. Morgan, E.C.; Lackner, M.; Vogel, R.M.; Baise, L.G. Probability distributions for offshore wind speeds. *Energy Convers. Manag.* **2011**, *52*, 15–26. [CrossRef]
8. Cai, H.; Jia, X.; Feng, J.; Yang, Q.; Li, W.; Li, F.; Lee, J. A unified Bayesian filtering framework for multi-horizon wind speed prediction with improved accuracy. *Renew. Energy* **2021**, *178*, 709–719. [CrossRef]
9. Li, L.; Liu, Y.-Q.; Yang, Y.-P.; Han, S.; Wang, Y.-M. A physical approach of the short-term wind power prediction based on CFD pre-calculated flow fields. *J. Hydrodyn. B* **2013**, *25*, 56–61. [CrossRef]
10. Zhang, K.; Qu, Z.; Dong, Y.; Lu, H.; Leng, W.; Wang, J.; Zhang, W. Research on a combined model based on linear and nonlinear features—A case study of wind speed forecasting. *Renew. Energy* **2019**, *130*, 814–830. [CrossRef]
11. Zhang, C.; Wei, H.; Zhao, X.; Liu, T.; Zhang, K. A Gaussian process regression-based hybrid approach for short-term wind speed prediction. *Energy Convers. Manag.* **2016**, *126*, 1084–1092. [CrossRef]
12. Dhiman, H.S.; Deb, D.; Foley, A.M. Bilateral Gaussian wake model formulation for wind farms: A forecasting based approach. *Renew. Sust. Energ. Rev.* **2020**, *127*, 109873. [CrossRef]
13. Karakuş, O.; Kuruoğlu, E.E.; Altınkaya, M.A. One-day ahead wind speed/power prediction based on polynomial autoregressive model. *IET Renew.* **2017**, *11*, 1430–1439. [CrossRef]
14. Tian, Z.; Wang, G.; Ren, Y. Short-term wind speed forecasting based on autoregressive moving average with echo state network compensation. *Wind Eng.* **2019**, *44*, 152–167. [CrossRef]
15. Liu, M.D.; Ding, L.; Bai, Y.L. Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. *Energy Convers. Manag.* **2021**, *233*, 113917. [CrossRef]
16. Liu, X.; Lin, Z.; Feng, Z. Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM. *Energy* **2021**, *227*, 120492. [CrossRef]
17. Cai, H.; Jia, X.; Feng, J.; Li, W.; Hsu, Y.M.; Lee, J. Gaussian Process Regression for numerical wind speed prediction enhancement. *Renew. Energy* **2020**, *146*, 2112–2123. [CrossRef]
18. Ak, R.; Li, Y.F.; Vitelli, V.; Zio, E. Adequacy assessment of a wind-integrated system using neural network-based interval predictions of wind power generation and load. *Int. J. Electr. Power* **2018**, *95*, 213–226. [CrossRef]
19. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [CrossRef]
20. Rani, R.H.; Victoire, T.A. Training radial basis function networks for wind speed prediction using PSO enhanced differential search optimizer. *PLoS ONE* **2018**, *13*, e0196871. [CrossRef]
21. Lu, P.; Ye, L.; Tang, Y.; Zhao, Y.; Zhong, W.; Qu, Y.; Zhai, B. Ultra-short-term combined prediction approach based on kernel function switch mechanism. *Renew. Energy* **2021**, *164*, 842–866. [CrossRef]
22. Dhiman, H.S.; Deb, D.; Guerrero, J.M. Hybrid machine intelligent SVR variants for wind forecasting and ramp events. *Renew. Sust. Energ. Rev.* **2019**, *108*, 369–379. [CrossRef]
23. Dhiman, H.S.; Anand, P.; Deb, D. Wavelet transform and variants of SVR with application in wind forecasting. In *Innovations in Infrastructure*; Springer: Singapore, 2018; Volume 757, pp. 501–511. [CrossRef]
24. Dhiman, H.S.; Deb, D. Machine intelligent and deep learning techniques for large training data in short-term wind speed and ramp event forecasting. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e12818. [CrossRef]
25. Dhiman, H.S.; Deb, D.; Balas, V.E. *Supervised Machine Learning in Wind Forecasting and Ramp Event Prediction*; Academic Press: Salt Lake City, UT, USA, 2020. [CrossRef]

26. Patel, P.; Shandilya, A.; Deb, D. Optimized hybrid wind power generation with forecasting algorithms and battery life considerations. In Proceedings of the IEEE Power and Energy Conference at Illinois (PECI), Champaign, IL, USA, 23–24 February 2017. [CrossRef]
27. Bai, Y.; Liu, M.-D.; Ding, L.; Ma, Y.-J. Double-layer staged training echo-state networks for wind speed prediction using variational mode decomposition. *Appl. Energy* **2021**, *301*, 117461. [CrossRef]
28. Xu, W.; Liu, P.; Cheng, L.; Zhou, Y.; Xia, Q.; Gong, Y.; Liu, Y. Multi-step wind speed prediction by combining a WRF simulation and an error correction strategy. *Renew. Energy* **2021**, *163*, 772–782. [CrossRef]
29. Zhu, X.; Liu, R.; Chen, Y.; Gao, X.; Wang, Y.; Xu, Z. Wind speed behaviors feather analysis and its utilization on wind speed prediction using 3D-CNN. *Energy* **2021**, *236*, 121523. [CrossRef]
30. Ma, Q.-L.; Zheng, Q.-L.; Peng, H.; Zhong, T.-W.; Qin, J.-W. Multi-step-prediction of chaotic time series based on co-evolutionary recurrent neural network. *Chin. Phys. B* **2008**, *17*, 536. [CrossRef]
31. Duan, J.; Zuo, H.; Bai, Y.; Duan, J.; Chang, M.; Chen, B. Short-term wind speed forecasting using recurrent neural networks with error correction. *Energy* **2021**, *217*, 119397. [CrossRef]
32. Wang, S.; Wang, J.; Lu, H.; Zhao, W. A novel combined model for wind speed prediction—Combination of linear model, shallow neural networks, and deep learning approaches. *Energy* **2021**, *234*, 121275. [CrossRef]
33. Saeed, A.; Li, C.; Gan, Z.; Xie, Y.; Liu, F. A simple approach for short-term wind speed interval prediction based on independently recurrent neural networks and error probability distribution. *Energy* **2022**, *238*, 122012. [CrossRef]
34. Liu, L.; Wang, J. Super multi-step wind speed forecasting system with training set extension and horizontal–vertical integration neural network. *Appl. Energy* **2021**, *292*, 116908. [CrossRef]
35. Xiong, D.; Fu, W.; Wang, K.; Fang, P.; Chen, T.; Zou, F. A blended approach incorporating TVFEMD, PSR, NNCT-based multi-model fusion and hierarchy-based merged optimization algorithm for multi-step wind speed prediction. *Energy Convers. Manag.* **2021**, *230*, 113680. [CrossRef]
36. Neshat, M.; Nezhad, M.M.; Abbasnejad, E.; Seyedali, M.; Lina, B.T.; Davide, A.G.; Bradley, A.; Markus, W. A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the Lillgrund offshore wind farm. *Energy Convers. Manag.* **2021**, *236*, 114002. [CrossRef]
37. Ahmad, T.; Zhang, D. A data-driven deep sequence-to-sequence long-short memory method along with a gated recurrent neural network for wind power forecasting. *Energy* **2022**, *239*, 122109. [CrossRef]
38. Zhang, Z.; Ye, L.; Qin, H.; Liu, Y.; Wang, C.; Yu, X.; Li, J. Wind speed prediction method using Shared Weight Long Short-Term Memory Network and Gaussian Process Regression. *Appl. Energy* **2019**, *247*, 270–284. [CrossRef]
39. Chen, Y.; Dong, Z.; Wang, Y.; Su, J.; Han, Z.; Zhou, D.; Zhang, K.; Zhao, Y.; Bao, Y. Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history. *Energy Convers. Manag.* **2021**, *227*, 113559. [CrossRef]
40. Tian, Z. Modes decomposition forecasting approach for ultra-short-term wind speed. *Appl. Soft Comput.* **2021**, *105*, 107303. [CrossRef]
41. Li, F.; Ren, G.; Lee, J. Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks. *Energy Convers. Manag.* **2019**, *186*, 306–322. [CrossRef]
42. Zhang, Y.; Chen, B.; Pan, G.; Zhao, Y. A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting. *Energy Convers. Manag.* **2019**, *195*, 180–197. [CrossRef]
43. Tian, Z.; Ren, Y.; Wang, G. Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM. *Energy Sources Part A Recovery Util. Environ. Eff.* **2018**, *41*, 26–46. [CrossRef]
44. Song, J.; Wang, J.; Lu, H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. *Appl. Energy* **2018**, *215*, 643–658. [CrossRef]
45. Ma, Z.; Chen, H.; Wang, J.; Yang, X.; Yan, R.; Jia, J.; Xu, W. Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction. *Energy Convers. Manag.* **2020**, *205*, 112345. [CrossRef]
46. Wang, J.; Li, Y. Multi-step ahead wind speed prediction based on optimal feature extraction, long short-term memory neural network and error correction strategy. *Appl. Energy* **2018**, *230*, 429–443. [CrossRef]
47. Liang, Z.; Liang, J.; Wang, C.; Dong, X.; Miao, X. Short-term wind power combined forecasting based on error forecast correction. *Energy Convers. Manag.* **2016**, *119*, 215–226. [CrossRef]
48. Liu, H.; Yang, R.; Wang, T.; Zhang, L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. *Renew. Energy* **2021**, *165*, 573–594. [CrossRef]
49. Liang, T.; Zhao, Q.; Lv, Q.; Sun, H. A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers. *Energy* **2021**, *230*, 120924. [CrossRef]
50. Li, C.; Tang, G.; Xue, X.; Saeed, A.; Hu, X. Short-term wind speed interval prediction based on ensemble GRU model. *IEEE Trans. Sustain. Energy* **2019**, *11*, 1370–1380. [CrossRef]
51. Feng, X.; Chen, J.; Zhang, Z.; Miao, S.; Zhu, Q. State-of-charge estimation of lithium-ion battery based on clockwork recurrent neural network. *Energy* **2021**, *236*, 121360. [CrossRef]
52. Luo, J.; Fu, Y. Dilated Recurrent Neural Network. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
53. Koutnik, J.; Greff, K.; Gomez, F.; Schmidhuber, J. A clockwork rnn. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014.

54. Xie, Y.; Zhang, Z.; Sapkota, M.; Yang, L. Spatial clockwork recurrent neural network for muscle perimysium segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 11–21 October 2016. [CrossRef]
55. Lin, C.; Wang, H.; Yuan, J.; Yu, D.; Li, C. Research on UUV obstacle avoiding method based on recurrent neural networks. *Complexity* **2019**, *2019*, 6320186. [CrossRef]
56. Achanta, S.; Godambe, T.; Gangashetty, S.V. An investigation of recurrent neural network architectures for statistical parametric speech synthesis. In Proceedings of the Sixteenth Annual Conference of the International Speech Communication Association, Dresden, Germany, 6–10 September 2015. [CrossRef]
57. Liu, W.; Gu, Y.; Ding, Y.; Lu, W.; Rui, X.; Tao, L. A Spatial and Temporal Combination Model for Traffic Flow: A Case Study of Beijing Expressway. In Proceedings of the 2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE), Beijing, China, 11–13 September 2020. [CrossRef]
58. Roberts, O.; Andreas, A. *United States Virgin Islands: St. Thomas & St. Croix (Data)*; NREL Report No. DA-5500-64451; NREL-DATA: Golden, CO, USA, 1997. [CrossRef]
59. NREL: Measurement and Instrumentation Data Center (MIDC). Available online: https://midcdmz.nrel.gov (accessed on 1 January 2022).
60. SHU DER AI Lab. Available online: https://github.com/SHU-DeepEnergyResearch/Time-Series-Prediction (accessed on 1 January 2022).

energies

MDPI

Article

Misfire Detection Using Crank Speed and Long Short-Term Memory Recurrent Neural Network

Xinwei Wang [1,2], Pan Zhang [3], Wenzhi Gao [3,*], Yong Li [3], Yanjun Wang [3] and Haoqian Pang [3]

[1] State Key Laboratory of Engine Reliability, Weifang 261061, China; wangxinw@weichai.com
[2] Weichai Power Co., Ltd., Weifang 261061, China
[3] State Key Laboratory of Engines, Tianjin University, Tianjin 300350, China; zhangpan@tju.edu.cn (P.Z.); li_yong@tju.edu.cn (Y.L.); wangyanjunz@tju.edu.cn (Y.W.); panghaoqian199817@gmail.com (H.P.)
* Correspondence: gaowenzhi@tju.edu.cn

Abstract: In this work, a new approach was developed for the detection of engine misfire based on the long short-term memory recurrent neural network (LSTM RNN) using crank speed signal. The datasets are acquired from a six-cylinder-inline, turbo-charged diesel engine. Previous works investigated misfire detection in a limited range of engine running speed, running load or misfire types. In this work, the misfire patterns consist of normal condition, six types of one-cylinder misfire faults and fifteen types of two-cylinder misfire faults. All the misfire patterns are tested under wide range of running conditions of the tested engine. The traditional misfire detection method is tested on the datasets first, and the result show its limitation on high-speed low-load conditions. The LSTM RNN is a type of artificial neural network which has the ability of considering both the current input in-formation and the previous input information; hence it is helpful in extracting features of crank speed in which the misfire-induced speed fluctuation will last one or a few cycles. In order to select the engine operating conditions for network training properly, five data division strategies are attempted. For the sake of acquiring high performance of designed network, four types of network structure are tested. The results show that, utilizing the datasets in this work, the LSTM RNN based algorithm can overcome the limitation at high-speed low-load conditions of traditional misfire detection method. Moreover, the network which takes fixed segment of raw speed signal as input and takes misfire or fault-free labels as output achieves the best performance with the misfire diagnosis accuracy not less than 99.90%.

Keywords: engine misfire; pattern recognition; fault detection; LSTM; time-frequency analysis

Citation: Wang, X.; Zhang, P.; Gao, W.; Li, Y.; Wang, Y.; Pang, H. Misfire Detection Using Crank Speed and Long Short-Term Memory Recurrent Neural Network. *Energies* **2022**, *15*, 300. https://doi.org/10.3390/en15010300

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 26 November 2021
Accepted: 29 December 2021
Published: 3 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Engine misfire is a phenomenon of no-burning in cylinder which may be caused by insufficient fuel injection, bad fuel quality, insufficient ignition energy, or mechanical failure, etc. Since misfire fault will cause abnormal engine running condition and air pollution, many researchers have been trying to put forward effective methods to achieve accurate and real-time misfire detection.

The techniques for engine misfire detection can be categorized according to the utilized sensor signals, which includes the method using engine body vibration signal [1], the method using acoustic signal [2], the method analyzing exhaust gas temperature [3], the method monitoring in-cylinder iron current [4], and the method using crank speed [5]. The method using engine body vibration signal could sample much information, since the vibration signal is sampled with high resolution and is related to in-cylinder combustion. However, a large amount of computation is required for processing vibration data. The method using acoustic signal has not solved the problem of noise interference in practical implementation. The method analyzing the temperature of exhaust gas is limited by the sensor's response time. The method monitoring in-cylinder iron current needs to modify

the engine body. The method using crank speed has been adopted by many researchers, since the crank speed can be sampled relatively easily and cannot be easily contaminated by uncorrelated noise.

The misfire detection methods based on crank speed can be categorized into physical model-based algorithms and data-driven diagnosis algorithms.

The model-based method is used to diagnose engine misfire by building the relationship between crank speed and in-cylinder pressure based on the engine dynamic model. Zheng et al. [5] designed a Luenberger sliding mode observer to estimate engine combustion torque based on experimental crank speed of a four-cylinder engine. Rizvi et al. [6] proposed a hybrid model for simulating the relationship between engine power and crank speed fluctuations. Misfire was detected by using Markov chain. Helm et al. [7] estimated engine torque based on a parametric Kalman filter. Misfire was detected by employing the estimated torque and an interacting multiple model algorithm. Hmida et al. [8] proposed the torsional model of crankshaft. The Lagrange method and Newmark algorithm were employed to derive the equations of motion. The appearance of sidebands around the acyclic frequency was adopted to detect misfire.

The model-based algorithm can lead to very accurate results if properly executed. Nevertheless, the method needs precise engine model parameters which is hard to gauge accurately. The damping is an example that cannot even be measured. Meanwhile, the complexity of model-based algorithm may not permit the real time implementation of the algorithm [9]. Therefore, this method has not been widely used in industrial application.

The data-driven diagnosis algorithms provide another way of misfire detection, in which the misfire related characteristics are extracted directly from crank speed instead of deriving the excitation torque or in-cylinder pressure. Misfire is detected by distinguishing the misfire related characteristics from fault free. The representative data-driven method is the engine roughness method which is proposed by Plapp et al. [10] and is still used in modern vehicles. However, this method is limited on high-speed low-load conditions when the number of engine cylinders is not less than six.

Another data-driven method is conducted by analyzing the typical frequencies of crank speed. Taraza et al. [11] utilized the lowest three harmonic orders of crank speed as an indicator for one-cylinder misfire detection. Geveci et al. [12] analyzed the first and the second harmonic components of crank speed under normal and cylinder 1# misfire conditions at various speeds and loads. This method is limited as well, since the speed spectrum of two-cylinder misfire fault may be confused with one-cylinder misfire patterns.

Over the past about twenty years, machine learning algorithms developed rapidly and have been exploited in misfire detection research field [13]. Compared with the algorithms that one or some human-designed indicators are calculated for misfire detection, the machine learning algorithm could extract more fault features from one signal or could process many signals at the same time. Not only the crank speed, but also the engine vibration and in-cylinder pressure have been used in the machine learning algorithm as reported in the literature.

Li et al. [14] utilized the crank speed and the techniques including the empirical mode decomposition, kernel independent component analysis, Wigner bispectrum and support vector machine (SVM) for detecting misfire of a marine diesel engine. Chen and Randall [15] designed a misfire detection method which consists of three stages: fault detection, fault localization and fault severity identification. This method was achieved by using the lowest four harmonic orders of crank speed and a fully connected artificial neural network (ANN). Jung et al. [16] distinguish misfire and fault-free conditions using crank speed and Kullback–Leibler divergence. SVM was utilized as the automatic classification tool. Gani and Manzie [17] also employed the SVM technique and crank speed for classifying normal condition, intermittent misfire and continuous misfire in cylinder 6# of engine. The accuracy approached 100% in test dataset. In the work of Sharma et al. [18], the statistic features of vibration signals, like standard deviation, kurtosis, median and so on, were selected as fault features for misfire detection. The decision tree algorithm was employed for fault

classification. As reported by Moosavian et al. [19], wavelet denoising technique, ANN, least square support vector machine, and D–S evidence theory were applied for misfire detection. The final classification accuracy of 98.56% was achieved by using acoustic and vibration signal under idle condition. Gu et al. [20] utilized multivariate empirical mode decomposition and SVM techniques for a twelve-cylinder diesel engine misfire detection. Qin et al. [1] designed a deep twin convolutional neural network with multi-domain inputs for misfire detection. Since the vibration signal was employed, the authors also studied the algorithm's performance when there was strong environmental noise on the vibration. Jafarian et al. [21] employed vibration signals from four sensors placed on the engine for misfire detection. The fast Fourier transform (FFT) was used for feature extraction; the ANN, SVM, and k-nearest neighbor (kNN) algorithms were used for classification. Liu et al. [22] took many signals, including engine speed, exhaust temperature, and fuel consumption, as the inputs of ANN for misfire detection. Bahri et al. [23] detected misfire of a homogeneous charge compression ignition engine by using in-cylinder pressure and ANN model.

It can be seen that the mentioned data-driven diagnosis algorithms mainly focused on the classical feature extraction methods, such as human-designed threshold, FFT, wavelet transform and empirical mode decomposition, and traditional pattern recognition methods, such as fully connected ANN and SVM. Thus, the algorithm performance often depends on the proper selected features and the domain expertise in engine misfire. In addition, the mentioned algorithms based on machine learning either mainly considered a few engine speed and load conditions, or only considered a few one-cylinder misfire types. In practice, the fault features would change with the engine running conditions or different misfire types, and for a data-driven algorithm especially a machine learning algorithm, the sample size is an important factor for the algorithm training. Therefore, there is still some research work needs to do for applying the machine learning based algorithms into actual industrial scenario.

Recurrent neural network (RNN) is a type of neural network which is good at processing sequence data. Sequences may be of finite or countably infinite length, and may be temporal or non-temporal. Examples of time-indexed data include the audio recordings which are sampled at fixed intervals. In fact, RNNs are frequently applied to sequences whose meaning are directly related to the data order but no explicit notion of time [24]. Engine crank speed is a type of sequence that the prior motion will affect the later motion. For example, assuming the firing order of a six-cylinder engine is 1-5-3-6-2-4, if misfire occurs in the first cylinder, not only the instantaneous speed variation of the first cylinder changes, but also the fourth cylinder. Therefore, RNN is hopeful for misfire detection. For the earlier RNN, it was difficult to handle the problems of vanishing and exploding gradient that occurred when training RNN across many steps [25]. Therefore, in this paper, the RNN with long short-term memory (LSTM) [26] which overcomes the training difficulties is utilized.

Compared with misfire detection studies in literature, the main contribution of this paper is as follows.

- A new misfire detection method that is based on LSTM RNN is proposed.
- Datasets for network training and testing are acquired in wide range of speed and load conditions of the tested diesel engine, which ensures the diversity of datasets and makes the network more applicable.
- For a six-cylinder engine, limited studies [12,16,22,27] have considered the detection of two-cylinder misfire faults which include more misfire types and may disturb the detection of one-cylinder misfire and even cause misdiagnosis. In this paper, besides the one-cylinder misfire faults, all the fifteen two-cylinder misfire faults are considered as well.

The rest of the paper is organized as follows. In Section 2, the experiment setup and diesel engine rig tests are introduced. The speed characteristics under misfire and the limitation of traditional misfire detection method are described in Section 3. The scheme

of misfire diagnosis and the LSTM algorithm are introduced in Section 4. In Section 5, the experimental results are analyzed and discussed. Finally, conclusions are given in Section 6.

2. Experimental Equipment and Data Acquisition

2.1. Test Rig Setup and Data Acquisition System

The test engine was a four-stroke, six-cylinder-inline diesel engine. In order to adjust the fuel injection parameter conveniently, a diesel engine with electronic unit pump was employed. In addition, with larger number of cylinders, an engine would operate steady, so the fault features of misfire would become weaker relatively [28]. Thus, a six-cylinder engine was selected. The basic technical data of engine is shown in Table 1. A hydraulic dynamometer was connected to the engine for providing external load. A flexible shaft coupling was mounted to connect the engine crank shaft and the dynamometer. Figure 1 shows the picture of the whole test-rig.

Table 1. Engine specifications.

Parameter	Value
Engine type	CY6BG332
Configuration	Six-cylinder, inline, four-stroke
Air intake	Turbocharged, intercooled
Firing order	1-5-3-6-2-4
Fuel injection system	Electronic unit pump
Total displacement	5.95 L
Compression ratio	18.5
Rated power	88 kW @ 2200 r/min
Maximum torque	450 Nm @ 1000–1800 r/min

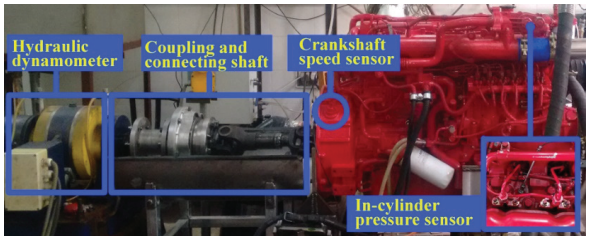


Figure 1. The test-rig.

A Kistler high-temperature pressure piezoelectric sensor, Type 6058A, was mounted in cylinder 1# through the glow plug hole for verifying misfire occurrence in cylinder chamber. A magnetic sensor, which was mounted opposite to the teeth on the flywheel was used to measure the angular speed of the crankshaft. The sensors' signals were synchronously sampled and primarily processed using Siemens LMS SCM05 system with 24-bit ADC resolution and a maximum sampling rate of 102.4 kHz.

2.2. Test Description

The measurements were conducted over the engine speed range 800–2200 r/min with interval 100 r/min, at different load levels, as shown in Figure 2. For each engine speed and load value, the measurements were operated under normal, one-cylinder misfire and two-cylinder misfire conditions. The misfire types are shown in Table 2. Including normal condition, the total fault types are 22. The misfire condition was achieved by setting the injection parameter zero on the programmable electronic control unit.

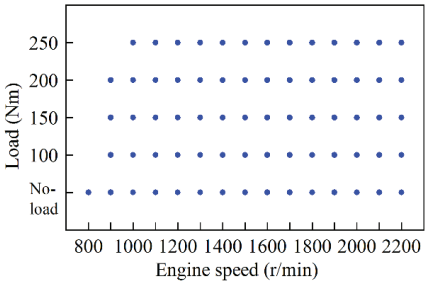


Figure 2. Engine operating conditions.

Table 2. All the misfire types.

Categories	Misfire Types	
Normal condition	Fault-free	
One-cylinder misfire	1#, 2#, 3#, 4#, 5#, 6#	
Two-cylinder misfire	Two consecutive cylinders.	1#5#, 3#5#, 3#6#, 2#6#, 2#4#, 1#4#;
	Two cylinders with one-cylinder interval.	1#3#, 5#6#, 2#3#, 4#6#, 1#2#, 4#5#;
	Two cylinders with two-cylinder interval.	1#6#, 3#4#, 2#5#.

The tests were conducted in an intensive engine running speeds and loads, and varied in a wide range. Partial data were used for network training and the rest were used for network testing. The size of training dataset had been set from large to small until an optimal size was achieved.

3. The Speed Characteristics under Misfire and the Limitation of Traditional Misfire Detection Method

When misfire occurs, the instantaneous engine crankshaft speed will drop and the subsequent speed will rise up compared with the normal conditions. The variation of the whole speed curve will be larger. An example is shown in Figure 3a, under the running condition of 1000 r/min and 100 Nm, when a misfire occurs in cylinder 1#, as the dash curve indicates, the speed becomes different from the normal condition. When the engine speed is high and the load is low, as shown in Figure 3b, the variation rule of instantaneous crankshaft speed becomes unclear, and the difference between normal and misfire condition also becomes indistinguishable. Moreover, when two-cylinder misfire occurs, the fault features expressed from engine speed curve is easily confused with that of one-cylinder misfire condition, especially under the high-speed and low-load conditions. Figure 4a,b show the comparison between speed curves of one-cylinder misfire and two-cylinder misfire.

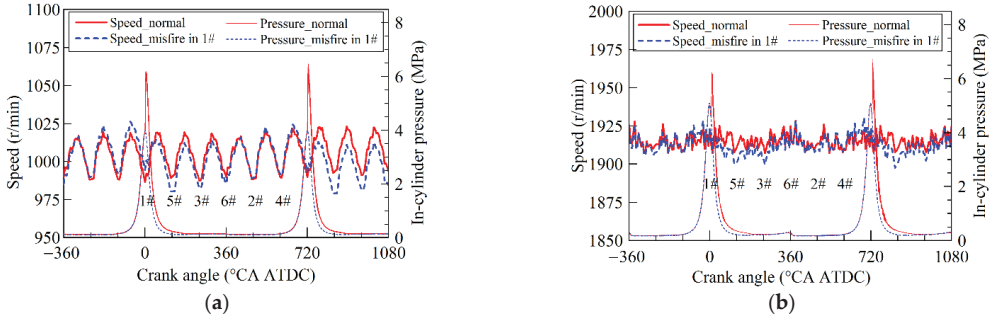


Figure 3. In-cylinder pressure and corresponding engine speed under normal and cylinder 1# misfire conditions. (**a**) Under 1000 r/min and 100 Nm. (**b**) Under 1900 r/min and 100 Nm.

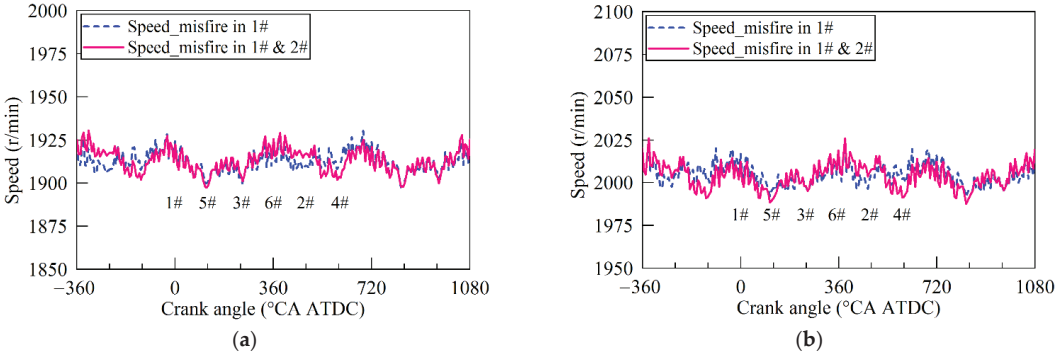


Figure 4. Instantaneous engine speed comparison between cylinder 1# misfire and cylinders 1# and 2# misfire conditions. (**a**) Under 1900 r/min and 100 Nm. (**b**) Under 2000 r/min and no-load condition.

Therefore, for detecting misfire accurately, the fault features that can reduce or eliminate the interference from engine speed and load should be found. One way to eliminate the impact of engine running range is to divide it into small blocks and then find fault features for each block [27]. However, this will increase workload when the engine running range is large. The better way is to find or design an algorithm that can extract useful feature or can learn more features for the whole engine running conditions.

An example of the traditional methods which is called engine roughness method is introduced as below. This method calculates a misfire indicator that is based on the difference of two consecutive angular accelerations. Equation (1) presents the calculation of the indicator [10].

$$S_i = (T_{i+1} - T_i)/T_i^3 \quad (1)$$

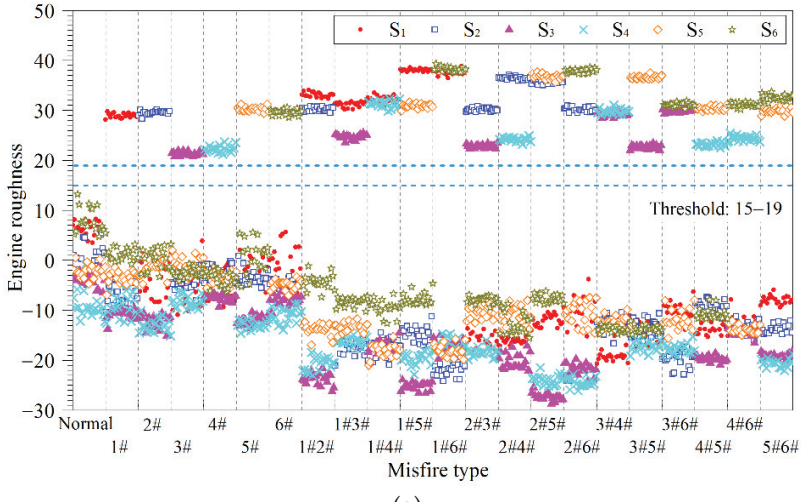
where S_i is engine roughness of the ith cylinder. T_i is the time period from ignition of the ith cylinder to ignition of the next cylinder in firing order.

Figure 5 shows the results of indicator S under different engine speeds and different misfire patterns. When misfire occurs in cylinder i, the corresponding S_i will become larger than the predefined threshold. An example is shown in Figure 5a, the threshold can be defined in range 15–19, and when S_i is detected larger than the threshold, it is thought the misfire happened in the cylinder i. This method is limited at the high-speed low-load conditions. As shown in Figure 5b, under 1900 r/min and no-load condition, it is hard to determine the threshold, and the two-cylinder misfire modes are easily confused with one-cylinder misfire modes.

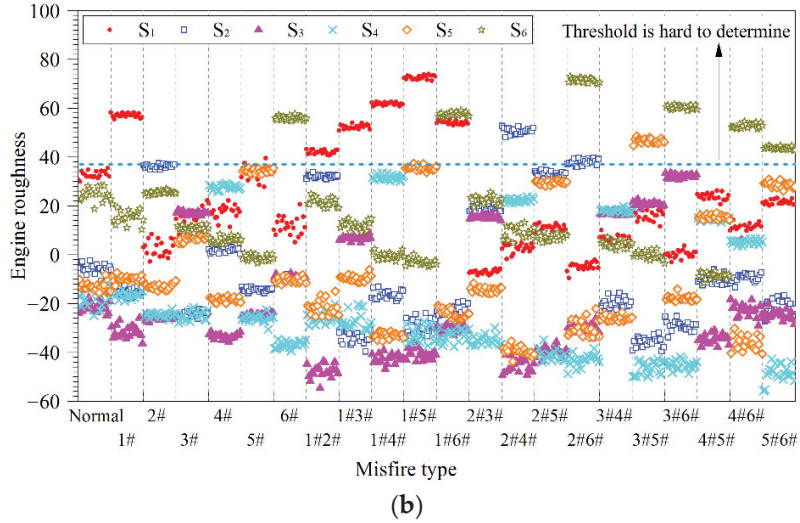
The unsatisfied results appeared at high-speed and low-load conditions are caused by the background noise which has approximately the same order of magnitude with the value related to the misfire presence. The reasons for relatively higher background noise are mainly from the different burning behaviors caused by the systematic nonuniformity. In addition, with the speed increasing and load decreasing, the signal to noise ratio will decrease since the useful features caused by misfire will decrease. Figure 6 presents the standard deviation of crankshaft speed under 800 r/min and no-load conditions. The standard deviation is calculated once per cycle, the points in Figure 6 which are shown in the form of mean value and standard deviation are calculated from 200 cycles of data. The results clearly show that when misfire occurs, the amplitude of speed variation will increase, and this is helpful for extracting misfire features. However, when engine speed becomes higher and load becomes lower, the amplitude of speed variation decreases, and the amplitude difference between normal and misfire patterns also decreases, as shown in Figure 7. Then, the signal to noise ratio decreases. The limitation of the engine

roughness method proves that it is hard to use too few features to achieve a perfect fault detection result.

Since LSTM RNN is good at learning features of sequences, the LSTM RNN is utilized in this paper to detect misfire and to overcome the limitation of traditional algorithm.



(**a**)



(**b**)

Figure 5. Engine roughness under normal and misfire conditions. (**a**) Under 1200 r/min and no-load condition. (**b**) Under 1900 r/min and no-load condition.

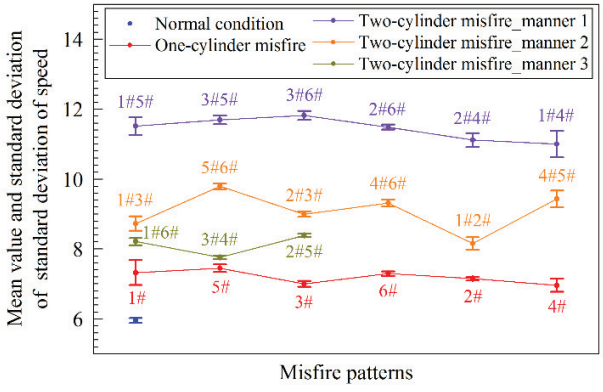


Figure 6. Mean values and standard deviations of the standard deviation of crankshaft speed under 800 r/min and no-load condition with different misfire patterns. Manner 1 represents misfire of two consecutive cylinders in firing order; manner 2 represents misfire of two cylinders with one-cylinder interval; manner 3 represents misfire of two cylinders with two-cylinder interval.

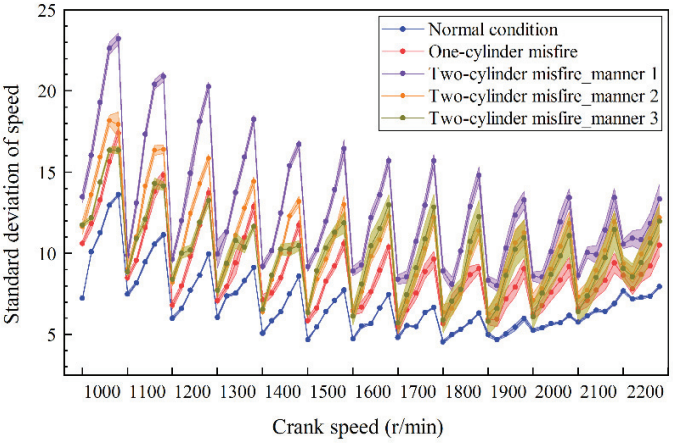


Figure 7. Standard deviation of one cycle speed under different speeds and loads with different misfire patterns. Fill areas mean error bars. Under each speed, the five points from left to right in a curve represent the five loads which are no-load, 100 Nm, 150 Nm, 200 Nm and 250 Nm. Moreover, manner 1, manner 2 and manner 3 have the same meaning with those in Figure 6.

4. The LSTM RNN

The classical artificial neural networks are design to extract features from datasets whose sub-samples are independent with each other. In some application scenarios, like natural language processing, the meaning of a whole sentence is dependent on the meaning and order of the previous and later words. RNNs are designed to be applied in this kind of research field. RNNs are connectionist models that capture the meaning of sequences via cycles in the network. Basic architecture of an RNN is shown in Figure 8, which is an unfold architecture.

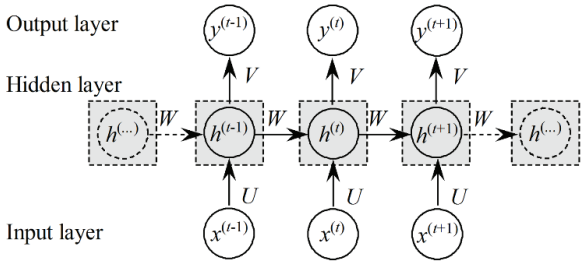


Figure 8. Basic architecture of an RNN module, showing shared parameters.

As presented in Figure 8, the forward pass of an RNN module looks the same as that of a multi-layer perceptron which has a single hidden layer. The main difference is that the activations of the hidden layer are from both the current input layer and the hidden layer activations from the previous step, as described in Equation (2) [29]. Equation (3) calculates the output value or vector. Thus, an RNN will map the input sequences into output.

$$h^{(t)} = f\left(b + Wh^{(t-1)} + Ux^{(t)}\right) \tag{2}$$

$$y^{(t)} = Vh^{(t)} \tag{3}$$

where W, U and V are the weight matrices. b, x, h, f, and y donate the bias vector, input vector, hidden layer vector, activation function and the output vector, respectively.

The classic RNN has the problem of a vanishing gradient [30]. In addition, sometimes gradient explosion will also occur. This is because the error surface is either very flat or very deep after updating weights in many time steps. This problem is also called the long-term dependency problem. One effective way to solve this problem is using gating mechanism to control the information passing path, such as LSTM.

LSTM RNN has the basic structure of RNN, which is a chain of repeating modules. The main difference of an LSTM RNN from other RNNs is the structure of the module, which is marked with shadow area in Figures 8 and 9. In a module of LSTM RNN, three gates are designed to control the output.

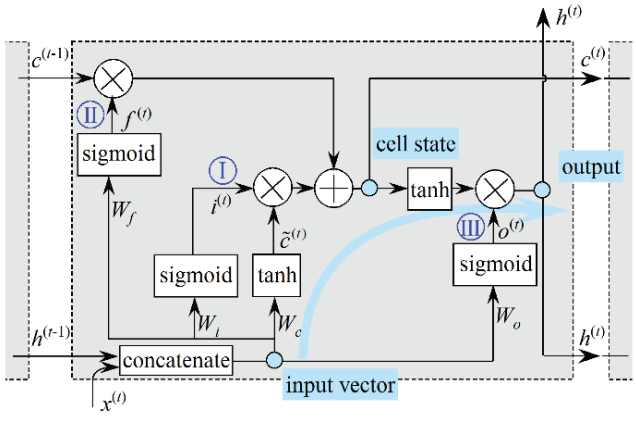


Figure 9. LSTM schematic. I, input gate; II, forget gate; III, output gate.

The main line of an LSTM module is the calculations of input vector, cell state and output, as indicated by the blue dot and arrow in Figure 9. First of all, the input vector of LSTM module is acquired by concatenating the outputs of the previous module and the current inputs. Secondly, two gates are designed to adjust the cell state. As shown

in Figure 9, the input gate is applied to decide whether the current inputs will be used to update the cell state. By the same principle, the forget gate is applied to adjust the proportion of previous cell state in the current one. This makes an LSTM module have the memory function. Then the cell state will be updated and stored for the LSTM module of next step. Next, an output gate is designed to adjust the output of the updated cell which has been rescaled by a tanh activation function firstly. Finally, the output will be transported to the next layer and the next module.

Equations (4) and (5) show the calculation of new candidate vector $\widetilde{c}^{(t)}$ and input gate vector $i^{(t)}$; Equation (6) calculates the forget gate vector $f^{(t)}$; the cell state $c^{(t)}$ can be updated by Equation (7); the output gate vector $o^{(t)}$ is calculated by Equation (8); and the final output $h^{(t)}$ will be acquired by Equation (9) [24]. The output vectors of these three gates are all the values between 1 and 0, which will make the outputs of corresponding layer change from original value to 0.

$$\widetilde{c}^{(t)} = \tanh\left(W_c \cdot [h^{(t-1)},\ x^{(t)}] + b_c\right) \tag{4}$$

$$i^{(t)} = \text{sigmoid}\left(W_i \cdot [h^{(t-1)},\ x^{(t)}] + b_i\right) \tag{5}$$

$$f^{(t)} = \text{sigmoid}\left(W_f \cdot [h^{(t-1)},\ x^{(t)}] + b_f\right) \tag{6}$$

$$c^{(t)} = f^{(t)} \circ c^{(t-1)} + i^{(t)} \circ \widetilde{c}^{(t)} \tag{7}$$

$$o^{(t)} = \text{sigmoid}\left(W_o[h^{(t-1)},\ x^{(t)}] + b_o\right) \tag{8}$$

$$h^{(t)} = o^{(t)} \circ \tanh(c^{(t)}) \tag{9}$$

where sigmoid and tanh are activation functions. $x^{(t)}$ is the input vector from training or testing dataset. $h^{(t-1)}$ and $h^{(t)}$ are the current and previous outputs of LSTM module, respectively. $[h^{(t-1)},\ x^{(t)}]$ means concatenating $h^{(t-1)}$ and $x^{(t)}$. b_c, b_i, b_f, and b_o are biases. W_c, W_i, W_f, and W_o are weight matrices. $\circ$ means Hadamard product. When the network is trained, b_c, b_i, b_f, and b_o are initialized with ones. For W_c, W_i, W_f, and W_o, each weight matrix is the concatenation of two matrices which are corresponding to $h^{(t-1)}$ and $x^{(t)}$, respectively; accordingly, the two parts of a weight matrix are initialized, respectively. In this work, both the two parts of each weight are initialized as uniform distribution which is shown in Equation (10) [31].

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}\right] \tag{10}$$

where n_j and n_{j+1} is the element number layer j and $j + 1$, respectively.

This is the key mechanism of LSTM RNN. The network training is also based on back propagation through time (BPTT) strategy and gradient descent algorithm. Up to now, there have been many types of variants on the LSTM, such as adding peephole connections, using coupled forget and input gates, gated recurrent unit, depth gated RNNs and so on. As reported in Greff's work [32], the result of comparing these popular LSTM variants shown that there were not significant differences among them. Therefore, the standard LSTM RNN is adopted in this paper.

5. Signal Processing and Results Analysis

5.1. Network Training Strategy

The experiments have been described in Section 2.2. There are 70 different engine running speeds and loads conditions in total. For each condition, 22 misfire types were conducted. Under each fixed speed, load and misfire type condition, 200 cycles data were sampled. Thus, the number of total datasets is 308,000 ($22 \times 70 \times 200 = 308{,}000$), and one dataset corresponds to one engine power cycle which contains 120 speed data point.

The datasets were acquired in a dense speed and load range. However, for industrial application, it would be better to use small number of datasets to train a well-performed neural network. Five division modes of training and testing datasets were attempted, as described in Table 3. The arrangement of mode 1 will be attempted firstly, and if the test result is higher than 90%, the rest arrangements will be tested. The arrangements of modes 2_a, 2_b, 2_c, and 2_d are shown in Figure 10, in which the training datasets are the conditions in shadow, the rest datasets are for testing. The arrangements in Figure 10a, c will be attempted first, and if the test results are higher than 90%, the arrangements in Figure 10b,d will be tested.

Table 3. Description of the training and testing datasets arrangement.

	Description
Mode 1	All the datasets are arrange randomly, then 70% datasets are chosen for network training and the rest for network testing.
Mode 2_a	As shown in Figure 10a.
Mode 2_b	As shown in Figure 10b.
Mode 2_c	As shown in Figure 10c.
Mode 2_d	As shown in Figure 10d.

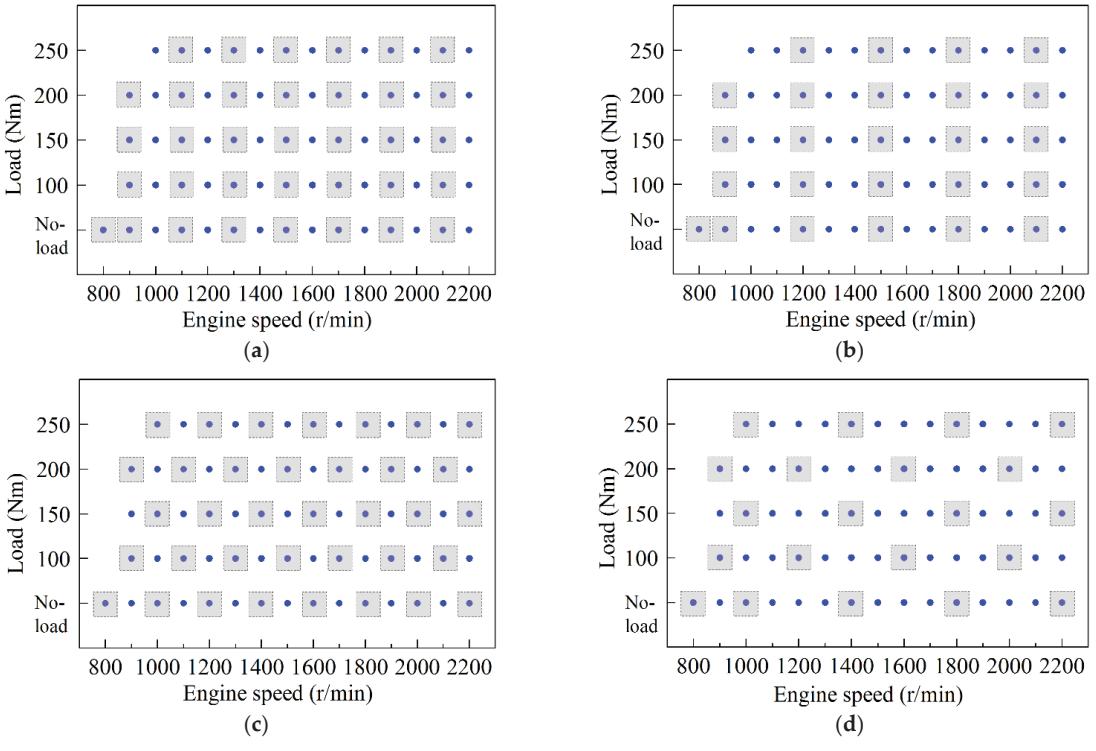


Figure 10. Different arrangements of training and testing datasets. (**a**) Arranging training data in 100 r/min interval. (**b**) Arranging training data in 200 r/min interval. (**c**) Arranging training data in dense speed and load interval. (**d**) Arranging training data in sparse speed and load interval. The engine running conditions in grey shadows are used for network training, the rest for network testing.

In order to achieve a better performance of the LSTM RNN, four different structures of input layer and output layer have been tested.

i. The first structure takes the original speed sequence as input. As shown in Figure 11a, one LSTM cell has one input element and one output element. For each engine cycle there are 120 speed points, and 120 output elements correspondingly. The output types are the normal and misfire types in Table 2.
ii. For the second structure, inputs are the same as those in the first structure. For each cycle, there is only one output element at the last LSTM cell. As shown in Figure 11b, when the 120 input elements have been calculated, one detection result will be output. The output types are the normal and misfire types in Table 2.
iii. In the third structure, one LSTM cell processes 20 raw speed points which correspond to the interval of ignition from one cylinder to the next. As shown in Figure 11c, the output of one LSTM cell consists of two categories: normal and misfire, which is different from those in the first and second structures.
iv. The inputs of the fourth structure are not the raw speed data, but the lowest 20 real and 20 imaginary parts of the frequency-domain results of instantaneous speed, as shown in Figure 11d. The output types are the normal and misfire types in Table 2.

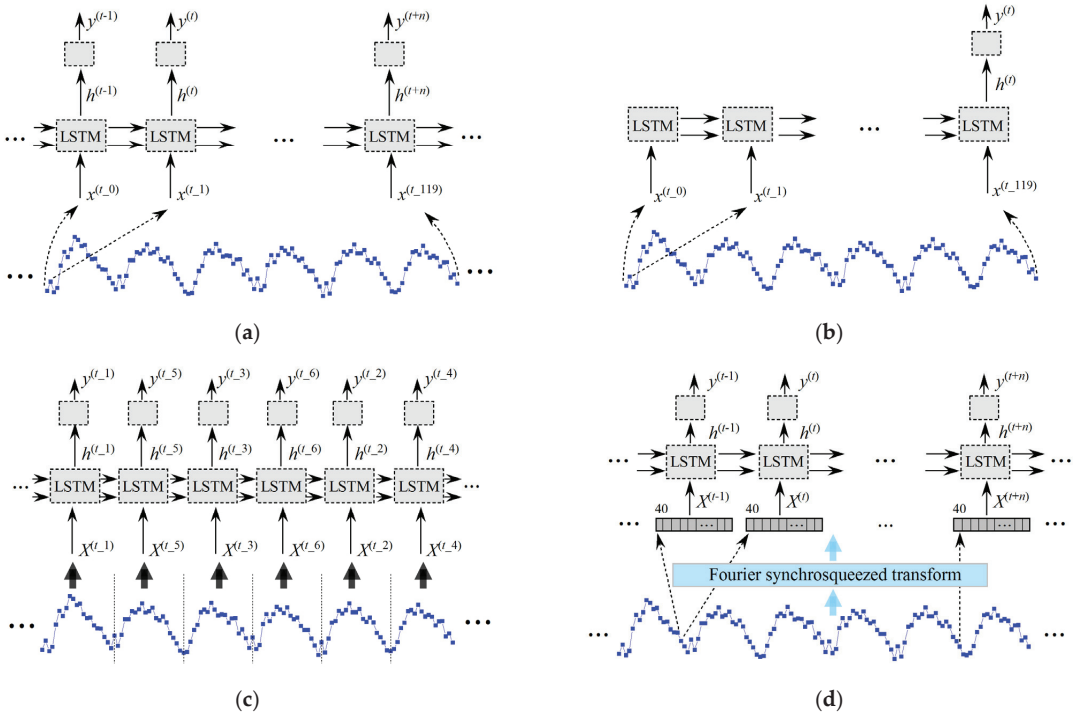


Figure 11. Different types of LSTM RNNs utilized in this paper. (**a**) LSTM with sequence inputs and sequence outputs. (**b**) LSTM with sequence inputs and last outputs. (**c**) LSTM with segment inputs and sequence outputs. (**d**) LSTM with segment inputs and sequence outputs.

The basic parameters for these four types of networks are summarized in Table 4.

Table 4. The basic parameters of networks utilized.

	The 1st Network	The 2nd Network	The 3rd Network	The 4th Network
Preprocessing	None	None	None	Fourier synchrosqueezed transform
Input mode of LSTM	Sequence	Sequence	Sequence	Sequence
Input size	1	1	20	40
Dimensionality of LSTM cell state (Number of hidden layer elements)	20	3, 5, 10, 20, 40, 80	3, 5, 10, 20, 40, 80	3, 5, 10, 20, 40, 80
Output mode of LSTM	Sequence	Last	Sequence	Sequence
Output size of fully connected layer	22	22	2	22
Postprocessing	None	None	Combining per cycle	None

5.2. Results Analysis

5.2.1. The First Network Structure

For the first structure which is designed as Figure 11a, one input group consists of 10 datasets, that is 1200 (120 × 10) speed data points. One output group has the same length with the corresponding input. The network consists of one input layer, one LSTM layer, and one output layer. The initial learn rate is 0.01 and the learn rate drop period is 3. The adaptive moment estimation method is adopted for network training. The number of elements of hidden layer is 20. When the training and testing datasets are arranged as mode 1, the final training and testing accuracies are 17.35% and 15.36%, respectively. Since the accuracies are not high, no more tests are attempted.

5.2.2. The Second Network Structure

The second structure is designed as Figure 11b. One input group consists of one dataset which is 120 speed data points. Only the last LSTM cell outputs prediction result. The network consists of one input layer, one LSTM layer, and one output layer. The initial learn rate, the learn rate drop period, and the training algorithm are the same as those in the first structure.

When the training and testing datasets are arranged as mode 1 (described in Table 3), the element numbers of LSTM layer, which are 3, 5, 10, 20, 40 and 80, have been tested. The corresponding training and testing accuracies are drawn in Figure 12. It is thus clear that when the training datasets are arranged as mode 1, 5 elements are enough for the network training, and the corresponding training and testing accuracies are 99.23% and 99.20%. Since the accuracies are very high, the datasets arranged in modes 2_a, 2_b, 2_c and 2_d are tested and the results are shown in Figure 13. It can be seen that utilizing the training datasets in a sparser manner, the acceptable performance can also be acquired. It seems that with less training data, it will be easier to train a network that can achieve the accuracy higher than 95%, such as the network trained in modes 2_b and 2_d.

5.2.3. The Third Network Structure

The third structure is designed as Figure 11c. One input layer consists of 20 elements which correspond to the interval of one-cylinder working. The output of one LSTM cell consists of two categories which are normal and misfire, and each LSTM cell has a corresponding output. The output indicates whether the current powering cylinder is fault-free. The initial learn rate, the learn rate drop period, and the training algorithm are the same as those in the first method.

The training and testing datasets are arranged as Table 3. The training strategy is also by changing the numbers of hidden layer elements, which are 3, 5, 10, 20, 40 and 80. The training and testing results under mode 1 are summarized in Figure 14. The training and testing results under modes 2_a, 2_b, 2_c, and 2_d are summarized in Figure 15. Since the outputs are calculated for each cylinder, it is unable to compare the original results with

other methods. Therefore, when we calculate the accuracy, the results for one cylinder are converted to that for one cycle. The concrete method is to group every six consecutive outputs from cylinder 1# and tag each group according to the fault cylinders. If three or more fault cylinders are detected in one cycle, the result will be categorized as a fault prediction. Compared with the second network structure, the third network structure has higher accuracy with the same number of hidden layer elements. For most of the five data division modes, 95% accuracy can be achieved with no more than 5 hidden layer elements.

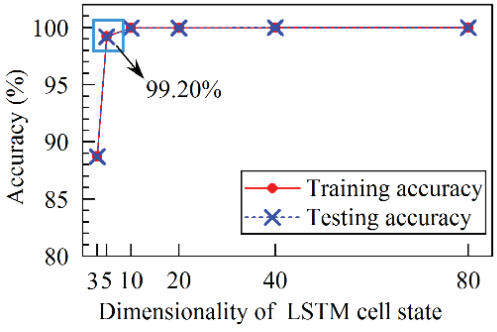


Figure 12. Training and testing accuracies of the second network structure and division mode 1. The cyan circle marks the accuracy that is more than 95% with the lowest hidden element number.

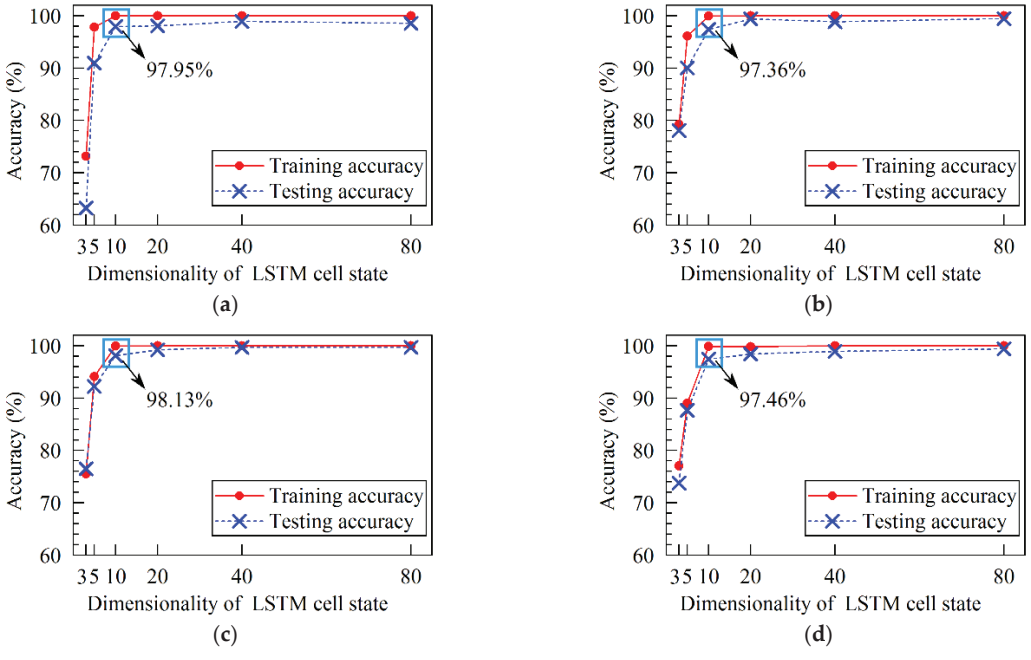


Figure 13. Training and testing accuracies of the second network structure. (**a**) Mode 2_a. (**b**) Mode 2_b. (**c**) Mode 2_c. (**d**) Mode 2_d. The cyan boxes mark the accuracies that are more than 95% with the lowest number of hidden layer elements.

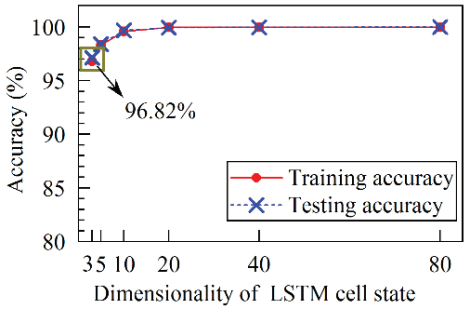


Figure 14. Training and testing accuracies of the third network structure and data division mode 1. The dark yellow circle marks the accuracy that is more than 95% with the lowest hidden element number.

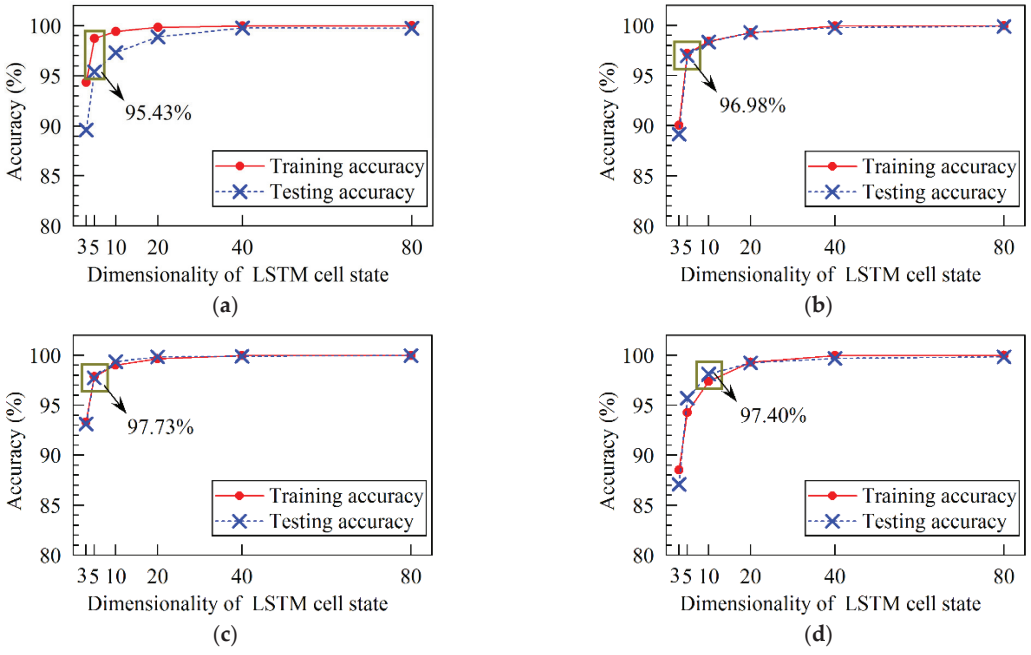


Figure 15. Training and testing accuracies of the third network structure. (**a**) Mode 2_a. (**b**) Mode 2_b. (**c**) Mode 2_c. (**d**) Mode 2_d. The dark yellow circles mark the accuracies that are more than 95% with the lowest number of hidden layer elements.

5.2.4. The Fourth Network Structure

The fourth structure is designed as Figure 11d. One input layer consists of 40 elements which are the lowest 20 real and 20 imaginary parts of the frequency-domain results of the speed signal. For the normalization of inputs, the 20 real parts and the 20 imaginary parts are normalized, respectively. The frequency-domain results are acquired by transforming the raw speed using Fourier synchrosqueezed transform algorithm [33]. As a type of time-frequency analysis method, the Fourier synchrosqueezed transform algorithm could acquire the instantaneous frequency-domain information more precisely than short-time Fourier transform. This advantage is helpful for abstracting fault features from the speed signal. Since there are 60 teeth of the flywheel, the sample rate is set as 60. The data length of one cycle is 120, and the data length of 2's power could achieve high computational

speed; thus, the truncated data length is set as 128. A Kaiser window is utilized for reducing spectral leakage. Considering that the output is in the manner of sequence which means each LSTM cell has an output, the accuracy is calculated according to the last output of one cycle. Figures 16 and 17 present the training and testing results using the fourth structure.

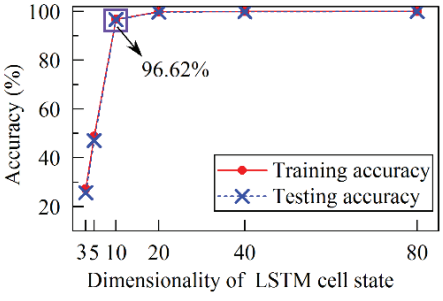


Figure 16. Training and testing accuracies of the fourth network structure and data division mode 1. The purple circle marks the accuracy that is more than 95% with the lowest hidden element number.

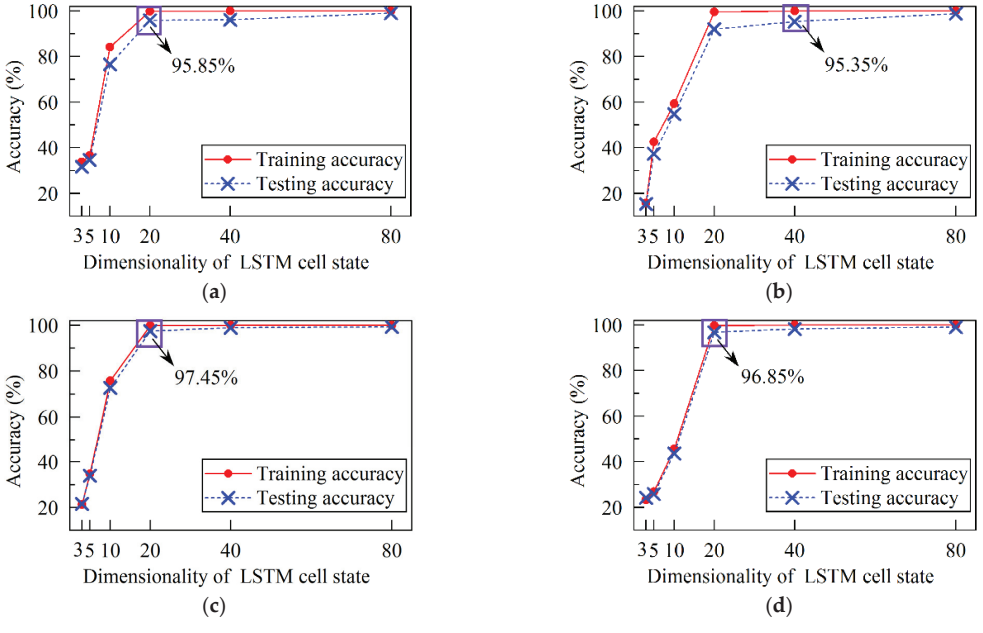


Figure 17. Training and testing accuracies of the fourth network structure. (**a**) Mode 2_a. (**b**) Mode 2_b. (**c**) Mode 2_c. (**d**) Mode 2_d. The purple circles mark the accuracies that are more than 95% with the lowest hidden element numbers.

Like the previous methods, the main variables are also the data division mode and the number of hidden layer elements. The results show that it is easy to achieve high accuracies which are more than 95% when the training data and testing data are arranged in mode 1. However, when the data are arranged as modes 2_a, 2_b, 2_c and 2_d, 20 or more hidden layer elements are needed for achieving 95% accuracy.

5.2.5. Results Comparison

(a) Through the above analysis, it is known that it acceptable training and testing accuracies can be achieved by using LSTM RNN.

(b) Compared with other network structures, the first network structure cannot achieve a satisfactory accuracy and is not studied further. The second, third and fourth network structures can achieve an accuracy which is more than 95%.

(c) From the point of view of data division mode, under mode 1, the second, third and fourth network structures can achieve accuracy more than 95%. The performance of these three structures is good, since 10 hidden layer elements have made them meet requirements and even 3 elements are enough for the third network. Under modes 2_a, 2_b, 2_c, and 2_d, the training and testing datasets are divided in the rule that they have no intersection on engine speed and load conditions. Under these four modes, even all the three methods could achieve accuracy more than 95%, to our minds, the third method performed the best since it could reach 95% accuracy with the least number of hidden layer elements.

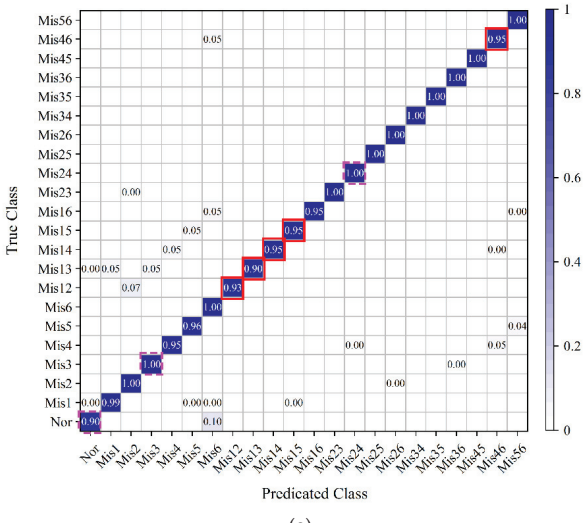
(d) In addition, when we tested the third network on data division mode 2_d with 5 and 10 hidden layer elements, it was found that the testing accuracy was higher than the training accuracy, as shown in Figure 15d. Taking the network structure with 10 hidden layer elements as an example, the confusion matrices are shown in Figure 18. The main increase of testing accuracies is marked by red boxes. It can be seen that many accuracies have increased to more than 95%. The increment may be caused by the testing data that belongs to the same fault category with the training data, but is sampled under different engine running speed or load. The network with 5 or 10 hidden layer elements just grasped the fault features of the training data unevenly. Therefore, the increase of total accuracy can be seen as a random result. However, the detection accuracy on the normal condition has decreased to 85%. Since the learning ability of network is determined when the network structure is determined, if there happened some excellent results on partial conditions, there must be some bad results correspondingly. The difference between excellent and bad results may be big or small, even the total accuracies are almost the same. In this case, the bad results occurred on the normal conditions. The accuracy of 85% is unacceptable on one hand, on the other hand, the misdiagnosis should be avoided on normal conditions since the normal conditions are the most common conditions for a vehicle. The results also indicate that both the total accuracy and the accuracy of each fault category should be acceptable especially under the condition of large datasets.

(e) In industrial applications, the smaller training dataset would be better for reducing workload for a calibration engineer. The best performance of the second, third and fourth networks under modes 2_a, 2_b, 2_c, and 2_d are summarized in Table 5. These accuracies are acquired with 40 or 80 hidden layer elements. By comparing the testing accuracies, the third network performs the best. Nevertheless, the difference between each network and each data division mode are not large. Both the training data size and the testing accuracy are the factors we considered, thus the third networks with data division mode 2_b and data division mode 2_c are recommended, and the numbers of hidden layer elements are 80 both. For further analysis, the confusion matrices of the training and testing results are drawn in Figures 19 and 20. Some conclusions can be drawn as follows.

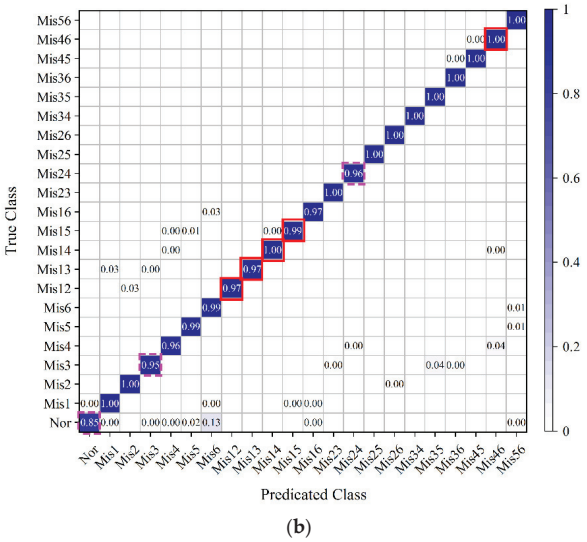
- It can be seen that the distribution of misdiagnosed results is scattered. The number of most misdiagnosed results do not exceed 10.
- In Figure 19, the main misdiagnosis results occur on normal, cylinder 3# misfire, cylinders 1# and 3# misfire, cylinders 4# and 5# misfire, and cylinders 4# and 6# misfire conditions. In Figure 20, the main misdiagnosis results occur on normal, cylinder 3# misfire, cylinders 4# and 5# misfire, and cylinders 4# and 6# misfire conditions. The misdiagnosed results are related to the true results, for example,

when misfire occurs in cylinder 3#, the predicted result is misfire in cylinders 3# and 5#.

- The most common running condition for an engine is normal condition. By observing the results in Figures 19 and 20, the worse misdiagnosed case is for normal condition which is presented in Figure 19b. However, the detection accuracy on this normal condition is 98.91% (8902 ÷ 9000 = 98.91%), which is a relatively high detection accuracy. Meanwhile, since the network performs well on the worst misdiagnosis case, it can be concluded that the detecting accuracy for each type of fault is acceptable.



(**a**)



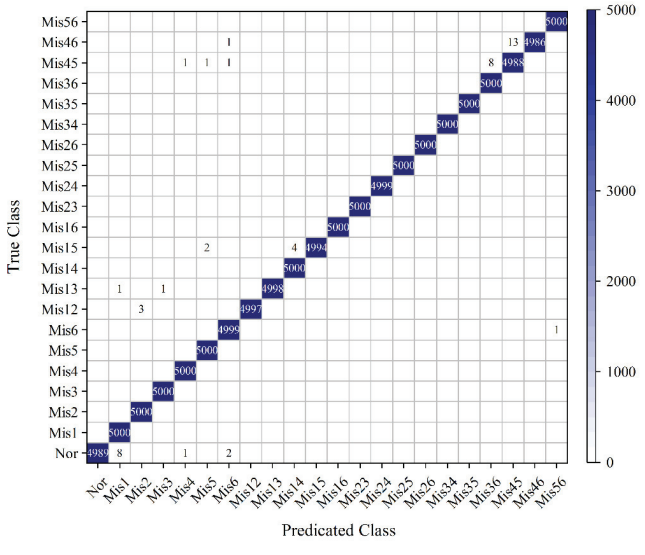
(**b**)

Figure 18. Training and testing results by using the third network and data division mode 2_d. (**a**) Training result. Total accuracy is 97.40%. (**b**) Testing result. Total accuracy is 98.12%. The red boxes mark the accuracies that increase more than 4%. The magenta boxes mark the accuracies that decrease more than 4%.

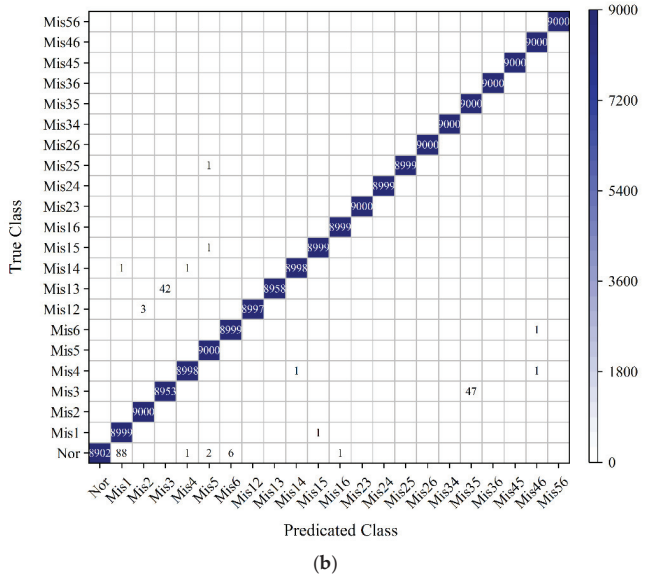
Table 5. The best accuracies for different networks under different data division modes.

	Mode 2_a (%)		Mode 2_b (%)		Mode 2_c (%)		Mode 2_d (%)	
The second network	99.99	98.93	99.98	99.44	99.99	99.69	100	99.37
The third network	99.98	99.76	**99.95**	**99.90**	**99.98**	**99.96**	99.99	99.83
The fourth network	99.99	99.14	99.97	98.83	99.97	99.33	100	99.06

For each mode, the left column is training accuracy, the right column is testing accuracy.

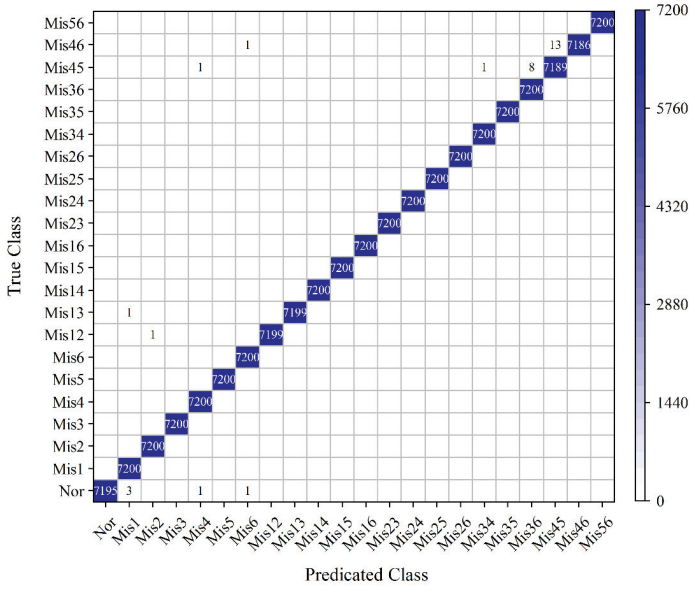


(**a**)

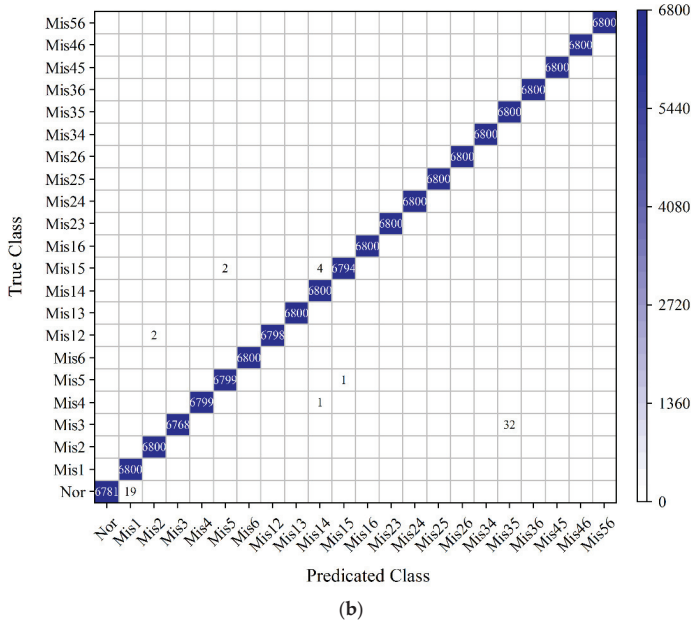


(**b**)

Figure 19. Confusion matrices of the results acquired with the third network and data division mode 2_b. (**a**) Training result. Total accuracy is 99.95%. (**b**) Testing result. Total accuracy is 99.90%.



(**a**)



(**b**)

Figure 20. Confusion matrices of the results acquired with the third network and data division mode 2_c. (**a**) Training result. Total accuracy is 99.98%. (**b**) Testing result. Total accuracy is 99.96%.

5.3. Comparison with Similar Research Efforts in the Literature

Table 6 provides a comparison of the results of this paper and some similar research efforts in the literature. Considering different application demands, the researchers conducted their studies on different types of engines. The main differences among these research works are the selection of engine running speed, running load, misfire types and

the fault detection algorithms. Although different research objects may lead to a different performance of an algorithm, such as that the four-cylinder engine has more clear fault features than the six-cylinder engine with the same displacement, the detection accuracy will still prove the effectiveness of an algorithm. On the whole, the accuracies reported in Table 6 are all relatively high, the comparison confirms the good performance of the algorithm utilized in this paper. In addition, many misfire types have been tested in this paper, which means more classification labels are needed for the algorithm, this also proves the effectiveness of the LSTM RNN algorithm. From the point of view of machine learning algorithm application, it is helpful for evaluating the network effectiveness if the datasets for network training and testing are sampled under different engine speed or load conditions. For example, if the network is trained on 1000 r/min and 1200 r/min, and it performs well on 1100 r/min, it is reasonable to infer that the network will perform well on 1150 r/min; however, if both the network training and testing are conducted under 1000 r/min and 1200 r/min, it is hard to evaluate the network performance on 1100 r/min or 1150 r/min. Compared with some research works in Table 6, the algorithm proposed in this paper are tested on the engine running conditions that are different from those for network training, which proves the feasibility of the algorithm.

Table 6. Comparison of our results with the similar works in the literature.

Similar Works	Details
Qin et al. [1]	Engine type: four-cylinder diesel engine Signal: vibration Speed: 1300 r/min, 1800 r/min, 2200 r/min Load (Nm): not mentioned Misfire type: 1#, 2#, 3#, 4#, 1#2#, 2#3#, 2#4# Methods: a deep twin convolutional neural network Accuracy: >97.019%
Jafarian et al. [21]	Engine type: four-cylinder engine Signal: vibration Speed: 2000 r/min Load (Nm): not mentioned Misfire type: 1#, 2#, 1#2# Methods: FFT for feature extraction; ANN, SVM, and kNN for classification Accuracy: >97%
Moosavian et al. [19]	Engine type: four-cylinder gasoline engine Signal: vibration, sound Speed: idle speed (867 r/min) Load (Nm): no load Misfire type: 1#, 2# Methods: wavelet denoising, ANN, least square support vector machine, Dempster–Shafer evidence Accuracy: 98.56%
Jung et al. [16]	Engine type: six-cylinder engine Signal: crank speed Speed: 500–4500 r/min, (step is 500 r/min) Load (Nm): not mentioned Misfire type: 1#, 2#, 3#, 4#, 5#, 6# Methods: model-based algorithm Accuracy: no quantitative result. The algorithm performs well except low load and speed conditions.

Table 6. *Cont.*

Similar Works	Details
Zheng et al. [5]	Engine type: four-cylinder gasoline engine Signal: crank speed Speed: 800–1150 r/min Load: not mentioned Misfire type: 1#, 2#, 3#, 4#, 1#3#, 2#4#, 1#4#, 2#3# Methods: state observer for combustion torque estimation; ANN for classification Accuracy: >52/54 (96.30%)
Boudaghi et al. [34]	Engine type: four-cylinder gasoline engine Signal: crank speed Speed: 1250–4000 r/min Load: 10–50% Misfire type: 1#, 2#, 3#, 4#, 1#3#, 2#4#, 1#4#, 2#3# Methods: extracting physics-based parameter Accuracy: >94%
Shahida et al. [35]	Engine type: twelve-cylinder diesel engine Signal: crank speed Speed: 720 r/min Load: 0–100% Misfire type: A1#, A6# Methods: one-dimensional convolutional neural network Accuracy: >99.7%
This paper	Engine type: six-cylinder diesel engine Signal: crank speed Speed: 800–2200 r/min, (step is 100 r/min) Load: no-load to 250 Nm Misfire type: 1#, 2#, 3#, 4#, 5#, 6#,1#2#, 1#3#, 1#4#, 1#5#, 1#6#, 2#3#, 2#4#, 2#5#, 2#6#, 3#4#, 3#5#, 3#6#, 4#5#, 4#6#, 5#6# Methods: LSTM RNN Accuracy: >99.90%

6. Conclusions

In this paper, an LSTM RNN based approach for engine misfire detection is proposed.

The traditional misfire detection method has limitations on the high-speed and low-load engine operating conditions. Hence, the traditional misfire detection method is conducted on the datasets to verify its feasibility first; and the reason of the limitation, that one threshold is insufficient to extract the fault feature when the background noise is high, is concluded. In order to extract the fault features extensively and effectively, unlike previous works, the LSTM RNN is a powerful technique on sequence signal processing is utilized to detect misfire. In addition, for the sake of ensuring the feasibility of proposed algorithm, two-cylinder misfire faults are tested beside one-cylinder faults, and a wide range of engine working speed and load conditions which including the high-speed and low-load conditions are tested.

The LSTM RNNs are designed according to the characteristic of speed signal. Four kinds of input layer structures are designed. These inputs contain instantaneous raw speed signal, a fixed segment of raw speed signal, and the extracted real and imaginary parts of speed signal. Moreover, five data division modes are attempted to explore the optimal training data size. These training datasets can be categorized into two parts: the training data that has running condition intersection with the testing data, and the training data that has no running condition intersection with the testing data. The testing results show that the sequence-input-sequence-output LSTM RNN which utilizes raw speed data could not achieve acceptable detecting accuracy. The second, third and fourth LSTM RNNs could achieve accuracies more than 98%. The best performance is achieved by the third LSTM RNN with data division mode 2_c, and the testing accuracy is 99.96%. Meanwhile, the

third LSTM RNN with data division mode 2_b is also recommended, because it has the relatively high testing accuracy 99.90% and small training data size as well.

In this study, misfire detection is conducted on complete misfire conditions. It is also significant that misfire fault could be detected when it is not severe. Therefore, in further research, the slight misfire fault including partial misfire will be utilized to improve the detection sensitivity of the proposed algorithm. In addition, future work will include developing hardware for misfire detection of this engine as well. The LSTM RNN models developed in this study will then be written into the hardware to provide misfire information.

Author Contributions: Conceptualization, W.G. and P.Z.; methodology, X.W. and P.Z.; software, P.Z.; validation, W.G.; formal analysis, P.Z.; investigation, P.Z., Y.L., Y.W. and H.P.; resources, X.W. and W.G.; data curation, P.Z.; writing—original draft preparation, P.Z.; writing—review and editing, W.G.; visualization, P.Z.; supervision, W.G.; project administration, X.W. and W.G.; funding acquisition, X.W. and W.G. All authors contributed to this work by collaboration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to convey special gratitude to the State Key Laboratory of Engine Reliability for sponsoring this research (No. skler-202010).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

LSTM	long short-term memory
RNN	recurrent neural network
ANN	artificial neural network
SVM	support vector machine
kNN	k-nearest neighbor
FFT	fast Fourier transform
BPTT	back propagation through time

References

1. Qin, C.J.; Jin, Y.R.; Tao, J.F.; Xiao, D.Y.; Yu, H.G.; Liu, C.; Shi, G.; Lei, J.B.; Liu, C.L. DTCNNMI: A deep twin convolutional neural networks with multi-domain inputs for strongly noisy diesel engine misfire detection. *Measurement* **2021**, *180*, 109548. [CrossRef]
2. Jafari, M.; Borghesani, P.; Verma, P.; Eslaminejad, A.; Ristovski, Z.; Brown, R. Detection of misfire in a six-cylinder diesel engine using acoustic emission signals. In Proceedings of the ASME 2018 International Mechanical Engineering Congress and Exposition, Pittsburgh, PA, USA, 9–15 November 2018.
3. Chung, Y.; Bae, C.; Choi, S. Application of a wide range oxygen sensor for the misfire detection. In Proceedings of the International Fuels & Lubricants Meeting & Exposition, Dearborn, MI, USA, 3–6 May 1999. SAE Paper No. 1999-01-1485.
4. Fan, Q.; Bian, J.; Lu, H.; Tong, S.; Li, L. Misfire detection and re-ignition control by ion current signal feedback during cold start in two-stage directinjection engines. *Int. J. Engine Res.* **2012**, *15*, 37–47. [CrossRef]
5. Zheng, T.; Zhang, Y.; Li, Y.; Shi, L. Real-time combustion torque estimation and dynamic misfire fault diagnosis in gasoline engine. *Mech. Syst. Signal Proc.* **2019**, *126*, 521–535. [CrossRef]
6. Rizvi, M.A.; Bhatti, A.I.; Butt, Q.R. Hybrid model of the gasoline engine for misfire detection. *IEEE Trans. Ind. Electron.* **2011**, *58*, 3680–3692. [CrossRef]
7. Helm, S.; Kozek, M.; Jakubek, S. Combustion torque estimation and misfire detection for calibration of combustion engines by parametric Kalman filtering. *IEEE Trans. Ind. Electron.* **2012**, *59*, 4326–4337. [CrossRef]
8. Hmida, A.; Hammami, A.; Chaari, F.; Ben Amar, M.; Haddar, M. Effects of misfire on the dynamic behavior of gasoline engine crankshafts. *Eng. Fail. Anal.* **2021**, *121*, 105149. [CrossRef]
9. Wong, P.K.; Zhong, J.; Yang, Z.; Vong, C.M. Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis. *Neurocomputing* **2016**, *174*, 331–343. [CrossRef]

10. Plapp, G.; Klenk, M.; Moser, W. Methods of on-board misfire detection. In Proceedings of the International Congress and Exposition, Detroit, MI, USA, 26 February–2 March 1990. SAE Paper No. 900232.
11. Taraza, D.; Henein, N.A.; Bryzik, W. The frequency analysis of the crankshaft's speed variation: A reliable tool for diesel engine diagnosis. *J. Eng. Gas Turbines Power* **2001**, *123*, 428–432. [CrossRef]
12. Geveci, M.; Osburn, A.W.; Franchek, M.A. An investigation of crankshaft oscillations for cylinder health diagnostics. *Mech. Syst. Signal Proc.* **2005**, *19*, 1107–1134. [CrossRef]
13. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Proc.* **2020**, *138*, 106587. [CrossRef]
14. Li, Z.; Yan, X.; Yuan, C.; Peng, Z. Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed. *J. Mech. Sci. Technol.* **2012**, *26*, 2413–2424. [CrossRef]
15. Chen, J.; Bond Randall, R. Improved automated diagnosis of misfire in internal combustion engines based on simulation models. *Mech. Syst. Signal Proc.* **2015**, *64–65*, 58–83. [CrossRef]
16. Jung, D.; Eriksson, L.; Frisk, E.; Krysander, M. Development of misfire detection algorithm using quantitative FDI performance analysis. *Control. Eng. Pract.* **2015**, *34*, 49–60. [CrossRef]
17. Gani, E.; Manzie, C. Misfire-misfuel classification using support vector machines. *Proc. Inst. Mech. Eng. Part D-J. Automob. Eng.* **2007**, *221*, 1183–1195. [CrossRef]
18. Sharma, A.; Sugumaran, V.; Babu Devasenapati, S. Misfire detection in an IC engine using vibration signal and decision tree algorithms. *Measurement* **2014**, *50*, 370–380. [CrossRef]
19. Moosavian, A.; Khazaee, M.; Najafi, G.; Kettner, M.; Mamat, R. Spark plug fault recognition based on sensor fusion and classifier combination using Dempster–Shafer evidence theory. *Appl. Acoust.* **2015**, *93*, 120–129. [CrossRef]
20. Gu, C.; Qiao, X.Y.; Li, H.; Jin, Y.; Castejón, C. Misfire fault diagnosis method for diesel engine based on MEMD and dispersion entropy. *Shock Vib.* **2021**, *2021*, 9213697. [CrossRef]
21. Jafarian, K.; Mobin, M.; Jafari-Marandi, R.; Rabiei, E. Misfire and valve clearance faults detection in the combustion engines based on a multi-sensor vibration signal monitoring. *Measurement* **2018**, *128*, 527–536. [CrossRef]
22. Liu, B.; Zhao, C.; Zhang, F.; Cui, T.; Su, J. Misfire detection of a turbocharged diesel engine by using artificial neural networks. *Appl. Therm. Eng.* **2013**, *55*, 26–32. [CrossRef]
23. Bahri, B.; Aziz, A.A.; Shahbakhti, M.; Muhamad Said, M.F. Understanding and detecting misfire in an HCCI engine fuelled with ethanol. *Appl. Energy* **2013**, *108*, 24–33. [CrossRef]
24. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A critical review of recurrent neural networks for sequence learning. *arXiv* **2015**, arXiv:1506.00019.
25. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef] [PubMed]
26. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
27. Song, Q.; Gao, W.; Zhang, P.; Liu, J.; Wei, Z. Detection of engine misfire using characteristic harmonics of angular acceleration. *Proc. Inst. Mech. Eng. Part D-J. Automob. Eng.* **2019**, *233*, 3816–3823. [CrossRef]
28. Klenk, M.; Moser, W.; Mueller, W.; Wimmer, W. Misfire detection by evaluating crankshaft speed—A means to comply with OBDll. In Proceedings of the SAE International Congress and Exposition, Detroit, MI, USA, 26 February–5 March 1993. SAE Paper No. 930399.
29. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; pp. 367–399.
30. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.
31. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS 2010), Sardinia, Italy, 13–15 May 2010.
32. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef]
33. Thakur, G.; Wu, H.T. Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples. *SIAM J. Math. Anal.* **2011**, *43*, 2078–2095. [CrossRef]
34. Boudaghi, M.; Shahbakhti, M.; Jazayeri, S.A. Misfire detection of spark ignition engines using a new technique based on mean output power. *J. Eng. Gas Turbines Power* **2015**, *137*, 091509. [CrossRef]
35. Shahid, S.M.; Ko, S.; Kwon, S. Real-time abnormality detection and classification in diesel engine operations with convolutional neural network. *Expert Syst. Appl.* 2021; *in press*.

sustainability

MDPI

Article

The Efficiency of Drones Usage for Safety and Rescue Operations in an Open Area: A Case from Poland

Norbert Tuśnio [1,*] and Wojciech Wróblewski [2]

[1] Faculty of Safety Engineering and Civil Protection, The Main School of Fire Service, 01-629 Warsaw, Poland
[2] Internal Security Institute, The Main School of Fire Service, 01-629 Warsaw, Poland; wwroblewski@sgsp.edu.pl
* Correspondence: ntusnio@sgsp.edu.pl

Abstract: The use of unmanned aerial systems (UAS) is becoming increasingly frequent during search and rescue (SAR) operations conducted to find missing persons. These systems have proven to be particularly useful for operations executed in the wilderness, i.e., in open and mountainous areas. The successful implementation of those systems is possible thanks to the potential offered by unmanned aerial vehicles (UAVs), which help achieve a considerable reduction in operational times and consequently allow a much quicker finding of lost persons. This is crucial to enhance their chances of survival in extreme conditions (withholding hydration, food and medicine, and hypothermia). The paper presents the results of a preliminary assessment of a search and rescue method conducted in an unknown terrain, where groups were coordinated with the use of UAVs and a ground control station (GCS) workstation. The conducted analysis was focused on assessing conditions that would help minimise the time of arrival of the rescue team to the target, which in real conditions could be a missing person identified on aerial images. The results of executed field tests have proven that the time necessary to reach injured persons can be substantially shortened if imaging recorded by UAV is deployed, as it considerably enhances the chance of survival in an emergency situation. The GCS workstation is also one of the crucial components in the search system, which assures image transmission from the UAV to participants of the search operation and radio signal amplification in a difficult terrain. The effectiveness of the search system was tested by comparing the arrival times of teams equipped with GPS and a compass and those not equipped with such equipment. The article also outlined the possibilities of extending the functionality of the search system with the SARUAV module, which was used to find a missing person in Poland.

Keywords: unmanned aerial systems; search and rescue operations; missing people; data transmission devices; automatic flight

Citation: Tuśnio, N.; Wróblewski, W. The Efficiency of Drones Usage for Safety and Rescue Operations in an Open Area: A Case from Poland. *Sustainability* **2022**, *14*, 327. https://doi.org/10.3390/su14010327

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 16 November 2021
Accepted: 26 December 2021
Published: 29 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Every year ca. 20,000–30,000 people go missing in Poland. Most of them are found on the day of their disappearance, but approximately 4000 continue to be missing over a longer period [1]. As it comes to search and rescue operations, the key factor in this type of situation is the time required to reach the missing person. Until recently, specialist search and rescue groups used specially trained dogs or modern basic technological solutions, such as thermal imaging. Nowadays, units delegated to such actions tend to deploy a wide assortment of technological solutions, which was presented in [2] for three different operational paradigms supporting this type of action in the field, and factors that affect the determination of the best paradigm in the human–robot system were specified, which include the following.

- Sequential operations—a strategy appropriate for search operations that need to be executed in a difficult terrain, with restricted mobility, in situations with limited

available human resources operating in the field and where the likelihood of the missing person's presence is evenly distributed over a large area;
- Remote-controlled operations—suitable for situations where the mission manager has more access to relevant information from the field than from the base station;
- Base-directed operations—appropriate for areas that offer high mobility for the field search team, but not enough information that would be required to conduct a hasty search.

The location of a missing subject is crucial to allow selecting the most advantageous paradigm under the given circumstances and modelling the behaviour of missing persons in a given situation is just as important. The technologies used so far greatly contribute to reducing the time required to reach missing subjects, which is particularly significant in the case of injuries caused by falls, hypothermia, dehydration and chronic illnesses. However, those technologies need to be continuously improved to reduce the mission times to a minimum.

In Poland search and rescue operations involve a lot of specialised units, among others also the State Fire Service, which carries such operations in the basic, specialist and specialised fields of international humanitarian assistance [3]. In Poland, in situations of disappearances taking place in an open ground, the leading entity is the police, which, on the basis of a concluded agreement, cooperate with search and rescue groups both in the structures of the state and volunteer fire brigades and other civil associations, e.g., Mountain Voluntary Rescue Service [4].

Unmanned aerial systems supported by devices that provide communication and data transmission from UAVs are increasingly frequently used in search and rescue operations for missing persons. This technology is still being validated by fire protection units, among others. Consequently, an attempt was made to develop a solution to support this type of action in the open ground, which involved the Main School of Fire Service (MSFS) and Volunteer Fire Department (VFD) Niegoszowice. The undertaken analysis demonstrates the effectiveness of actions taken by rescuers using UAS and supported by a GCS workstation, which assures a local Wi-Fi network and web server, internet access, video transmission from UAVs to mobile devices and a powerful radio station. The study takes into consideration not only UAVs as such, but also the environment related to their operation and coordination of ground groups and modern devices for image transmission and communication assurance. Until now, the use of UAVs has consisted of terrain observation and image analysis, including thermal imaging. A part of the performed tasks included producing an orthophotomap composed of images taken from one or many UAVs. An analysis of the orthophotomap made it possible to determine the location coordinates of the missing subject. The effectiveness of the use of UAVs in search and rescue operations was proven (Tychy [5], Lincolnshire and Fort Wayne [6]). However, this is not the full extent of their capabilities, and the advancements in this technology confirm the increasingly frequently implemented direction, and namely the use of automated systems based on artificial intelligence [7]. One such solution is the SARUAV software, which facilitates and speeds up search and rescue operations in which UAVs are used [8]. The system is a very useful tool that can be crucial in many search and rescue operations. Mountain Volunteer Rescue Service successfully used the system during a search operation for a missing person. The incident took place on 29 June 2021 [9].

The literature indicates that similar solutions exist in the area under study, such as Loc8 software or the MOBNET system. The differences between these and SARUAV are shown in Table 1.

Table 1. Comparison of the capabilities of the SARUAV system and other solutions with similar functionality.

Item	System (Software or Specialist UAV)	Description of System	Source
1.	Search and Rescue with Unmanned Aerial Vehicle	Algorithm for searching for missing persons in undeveloped areas. The system has 2 modules for the following: 1. Determining the maximum human walking range in a given time. 2. Automatic image analysis and identification of potential human locations.	[8]
2.	Loc8: Image Scanning Software for Search and Rescue	Software used to scan images or videos after entering colour that is being searched for (e.g., clothing). Finding a match is signalled by an alarm and indication of the location. The target is then verified and rescuers are dispatched to the location. It can even analyse satellite images.	[10]
3.	Mobile network for people's location in natural and man-made disasters	A system for locating victims during natural disasters and crisis situations, such as earthquakes, hurricanes or major blizzards. The basic assumption of its operation is related to the fact that the searched person has a working and switched-on mobile phone.	[11]
4.	Multi-task UAV	A rotary wing flying platform designed for flight in mountainous terrain at negative temperatures, high altitudes and strong winds. Equipped with an avalanche detector, cameras (daylight and thermal imaging) and various payloads (rescue kits, special explosive cartridge for controlled triggering of an avalanche). Capable of fully autonomous flight and terrain search.	[12]
5.	MAGI: Multistream aerial segmentation of ground images	A fast image recognition algorithm with which, thanks to the hardware used, real-time performance can be achieved. The model is suitable for operations where time is critical, such as fire detection and search and rescue operations.	[13]
6.	RGDiNet: efficient onboard object detection	A multimodal platform for real-time object detection that can be mounted on a UAV and which is insensitive to changes in the brightness of the surrounding environment.	[14]

The analysis shows that there are several image recognition systems on the market, but only SARUAV works offer a ready-made map, which makes it possible to determine the area that could be reached by the missing subject, based on their speed and taking into account terrain difficulties.

There is also a trend to build dedicated UAS appropriately to the type of threat, but the optimal solution in search and rescue operations seems to be to use any UAV to take pictures and then send the rescue teams to the place pre-designated by the software. This is how both the SARUAV and Loc8 software are designed.

Consequently, the aim of this publication is to analyse the process of cooperation between particular services and the effects of search and rescue activities, as well as to estimate the conditions for minimising the time of arrival of rescuers to the injured person in open areas. An outline is made of UAVs and IT technologies (such as GCS workstation), which are capable of supporting the search for missing persons and should be in use by services responsible for safety and civil protection.

In the experiment carried out in Nowy Dwór Mazowiecki, during the testing phase, selected systems supporting search and rescue operations (UAVs, and command suitcase) were used. For formal reasons, the SARUAV system was not used for testing, and a concept was presented that extended the functionality of the coordinated search and rescue group with the SARUAV system. We found some license restrictions in one province due to the lack of availability of appropriate digital maps. For each launch of the system, it is necessary to prepare dedicated spatial data, and the automatic detection of people

works in a specific area "purchased" by the recipient of the system. Evidence of the effectiveness of the presented concept with the SARUAV model was demonstrated in the form of positive results of finding a living person and may constitute an effective extension of the functionality of the system used in search and rescue operations [15]. The solution uses a nested k-means algorithm to detect people in aerial photos from close proximity [16]. The software uses a variety of non-statistical detection methods [15]. The SARUAV system was trained on pictures of people wearing clothes of different colours and patterns (including smaller objects: children and dogs on the pictures). The training sets were made with the engagement of the study participants. The SARUAV system effectively detects the figures of any dressed adults and children, and the presence of dogs (it does not detect big animals). The key parameter for the SARUAV system is the ground sample distance (GSD), which depends on the flight altitude and the characteristics of the camera. It is best to fly so that the GSD is less than 3 cm / px. The two detection algorithms work in parallel, only on nadir and RGB images. The conducted near infrared (NIR) tests gave less satisfactory detection results. From the perspective of the algorithms used, the far-infrared imaging has insufficient resolution.

2. Methods

The conducted research focused on the characteristics, specifics and prospects of deploying UAVs in rescue operations. Tests related to the usage of UAVs in rescue operations, prepared by the VFD, took place on 8 July 2021 at the Base of Training and Rescue Innovation at the Main School of Fire Service in Nowy Dwór Mazowiecki (Figure 1).

Figure 1. Practice area on the training grounds in Nowy Dwór Mazowiecki. Source: Google Maps.

Hardware and software components of the VFD concept system (listed in the Appendix A) were subjected to testing, as well as the possibility of using UAVs and the GCS workstation (Figure 2).

The MSFS took part in simulated operations. The participants of those simulations coordinated actions, using preview and the public address system integrated with UAVs. The tests verified the complementarity of the information exchanged, which can be used in operations, along with the level of interoperability between the SFS and the VFD.

The participants were divided into two groups with Alpha and Delta identifiers. Starting from different locations on the training ground, they were to reach a specific point according to commands received via radio from the head of the rescue operation (HRO), who directed them by comparing an orthophotomap of the terrain with their position indicated by the drones. Additionally, communication failures were simulated in the Alpha group. The HRO could only transmit commands and messages via the speaker mounted on the DJI Mavic 2 Enterprise Dual drone. However, there was no return radio channel,

and the accuracy of the executed commands was judged by the HRO on the basis of the drones' image. Two parts of this phase of the exercise were executed. In the first one, the participants of the study went into action without prior preparation and without having at their disposal any locating equipment (GPS, compass). It was not easy to coordinate their work, and the commands, directions, reference points proved to be ambiguous. Despite these difficulties, both teams were successfully led to the set-out point. At the second attempt, the participants of the exercise already knew the terrain and were equipped with locating devices; therefore, the time of reaching the target was significantly shortened.

Figure 2. GCS workstation designed and built by VFD members. Photo: N. Tuśnio.

The experimental methods included the study of the following:

- Resilience of the radio communication system to interferences and harsh terrain conditions;
- Resistance to communication interference between UAV and the control device;
- Elements that can excite frequencies that could affect the work of UAV during operations;
- Elements that reduce the risk of losing control to an acceptable level of risk;
- The quality of images from UAV cameras and their usefulness in operations;
- The audibility and intelligibility of messages emitted from the UAV integrated public address system;
- The time required to search for a missing person without and with the system.

Tests using UAVs: AUTEL EVO II and DJI Mavic 2 Enterprise Dual were executed according to the guidelines provided in [17–19] and in internal VFD working arrangements. A simulation was carried out of search and rescue operations, during which teams were coordinated with the use of UAVs. A prototype GCS workstation was used, which allowed real-time viewing of UAVs on the HRO workstation [20].

In the first stage of the exercises, a precise orthophotomap of the area was produced. For this purpose, a UAV AUTEL EVO II with a camera of 8K resolution was used, which performed a flight at an altitude of 90 m in the mode of serial nadir images (camera directed perpendicularly to the surface of the earth) [21]. The flight route was prepared in the ladder method. This method allows a regular coverage of the area with shots overlapping each other in 50%–70% to enable the establishment of a very accurate orthophotomap of a given site, which can then serve as a basis for further operational activities. A similar method was successfully used during the Biebrza National Park fire in Poland in 2020 [22].

Three exercise scenarios were carried out as part of the research: 1—creation of orthophotomap in field conditions; 2—coordination of ground groups; and 3—terrain

observation with using a thermal imaging camera. In each case, three people controlled the flight operation (UAV operator, observer technician, and team leader).

A time measurement technique allowed recording the experiment on a video camera and then reading out the individual task times.

On the day of the experiment, the weather conditions in Nowy Dwór Mazowiecki were as follows: cloudless, light rainfall (1 mm), daytime temperature 29 °C, pressure 1019.5 hPa, and wind speed 20.25 km/h, direction south-east.

A search algorithm involving HRO targeting teams equipped with GPS and a compass was used, and tests were executed of the effectiveness of operations without this equipment and in the event of communication failure.

2.1. Structure and Organisation of the UAVs Section

The organisational structure of the UAVs section during operational activities arises not only from aviation law regulations (the specific role of the UAV operator), but also from actual operational needs.

The UAV operator is the person responsible for the execution of each flight and makes the final decision of whether a flight can be performed at a given location. It is his/her responsibility to carry out the technical inspection of the UAV before take-off, control it and assure safety of the mission performed.

The responsibility of a technician, also acting as an observer, is the preparation of other necessary equipment including running the GCS workstation, establishing communications with other units and observing the UAV flight. During the exercise, he/she remains with one of the cadet teams to assess their performance. The role of the technician is also to secure the area of flight operations to make sure that especially take-offs and landings are done in a manner safe for bystanders.

The team leader acts primarily as a liaison between the UAVs unit and the HRO (or other services). His/her task is to be a kind of information filter so that the operators can fully concentrate on their tasks. He/she is also responsible for coordinating the activities of various operators who, controlling UAVs in the same space, must correlate their actions so as to ensure an adequate level of flight safety.

An important issue when conducting operations involving the use of unmanned aerial systems is the time of data transmission. A method is adopted in the organisation of operational activities that introduces parallel actions upon arrival in the area of operation in the following operational algorithm:

1. Operators prepare the UAVs for launch procedure in the shortest operational time possible while maintaining relevant safety rules.
2. The technician prepares GCS workstation and activates communication systems including the LTE/5G Wi-Fi mast.
3. The flight team leader establishes a detailed action plan with the other participants involved in the operation and supervises the completeness of the launch procedures.

All this allows launching of the UAV within 3 min of arrival in the operational area. The first image in the GCS workstation can be obtained as early as in the 6th minute, and the image in the mobile devices of the participants of the operation in the 8th minute after the start of the operation.

2.2. Equipment Resources

The VFD unit is equipped with the following technology and software (Table 1 in the Annex A lists the equipment used during the exercises):

1. Two UAVs: DJI Mavic 2 Enterprise Dual and AUTEL EVO II.
2. PIX4Dreact—the application enables the mapping of the area of action (making an orthophotomap). On the basis of the UAV flight, an accurate and up-to-date situational map is established in field conditions.

3. Live preview RTMP—the image from the drones is visible on monitors (laptop, GCS workstation) and on smartphones of the participants. It is also possible to transmit the image via the Internet.
4. Internet access—a Wi-Fi MESH network with Internet access is set up. This enables information to be passed on quickly to rescuers.
5. LOCAL WWW—the GCS workstation has its own web server, which makes it possible to record and present information that is relevant for the operation (e.g., search report).
6. SARUAV—supports the search for missing persons through numerical modelling of movements and analysis of images from UAVs.
7. Thermal imaging camera—one of the UAVs equipped with it, together with a telemetry system, allows precise identification of temperatures in the field of view of the camera.
8. NFRS radio communications—high-power mobile radio with high-performance antenna that allows maintaining communications in difficult conditions.
9. Internal radio communication—communication within the group of UAV operators takes place on dedicated equipment and frequencies so as not to cause interference or interference with other services.

Worthy of particular attention is the GCS workstation developed by VFD members (see Figure 3 for a working diagram), which allows the following:

1. Multiplication (replication) of drone images and their transmission to other devices both on the Internet and on a local network.
2. Creating a local Wi-Fi network to provide access to the Internet and to local information as well as video from UAVs.
3. Communication set-up also in difficult terrain conditions, where ordinary mobile or portable radios are unable to cope with communication requirements (it is equipped with a mobile radio in the SFS radio communication standard with an antenna of high energy gain).

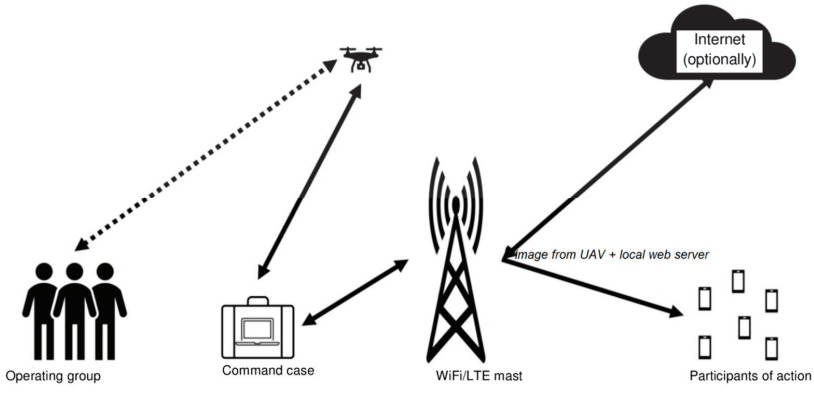


Figure 3. Operational diagram of an operational group equipped with UAV and GCS workstation.

A key issue in the assembly of the GCS workstation is the proper separation of all connections and the elimination of any potential radio interference.

Both the Wi-Fi network and UAVs communication operate in the 2.4 and 5 GHz standards. It is therefore important to choose the optimal slot (channel) for connections to avoid possible interference with the connection between UAVs and control equipment. This may be done by using high quality Ubiquiti network devices, which constantly verify the network occupancy.

All cables and connections used in the GCS workstation need to be shielded to minimise potential interference. A consistent power source and high quality converters and voltage stabilisers guarantee the stable operation of the entire solution.

2.3. Selected Scenarios

Scenario 1. Development of an orthophotomap in field conditions.

- The concentration point acts as a field command centre.
- Site mapping action is carried out to plan and support subsequent activities.
- The action area is not known to the rescuers beforehand. It was necessary to produce an orthophotomap of the area in field conditions quickly and precisely, which was essential for further actions.
- Performing a ladder flight, creating a series of photos for processing in PIX4Dreact application—UAV AUTEL EVO II was used.
- Video transmission from the UAV to the concentration point and live video retransmission to the participants' mobile devices—use of GCS workstation system.
- Estimation of distances between sites and planning access from three sides by different ground teams—using the orthophotomap created beforehand.
- Activities of two-person ground teams when choosing different routes to reach—based on the developed orthophotomap.

Scenario 2. Coordination of ground teams.

- The concentration point acts as a field command centre.
- Coordination of ground teams—follow-up action after Scenario 1.
- Ongoing monitoring of the passage of two ground teams; the Commander may relay commands via radio from the command centre to the ground teams, e.g., to modify the route of the UAVs.
- Simulation of radio communication failure in one of the groups (radio failure), transmission of the command from the Commander to the ground group using the integrated speaker system UAV DJI Mavic 2 Enterprise Dual (Figure 4).

Figure 4. DJI Mavic 2 Enterprise Dual—UAV with mounted speaker. Photo: K. Orzepowski.

- Arrival of the first team to the operation site (selected object on the training grounds).

Scenario 3. Ground observation using a thermal imaging camera.

- The concentration point serves as a field command centre.
- A fly-around of the operations site, carrying out observations with the use of a thermal imaging camera, live viewing visible in the field command centre—use was made of a DJI Mavic 2 UAV.
- Indication by the commander by radio of particularly dangerous places to conduct operations.
- Assuming static positions by other available UAVs in designated places over the operation site, live transmission of images of implemented actions taken from different perspectives, continuous monitoring by each UAV of the assigned area (person or object).

3. Results

The solution adopted in Scenario 1 allowed the development of a precise, up-to-date map of operations on the entire area of over 25 ha in less than 20 min (this time should be shorter than the flight duration of the available drone). This turned out to be greatly helpful in conducting real rescue actions in an unknown area. It made it possible to determine the required forces, the route of access, key locations, etc., in a remote and therefore safe way. This method also allows assessing the differences in the image of a given area, e.g., following fires, floods, hurricanes or other disasters, which is also extremely important during search operations after building disasters caused by earthquakes.

Two battle groups, designated Alpha and Delta, were deployed to perform the task in Scenario 2. Rescuers from Section A and Section B were deployed in two locations characterised by a differentiated terrain. Both the A section and B section were tasked to reach the simulated incident site. The HRO directed the rescue teams using radio communications based on a photomap of the terrain and imagery transmitted from UAVs. A simulation was made of radio communication failure, and the integrated sound system in the DJI Mavic 2 Enterprise Dual UAV was used as a surrogate environment. However, there was no return radio channel, and the accuracy of the executed commands was assessed by the HRO based on the UAVs image. Therefore, this type of communication was unidirectional and only suitable for use in emergency situations.

Two variants of this particular phase of the exercise were carried out. In the first variant, the rescuers started executing the assigned actions without previous preparation and without locating equipment (GPS, compass). The coordination of their work turned out to be difficult (as only the direction of the needed turn had been specified) and the commands, directions and reference points proved to be quite ambiguous. Despite these difficulties, teams A and B were successfully led to the planned point.

In the second variant, the HRO and rescue teams were equipped with GPS and a compass. This allowed reaching the destination in a much shorter time, and both command and coordination proved to be more efficient.

The implementation of Scenario 3 that involved using a thermal imaging camera made it possible to further improve the effectiveness of search and rescue operations for missing persons carried out in the wilderness. In such situations, however, there are some limitations due to the fact that the camera is unable to scan solid materials. It is therefore only possible to look for the outlines of objects that actually emit heat. The thermogram only shows small, warmer spots that may indicate a person standing behind bushes and trees (Figure 5).

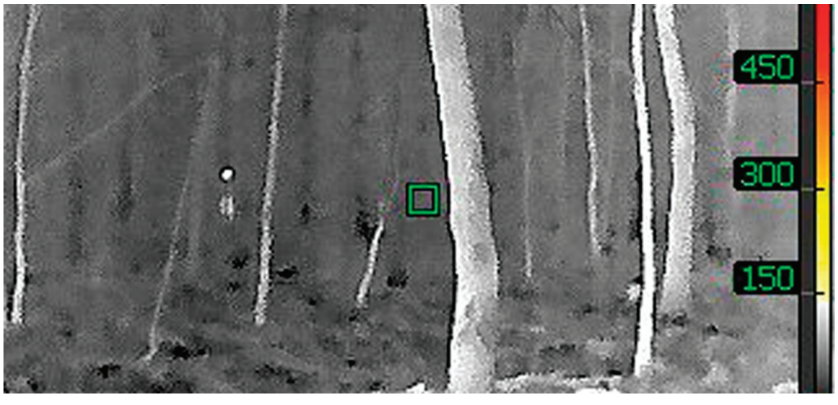


Figure 5. Small warmer spots visible in the thermogram (to the left of the green square), which may indicate a person standing behind bushes and a tree. Photo: W. Pruss.

The above scenarios show that the time necessary to reach a target (e.g., a lost subject) depends to a large extent on the supporting technologies used. This means that the applied solutions should be extended with new functionalities. Support in this regard can be provided by the SARUAV software, which provides a real possibility of the optimal use of drones in the search for missing people. The system enables this technology to be used quickly, efficiently and inexpensively, without the need of involving an additional group of people or other resources in the search [23].

The most important functionality of the SARUAV system is the automatic detection of people.

For example, it is possible to survey an area of ca. 3 ha with great accuracy, using a drone in approximately 10 min by taking about 150 aerial photographs with very high lateral and longitudinal coverage and high field resolution of the images (less than 2 cm/px) [23].

This set of images is automatically processed by the SARUAV system in 2–3 min, and false positives can be filtered out in further 3–5 min in a verification panel designed for this purpose.

By using two detectors with different methodological underpinnings, the system maximises the probability of human detection in an unattended mode and minimises the number of false negative readings [23].

The time needed for an automatic analysis of aerial photographs and evaluation of its results was under 10 min, so the full process of implementing the flight and detection described above came down to ca. 20 min, and two people were sufficient to operate it (i.e., the drone operator and the analyst operating the computer).

An analogical search of a 3-hectare area conducted by a team of rescuers, for example, with the use of the quick-three method (this method assumes that rescuers are divided into groups of three, and one of the rescuers in each group is equipped with a device with a GPS receiver), would take much longer, and the task would have to be carried out by a significantly higher number of people, which entails an increase in the cost of the action.

The system estimates the field coordinates of indicated people with a high accuracy and allows the generation of clear map reports that can be displayed on any stationary or mobile device. After the detection, the analyst transmits the map report to the rescuers, giving them a chance to quickly reach the location of the missing person.

Full field operability means that the detection process can take place entirely in the field, the calculations are performed on a mid-range laptop computer that does not need internet access and communication with high computing power servers.

What is more, the SARUAV solution is not related to any specific UAV platform, so its use in the work of rescuers does not represent a substantial financial barrier. On the contrary, using a SARUAV system can become an element that allows savings if it is a tool that complements standard search methods [23].

However, the newly introduced software does not have the capability of analysing infrared images and videos or to work at night.

The two SARUAV detection algorithms operate in parallel, only on nadir and RGB images. Tests conducted in the near infrared (NIR) gave less satisfactory detection results. It also seems to be inefficient to work on images taken in the far infrared–thermal imaging, as the resolution is insufficient from the viewpoint of algorithms.

SARUAV detection algorithms work on high-resolution RGB imagery, and consequently, their application for night-time images is somewhat limited. Even if night flights were possible from a legal perspective, the images acquired during such a flight are often of insufficient quality for detection to be possible.

Exceptions may be made for areas that are very well lit by artificial light, or for flights with an additional powerful light source mounted on the aerial unmanned vehicle. Such a reflector could be directed downward to illuminate the terrain being photographed, and flights should take place at an altitude that allows appropriate illumination while

contemporaneously maintaining all safety conditions. SARUAV already carried out the first positive tests of this solution.

The cost of the equipment used is as follows:

SARUAV is an innovative IT system to support the search for missing persons—an annual licence for one province costs EUR 1869, including VAT.

The AUTEL EVO II UAV drone is a cost ranging from EUR 1500 to 7000, including VAT.

DJI Mavic II Enterprise Dual UAV drone costs around EUR 3000, including VAT.

To be able to point out a practical application of the above-described software, ten full search flights were analysed, which involved between 14 and 145 nadir aerial images. A review was then carried out of a total of 668 images taken with the use of UAV-mounted cameras. These included missions carried out in both lowland and upland sites. The areas monitored varied in coverage: from simple detections in grasslands and wastelands, to pinpointing people in areas with varying coverage over a small area, to difficult detections in mixed forests outside the vegetative period.

Table 2 provides a tabular summary of the results (the human detector was not familiar with the 668 images previously entered).

Table 2. Effectiveness of the SARUAV system depending on relief and land cover.

Flight No.	Land Relief	Land Cover [1]	Number of Photos	Number of Persons		Effectiveness [%]
				Actual	Detected	
1	upland	a	37	3	3	100
2	upland	a	124	1	1	100
3	upland	a	98	1	1	100
4	upland	b	115	8	7	87.5
5	lowland	c	20	7	7	100
6	lowland	c	20	7	6	85.7
7	upland	d	14	3	3	100
8	upland	d	18	3	3	100
9	lowland	d	77	6	6	100
10	upland	e	145	31	30	96.8
		Σ	668	70	67	100

[1] a—temperate broadleaf and mixed forest (lack of leaves outside the vegetative season), b—temperate broadleaf and mixed forest, meadow, developed areas, football field, c—fallow, low vegetation, single trees without leaves, d—meadow, e—temperate broadleaf and mixed forest (lack of leaves outside the vegetative season), meadow, castle. Source: SARUAV [23].

It is recommended that in operational activities, photogrammetric flights with high lateral and longitudinal coverage (at least 60/80%) be executed with the use of the SARUAV system to serve as a basis of further analyses. This means that the person being searched for can be registered in multiple images. The reason why this solution is proposed is because it reduces the probability that a person could remain undetected if he/she is not visible from certain camera positions (e.g., at the border of a meadow and a forest).

The column 'effectiveness' contains a summary of the percentage of detected persons in relation to the actual number of persons out in the wilderness, where the detection of a person is considered to be effective if the SARUAV system detects automatically at least one image covering his/her location in the field (or potentially covering if the person is obscured in some images).

Ref. [24] specified cases of application of the developed methodology in operational conditions that resemble real ones. As part of the research work, several scenarios were worked out of searching for missing persons in mountain and lowland environments, taking into consideration sites with different characteristics (exposed, covered with vegetation or snow). At this point, it should be emphasised that the system was subjected to critical verification by its testing in simulated operational conditions similar to actual ones. The identified problems provided invaluable help in developing the final form of the system (for the production phase). The most valuable conclusion that arises from this study is that

the use of the system makes it possible for rescuers to find a missing person in an open area within just 1 hour. The conclusions formulated at the end of this publication provide a very specific summary of the design and practical functionality of the SARUAV system.

4. Discussion and Summary

4.1. Drone Module

Technological advancements make it possible to deploy UAVs during various types of search and rescue operations, as well as during natural disasters. As demonstrated by actions executed by the police, firefighters and voluntary organisations, the use of UAVs in search operations has contributed to reducing the time needed to find missing persons [8]. The use of UAVs in search and rescue operations can help achieve a significant reduction in the number of victims of various types of accidents, and through the use of avalanche victim detectors, they can also provide support during search operations after avalanches have descended [25].

Adequate technological equipment of drone modules, as well as good coordination in operational activities and developed cooperation of various services (state and voluntary), will contribute to increasing the effectiveness of search operations. It should also be pointed out that cooperation between different formations should include an increase in the number of specialist training courses to deepen theoretical and practical knowledge (along with an analysis of actions taken), as well as joint search manoeuvres in various areas and terrains. The practical use of knowledge and experience could significantly affect the system of searching for missing persons and increase the effectiveness of operations.

4.2. UAV Ground Control Station

Members of the VFD designed and executed a GCS workstation dedicated to actions involving the use of rotary wing UAVs. The tested system allows streamlining the communication between HRO, UAV section and search participants, and thanks to the possibility of radio signal amplification and the transmission of necessary data, it is also appropriate for searching people in mountainous areas. Before now, when this solution was not available yet, problems arose in obtaining radio communication due to the demanding and difficult terrain owing to the presence of rocks and ridges, as well as deposits of various deposits. Since radio communication was supported by relay stations, a significant improvement in signal quality has been achieved.

4.3. SURUAV System

The undertaken operational activities could be supported by the use of SARUAV software, the effectiveness of which was proven during a rescue operation conducted in a mountainous terrain with large forest complexes in the Low Beskid (June 2021). While searching for a lost male subject aged 65 years (the person in need of assistance was ailing and was not carrying a phone) first of all, traditional search and rescue methods were used. After the lapse of more than 20 h, the decision was made to provide support by using a specialised drone and the SARUAV system. As a result of this solution, after 4 additional hours, the rescue team was able to find the lost subject.

Figure 6 shows a photo in which the SARUAV software identified the missing person (this would have been very difficult when visually reviewing multiple photos, and the further difficulties were such factors as tall, two-metre-high grass and the passage of a storm, after which the dogs lost the trail).

The software automatically detects people in aerial photos by indicating places where a person could potentially be—it carries out aerial human detection. The algorithm sets out the search site, where the UAV takes hundreds of photos. Thanks to the software, several hundred photos are analysed in only a few minutes, which would be absolutely impossible for the human eye to analyse, particularly given such quantities, maintaining this kind of speed and accuracy—it would simply not be able to notice the missing person. When a person is detected by the SARUAV system, a location pin is placed on the map,

and the UAV operator then knows exactly where to direct the rescuers. Such a solution is innovative on a global scale because this technology allows a significant reduction in time needed to find a missing person during search and rescue operations, which clearly has an advantageous impact on the further health and life of the victim. In its analysis, the SARUAV system can take into account such inconspicuous elements as a hat lying on the ground (which can be an excellent indication in the case of a real search) or an image of a dog running alone (its owner may be present close by). Works are underway to adapt the system for search operations on water.

Figure 6. Image processed by the SARUAV application indicating the missing person. Source: [15].

Exercises conducted at the MSFS Field Training and Rescue Innovation Base have shown that access to specialised equipment and modern solutions is a guarantee of a shorter time that would be required to achieve the intended objective.

The use of UAVs is proven to facilitate efficient mapping of the search area, followed by coordinated operations using a GPS and compass module, and in an emergency situation by relaying messages and commands from the HRO to teams and sections via a speaker integrated into the drone.

One of the most common technologies in AI, the so-called neural networks, can be used to support this type of activity. The algorithm has several inputs via which information is received, and an "inference module" that generates an output signal on the basis of input information and their weights. The processed information is then directed to the output and passed on. In this way, an image seen by a camera, for example, can be compared with a previously created pattern. These patterns can be further developed, and machines can be taught new patterns through the process of machine learning.

Such a solution that involves the SARUAV system is already available for implementation by the police, fire and border guards, as well as mountain and water rescue units. Its functioning requires basic information, such as the last known location of the lost person.

The availability of the SARUAV module during tests and real search actions would make it possible to extend the activities performed by an automatic analysis of aerial photographs and searching for a missing person on them. The system has high detection efficiency and successfully distinguishes human silhouettes from other elements of the environment, such as animals, vegetation or rocks [26].

5. Recommendations

1. The application of the above outlines the organisational scheme of the UAV sections along with dedicated solutions to support equipment communication, and team liaison allows UAVs to be launched within 3 min of arrival in the operational area. The first image in the GCS workstation can be obtained already in the 6th minute, and

the image in the mobile devices of the participants of the operation in the 8th minute after the commencement of the operation.

2. The use of SARUAV software allows replacing human labour associated with time-consuming analyses of images taken from the air with automatic image recognition. In combination with the possibility of covering the area of operations by a larger number of UAVs, this provides an extremely efficient system for the search for missing persons.
3. Search and rescue activities should be oriented at assuring that the ability to carry out rescue operations on a basic level could become universal for all NFRS entities [27].
4. The nationwide cooperation of the police with search and rescue groups and other entities should be further intensified and developed, involving, inter alia, the launching of joint undertakings (training, sham search operations) the purpose of which comprises exchanging experience, mutual requirements and consolidation of knowledge in the search for missing persons [28].
5. In the area of scientific development in the context of search and rescue operations, it is necessary to conduct research to reduce the time it takes to find a missing person. In this respect, there is a need to improve the technical parameters of components of search coordination systems:
    - UAVs—increase in flight duration;
    - Data transmission devices—increased speed of information transfer (including high quality images);
    - GCS workstation—extending capabilities by subsequent modules (currently: ensuring stable communication in the official SFS radio bands, creation of own Wi-Fi network in the MESH system, preview on the monitor and transmission of the image from UAVs to the participants of the action in possession of any mobile device, acting as a web server with important data concerning the conducted action, and strengthening of mobile telephony signal).
6. The conducted research related to the organisation of search and rescue operations confirmed the following:
    (a) The need to develop cooperation between rescue parties and improve coordination of actions undertaken on a previously unknown terrain;
    (b) The benefits of setting up a command post in the vicinity of the incident and starting to support the HRO by viewing the situation from above;
    (c) In the case of a lack of communication, the solution to the problem turned out to be GCS workstation being on the equipment of the UAVs VFD section, which uses antennas that are much more robust.

6. Conclusions

Unmanned aerial vehicles (UAVs), commonly known as drones, are becoming increasingly common and keep gaining new functionalities. Currently, they are one of the most innovative elements in the activities of rescue services, including the State Fire Service (SFS).

On 8 July 2021, on the premises of the MSFS Field Training and Rescue Innovation Base in Nowy Dwór Mazowiecki, tests were conducted to verify the characteristics, specifics and prospects of using drones in rescue operations. During the exercises, the hardware and software used by VFD were tested as well as the possibilities of using UAVs and GCS workstation.

The scheduled exercise scenarios included terrain mapping—producing a map under field conditions, coordination of ground groups, evaluation of the terrain (object fire) with the use of a thermal imaging camera, and precise observation of a designated object (region and person).

The research participants were involved in various simulated actions. They coordinated the action, using mounted speakers and the UAVs preview. The aim of joint exercises was to achieve an improvement of techniques and skills of coordination of actions and cooperation.

Moreover, the article also presents aspects related to the use of unmanned aerial systems supported by image recognition software based on artificial intelligence algorithms. The SARUAV system of searching for missing persons, dedicated in particular to open spaces, the main part of which is the algorithm of detecting silhouettes of people from images acquired from UAVs, is being used in Poland by 4 VFD units and 2 MVRS groups. The SARUAV solution was developed and tested in cooperation with the MVRS Jurassic Group. Field tests of the system have pointed to a high performance of the algorithm, and the results were published in [24,29,30]. The system received very positive feedback from, among others, MVRS, but also from the FlyTech UAV company from Krakow, which produces professional BIRDIE UAVs that can be used in the SAR rescue system.

The system was also proven effective in a real-life situation, where time was of the essence.

Author Contributions: Conceptualisation, N.T. and W.W.; methodology, W.W.; validation, W.W. and N.T.; formal analysis, N.T.; investigation, N.T.; resources, W.W.; writing—original draft preparation, N.T.; writing—review and editing, W.W.; visualisation, N.T.; supervision, W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. List of equipment used in the exercises [1].

Type of Equipment	Additional Equipping	Operating Limitations
UAV AUTEL EVO II	4 battery packs Live Deck transmission systemControl unit Battery charging station	Weather without precipitation Temperature from −10 to +40 °C Flight duration 25–38 min on one battery (depending on weather) Wind speed < 40 km/h Battery charging time approx. 80 min
UAV DJI Mavic II Enterprise Dual	3 battery packs Spotlight Loudspeaker Thermal imaging camera Battery charging station	Weather without precipitation Temperature from −10 to +40 °C Flight duration 18–25 min on a single battery (depending on weather conditions) 18 min—using the spotlight 20 min—using the loudspeaker Wind speed < 40 km/h Battery charging time approx. 80 min
Ground Control Station Workstation	Motorola transceiver 4600e (mobile) + antenna with high energy gain Wi-Fi network LTE network 19” monitor for viewing live from UAV AUTEL Redistribution of images to mobile devices Power distributor for 4 socketsMast (tripod) for installation of antennas—at a distance of up to 10 m from GCS workstation	A 230 V alternating current source is required Power (depending on the number of receivers) approx. 1000–1500 W Weather without precipitation or use in a sheltered location Temperature from −18 to +45 °C

Table A1. *Cont.*

Type of Equipment	Additional Equipping	Operating Limitations
ACER Laptop	SD card reader Wireless mouse 230 V power supply (included in estimated power of GCS workstation) Ability to view live image from UAV	Weather without precipitation or use in a sheltered location Temperature from −10 to +40 °C
Samsung Tablet A10—for the operation UAV	USB-C power adapter	Weather without precipitation
5 pcs. Motorola GP360 Radiotelephone for communication within the group of UAV operators	Chargers	Temperature −18 to +45 °C

[1] Source: [20].

References

1. Duda, M. Disappearances of People in Poland—Report on Criminological Research. Online Seminar on 24 March 2021. Department of Criminal Law and Criminology, University of Bialystok, Białystok, Poland. Available online: https://www.prawo.uwb.edu.pl/nowosci/aktualnosci/zaginiecia-osob-w-polsce-raport-z-badan-kryminologicznych/7a4c2b25 (accessed on 1 April 2021).
2. Goodrich, M.A.; Cooper, J.L.; Adams, J.A.; Humphrey, C.; Zeeman, R.; Buss, B.G. Using a Mini-UAV to Support Wilderness Search and Rescue: Practices for Human-Robot Teaming. In Proceedings of the 2007 IEEE International Workshop on Safety, Security and Rescue Robotics, Rome, Italy, 27–29 September 2007; pp. 1–6. [CrossRef]
3. Drosio, W.; Podlasiński, R.; Pastuszka, Ł. An Analysis of the Equipment Used by Search and Rescue Groups at Home and Abroad. *Saf. Fire Tech.* **2017**, *46*, 2.
4. Mencel, P. The reality of the functioning of search and rescue groups in Poland. *Nowa Kodyfikacja Prawa Karnego* **2020**, *56*, 177–189. [CrossRef]
5. Found by a Police Pilot Drone. Available online: https://policja.pl/pol/aktualnosci/205965,Odnaleziona-przez-pilota-policyjnego-drona.html (accessed on 28 July 2021).
6. Police Used a Drone to Find a Rape Victim. Available online: https://fotoblogia.pl/12951,policja-wykorzystala-drona-zeby-odnalezc-ofiare-gwaltu (accessed on 28 July 2021).
7. SARUAV—Polish Software That Saves Human Lives. Available online: https://aeromind.pl/SARUAV-POLSKIE-OPROGRAMOWANIE-KTORE-RATUJE-LUDZKIE-ZYCIA-blog-pol-1626185881.html (accessed on 3 November 2021).
8. Jurecka, M.; Niedzielski, T. Searching for Lost Persons in the Wilderness. In *Review of Applied Methods. Treatise*; Institute of Geography and Regional Development of the University of Wroclaw: Wrocław, Poland, 2020.
9. A Lost Person Is Found!—Spectacular Success of Drones and the SARUAV System. Available online: https://uni.wroc.pl/en/a-lost-person-is-found-spectacular-success-of-drones-and-the-saruav-system/ (accessed on 19 July 2021).
10. Weldon, W.T.; Hupy, J. Investigating Methods for Integrating Unmanned Aerial Systems in Search and Rescue Operations. *Drones* **2020**, *4*, 38. [CrossRef]
11. Półka, M.; Ptak, S.; Kuziora, Ł. The Use of UAV's for Search and Rescue Operations. *Procedia Eng.* **2017**, *192*, 748–752. [CrossRef]
12. Silvagni, M.; Tonoli, A.; Zenerino, E.; Chiaberge, M. Multipurpose UAV for search and rescue operations in mountain avalanche events. *Geomat. Nat. Hazards Risk* **2017**, *8*, 18–33. [CrossRef]
13. Avola, D.; Pannone, D. MAGI: Multistream Aerial Segmentation of Ground Images with Small-Scale Drones. *Drones* **2021**, *5*, 111. [CrossRef]
14. Kim, J.; Cho, J. RGDiNet: Efficient Onboard Object Detection with Faster R-CNN for Air-to-Ground Surveillance. *Sensors* **2021**, *21*, 1677. [CrossRef] [PubMed]
15. Niedzielski, T.; Jurecka, M.; Mizinski, B.; Pawul, W.; Motyl, T. First Successful Rescue of a Lost Person Using the Human Detection System: A Case Study from Beskid Niski (SE Poland). *Remote Sens.* **2021**, *13*, 4903. [CrossRef]
16. Niedzielski, T.; Jurecka, M.; Stec, M.; Wieczorek, M.; Miziński, B. The nested k-means method: A new approach for detecting lost persons in aerial images acquired by unmanned aerial vehicles. *J. Field Robot.* **2017**, *34*, 1395–1406. [CrossRef]
17. Fonio, C.; Widera, A. (Eds.) *Trial Guidance Methodology Handboo*; DRIVER+ (Driving Innovation in Crisis Management for European Resilience): Brussels, Belgium, 2020.
18. JARUS (Joint Authorities for Rulemaking on Unmanned Systems). SORA Methodology (Specific Operations Risk Assessment). Available online: http://jarus-rpas.org/content/jar-doc-06-sora-package (accessed on 3 November 2021).

19. European Union Aviation Safety Agency (EASA). Easy Access Rules for Unmanned Aircraft Systems (Regulations (EU) 2019/947 and (EU) 2019/945). Available online: https://www.easa.europa.eu/document-library/easy-access-rules/easy-access-rules-unmanned-aircraft-systems-regulation-eu (accessed on 3 November 2021).
20. Piwowoński, P.; Robak, M.; Kuflik, A.; Ziemba, J.; Górecki, W.; Fellner, A.; Fellner, R. Proposed plan of exercises and demonstration on 07/09/2021. *Niegoszowice* **2021**. Unpublished Document.
21. Niedzielski, T.; Miziński, B.; Jurecka, M. Application of the SARUAV System During the Search Operation in Nowogrodziec. In Proceedings of the DroneTech World Meeting, Toruń, Poland, 6–7 November 2020.
22. How Firefighters Used Drones to Fight the Fire in Biebrza. Available online: https://terazpolska.pl/pl/a/Jak-strazacy-wykorzystali-drony-do-walki-z-pozarem-na-Biebrzy (accessed on 28 July 2021).
23. SARUAV Facebook Page. Available online: https://m.facebook.com/SARUAVPL/ (accessed on 3 November 2021).
24. Niedzielski, T.; Jurecka, M.; Miziński, B.; Remisz, J.; Ślopek, J.; Spallek, W.; Świerczyńska-Chlaściak, M. A real-time field experiment on search and rescue operations assisted by unmanned aerial vehicles. *J. Field Robot.* **2018**, *35*, 6. [CrossRef]
25. Merkisz, J.; Nykaza, A. Prospects for the development and use of unmanned aerial vehicles in rescue services. *Autobusy Tech. Eksploat. Syst. Transp.* **2016**, *6*, 291–296.
26. The SARUAV System Is Now Ready. Drones Will Help You Look for Missing People. Available online: https://uni.wroc.pl/system-saruav-juz-gotowy-drony-pomoga-szukac-zaginionych-ludzi/ (accessed on 3 November 2021).
27. Wentkowska, A. Search for Missing Persons. System and Methods of Operation in the Public Security Services' Procedures. Office of the Commissioner for Human Rights, Warsaw. 2016. Available online: https://bip.brpo.gov.pl/sites/default/files/Poszukiwania%20os%C3%B3b%20zaginionych%20-%20Aleksandra%20Wentkowska%20-monografia%202016.pdf (accessed on 3 November 2021).
28. Department of Order and Internal Security. Search for Missing Persons. Information on the inspection results. Supreme Chamber of Control, Warsaw. 2015. Available online: https://www.nik.gov.pl/plik/id,8333,vp,10395.pdf (accessed on 3 November 2021).
29. Jurecka, M.; Niedzielski, T. A procedure for delineating a search region in the UAV-based SAR activities. *Geomat. Nat. Hazards Risk* **2017**, *8*, 1. [CrossRef]
30. Niedzielski, T.; Jurecka, M. Can Clouds Improve the Performance of Automated Human Detection in Aerial Images? *Pure Appl. Geophys.* **2018**, *175*, 3343–3355. [CrossRef]

applied sciences

MDPI

Article

A Spectrum Correction Method Based on Optimizing Turbulence Intensity

Wenwu Yi [1,2], Ziqi Lu [1,2], Junbo Hao [1,2], Xinge Zhang [1,2], Yan Chen [1,2,*] and Zhihong Huang [1,2]

[1] Key Laboratory of Intelligent Manufacturing Technology, Ministry of Education, Shantou University, Shantou 515063, China; 19wwyi@stu.edu.cn (W.Y.); 20zqlu@stu.edu.cn (Z.L.); 19jbhao@stu.edu.cn (J.H.); 19xgzhang@stu.edu.cn (X.Z.); zhhuang@stu.edu.cn (Z.H.)

[2] Institute of Energy Science, Shantou University, Shantou 515063, China

* Correspondence: ychen@stu.edu.cn; Tel.: +86-138-2968-2901

Abstract: Based on the classical spectral representation method of simulating turbulent wind speed fluctuation, a harmonic superposition algorithm was introduced in detail to calculate the homogeneous turbulence wind field simulation in space. From the view of the validity of the numerical simulation results in MATLAB and the simulation efficiency, this paper discussed the reason for the bias existing between three types of turbulence intensity involved in the whole simulation process: simulated turbulence intensity, setting reference turbulence intensity, and theoretical turbulence intensity. Therefore, a novel spectral correction method of a standard deviation compensation coefficient was proposed. The simulation verification of the correction method was carried out based on the Kaimal spectrum recommended by IEC61400-1 by simulating the uniform turbulent wind field in one-dimensional space at the height of the hub of a 15 MW wind turbine and in two-dimensional space in the rotor swept area. The results showed that the spectral correction method proposed in this paper can effectively optimize the turbulence intensity of the simulated wind field, generate more effective simulation points, and significantly improve the simulation efficiency.

Keywords: uniform wind field simulation; turbulence intensity; deviation of standard deviation; spectral representation

Citation: Yi, W.; Lu, Z.; Hao, J.; Zhang, X.; Chen, Y.; Huang, Z. A Spectrum Correction Method Based on Optimizing Turbulence Intensity. *Appl. Sci.* **2022**, *12*, 66. https://doi.org/10.3390/app12010066

Academic Editors: Luis Hernández-Callejo, Sara Gallardo Saavedra and Sergio Nesmachnow

Received: 20 October 2021
Accepted: 6 December 2021
Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The random turbulent wind is one of the critical factors causing the fatigue load of wind turbine blades. Aerodynamic analysis of wind turbines under the influence of turbulence is a vital basis for designing the ultimate load and fatigue load [1]. Meanwhile, with the trend of large-scale and flexible blades, the coupling of nonlinear aerodynamic and structural problems with the environment becomes more complex. Therefore, it is of great significance to establish a pulsating wind field model with high adaptability to multiple design conditions and complicated incoming flow environments in line with engineering applications.

In engineering practice, considering the speed of the solution and data processing is important. While using the computational fluid dynamics (CFD) method to solve Navier-Stokes equations results in a high accuracy, this method is complex and heavily dependent on computer performance. Accordingly, the current research on turbulent wind field simulation tends to be based on classical stochastic process theory. The PSD (power spectral density) function is employed to simulate the time history of pulsating wind speed. Regularly, Harris spectrum, Von Karman spectrum, Simiu spectrum, and Kaimal spectrum are applied to power spectral density models [2–6]. Shinozuka proposed the harmonic synthesis method to settle the matter of the stationary Gaussian random process and non-stationary random process of wind speed time history simulation, introducing the double index frequency combined with FFT technology to achieve the ergodic properties of each state of the simulation curve [7,8]. Lagrange interpolation, Hermite interpolation,

PoD-based interpolation, and other interpolation methods [9–11] have significantly reduced the number of analog signal decompositions and have further optimized the operation efficiency. In essence, the power spectrum simulation method is a Monte Carlo numerical statistical method. To date, most studies have mainly concentrated on how to optimize the algorithm to improve the computational efficiency in order to solve more dense simulation points, or to verify the fitting degree between the power spectrum simulating fluctuating wind speed and the original spectrum [12–14]. However, few scholars have paid attention to the turbulence intensity of the simulated fluctuating wind field and the validity of the time history of the simulating point fluctuating wind speed.

Based on applying the spectral representative method to simulate wind speed, the significant standard deviation's bias, generated due to valuing a truncated simulated frequency interval and the simulation processing itself, respectively, is proposed to shorten the time required to complete the method by creatively introducing a compensation coefficient. The corrective method modifies the original power spectral density function to produce more effective simulated points in the uniform simulated wind field that can generate the simulated turbulence intensity, meeting the set requirements. Moreover, this paper also discussed the correction method's further development by expanding its usage to the two-dimensional level. The modified method would be verified by simulating the uniform fluctuating wind field of a 15 MW referenced wind turbine, which may lay the foundation for simulating a turbulent field that is applied to a larger scale wind turbine.

2. Materials and Methods

2.1. Turbulent Wind Field Model

In a turbulent wind field, the wind speed can be decomposed into average wind speed and fluctuating wind speed. Thus, the simulated three-dimensional wind speed at any simulation point is the linear superposition of the average wind speed and fluctuating wind speed in longitudinal, transverse, and vertical directions [2], which can be calculated by:

$$\widetilde{V}(z,t) = \overline{U}(z) + v(z,t) \tag{1}$$

where, $\overline{U}(z)$ denotes the average wind speed, $v(z,t)$ denotes the fluctuating wind speed of a simulated point, z denotes simulated height, and t denotes time.

When setting the average wind speed, the influence of wind shear must be considered for large wind turbines whose installation height is usually higher than 100 m. Commonly used wind shear models include exponential model and logarithmic model, and the stable modified logarithmic model was used in this paper [8], which can be calculated by:

$$u(z) = \frac{u^*}{\kappa}\left(ln\frac{z}{z_0} - \psi\right) \tag{2}$$

where, u^* denotes friction velocity, it can be represented as $u^* = 0.045V_{ref} - 0.012$, V_{ref}, which means the reference mean wind speed; κ denotes Von Karman constant, under neutral atmospheric conditions, $\kappa = 0.4$; z_0 denotes terrain roughness parameter, which can be calculated as $z_0 = A_c\frac{(u^*)^2}{g}$, where $A_c = 0.034$ when the terrain type is an offshore area, g denotes the acceleration of gravity; and ψ denotes the stability function when in the neutral condition, its value is 0.

Pulsating wind speed is regarded as a stationary Gaussian random process. According to the stochastic process theory, the power spectral density function combines with the coherence function, which is set to describe the wind speed correlation between two different points that may generate different wind speeds during a period in a stochastic wind field while a smaller separation distance of any two points will bring a greater correlation, is used to simulate the pulsating wind speed time history [15].

2.2. Basic Algorithm for Simulating Fluctuating Wind Speed Time History

Considering the simulation process at a certain point $v_j(t), (j = 1, 2, 3, \ldots, m)$ in the turbulent wind field in space, the basic algorithm for simulating and solving the one-dimensional M-variable turbulent wind field $\{v_j(t)\}$ is described, and its algorithm flow is shown as follows:

(1) According to the sampling theorem [16], the simulation parameters are set in the frequency domain, including:
   - Sampling frequency f_s, in order to ensure that the analog signal can be reconstructed accurately without aliasing, f_s is required to be greater than two times the value of F_{max}, F_{max} indicates the upper limit of the analog frequency range, which denotes also the cut-off frequency in this paper;
   - The initial frequency F_{min} and cutoff frequency F_{max}, which are selected by considering the dimensionless power spectral density function image and the influence of truncation error in the simulation frequency range on the simulation variance [17], the frequency step size is denoted by $df = \frac{F_{max} - F_{min}}{N}$, where N represents the frequency sampling number.

(2) Set simulation parameters in the time domain, including:
   - Sampling interval dt, $dt = 1/f_s$;
   - Analog time points M, generally $M = 2N$.

(3) Set the basic parameters of the simulated wind field, including simulation points m, simulation height z, and spacing between simulation points dr, etc.

$$dr = \left(\frac{z}{m-1}\right) \cdot \sqrt{(x_k - x_j)^2 + (y_k - y_j)^2} \tag{3}$$

where, x_k, x_j, y_k, y_j represent the abscissa and ordinate values of simulation point k and simulation point j, respectively.

(4) Calculate the cross-spectral density matrix of each frequency sampling point $\{S(f_n)\}$, $(n = 1, 2, 3, \cdots, N)$:

$$S(f_n) = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1m} \\ S_{21} & S_{22} & \cdots & S_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mm} \end{bmatrix} \tag{4}$$

where, $S(f)$ denotes the power spectral density function, S_{jj} denotes the autocorrelation spectrum of corresponding points, and $S_{kj}(k \neq j)$ denotes the cross-correlation spectrum between two simulated points, $(k = 1, 2, 3, \cdots, j)$, which can be calculated as:

$$S_{kj}(f_n) = \sqrt{S_k(f_n) S_j(f_n)} \cdot coh(k, j; f) \tag{5}$$

Since the cross-correlation function of the fluctuating wind speed in one-dimensional space does not involve the change of height, the cross-correlation spectrum between the simulated point j and other simulated points is consistent, and the cross-correlation spectrum of the fluctuating wind speed in one-dimensional space can be obtained by the following formula:

$$S_{kj}(f_n) = S_j(f_n) \cdot coh(k, j; f) \tag{6}$$

(5) Combined with the double-indexed frequency method, the Cholesky decomposition method is performed on the cross-spectral density matrix of sampling points at each frequency, and the decomposed lower triangular matrix is obtained, that is:

$$S(f_{kn}) = H(f_{kn}) H^T(f_{kn}) \tag{7}$$

$$f_{kn} = (n-1)df + (k/m)df \tag{8}$$

where, f_{kn} denotes the double-indexed frequency; $H(f_{kn})$ denotes the decomposed lower triangular matrix.

(6) Random phase obedience ϕ_{kn} is introduced and subjected to independent random distribution among intervals $(0, 2\pi)$, carrying out the Fast Fourier transformation, namely:

$$G_{kn}(t) = FFT(B_{kn}) \tag{9}$$

$$B_{kn}(ndf) = \begin{cases} H(ndf + (k/m)df)e^{i\phi_{kn}}, (0 \le n < N) \\ 0, (N \le n < M) \end{cases} \tag{10}$$

(7) Employing harmonic superposition method to generate the wind speed time history of the simulated point j samples, and $v_j(t)$ can be calculated by Equation (11):

$$v_j(p \cdot dt) = \sqrt{2df}Re(\sum_{k=1}^{j} G_{jk}exp(i\frac{k}{m} \cdot 2\pi df \cdot p \cdot dt) \tag{11}$$

where, $p = 1, 2, 3, \cdots, (M \cdot m - 1)$.

2.3. Power Spectrum Correction Method

In the practical application of simulating a uniform pulsating wind field, since the dimensionless power density spectrum of one-dimensional space accords with Gaussian distribution (Figure 1 shows the longitudinal dimensionless wind speed power density spectrum curve of several classical spectra), the relationship between the simulating pulsating wind speed variances σ^2 and the power spectrum density function $S(f)$ is defined as below [18,19]:

$$\sigma^2 = \int_0^\infty S(f)df \tag{12}$$

Therefore, when the analog frequency range $[F_{\max,}\ F_{min}]$ is a segment interval and is cut out from the spectral density function originally defined interval $(0, \infty)$, theoretically, the variance of the simulated fluctuating wind speed ${\sigma_{theory}}^2$ is defined as:

$${\sigma_{theory}}^2 = \int_{F_{min}}^{F_{max}} S(f)df \tag{13}$$

The defined standard deviation of the fluctuating wind speed σ can be determined by an empirical formula in the simulation algorithm (e.g., Equation (20) in Section 3.1), according to the calculation formula of turbulence intensity $I_T = \sigma/v(t)$, choosing the segment interval $[F_{\max,}\ F_{min}]$ will cause the simulated results to produce truncated standard deviation, which will inevitably affect the numerical error of the simulated turbulence intensity due to the existence of the deviation of σ and σ_{theory}. The deviation between the two can be calculated as:

$$\varepsilon = \left| \frac{\sigma_{theory} - \sigma}{\sigma} \right| \tag{14}$$

The standard deviation's deviation ε affects the simulated turbulence intensity to a certain extent, and the smaller the deviation is, the smaller the influence is.

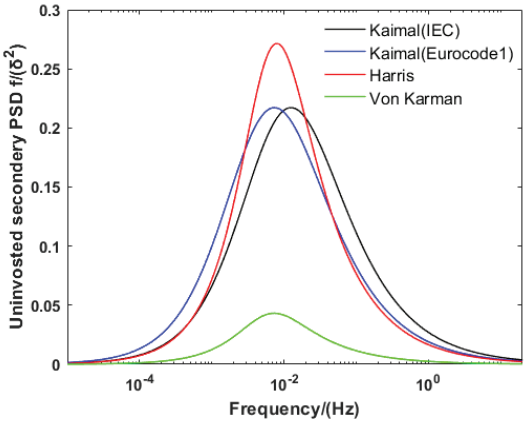


Figure 1. The longitudinal dimensionless velocity of the wind power density spectrum curve.

It can be seen from Equations (12)–(14) that when increasing the proportion of the interval length in the simulated frequency interval, the theoretically calculated standard deviation's deviation will decrease. However, it can be seen from the simulation algorithm introduced in Section 2.2 that the frequency step has a significant impact on the simulation results. On the other hand, according to the formula $df = \frac{F_{max}-F_{min}}{N}$, after determining the basic range of the simulation frequency interval, a slight change caused by the minor increase in cutoff frequency relative to the magnitude of N is almost negligible. When the ratio of the turbulence integral length to the average wind speed $X_k = L_k/\overline{U}$ is constant, the upper limit of the interval expands to the right indefinitely, and the deviation will not significantly decrease [10].

This is also consistent with the simulation conclusion in this paper: under the same simulation conditions, when the initial frequency changes by one step unit, the impact on the standard deviation's deviation ε is significantly greater than that when the cutoff frequency changes by one step unit (see Section 3.2 for details). In addition, the simulation results also show that the theoretical truncation bias ε caused by the truncation of the simulation frequency range will affect the numerical value of the simulated object in the simulation process, which leads to the simulation bias $\hat{\varepsilon}$. Simulation deviation is not only related to truncation deviation, but also to the setting of other simulation parameters and the application of interpolation methods. The simulated deviation is a manifestation of the error generated by simulating the turbulence intensity, which can be expressed by the following calculation formula:

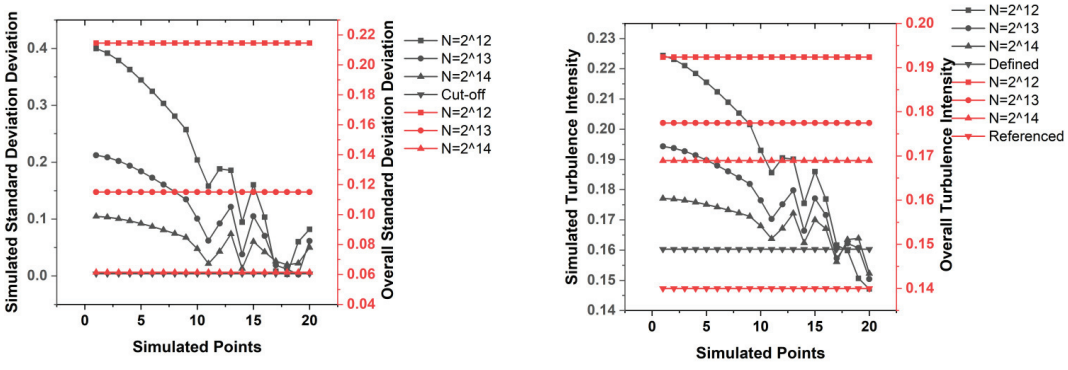
$$\hat{\varepsilon} = \left| \frac{\hat{\sigma} - \sigma}{\sigma} \right| \tag{15}$$

where, $\hat{\sigma}$ denotes the standard deviation of actual simulated wind speed.

Considering df $= \frac{F_{max}-F_{min}}{N}$, if the magnitude of the frequency sample N is increased by an order (usually N is an exponential form with base 2), the simulation bias can indeed be reduced, but at the same time the computational memory will be doubled and the simulation speed will be seriously slowed down. Take the analog frequency range $[F_{\max,}\, F_{min}] = [10^{-5}, 12]$, and the other simulation conditions are consistent with the case of one-dimensional simulation in Section 3. The simulation experiment results are shown in Table 1.

Table 1 and Figure 2 show that increasing the scale N causes the whole standard deviation bias of the turbulent wind field simulation to almost linearly reduce, or even when increased to $N = 2^{14}$, the overall standard deviation's bias of the turbulent wind field has reached the truncation error, but with a significant disadvantage of longer running time, and a great computing cost if more simulated points are set in the turbulent wind

field space. On the other hand, this result is not very ideal. When increasing the magnitude of N, the simulated turbulence intensity of more simulation points falls outside the range of reference turbulence intensity and the defined turbulence intensity and needs to be removed, resulting in an unnecessary waste of computing resources.



(**a**) (**b**)

Figure 2. (**a**) The deviation of standard deviation at a different frequency sampling magnitude; (**b**) Simulated turbulence intensity at a different frequency sampling magnitude.

Table 1. The relationship between the frequency sampling points scale and simulated deviation.

Test No.	Test 1	Test 2	Test 3
magnitude of N	$N = 2^{12}$	$N = 2^{13}$	$N = 2^{14}$
$\hat{\varepsilon}_{min}$	0.0026	0.0026	0.0131
$\hat{\varepsilon}_{ave}$	0.2146	0.1151	0.0615
$\hat{\varepsilon}_{max}$	0.3998	0.2122	0.1049
CPU time	40s	84s	212s

To sum up, the correction method should realize the minimum truncation deviation ε_{min} in theory and the minimum simulation deviation $\hat{\varepsilon}_{min}$ generated in the actual simulation process, without adding more frequency sampling points to reduce the simulation efficiency. In addition, the method should generate as many effective simulation points as possible (effective simulation points are defined as the simulation points whose simulated turbulence intensity $\hat{I}_T$ falls within the range of reference turbulence intensity I_{ref} and defined turbulence intensity I_T).

Consider the error values of truncation bias ε and simulation bias $\hat{\varepsilon}$:

$$\Delta\varepsilon = \hat{\varepsilon} - \varepsilon \tag{16}$$

Introduce compensation coefficient β:

$$\beta = (1 + \Delta\varepsilon)^2 \tag{17}$$

If the variance σ^2 in the PSD function is substituted into the compensation coefficient β, then the original factor will be corrected $\frac{\sigma^2}{\beta}$.

3. Results and Discussion

Given the IEA 15 MW (NREL) wind turbine [20], its basic parameters are shown in Table 2, and frequency domain simulation parameter settings as shown in Table 3.

Table 2. 15 WM wind turbine parameters.

capacity	15 MW
wind turbine diameter D	240 m
hub height Z	150 m
the reference mean wind speed V_{ref}	10.59 m/s
reference turbulence intensity I_{ref}	0.14

Table 3. Frequency domain simulation parameters.

	One-Dimensional Space	Two-Dimensional Space
number of simulated points m	40	49
simulated point spacing dr	$6(k-j)$	Calculate according to the Formula (3) in Section 2.2
frequency sampling number N	2^{12}	2^{12}
analog time points M	2^{13}	2^{13}

3.1. Power Density Spectrum and Coherence Function Selection

Kaimal spectrum is widely used to describe the spectral density during wind speed [21], and its simulation results are more conservative when applied to the fatigue load design of wind turbines. The function expression of Kaimal spectrum in IEC61400-1 4th Edition [18] is as below, $X_k = L_k/\overline{U}$ represents the ratio of the turbulence integral length to average wind speed ($k = 1, 2, 3$ represents the longitudinal, transverse and vertical component spectra, respectively):

$$S(f) = {\sigma_k}^2 \frac{4X_k}{(1+6fX_k)^{5/3}} \tag{18}$$

IEC61400-1 uses an exponential form of coherence function, which is expressed as follows:

$$Coh(r,f) = exp[-12((\frac{fr}{u})^2 + \left(0.12\frac{r}{L_k})^2\right)^{0.5}] \tag{19}$$

In Equation (18), according to IEC61400-1, the parameters of the longitudinal component are calculated as follows:

(1) Wind speed definition standard deviation σ_1:

$$\sigma_1 = I_{ref}\left(0.75\overline{U} + b\right);\ b = 5.6m/s \tag{20}$$

where, $I_{ref} = 0.14$, determined by 15 MW wind turbine design grade Class IB; $\overline{U} = 10.59\ m/s$.

(2) Integral scale parameter L_1:

$$L_1 = 8.1\Lambda_1 \tag{21}$$

The longitudinal integral scale parameters at the height of the hub Z are:

$$\Lambda_1 = \begin{cases} 0.7Z, (Z \le 60\ m) \\ 42\ m, (Z \ge 60\ m) \end{cases} \tag{22}$$

The horizontal and vertical components of turbulence spectrum parameters simulating a three-dimensional fluctuating wind speed are shown in Table 4.

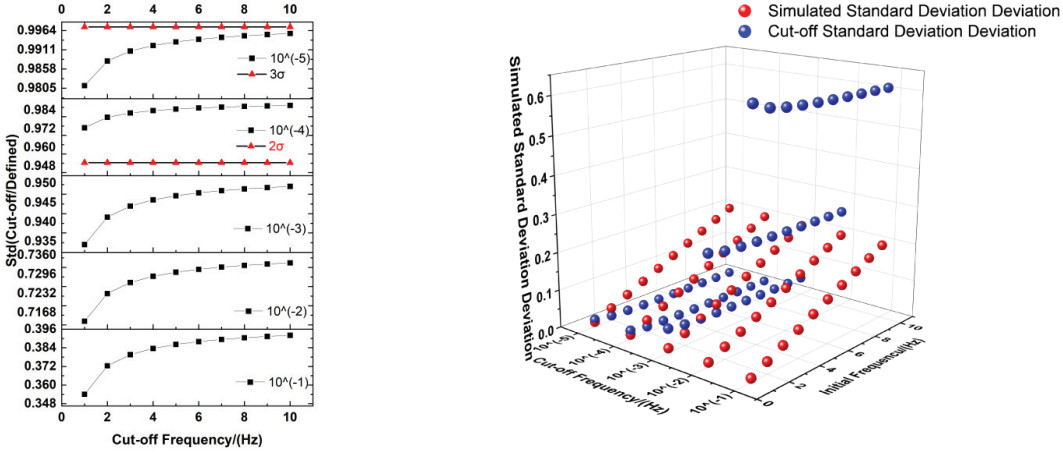
Table 4. Turbulence spectrum parameters of horizontal and vertical components.

Main Parameters	σ_k	L_k
the transverse (k = 2)	$0.8\sigma_1$	$2.7\Lambda_1$
the vertical (k = 3)	$0.5\sigma_1$	$0.66\Lambda_1$

3.2. Determine the Analog Frequency Range $[F_{max}, F_{min}]$

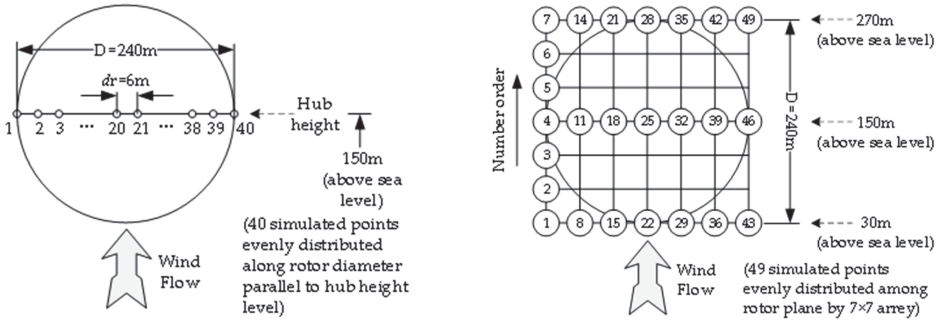
By setting the initial frequency $F_{min} \in [10^{-5}, 10^{-1}]$, setting the cut-off frequency $F_{max} \in [1, 10]$ (step by 10^1 units and 1 unit respectively to form an 5×10 analog frequency interval matrix), and performing integral operations one by one according to Equation (13), the ratio of the standard deviation of the truncated interval integral to the standard deviation of the fully defined interval integral is $\eta = \sigma_{theory}/\sigma$, the result of which is shown in Figure 3a. By combining this with the 3σ principle, it can be found that when the initial frequency $F_{min} < 10^{-4}$ and cut-off frequency is given any value, the standard deviation ratio η will always fall within a confidence interval of $+2\sigma$, as the cutoff frequency increases, η infinity goes to $+3\sigma$, therefore, it achieves a value of $F_{min} = 10^{-4}$.

Under the same simulation conditions, the simulation frequency interval element from the matrix described above was substituted, one by one, into the uniform turbulent wind field simulation algorithm, for simulating fluctuating wind speed of time history in one-dimensional space (at hub height), accompanied with the standard deviation's deviation and truncation deviation. The comparison between the two is shown in Figure 4b. The figure shows that when $F_{min} = 10^{-4}$, the simulation bias $\hat{\varepsilon}$ does not decrease significantly with the increase of the simulation frequency range, the truncated bias step rises when $F_{max} > 6$, instead. To weaken the distortion of the analog signal, the cutoff frequency $F_{max} = 5$ was selected to simulate the uniform turbulent wind field in one-dimensional space (at the height of the hub) and two-dimensional space (at the swept surface of the wind wheel), respectively, (the distribution of simulated points of these two types of spatial wind field simulation are shown in Figure 4a,b, as above) considering the truncation error and simulation error comprehensively, and the compensation coefficient was determined.



(**a**) (**b**)

Figure 3. (**a**) The ratio of cut-off standard deviation and defined standard deviation along with the cut-off frequency change curve; (**b**) Standard deviation scatter distribution under different cut-off frequency intervals.



(**a**) (**b**)

Figure 4. (**a**) One-dimensional spatial simulation point distribution diagram; (**b**) Two-dimensional spatial simulation point distribution diagram.

n

3.3. Compensation Coefficient β Correction Method

According to Section 3.2, the time history simulation of the fluctuating wind speed in a turbulent wind field is performed under the condition that the simulated frequency range is determined as $[F_{\max,}\ F_{min}] = [10^{-4}, 5]$. The main parameters are shown in Table 5, and the distribution of the standard deviation error and compensation coefficient at each simulated height in space is obtained, as shown in the Figure 5.

The longitudinal Kaimal spectrum is modified by considering the compensation coefficient β:

$$S^0(f) = \frac{{\sigma_1}^2}{\beta} \frac{4X_1}{(1+6fX_1)^{5/3}} \tag{23}$$

where, $X_1 = L_1/\overline{U}$.

Table 5. The compensation coefficient of relevant parameters.

Object Parameters	Numerical Value
fully defined standard deviation σ	2.2727
truncated standard deviation σ_{theory}	2.2476
the standard deviation of wind field simulation $\hat{\sigma}$	2.4883
truncated standard deviation's bias ε	0.0110
simulated standard deviation's bias $\hat{\varepsilon}$	0.0949

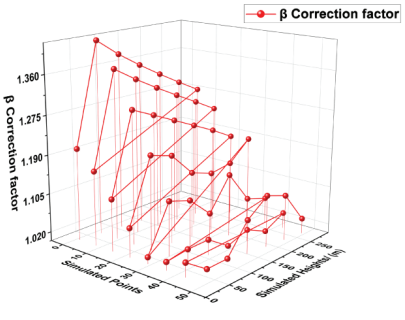


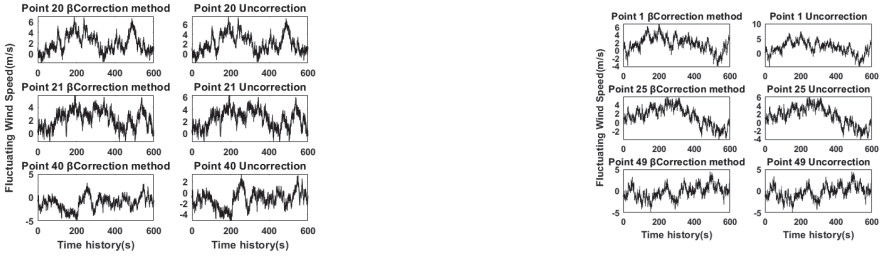
Figure 5. The distribution of correction factor of simulated points for two-dimensional space.

After compensation and correction, the main parameters of one-dimensional space are shown in Table 6.

Table 6. Parameters' contrast between being corrected and uncorrected.

Mock Generated Objects		Uncorrected	Corrected
simulated standard deviation $\hat{\sigma}$	Maximum	2.7103	2.4802
	Minimum	2.1604	1.9818
	Average	2.4883	2.2954
simulated standard deviation's bias $\hat{\varepsilon}$	Maximum	0.1925	0.1280
	Minimum	0.0024	0.0031
	Average	0.1020	0.01
simulated turbulence intensity $\hat{I}_T$		0.1755	0.1619

The modified simulated point fluctuating wind speed diagram is shown as Figure 6.



(**a**) (**b**)

Figure 6. 10 min fluctuating wind speed time history diagram of points: (**a**) One-dimensional space for points 20, 21, 40; (**b**) Two-dimensional space for points 1, 25, 49.

A comparison of the power spectral density at simulated points and power spectral density related to simulated wind speed before and after modification is shown in Figure 7. Among which, the line of the 'Corrected Spectrum Simulated' represents the power density spectrum of the fluctuating wind speeds that are generated by the corrected Kaimal spectrum; the line of the 'Corrected Spectrum' represents the corrected Kaimal power density spectrum itself; the line of the 'Original Simulated Spectrum' represents the power density spectrum of the fluctuating wind speed that are generated by the original Kaimal spectrum; the line of the 'Original Spectrum' represents the original Kaimal power density spectrum itself.

Comparison of the standard deviation, standard deviation's bias, and turbulence intensity of the simulated wind speed before and after modification is shown in Figure 8. Among which, in Figure 8a,c,e, the line of the 'Corrected Simulation' represents the standard deviation/standard deviation's deviation/turbulence intensity of the fluctuating wind speed that is simulated by corrected Kaimal spectrum and generated by 40 points in a one-dimensional simulated wind field; the line of the 'Original Simulation' represents the standard deviation/standard deviation's deviation/turbulence intensity of the fluctuating wind speed that is simulated by the Kaimal spectrum and generated by 40 points in a one-dimensional simulated wind field; the line of the 'Corrected mean value' represents the mean value of the simulation results from the corrected Kaimal spectrum among 40 points in a one-dimensional simulated wind field; the line of the 'Original mean value' represents the mean value of the simulation results from the Kaimal spectrum among 40 points in a one-dimensional simulated wind field; in Figure 8e, the line named 'Defined' represents the defined turbulence intensity I_T; the line named 'Referenced' represents the reference turbulence intensity I_{ref}; in Figure 8b,d,f, the surface named 'Corrected'

represents the standard deviation/standard deviation's deviation/turbulence intensity of the fluctuating wind speed of the points varying along with the simulated heights generated by the corrected Kaimal spectrum in a two-dimensional wind field; the surface named 'Uncorrected' represents the standard deviation/standard deviation's deviation/turbulence intensity of the fluctuating wind speed of the points varying along with the simulated heights generated by the Kaimal spectrum in a two-dimensional wind field; and the surface named 'Defined' represents the defined standard deviation/turbulence intensity of the fluctuating wind speed of the points varying along the simulated heights.

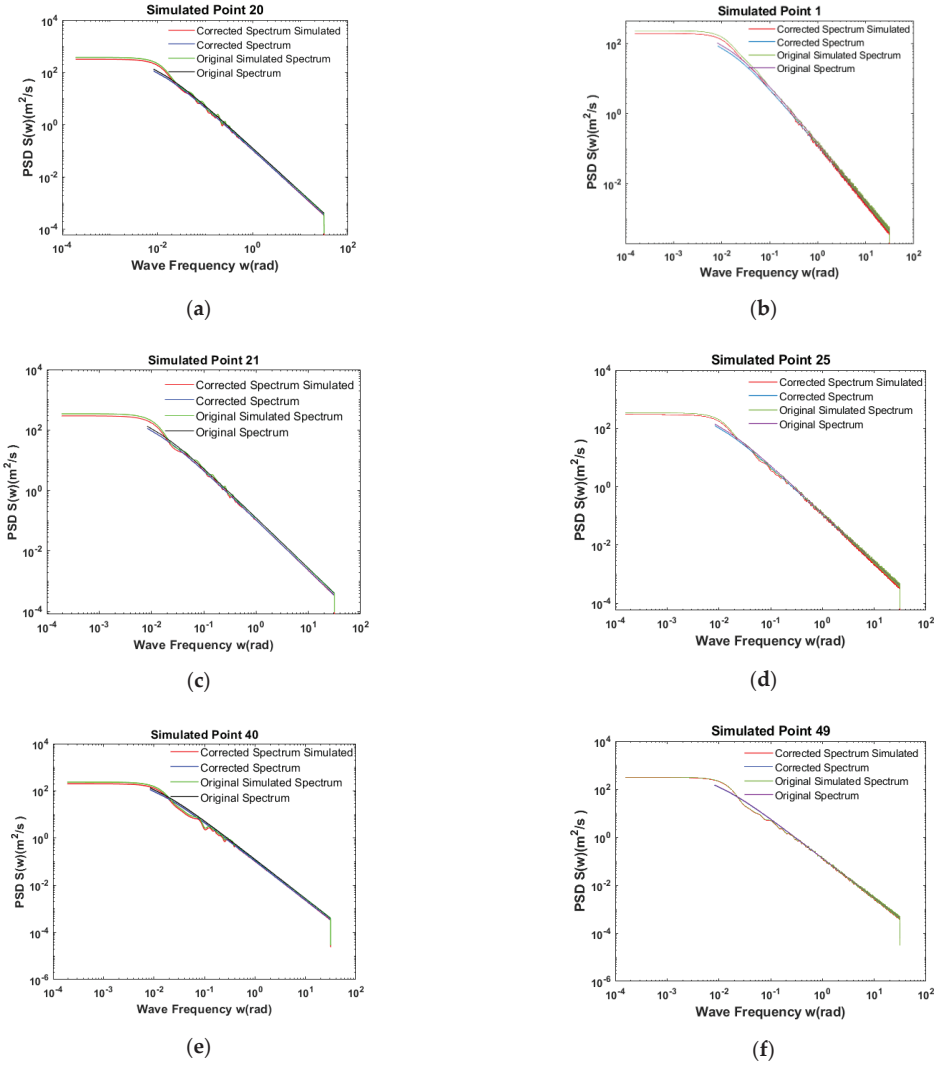


Figure 7. The power density spectrum fitting: (**a**) One-dimensional simulated point 20 being corrected and uncorrected; (**b**) Two-dimensional simulated point 1 being corrected.; (**c**) One-dimensional simulated point 21 being corrected and uncorrected; (**d**) Two-dimensional simulated point 25 being corrected; (**e**) One-dimensional simulated point 40 being corrected and uncorrected; (**f**) Two-dimensional simulated point 49 being corrected.

Concerning the longitudinal one-dimensional space spectrum correction method and the lateral spectrum and the vertical spectrum parameters in Table 4, 20 evenly distributed simulation points are set to verify whether the compensate correction method suggested is applicable for the lateral spectrum and the vertical spectrum. A one-dimensional simulation of three-dimensional wind speed is acquired, and the three-dimensional fluctuating wind velocity contour map and 3 d surface figure of simulation point 10 are shown in Figure 9.

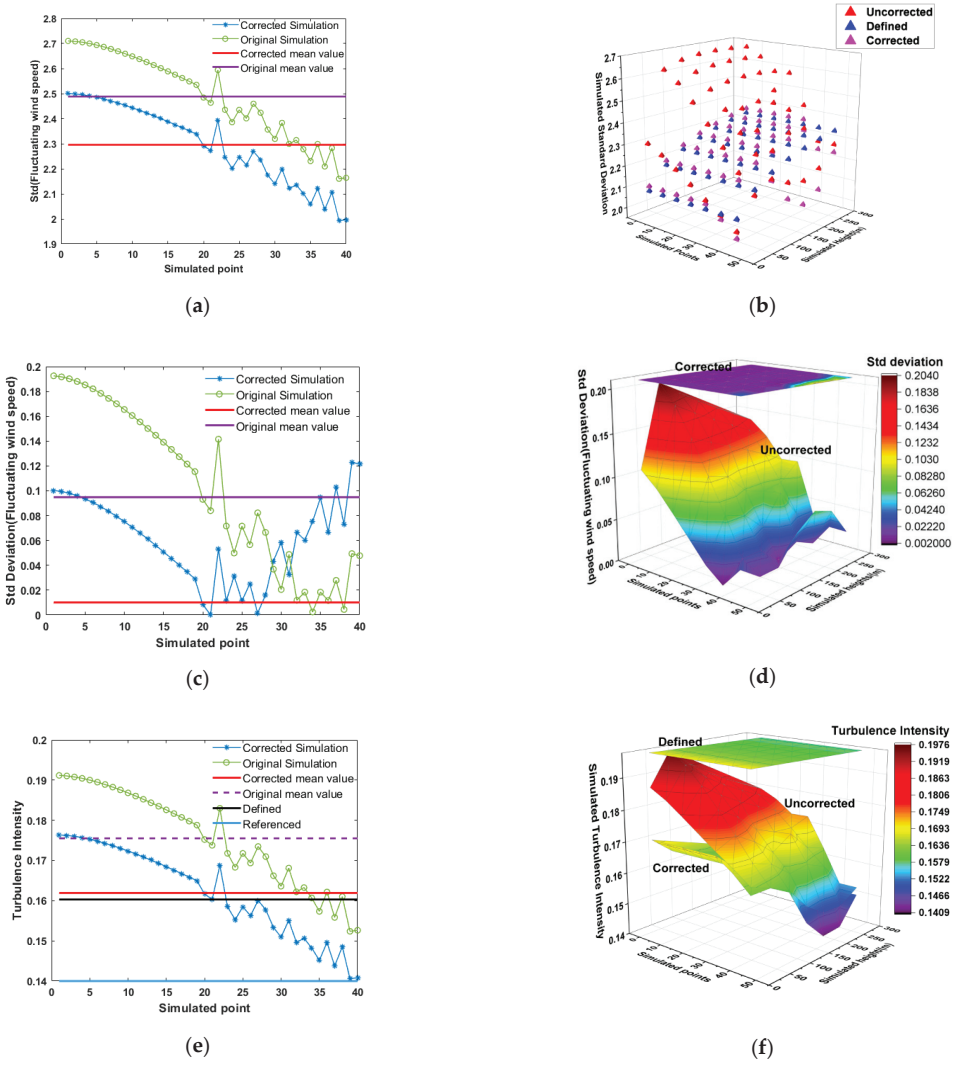
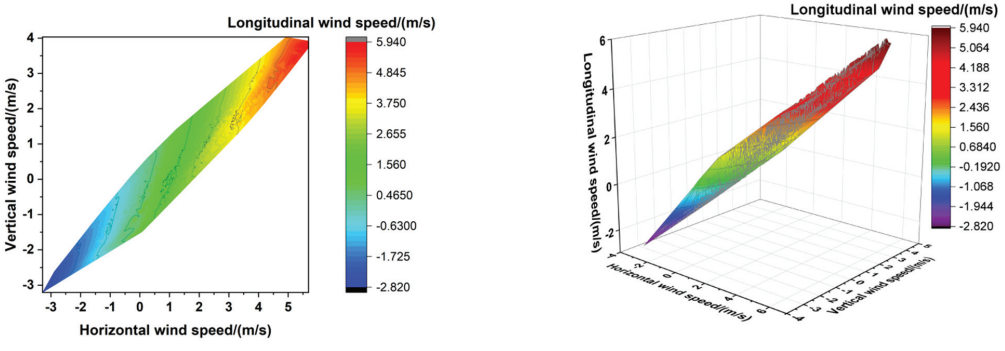


Figure 8. The contrast of standard deviation of simulated points between being corrected and uncorrected: (**a**) One-dimensional space; (**b**) Two-dimensional space; The contrast of standard deviation deviation of simulated points between being corrected and uncorrected: (**c**) One-dimensional space; (**d**) Two-dimensional space; The contrast of turbulence intensity of simulated points between being corrected and uncorrected: (**e**) One-dimensional space; (**f**) Two-dimensional space.



(**a**) (**b**)

Figure 9. (**a**) The three-dimensional fluctuating wind velocity contour map of point 10; (**b**) Standard The three-dimensional fluctuating wind velocity 3D surface of point 10.

The above results show that:

(1) The standard deviation $\hat{\varepsilon}$ generated by the Kaimal spectrum simulation of a one-dimensional turbulent wind field decreases from 0.102 to 0.01, and the reduction is 9.2 times, and the error of truncation deviation ε is only 10%;

(2) The numerical values of the Kaimal power density spectrum with a modified compensation coefficient β simulating the overall turbulence intensity $\hat{I}_T$ of the spatial turbulent wind field are closer to the defined turbulence intensity I_T, which is defined as the ratio of the fully defined standard deviation to the average wind speed $I_T = \sigma/\overline{U}$. In the one-dimensional wind field, the corrected Kaimal spectrum can provide more simulated points that are conservative, according to Figure 8a,c, and the mean value of standard deviation has declined by 0.2 point while the standard deviation's deviation has dropped over 84.2%. Moreover, it can be drawn from Figure 8e that the turbulence intensity value of 45% (18/40) of the simulated points set in the simulation process finally reach the interval of [Referenced, Defined] after the spectrum is corrected, compared with that of the uncorrected spectrum simulation result, which has increased 35%. Therefore, the rate of efficient simulation points in the wind field has obtained a significant promotion. As can be seen from Figure 8e,f, the overall turbulence intensity of the compensated corrected turbulent wind field (referring to the "corrected mean value" line and "Defined" surface, respectively, in Figure 8) is very close to the turbulence intensity defined by the wind field;

(3) The Kaimal power density spectrum simulation with the modified compensation coefficient β generates a more simulated point of fluctuating wind speed turbulence intensity that is distributed within the IEC reference turbulence intensity and defined turbulence intensity range, which can be effectively used in load design evaluation;

(4) According to Figure 7, the line of the 'Corrected Spectrum Simulated' can always fit well with the line of the 'Corrected Spectrum', just like the line of the 'Original Simulated Spectrum' and the line of the 'Original Spectrum' does, which indicates that the fluctuating wind speed generated by the Kaimal power density spectrum simulation with the modified compensation coefficient β can perfectly fit the modified self-power spectrum and meet the requirements of engineering applications.

4. Conclusions

(1) The truncation standard deviation's bias ε that is generated due to the simulated frequency range truncating from the fully defined range affects the simulated turbulent intensity $\hat{I}_T$ of the fluctuating wind speed. Reducing the error between the truncation standard deviation σ_{theory} and the defined standard deviation σ can effectively reduce the effect and take the simulated turbulence intensity $\hat{I}_T$ closer to the defined turbulence intensity I_T;

(2) Under the same simulation conditions, the influence of the initial frequency F_{min} on the truncation standard deviation's bias ε is significantly greater than that of the cutoff frequency F_{max}. When $F_{min} < 10^{-4}$, the ratio of truncation standard deviation σ_{theory} to defined standard deviation σ has fallen within the confidence interval of 2σ. Under this value, the value range of cutoff frequency F_{max} can be within [5,10], and the truncation deviation is small enough;

(3) The turbulence intensity of the turbulent wind field simulated by the Kaimal spectrum of IEC61400-1 and IEC referenced the exponential coherent function is more conservative and more consistent with the defined turbulence intensity;

(4) The calculation method of the compensation coefficient β proposed in this paper is not that precise, for its value would change with the simulated height and distance. Nevertheless, the correction methodology can be employed to any wind speed PSD model for wind speed time-history simulation in uniform space.

Author Contributions: Conceptualization, W.Y., Z.L.; Methodology, W.Y.; Data curation, W.Y., J.H.; Formal analysis, W.Y.; Investigation, W.Y., X.Z.; Project administration, Y.C.; Software, W.Y.; Supervision, Z.H.; Validation, W.Y.; Visualization, W.Y., J.H.; Writing—original draft, W.Y., Z.L.; Writing—Review and editing, W.Y., X.Z., Z.L., J.H.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 51976113.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bustamante, A.; Vera-Tudela, L.; Kuhn, M. Evaluation of wind farm effects on fatigue loads of an individual wind turbine at the EnBW Baltic 1 offshore wind farm. In Proceedings of the Wake Conference 2015, Visby, Sweden, 9–11 June 2015. [CrossRef]
2. Chen, Y.; Zhang, J.; Wang, N. Wind turbine wind field models study numerical simulation of turbulence wind field with MATLAB. *Acta Energiae Solaris Sinica* **2006**, *27*, 954–960.
3. Von Karman, T. Progress in the statistical theory of turbulence. *Proc. Natl. Acad. Sci. USA* **1948**, *34*, 530–539. [CrossRef] [PubMed]
4. Harris, R.I. The nature of wind. In *The Modern Design of Wind-Sensitive Structures*; Construction Industry Research and Information Association: London, UK, 1971.
5. Simiu, E. Wind spectrum dynamic along wind response. *J. Struct. Div.* **1974**, *100*, 203–209. [CrossRef]
6. Kaimal, J.C.; Wyngaard, J.C.; Izumi, Y.; Coté, O.R. Spectral characteristics of surface-layer turbulence. *Q. J. R. Meteorol. Soc.* **1972**, *98*, 563–589. [CrossRef]
7. Liang, J.; Chaudhuri, S.; Shinozuka, M. Simulation of nonstationary stochastic processes by spectral representation. *J. Eng. Mech.* **2007**, *133*, 616–627. [CrossRef]
8. Shinozuka, M. Simulation of multivariate and multidimensional random process. *J. Acoust. Soc. Am.* **1971**, *49*, 357–368. [CrossRef]
9. Ding, Q.; Zhu, L.; Xiang, H. An efficient ergodic simulation of multivariate stochastic process with spectral representation. *Probabilistic Eng. Mech.* **2011**, *26*, 350–356. [CrossRef]
10. Tao, T.; Wang, H. Reduced simulation of the wind field based on Hermite interpolation. *Eng. Mech.* **2017**, *34*, 182–188. [CrossRef]
11. Peng, L.; Huang, G.; Kareem, A.; Li, Y. An efficient space-time based simulation approach of wind velocity field with embedded conditional interpolation for unevenly spaced locations. *Probabilistic Eng. Mech.* **2016**, *43*, 156–168. [CrossRef]
12. Chen, X.; Chen, J.; Li, J. Numerical simulation of fluctuating wind velocity time series of offshore wind turbine. *Proc. CSEE* **2008**, *28*, 111–116.

13. Shen, G.H.; Huang, Q.Q.; Guo, Y.; Xing, Y.L.; Lou, W.J.; Sun, B.N. Simulation methods of fluctuating wind field and its application in wind-induced response of transmission lines. *Acta Aerodyn. Sinica* **2013**, *31*, 69–74. [CrossRef]
14. Zhang, W.; Ma, C.; Sun, X.; Ju, X.L.; Liu, Y.C. Simulation of wind field with spacial correlation based on wavelet analysis method. *Acta Aerodyn. Sinica* **2008**, *26*, 425–429.
15. Det Norske Veritas, Risø DTU National Laboratory. *Guidelines for Design of Wind Turbines*, 2nd ed.; China Machine Press: Beijing, China, 2011.
16. Proakis, J.G.; Manolakis, D.G. *Digital Signal Processing: Principle, Algorithms, and Applications*; Electronic Industry Press: Beijing, China, 2014.
17. Zhao, W.; Liao, M. A method for improving Kaimal spectrum and its algorithm implementation. *Mech. Sci. Technol. Aerosp. Eng.* **2013**, *32*, 1446–1450.
18. IEC 61400-1. *Wind Energy Generation Systems—Part 1: Design Requirements*, 4th ed.; International Electrotechnical Commission: Geneva, Switzerland, 2019.
19. Davenport, A.G. The spectrum of horizontal gustiness near the ground in high winds. *Q. J. R. Meteorol. Soc.* **1961**, *87*, 194–211. [CrossRef]
20. IEA Task 37 2020 IEA GitHub Repository. Available online: https://github.com/IEAWindTask37/IEA-15-240-RWT (accessed on 11 August 2021).
21. Hong, X.; Li, J. Stochastic Fourier Spectrum model and probabilistic information analysis for wind speed process. *J. Wind Eng. Ind. Aerodyn.* **2018**, *1*, 174. [CrossRef]

energies

MDPI

Article

Time-Averaged Wind Turbine Wake Flow Field Prediction Using Autoencoder Convolutional Neural Networks

Zexia Zhang [1], Christian Santoni [1], Thomas Herges [2], Fotis Sotiropoulos [3] and Ali Khosronejad [1,*]

[1] Department of Civil Engineering, Stony Brook University, Stony Brook, NY 11794, USA; zexia.zhang@stonybrook.edu (Z.Z.); christian.santoni@stonybrook.edu (C.S.)
[2] Wind Energy Technologies, Sandia National Laboratories, Albuquerque, NM 87185, USA; therges@sandia.gov
[3] Mechanical & Nuclear Engineering Department, Virginia Commonwealth University, Richmond, VA 23284, USA; sotiropoulosf@vcu.edu
* Correspondence: ali.khosronejad@stonybrook.edu

Abstract: A convolutional neural network (CNN) autoencoder model has been developed to generate 3D realizations of time-averaged velocity in the wake of the wind turbines at the Sandia National Laboratories Scaled Wind Farm Technology (SWiFT) facility. Large-eddy simulations (LES) of the SWiFT site are conducted using an actuator surface model to simulate the turbine structures to produce training and validation datasets of the CNN. The simulations are validated using the SpinnerLidar measurements of turbine wakes at the SWiFT site and the instantaneous and time-averaged velocity fields from the training LES are used to train the CNN. The trained CNN is then applied to predict 3D realizations of time-averaged velocity in the wake of the SWiFT turbines under flow conditions different than those for which the CNN was trained. LES results for the validation cases are used to evaluate the performance of the CNN predictions. Comparing the validation LES results and CNN predictions, we show that the developed CNN autoencoder model holds great potential for predicting time-averaged flow fields and the power production of wind turbines while being several orders of magnitude computationally more efficient than LES.

Keywords: convolutional neural network; wind turbine; wake flow predictions; large-eddy simulation

Citation: Zhang, Z.; Santoni, C.; Herges, T.; Sotiropoulos, F.; Khosronejad, A. Time-Averaged Wind Turbine Wake Flow Field Prediction Using Autoencoder Convolutional Neural Networks. *Energies* **2022**, *15*, 41. https://doi.org/10.3390/en15010041

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 15 November 2021
Accepted: 17 December 2021
Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In a wind farm, turbine wake interactions cause power losses and may increase fatigue loads on downwind wind turbines [1,2]. Therefore, the accurate prediction of turbine wakes is an important consideration in wind farm layout optimization, which can improve the efficiency of power production and reduce the overall levelized cost of energy. As a result, extensive efforts have been made on analytical and numerical models for the estimation of turbines wake [3–6].

Due to the simplicity and low computational cost, engineering models are widely used to predict wake flows and optimize wind farm power production, especially in industrial applications. The very first and extensively studied model was proposed by Jensen [7]. This model was derived from mass conservation, assuming a top-hat shape distribution of velocity deficit in the wake. However, the top-hat wake shape assumption is an oversimplification of the actual wake flow, which can be represented more accurately by a Gaussian distribution [8–10]. Furthermore, more complex real-life characteristics of wake flows have also been considered to improve the accuracy and flexibility of the Gaussian models, including the double-Gaussian type velocity profile of the near wake [11,12], three-dimensional effects [13,14], more accurate models for turbulence intensities [15], wind turbine yaw offset [16], atmospheric stability, and Coriolis force [17]. Although these models are efficient, the accuracy varies significantly from case to case [18,19], especially in the near wake region [12,20]. In addition, wake overlapping effects are not accurately described, as shown by Archer et al. [21].

Compared to engineering models, computational fluid dynamic (CFD) models can provide a better physics-based description of the dynamics of the turbine wakes, such as wake meandering [22–24] and the effects of atmospheric stability [25]. Moreover, some CFD models can even take into account the effect of complex terrain topology [26–28] in addition to the turbine tower and the nacelle [29,30]. The prediction of the velocity deficit and turbulent kinetic energy using the CFD methods is more accurate than engineering models [31]. However, the CFD methods employing LES are computationally expensive, and their use in wind farm optimization is becoming prohibitively expensive.

The development of machine learning and artificial intelligence has encouraged researchers to explore data-driven models to predict the wake and power production of turbines in a wind farm. For example, Japar et al. [32] used five different machine-learning methods, i.e., linear regression, linear regression with feature engineering, nonlinear regression, artificial neural network (ANN), and support vector regression (SVR), to estimate wind turbine power production based on free stream wind speed, wind direction and the turbine position in the wind farm. Although the more elaborate models, i.e., ANN and SVR, have higher accuracy, they slightly deviate from the measured power production in the high wind speed case. Sun [33] developed an ANN to predict power production of wind farms that considers wake effects for varying wind direction, wind speed, and yaw angle. The trained model successfully predicted the power production and was used to optimize the yaw angle of each wind turbine. However, these power production models need a large amount of input parameter combinations for the training and may only be applied to a particular wind farm. When considering the effects of more parameters, for example, turbulence kinetic energy, the whole neural network has to be retrained.

Other data-driven machine learning models have focused on the prediction of the velocity deficit of the wake. Wilson et al. [34] used Random Forest and Multilayer Perceptron (MLP) models to interpolate and predict wind velocity in the turbine wake, but it cannot be applied in different wind fields. Ti et al. [35,36] developed an ANN to predict the velocity deficit and turbulent kinetic energy field in the turbine wake from the incoming wind velocity and turbulence intensity. However, the method requires a large amount of CFD simulation results for training. Yang [37] developed a neural network model to predict the instantaneous position of the meandering turbine wake, using the upwind velocity, turbine torque, and turbine thrust as input features. Zhang and Zhao [38] proposed a neural network combining different dimensionality reduction techniques to predict the velocity field of distributed fluid systems and applied it successfully to predict the flow field of both a single turbine and turbine arrays. King et al. [39] proposed a Gaussian Process (GP) model to correct wind turbine flow field predictions from low-fidelity models, e.g., RANS model, to high-fidelity models, e.g., LES model. Ali et al. [40] used a Long-Short Term Memory (LSTM) model to successfully predict wind velocity fluctuation at specific locations in the turbine wake for a long time period. Renganathan et al. [41] combined an MLP and GP with a Convolutional Neural Network (CNN) decoder to map the wind turbine operation parameters, such as inflow wind speed, turbulent intensity, turbine power generation, atmospheric-boundary layer (ABL) Richardson number, rotor angular speed, and pitch angle, to the wake flow field. In addition to wake reconstruction, data-driven methods have also been used to identify and characterize turbine wakes. Aird et al. [42] developed a mask Region based Convolutional Neural Network (R-CNN) model that identifies turbine wakes in Lidar scan images with high accuracy, even with some missing data points, and is also able to character wake shapes in its forming and dissipating.

Despite these contributions, the accuracy of the existing algorithms for velocity field predictions varies with flow conditions and wind farm layouts, limiting their application for wind farm optimization. In this study, we develop a CNN autoencoder model for generating 3D realizations of time-averaged turbulent wake flow of wind turbines at the Sandia National Laboratories Scaled Wind Farm Technology (SWiFT) site in Lubbock, Texas. The site includes three Vestas V27 wind turbines to investigate the performance of the downwind turbine versus the upwind turbine wake. SpinnerLidar measurements of the

upwind turbine wake at the SWiFT facility have been used to validate the LES results of our in-house virtual flow simulator code, Virtual Flow Simulator (VFS-Wind model). After these validation comparisons, a series of LESs of the SWiFT site were conducted to produce instantaneous and time-averaged flow field results to train and test the CNN. Subsequently, the so-trained CNN was employed to predict the time-averaged flow field of new wind conditions. A data augmentation technique is employed to handle the location sensitivity problems of the CNN. The CNN predictions for the validation test cases were compared against the simulation results of the separately done LES validation case not used in the CNN training. In addition, the predicted time-averaged flow field of wind turbines was used to predict the time-averaged power production of the wind turbines.

This paper is organized as follows. In Section 2, the governing equations of the numerical model and the computational details of the LES of the SWiFT site are presented. In Section 3, the CNN autoencoder algorithm is described, followed by the results and discussion in Section 4. Final remarks can be found in Section 5.

2. Numerical Methods

2.1. Governing Equations

Simulations of the SWiFT site are performed by the LES module of the in-house incompressible Navier-Stokes solver code–VFS-Wind model. In the VFS-Wind code, the incompressible turbulent flow is described by the filtered Navier-Stokes and continuity equation written in curvilinear coordinates given as [43]:

$$J\frac{\partial U^j}{\partial \xi^j} = 0, \tag{1}$$

$$\frac{1}{J}\frac{\partial U^i}{\partial t} = \frac{\xi_l^i}{J}\left(-\frac{\partial}{\partial \xi^j}\left(U^j u_l\right) + \frac{1}{\rho}\frac{\partial}{\partial \xi^j}\left(\mu\frac{g^{jk}}{J}\frac{\partial u_l}{\partial \xi^k}\right) - \frac{1}{\rho}\frac{\partial}{\partial \xi^j}\left(\frac{\xi_j^i p}{J}\right) - \frac{1}{\rho}\frac{\partial \tau_{lj}}{\partial \xi^j} + f_l\right), \tag{2}$$

where $J = \left|\partial\left(\xi^1, \xi^2, \xi^3\right)/\partial(x_1, x_2, x_3)\right|$ is the determinant of the Jacobian of the geometric transformation of the Cartesian coordinates $\{x_i\}$ and the generalized curvilinear coordinates $\left\{\xi^i\right\}$, $\xi_l^i = \partial\xi^i/\partial x_l$ are the transformation metrics, $g^{jk} = \xi_l^i\xi_l^k$ are the components of the contravariant metric tensor, u_i is the ith Cartesian velocity component, $U^i = \left(\xi_m^i/J\right)u_m$ is the contravariant volume flux, p is the pressure, ρ is the fluid density, μ is the dynamic viscosity of the fluid, τ_{ij} is the sub-grid stress tensor for LES, which is modeled using the dynamic Smagorinsky sub-grid scale (SGS) model [44], and f_l is the body force.

2.2. Numerics

In the VFS-wind code, the governing equations are discretized on a hybrid staggered/non-staggered grid in space. The convective terms are discretized using second-order accurate central differencing. For the divergence, pressure gradient, and viscous-like terms, the discretization used is the second-order accurate, three-point central differencing method [45]. The time derivatives are discretized using a second-order backward differencing scheme [46]. The discrete flow equations are time-integrated using an efficient, second-order accurate fractional step methodology in conjunction with a Jacobian-free, Newton-Krylov solver for the momentum equations and a GMRES solver enhanced with the multigrid method as a preconditioner for the Poisson equation.

2.3. Actuator Surface Model

The wind turbine blades and nacelle are modeled using the actuator surface method developed by Yang and Sotiropoulos [47]. This method describes the flow field on the background Cartesian mesh and the wind turbine on the Lagrangian mesh following the actuator surfaces. Velocities on the actuator surfaces are interpolated from the background mesh using the smoothed discrete delta function proposed by Yang et al. [48]. Actuator surfaces of wind turbine blades are represented by chord lines along the radial direction.

Lift and drag forces on blades are calculated similarly to the actuator line model, using the blade element momentum theory. For the nacelle actuator surfaces, the normal component of force is calculated by reconstructing the wall-normal velocity near the actuator surface to satisfy the non-penetration constraint, and the tangential forces are computed as a function of the local incoming velocity and a friction coefficient that parameterizes the effects of near-wall turbulence and the effects of surface geometry. Then the counterforces to the flow field (f_l in Equation (2)) are calculated by distributing the forces on the blades to the background mesh using the above-mentioned interpolation method. Details of the actuator surface method can be found in Yang and Sotiropoulos [47].

3. Computational Details of SWiFT Site Simulation

The SWiFT facility, located in Lubbock, Texas, is an experimental site supported by the U.S Department of Energy to investigate turbine wakes and turbine-turbine interactions. It is comprised of three research-scale wind turbines and two meteorological towers (METs). The layout of wind turbines is as shown in Figure 1b. The wind turbines, Vestas V27s, have rotor diameters of $D = 27$ m and hub heights of 32.1 m. A nacelle mounted Technical University of Denmark (DTU) SpinnerLidar is installed on turbine T1 to measure the turbine wake. The two METs are located upwind against the predominant wind direction to measure the atmospheric inflow. Details of the SWiFT facility can be found in Berg et al. [49].

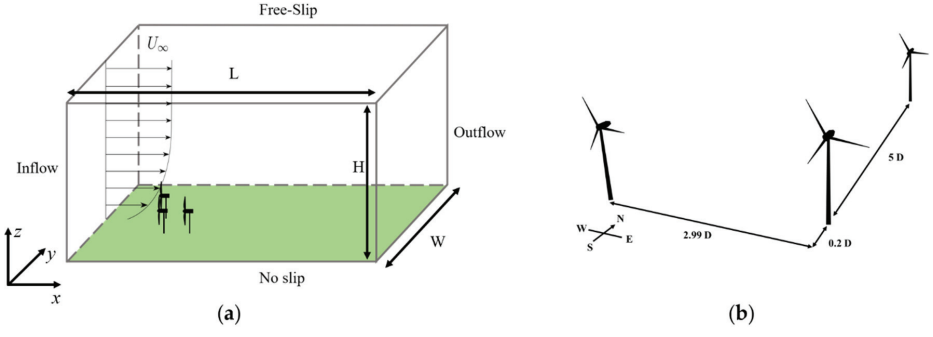


Figure 1. (**a**) is the diagram of the computational domain. (**b**) is the relative location of three turbines.

A series of LESs were performed to generate 3D flow fields of the SWiFT site for a variety of inflow conditions, which give rise to different turbine wake interaction configurations. The computational domain of the real-scale modeled SWiFT site is shown in Figure 1a. It has a length of $L = 80D$, a width of $W = 24.9D$, and a height of $H = 37D$. Free-slip boundary condition is applied to the top and periodic condition along the spanwise direction. A logarithmic law of wall boundary condition is applied to the ground, which is given by

$$u = u^*/\kappa \ \ln(z/z_o), \tag{3}$$

where u^* is the friction velocity, κ is the von Karman constant, and $z_o = 0.0037D$ is the surface roughness height. The outlet is given by the Neumann boundary condition, and the inlet is fed with a fully-developed turbulent flow generated by a precursor simulation as described below. Three wind turbines are located over $8D$ downwind from the inlet. Each turbine has a hub height of $h = 1.19D$ and a rotor diameter of $D = 27$ m. Turbine 2 is $2.99D$ west and $0.2D$ south from turbine 1. Turbine 3 is $5D$ north from turbine 1. The arrangement of the turbines is shown in Figure 1b. To generate different wake conditions, we conducted simulations for four wind directions (150°, 0°, 330°, and 274°, taking south as 0°) as shown in Figure 2. Although the wind directions do not perfectly align with the cardinal directions, for the sake of brevity, they would be referred to as North-East, South, South-West, and West, respectively, through the rest of the paper. The flow domain is rotated in different

wind directions to ensure the x-axis is always along the wind directions. In addition, five wind velocities are considered for each direction ($U_\infty = 7,\ 9,\ 11,\ 13,\ 15$ m/s).

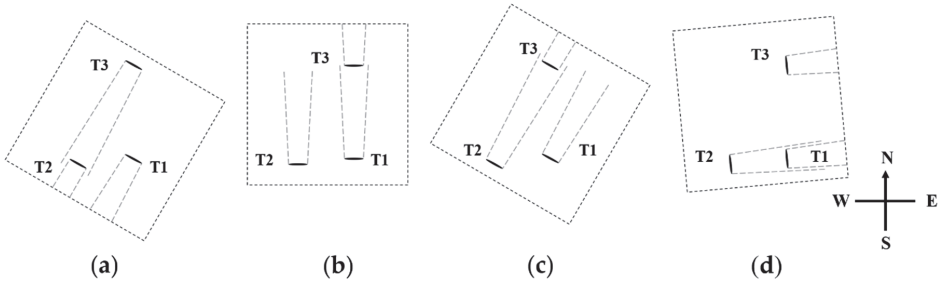


Figure 2. Diagram of wind turbine configurations in different wind directions; (**a**) north-east (150°), (**b**) south (0°), (**c**) south-west (330°), and (**d**) west (274°). Solid lines marked using T1, T2 and T3 represent the location of three turbines. Dashed lines represent the wake of each turbine.

The computational domain is discretized with a grid resolution of $\Delta x = 0.177D$ and $\Delta y = 0.089D$, along the windwise and spanwise directions, respectively. A constant resolution of $\Delta z = 0.089D$ was given along the wall-normal direction up to a height of $7.46D$ and a grid stretching up to the top of the domain with a final resolution of $\Delta z = 1.48D$. Therefore, a uniform grid is obtained in the bottom area that the turbines are located. The details of the computational domain are shown in Table 1.

Table 1. Geometrical and wind characteristics of the SWiFT site model simulations used for the training and validation of the CNN. U_∞ is the free-flow velocity. Re is Reynolds number. D is the diameter of the turbine rotor (=27 m). N_x, N_y, and N_z are the number of computational grid nodes in windwise, spanwise, and vertical directions, respectively. Δx, Δy, and Δz are the special resolution in windwise, spanwise, and vertical directions, respectively.

H	$37D$	$N_x \times N_y \times N_z$	$451 \times 143 \times 281$
W	$24.9D$	Δx	$0.177D$
L	$80D$	Δy	$0.089D$
U_∞ (m s^{-1})	7, 9, 11, 13, 15	Δz	0.089–$1.481D$
Re	5.7×10^8	Δt (s)	$2 \times 10^{-4}\ H/U_\infty$
D (m)	27		

Precursor simulation of a neutral atmospheric boundary layer has been performed to prescribe the velocity at the inlet of the SWiFT site simulations. The precursor simulation has the same grid size and boundary conditions as the SWiFT site numerical domain. The initial transient of the simulation was discarded. After the mean kinetic energy of the computational domain reached steady state, velocities at a plane located in the center of the channel were saved at a $\Delta t = 2.0 \times 10^{-4} H/U$. The time-averaged velocity profile and turbulent intensity of the precursor simulation are shown in Figure 3.

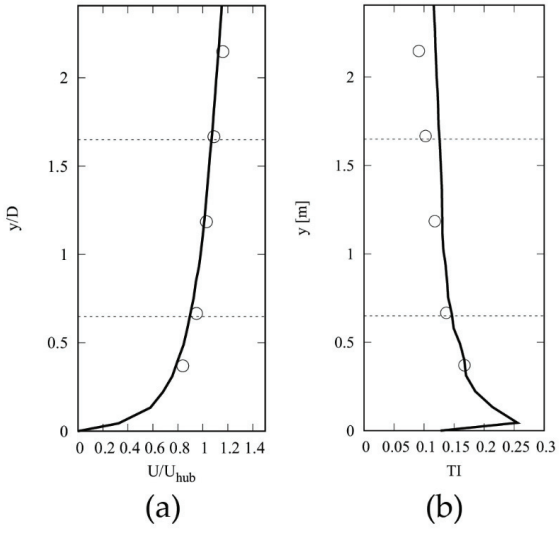


Figure 3. Ten-minute time-averaged (**a**) velocity profile and (**b**) turbulence intensity; (○) met-tower measurements and (–) precursor simulation.

4. Validation of the Computational Model

A numerical simulation of a single wind turbine has been performed to validate the actuator surface model of the SWiFT wind turbine (Vestas V27) [50]. Numerical results are compared against MET data, the turbine supervisory control and acquisition system (SCADA), and SpinnerLidar measurements performed at the SWiFT facility by Sandia National Laboratories and the National Renewable Energy Laboratory [51]. Details of the field experiment and data acquisition can be found in Herges et al. [51].

An initial comparison of the inlet velocity profile obtained from the precursor simulation and the measurements was performed. Measurements of the 10-min time-averaged velocity and turbulence intensity obtained from a MET 2.5D upwind from the wind turbine are compared against the precursor simulation in Figure 3. Although the velocity profile shows good agreement, the turbulence intensity of the precursor simulation shows to be slightly larger than that obtained from the met-tower measurements. Furthermore, the difference in the turbulence intensity seems to be amplified with height, which suggests a slight decrease in turbulent kinetic energy due to atmospheric stratification may have occurred. However, the difference in the turbulence intensity is negligible, around 1% at hub height and 2.7% at 2.15D.

The yaw misalignment of the wind turbine was recorded with the SCADA system. Due to the constantly changing wind direction and the turbine yaw controller not perfectly tracking the wind during the measuring campaign, a time-dependent yaw offset is prescribed to the wind turbine in the simulations to match the measured flow misalignment, shown in Figure 4. In addition, the high-frequency fluctuations of the measured yaw offset, due to the small-scale turbulent coherent structures, are filtered by applying a locally weighted scatterplot smoothing with a window size of 100s to the offset signal. Therefore, a smoother yawing is prescribed to the turbine in the numerical simulation.

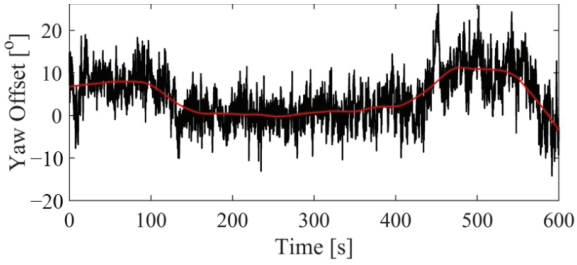


Figure 4. Yaw offset in time between the wind and the turbine; (–) SCADA measurement and (–) prescribed to the turbine on the numerical simulation.

The numerical results are compared against the 10-min line-of-sight velocity measurements in the wake of the T1 wind turbine obtained from the nacelle-mounted DTU SpinnerLidar device [51]. The SpinnerLidar performed 984 scans at a constant focus distance from the device every two seconds. In addition, measurements were performed at 1D to 5D behind the wind turbine by varying the focus distance. The SpinnerLidar cycled the focus distance from 1D to 5D every 30 s.

To mimic the SpinnerLidar measurements, the velocity field of the numerical results are decomposed into a line-of-sight velocity (V_{LOS}) with vertex, or origin, at the location of the nacelle. For higher fidelity in the comparison to the measurements, the simulated V_{LOS} field is sampled at the approximate spatial coordinates and time as the SpinnerLidar measurements. The scattered data from the numerical simulation and the SpinnerLidar are interpolated into a spherical surface mesh to compute the 10-min time-averaged V_{LOS}. A comparison of the horizontal line-of-sight velocity profiles at hub height is shown in Figure 5. Computed velocities in the wake of the turbine are slightly overestimated compared to the SpinnerLidar measurements, specifically close to the center of the rotor. Although the nacelle is modeled and avoids the formation of an unphysical jet in the center of the rotor (commonly observed in standard actuator line models), the momentum deficit seems slightly under-estimated. However, the numerical results from VFS-Wind show good agreement with the SpinnerLidar measurements.

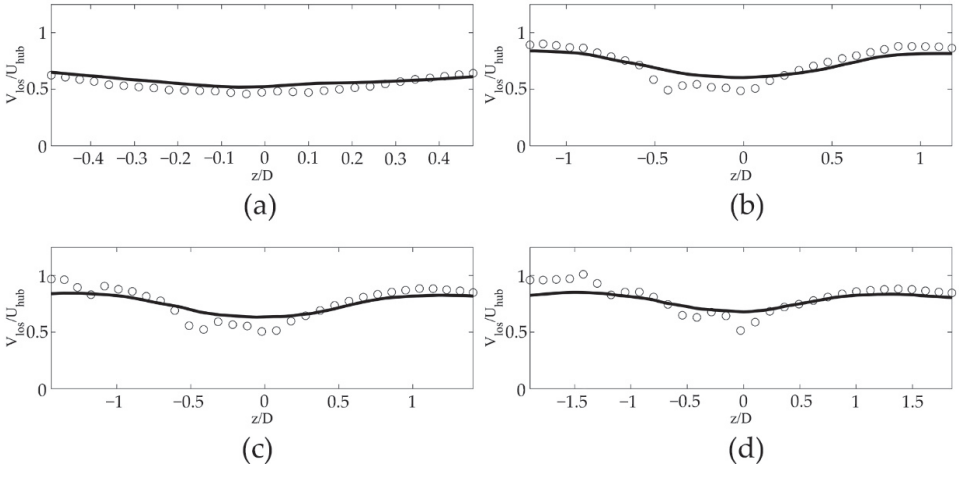


Figure 5. Time-averaged line-of-sight velocity: (○) SpinnerLidar and (–) numerical LiDAR from the LES; (**a**) 1D, (**b**) 2D, (**c**) 3D, and (**d**) 4D behind the wind turbine.

5. CNN Autoencoder Model

We employed a CNN autoencoder model to predict the time-averaged flow field by extracting the key flow field features from instantaneous LES results. The CNN algorithm was originally developed to handle image recognition and image classification tasks [52,53]. Therefore, the CNN has some inherent advantages in handling high-dimensional data: it generally consists of convolutional layers and down-sampling layers that can reduce dimensions of the input image and extract abstract features of the image; the weight sharing concept (it will be explained in the next paragraph) used in CNN allows it to handle high-dimensional data using less learnable parameters and avoid location sensitivity problem [52]. As a variant architecture, the CNN autoencoder (Figure 6) consists of an encoder, which extracts features from input data, and a decoder, which is an inverse of the encoder. As a result, such CNN would enable the image reconstruction from the extracted features. Because of its ability to reconstruct field data, the CNN autoencoder has become a popular tool in the field of fluid dynamics, as well [54–57]. For these reasons, we employed the CNN autoencoder model in this study.

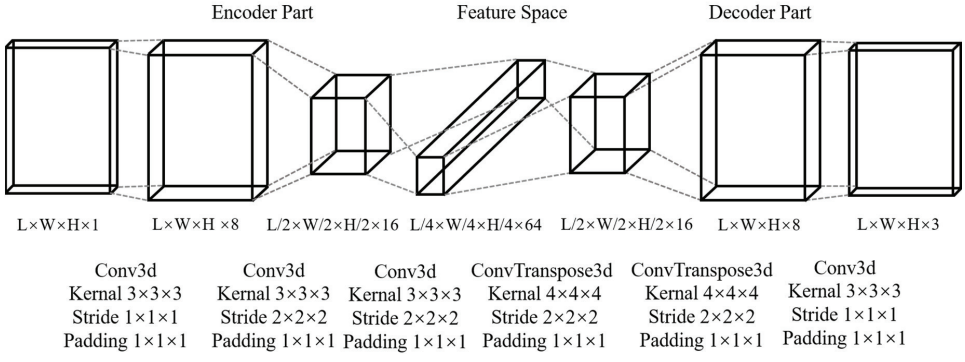


Figure 6. Schematic of the encoder-decoder CNN. Feature maps are depicted as solid boxes. Convolutional layers are depicted as gray dash lines. $L \times W \times H \times channels$ represents the dimensions of each feature map. L, W, and H represent the resolution of the input image in windwise, spanwise and vertical directions, respectively. The layer type, filter size, and stride size of each layer are shown below it. Strides represent the movement step-size of the convolutional filter.

The architecture of the CNN autoencoder used in this work is illustrated in Figure 6. The encoder part includes three 3-dimensional convolution layers for extracting features and down-sample the input data. Each convolution layer includes multiple channels corresponding to different features to be learned. The convolutional layer embedded with a nonlinear activation function and bias operates as follows [55]:

$$\boldsymbol{q}_i^l = \sigma\left(\boldsymbol{k}_i^l \otimes \boldsymbol{q}^{l-1} + b_i^l\right), \tag{4}$$

where $\boldsymbol{q}_i^l$ is the output of ith channel in the lth convolutional layer, σ is the rectified linear unit (ReLU) nonlinear activation function [58], $\boldsymbol{k}_i^l$ is the ith trainable convolutional filter, $\otimes$ is the convolution operator, $\boldsymbol{q}^{l-1}$ is the input of the lth convolutional layer and b_i^l is the ith bias.

In Figure 7, we demonstrate the concept of the convolution operation. In this figure, x is a 4×4 input image with paddings of zero, k is a 3×3 convolution kernel (or filter), while y is the output of the convolution operation. The convolution window is traversed through the padded input image in both horizontal and vertical directions, where the convolution operation with the filter is performed. The step size of each move is the stride—i.e., the step size from red square to orange square. In the convolution operation, the filter can extract features from the input image, and the learnable weights stay unchanged as the

convolution window moves. As a result, the weights of the filter are shared through the whole input image, and because of the weight sharing only one filter with nine weights is required to extract a feature through the entire padded image—instead of 16 different filters corresponding to each output element. In this approach, the filter will be independent of its location leading to fewer learnable weights and thus a more efficient training process. The convolution operation consists of an inner product between the convolution window (i.e., the red dashed box in the input of Figure 7) and the filter to generate the corresponding output element (e.g., the red dashed box in the output of Figure 7). Since the convolution operation reduces the image size, paddings of zeros are used to control the output size. For instance, the output size of the original 4×4 input image in Figure 7 (solid boxes in x) is 2×2 (solid boxes in y). However, with paddings, the output has the same size as the original input. In practice, the convolution operation can work on both 2- and 3-dimensional inputs.

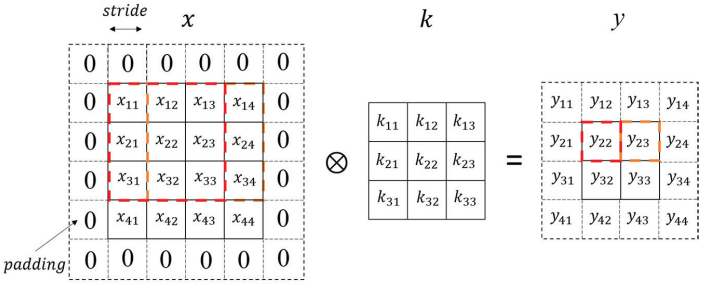


Figure 7. Schematic of the convolution operation. x is the input, k is the convolution kernel, y is the output. Zeros around the input are padding. The orange and red squares represent corresponding input and output cells of the first and second convolution operation, respectively.

Since the convolution operation only involves a linear transformation, an activation function is needed to provide the nonlinear transformation into the CNN. Compared to the sigmoid or hyperbolic tangent activation functions, a rectified linear unit (ReLU) in hidden layers can increase the computational efficiency of the machine learning algorithm [58]. The ReLU function is given by [58]

$$\sigma(\theta) = \max(0,\ \theta), \tag{5}$$

where θ is the result of convolution operation plus the linear bias.

The decoder contains two transpose convolution layers, which are the inverses of the convolution layers, and a convolution layer to up-sample the data and construct the flow field. The discrepancy between output and the target values during the training iterations is calculated using the mean square error (MSE) loss function. Then, the weights in convolutional filters and the biases are updated by the backpropagation algorithm to minimize the loss function [59].

To determine the parameters of the CNN architecture, i.e., the number of layers and channels, and the kernel and stride sizes, a series of parameter combinations are tested to ensure the highest accuracy of results and least number of learnable parameters. The padding sizes are elaborately determined to guarantee the correct output size.

6. Results and Discussion

We carried out LES of the SWiFT site under four different wind directions and five different wind velocities to train and validate the predictions of the CNN. First, we carried out the LES for all cases and discarded the initial two flow-through times—i.e., the duration of time it takes for an air particle to travel through the wind farm. Subsequently, the numerical simulations were continued until the first and second-order turbulence statistics

were fully converged. The convergence of the time-averaged flow field is determined using a time-history-analysis approach reported in Khosronejad et al. [60]. During the training process, the fully converged instantaneous flow fields were fed into the CNN at the input layer, while the time-averaged flow field was designated as the target of CNN at the output layer. We note that the samples used in the training and, consequently, validation processes are taken from smaller domains around each turbine. These subdomains are 6D long, 3D wide, and 1D high. The turbine is located 2D downwind from the inlet and centered along the spanwise direction (Figure 8).

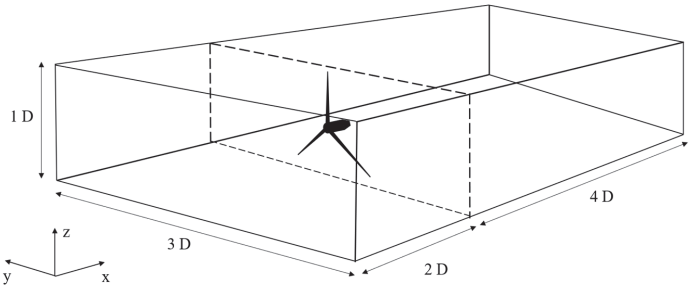


Figure 8. Schematic of the local area around turbine of CNNs sub-domain. The dashed line shows the plane of the rotor.

A schematic of the training procedure is shown in Figure 9. The input of each sample is composed of five neighboring snapshots with 1000 time-step intervals to convey the information of flow field fluctuations—induced by the wake meandering and large turbulent structures from the incoming flow. To examine the effect of the number of input snapshots on the accuracy of the trained CNN's predictions (for the time-averaged results), we conducted a systematic analysis in which we used a different number of input snapshots and calculated the computational errors. The computational error refers to the difference between the CNN's and LES results of the training and validation cases. Our findings for the relation between the number of input snapshots and the corresponding computational errors are shown in Figure 10. As seen in this figure, after five snapshots, the computational errors seem to plateau, suggesting that five snapshots are enough to reconstruct the time-averaged flow field. The target value of the CNN at the output layer is a statistically converged time-averaged flow field. Therefore, each training sample consists of five instantaneous snapshots and the time-averaged flow field.

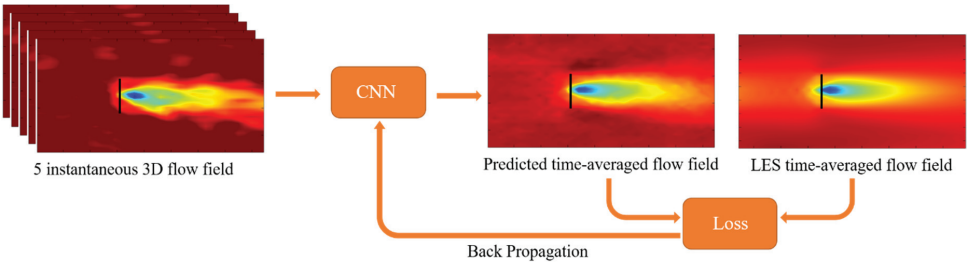


Figure 9. Schematic of the training procedure to develop the CNN. The instantaneous velocity fields are fed into the CNN as the input signal while the time-averaged windwise velocity field is enforced as the target of the output signal.

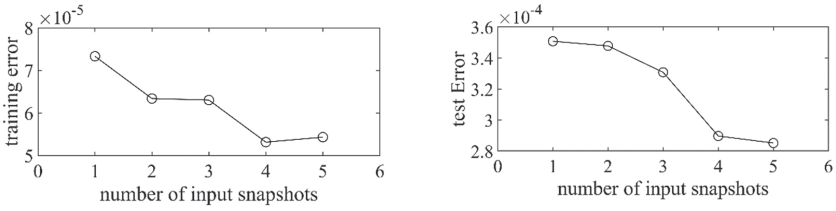


Figure 10. Mean square error as a function of the number of input snapshots; (**a**,**b**) present the computational errors of the CNN for the training and test dataset, respectively.

The case we considered for the training of the CNN includes the SWiFT site with the north-east wind direction and wind speed of $U_\infty = 7$ m/s. The learning rate of the CNN had an initial value of 0.001 with a decay rate of 0.7 in a step size of 500 training epochs. Overall, 99 samples were used in the training process which took 1500 epochs of iterations until the loss curve was fully converged. The prediction of the training case is compared with the LES result in Figure 11. The difference between CNN prediction and the LES result is presented in Figure 12. Velocity profiles from six spanwise cross-sections are compared in Figure 13. The high accuracy of the CNN prediction demonstrates that the CNN is well trained.

Then the CNN is validated using 19 cases. Similar to the training, five successive snapshots (i.e., five instantaneous flow field data taken from five successive time steps) are time-averaged to produce the limited time-averaged flow field and used as the input to the trained-CNN to generate the reconstructed time-averaged flow field. The CNN-predicted time-averaged flow fields are then compared against the time-averaged results of separately conducted LES. For brevity, we only present the comparison of the overlapped turbine wakes for the four cases, e.g., north-east with $U_\infty = 15$ m/s, south with $U_\infty = 13$ m/s, south-west with $U_\infty = 11$ m/s and west with $U_\infty = 9$ m/s. In Figure 14, we plot the windwise velocity contours from top views at the hub level and side views at the hub layer. In Figure 15, we plot the difference between the CNN predictions and LES results. Additionally, velocity profiles taken from six spanwise cross-sections are compared in Figure 16. In the north-east (Figures 14a, 15a and 16a) and south-west (Figures 14c, 15c and 16c) wind direction cases, the presented turbines are 6D downwind of the other turbines. As seen in these figures, the trained CNN has been able to accurately resemble the velocity deficit in both the upwind and downwind wakes. The largest discrepancy is observed for the cases with the south (Figures 14b, 15b and 16b) and west (Figures 14d, 15d and 16d) wind directions, with the latter showing the maximum discrepancy. The present turbines are 5D and 3D downwind of the other turbines in the south and west case, respectively. The difference in the LES versus the CNN computed velocity fields seems to be caused by the proximity of the wind turbines, i.e., the closer the two turbines, the greater the computational error of the CNN predictions. We note that the CNN results for the velocity fields upwind of the turbines (Figure 15b,d, and profiles I, II in Figure 16b,d) in both the south and west cases seem to be slightly over-estimated. However, the CNN could predict the wake of the centered turbine with great accuracy, suggesting that the relatively higher discrepancy upwind of the turbine is due to the location sensitivity of the trained CNN. Theoretically, because of the "weight sharing" feature of the CNN, a trained CNN should be able to reconstruct the turbine wake at any location within the domain. However, the turbines and their wakes have almost the same location in all the training samples and the upwind wake does not strongly affect the training subdomain. Therefore, the trained CNN deems sensitive to turbine location.

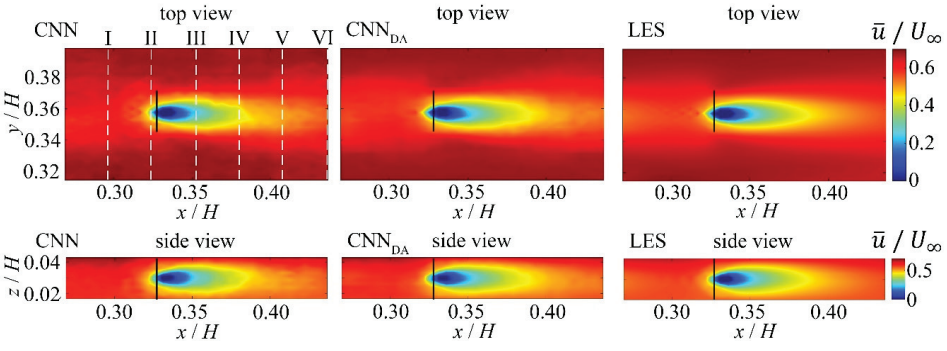


Figure 11. Contours of time-averaged velocity normalized with the free-flow velocity for wind turbines in the training case. Top view cross-sections are at hub-height and side-view cross-sections are at the rotor center. Contours are from the CNN, the CNN with data augmentation method (CNN_{DA}), and the LES results.

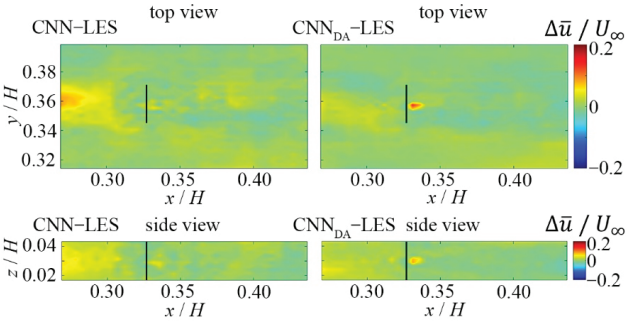


Figure 12. Contours of velocity difference between CNN and LES results normalized with the free-flow velocity for wind turbines in the training case. Top view cross-sections are at hub-height and side-view cross-sections are at the rotor center.

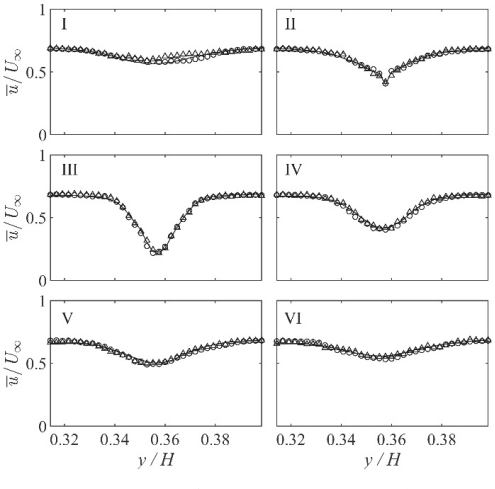


Figure 13. Velocity profiles along the spanwise direction at I, II, III, IV, V, and VI in Figure 11. (–) LES, (Δ) CNN, and (○) CNN with data augmentation.

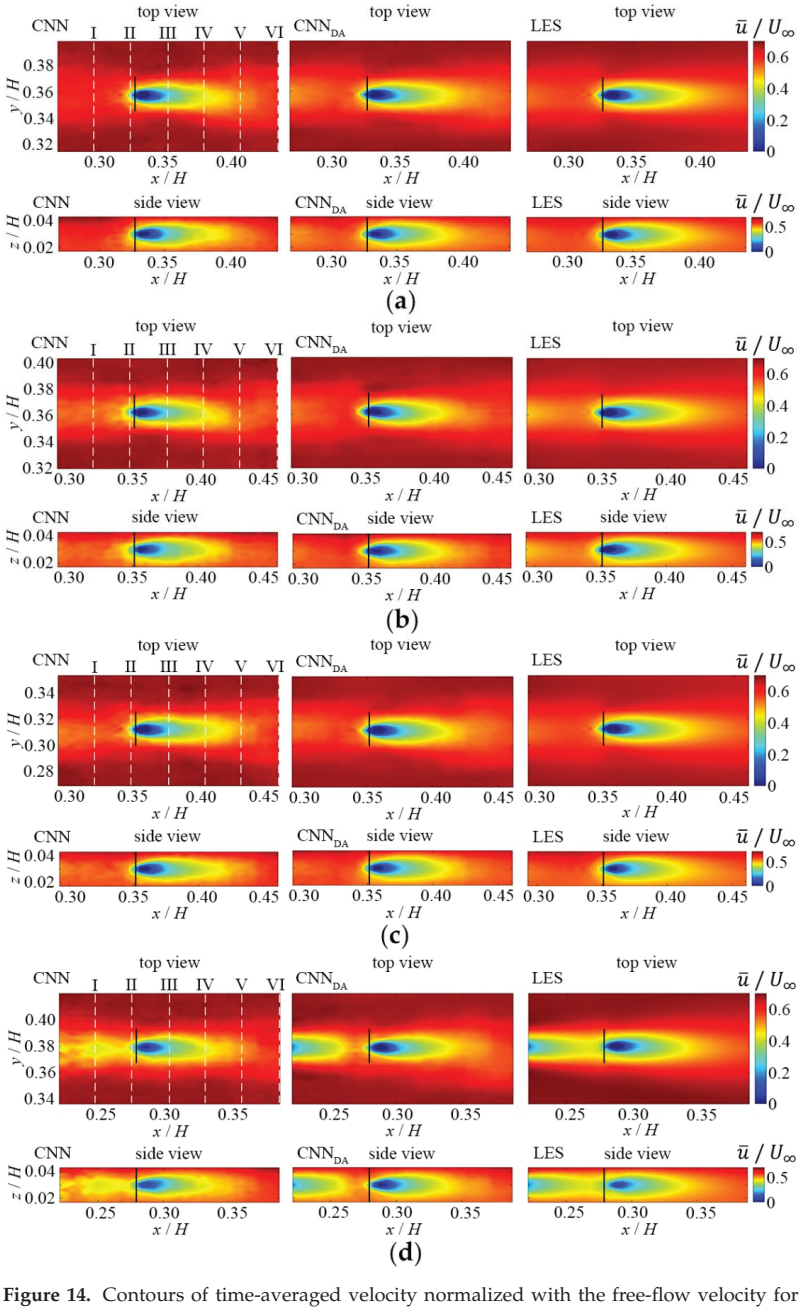


Figure 14. Contours of time-averaged velocity normalized with the free-flow velocity for wind turbines. (**a**) wind turbine 2 in north-east case; (**b**) wind turbine 3 in south case; (**c**) wind turbine 3 in south-west case; (**d**) wind turbine 1 in west case. Top view cross-sections are at hub-height and side-view cross-sections are at the rotor center. Contours are from the CNN, the CNN with data augmentation method (CNN_{DA}), and the LES results.

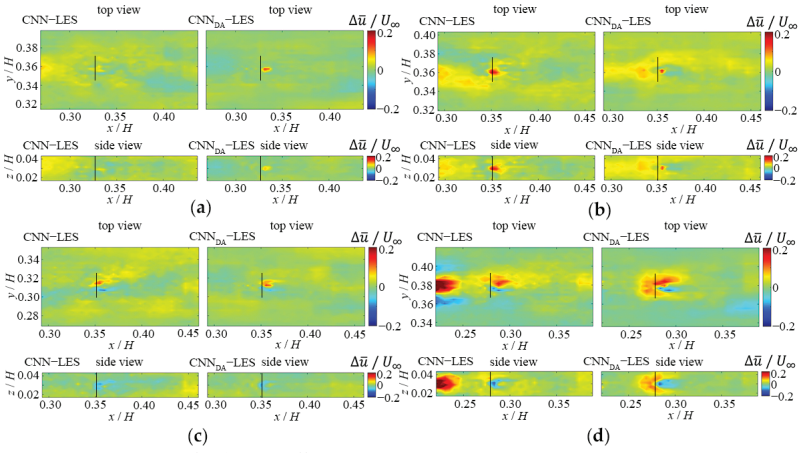


Figure 15. Contours of velocity difference between the CNN and LES results normalized with the free-flow velocity for wind turbines. (**a**) wind turbine 2 in the north-east case; (**b**) wind turbine 3 in the south case; (**c**) wind turbine 3 in the south-west case; (**d**) wind turbine 1 in the west case. Top view cross-sections are at hub-height, and side-view cross-sections are at the rotor center.

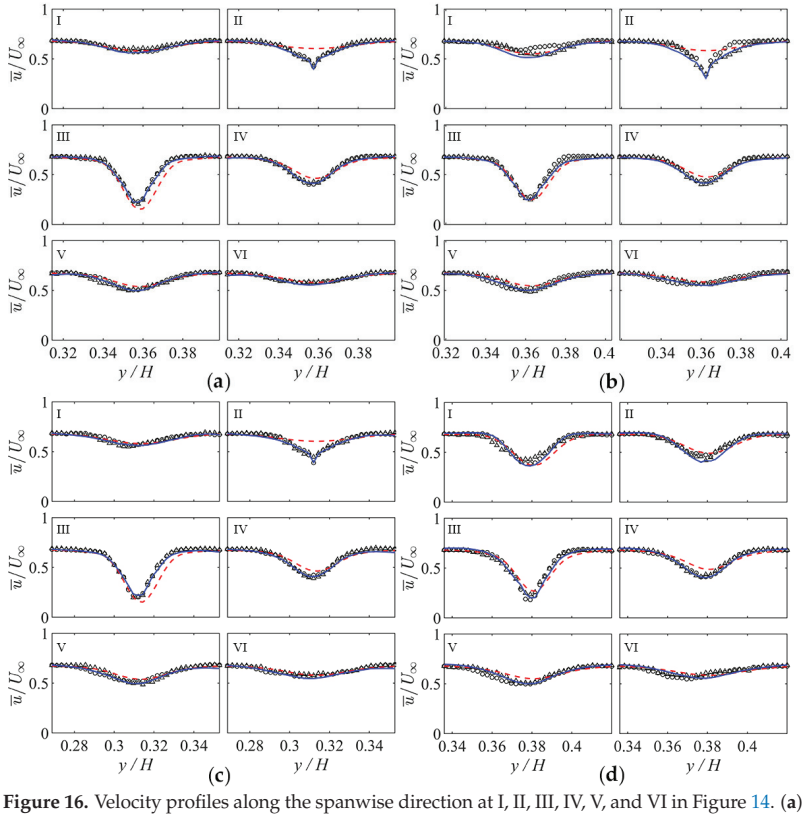


Figure 16. Velocity profiles along the spanwise direction at I, II, III, IV, V, and VI in Figure 14. (**a**) wind turbine 2 in the north-east case; (**b**) wind turbine 3 in the south case; (**c**) wind turbine 3 in the south-west case; (**d**) wind turbine 1 in the west case. (–) LES, (- -) Analytical model, (Δ) CNN, and (○) CNN with data augmentation.

To overcome the location sensitivity problem, we used the data augmentation technique. In this approach, instead of having a fixed location, the subdomains around the turbines are randomly moved up to $5D$ upwind. We also increase the size of training samples by a factor of five. Using this approach, the training samples contain the turbines and their wakes which are randomly scattered in each subdomain. The training and validation results of the CNN trained using the data augmentation technique (CNN_{DA}) are depicted in Figures 11–16. The velocity deficits in the upwind wakes of the south and west cases are predicted more accurately in comparison to the results without data augmentation (Figure 15b,d, and I in Figure 16b,d). However, the discrepancies of the velocities around the turbines are still large in Figure 15b,d, because in these cases, the interaction with upwind turbine's wake is stronger than the training case.

The analytical model developed by Bastankhah and Porté-Agel [9] is used to compare against the CNN model. In this analytical model, the velocity deficit in the wake of a turbine is described as [19]:

$$\frac{\Delta u}{u_0} = \left(1 - \sqrt{1 - \frac{C_T}{8\left(\frac{\sigma_y \sigma_z}{D^2}\right)}}\right) \exp\left(-0.5\left[\left(\frac{y}{\sigma_y}\right)^2 + \left(\frac{z}{\sigma_z}\right)^2\right]\right), \tag{6}$$

where the Δu is the velocity deficit in the wake, u_0 is the mean wind velocity perceived by the wind turbine, C_T is thrust coefficient, σ_y and σ_z are the wake widths in spanwise and vertical directions, respectively, which are given by [19]:

$$\frac{\sigma_y}{D} = k_y \frac{x}{D} + 0.2\sqrt{\frac{1 + \sqrt{1 - C_T}}{2\sqrt{1 - C_T}}}, \tag{7}$$

$$\frac{\sigma_z}{D} = k_z \frac{x}{D} + 0.2\sqrt{\frac{1 + \sqrt{1 - C_T}}{2\sqrt{1 - C_T}}}, \tag{8}$$

where k_y and k_z are the wake growth rates in spanwise and vertical directions, respectively. In the training and validation cases, $C_T = 0.8$. The wake growth rates $k_y = 0.075$ and $k_z = 0.075$ are obtained by fitting the training case. For the turbines in the wake, a wake superposition model is considered [61]:

$$u = U_{hub} - \sqrt{\sum_i^n \Delta u_i^2}, \tag{9}$$

where U_{hub} is the free stream velocity at the hub height, u is the velocity in the wake of current turbine, Δu_i is the wake velocity deficit of the ith turbine in stand-alone condition, n is the number of superposition wakes. The velocity profiles of the analytical model are presented in Figure 16. The analytical model has a good performance in far wakes. However, the velocity deficits are overpredicted in the near wake areas (III in Figure 16) and underpredicted in the near upwind areas (II in Figure 16). In comparison, the presented CNN_{DA} model seems to have a better performance than the analytical model in predicting the velocity field.

The discrepancy between the CNN_{DA} predictions and the LES time-averaged results were quantified using the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and the mean absolute relative error (MARE). These statistical error indices are defined as follows [62]:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\psi_{i(CNN)} - \psi_{i(LES)}\right)^2}{\sum_{i=1}^{N}\left(\psi_{i(CNN)} - \overline{\psi}_{i(CNN)}\right)^2}, \tag{10}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}\left|\psi_{i(CNN)} - \psi_{i(LES)}\right|}{N}, \tag{11}$$

$$\mathrm{RMSE} = \left(\frac{\sum_{i=1}^{N}\left(\psi_{i(CNN)} - \psi_{i(LES)}\right)^2}{N}\right)^{0.5}, \tag{12}$$

$$\mathrm{MARE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\psi_{i(CNN)} - \psi_{i(LES)}\right|}{\psi_{i(LES)}}, \tag{13}$$

where $\psi_{i(CNN)}$ is the predicted value (of time-averaged windwise velocity component) using the $\mathrm{CNN_{DA}}$, $\psi_{i(LES)}$ is the value obtained from the LES model, $\overline{\psi}_{i(CNN)}$ is the mean predicted value using the $\mathrm{CNN_{DA}}$, and N is the total number of samples, i.e., the total number of computational nodes in the subdomain surrounding the turbine.

The discrepancies between the $\mathrm{CNN_{DA}}$ predictions and the LES time-averaged results for the four cases are presented in Table 2. The $\mathrm{CNN_{DA}}$ predictions maintain high accuracy for the turbines located in the free flow for all wind conditions (i.e., various magnitudes and directions). The largest discrepancies are observed for the turbines located in the wake and, specifically, the west wind direction case, where the distance between the turbines is the smallest. Nevertheless, the R^2 for all cases are over 0.95 and the RMSE less than 3%, which is quite remarkable.

Table 2. Statistical error indices of the $\mathrm{CNN_{DA}}$ relative to the LES results for different wind direction cases.

Wind Direction	Turbine	R^2	MAE	RMSE	MARE
NE	T_1	0.99	0.0055	0.0097	0.01
	T_2	0.99	0.0066	0.0111	0.01
	T_3	0.99	0.0045	0.0080	0.01
S	T_1	0.98	0.0083	0.0147	0.02
	T_2	0.98	0.0080	0.0162	0.01
	T_3	0.97	0.0106	0.0149	0.02
SW	T_1	0.97	0.0102	0.0095	0.02
	T_2	0.98	0.0074	0.0143	0.01
	T_3	0.92	0.0166	0.0201	0.03
W	T_1	0.95	0.0140	0.0256	0.03
	T_2	0.95	0.0138	0.0267	0.03
	T_3	0.96	0.0116	0.0181	0.02

The computational efficiency of the proposed CNN model has a great advantage over the LES. For each case, the LES required over 8×10^4 CPU hours to generate the time-averaged flow fields. In comparison, although the proposed CNN model requires about 10 CPU hours for training, the well-trained model only requires 50 s to reconstruct the time-averaged flow field. Considering the LES requires 9.6×10^3 to generate the instantaneous flow fields for inputs, the total cost of the proposed CNN model is still 88% less than the LES.

Now we turn our attention to the possibility of using the proposed CNN to predict aerodynamic power production of the individual turbines using the predicted velocity fields as follows:

$$P_{aero} = \frac{1}{2}\rho A u^3, \tag{14}$$

where P_{aero} is the aerodynamic power production, $\rho = 1.225$ kg m^{-3} is the air density, A is the frontal rotor area, and u is the mean time-averaged windwise velocity component over the rotor area. In Figure 17, we compare the aerodynamic power productions of the four turbines using the LES results and $\mathrm{CNN_{DA}}$ predictions. Wind velocities of the four

cases range from 7 m/s to 15 m/s. Even though the velocity field discrepancies between LES and CNN_{DA} are satisfactory, all wind turbines across all wind velocities overestimated the predicted power productions. As expected, the largest discrepancy is at turbine 1 in the west wind direction case, where the velocity deficit around the rotor area is underestimated (for instance, II in Figure 16d).

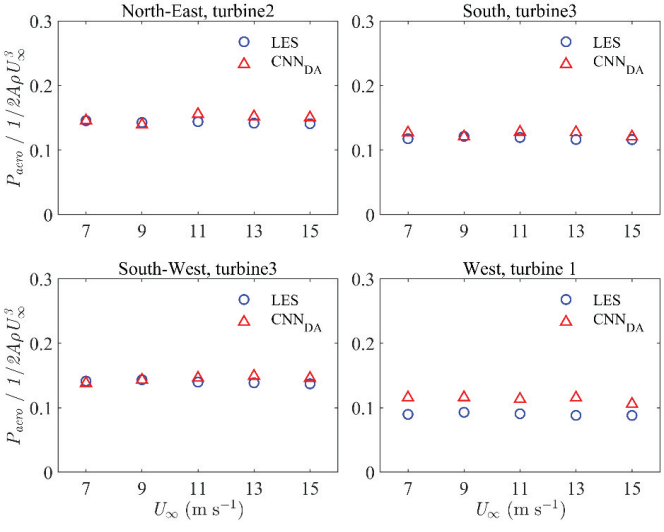


Figure 17. Power production of turbines located in the wake.

7. Conclusions

In this study, we examined the capability of the CNN autoencoder to reconstruct the time-averaged flow field around the wind turbines at the SWiFT facility and predict turbine power output. LES of the SWiFT facility with four different wind conditions (i.e., north-east, south, south-west, and west directions with wind speeds of 7, 9, 11, 13, and 15 m/s) were performed to generate training and validation data for the CNN. A six-layer CNN autoencoder was developed and trained using both instantaneous and time-averaged LES results around three individual turbines using the north-east wind direction case. The input of every sample is constructed using five instantaneous velocity fields to reflect the temporal variations of turbulent structures. Subsequently, the trained CNN is validated and compared with time-averaged results of the additional large-eddy simulations. Based on the findings of this study, the following conclusions can be drawn:

(1) The trained CNN can successfully predict the time-averaged flow field around individual turbines, while the data augmentation technique can effectively address the location sensitivity of the trained CNN. The predicted flow field clearly reflects the main features of the turbine wakes obtained from LES. The velocity profiles drawn from CNN predictions agree well with LES time-averaged results and the overall relative errors are no more than 3%. The presented model has a good generality in different wind speeds. However, wake overlapping will affect the accuracy of the predictions. Different turbine distances lead to different wake-turbine interaction effects and, thus, the flow structure near the turbine varies significantly. Since the CNN model is only trained using one turbine wake interaction case, the generality in different wake interaction cases is approvable. In a future study, we will consider more wake overlapping cases in the training dataset to enable better flow field predictions.

(2) The computational cost associated with the LES to generate the time-average flow field for each case was over 8×10^4 CPU hours, whereas the CNN required nearly 50 s to reconstruct the same flow field. Considering the training cost of about 10 CPU hours

and the cost of LES to produce instantaneous flow field (i.e., 9.6×10^3 CPU hours) for the inputs of the CNN, the total cost of the proposed CNN is 88% less than that of the LES. Therefore, the proposed CNN algorithms could enable reliable predictions of the wake flow field at a fraction of the cost required by the LES.

(3) The CNN predictions for the aerodynamic power productions were in good agreement with the LES results, except for the turbine located in the near wake of the upwind turbine owing to an underestimation of the velocity deficit within the wake. Overall, the comparisons between the LES results and CNN predictions of the SWiFT wind turbines demonstrate the potential of the developed CNN autoencoder for predicting time-averaged flow fields and the power production of wind turbines while being several orders of magnitude less computationally expensive than high-fidelity numerical simulations.

Author Contributions: Conceptualization, Z.Z., A.K.; methodology, Z.Z., A.K., T.H.; software, A.K., F.S.; validation, C.S., T.H.; formal analysis, Z.Z., A.K., C.S., F.S.; investigation, Z.Z., A.K., C.S.; resources, A.K., F.S.; data curation, C.S.; writing—original draft preparation, Z.Z., C.S.; writing—review and editing, A.K., F.S., T.H.; visualization, Z.Z., C.S.; supervision, A.K.; project administration, A.K., F.S.; funding acquisition, A.K., F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Offshore Wind Research and Development Consortium (NOWRDC) under agreement number 147503.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the size of the dataset.

Acknowledgments: The computational resources were provided by the Civil Engineering Department, Stony Brook Research Computing and Cyberinfrastructure, and the Institute for Advanced Computational Science at Stony Brook University. Additional computational resources were provided by the Minnesota Supercompting Institute at the University of Minnesota. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thomsen, K.; Sørensen, P. Fatigue loads for wind turbines operating in wakes. *J. Wind Eng. Ind. Aerodyn.* **1999**, *80*, 121–136. [CrossRef]
2. El-Asha, S.; Zhan, L.; Iungo, G.V. Quantification of power losses due to wind turbine wake interactions through SCADA, meteorological and wind LiDAR data. *Wind Energy* **2017**, *20*, 1823–1839. [CrossRef]
3. Blondel, F.; Cathelain, M. An alternative form of the super-Gaussian wind turbine wake model. *Wind Energy Sci.* **2020**, *5*, 1225–1236. [CrossRef]
4. Ge, M.; Wu, Y.; Liu, Y.; Li, Q. A two-dimensional model based on the expansion of physical wake boundary for wind-turbine wakes. *Appl. Energy* **2018**, *233–234*, 975–984. [CrossRef]
5. Regodeseves, P.G.; Morros, C.S. Unsteady numerical investigation of the full geometry of a horizontal axis wind turbine: Flow through the rotor and wake. *Energy* **2020**, *202*, 117674. [CrossRef]
6. Hornshøj-Møller, S.D.; Nielsen, P.D.; Forooghi, P.; Abkar, M. Quantifying structural uncertainties in Reynolds-averaged Navier–Stokes simulations of wind turbine wakes. *Renew. Energy* **2020**, *164*, 1550–1558. [CrossRef]
7. Jensen, N.O. *A Note on Wind Generator Interaction*; Risø National Laboratory: Roskilde, Denmark, 1983.
8. Hamedi, R.; Javaheri, A.; Dehghan, O.; Torabi, F. A semi-analytical model for velocity profile at wind turbine wake using blade element momentum. *Energy Equip. Syst.* **2015**, *3*, 13–24. [CrossRef]
9. Bastankhah, M.; Porté-Agel, F. A new analytical model for wind-turbine wakes. *Renew. Energy* **2014**, *70*, 116–123. [CrossRef]

10. Gao, X.; Yang, H.; Lu, L. Optimization of wind turbine layout position in a wind farm using a newly-developed two-dimensional wake model. *Appl. Energy* **2016**, *174*, 192–200. [CrossRef]
11. Keane, A.; Aguirre, P.E.O.; Ferchland, H.; Clive, P.; Gallacher, D. An analytical model for a full wind turbine wake. *J. Phys. Conf. Ser.* **2016**, *753*, 032039. [CrossRef]
12. Keane, A. Advancement of an analytical double-Gaussian full wind turbine wake model. *Renew. Energy* **2021**, *171*, 687–708. [CrossRef]
13. Sun, H.; Yang, H. Study on an innovative three-dimensional wind turbine wake model. *Appl. Energy* **2018**, *226*, 483–493. [CrossRef]
14. Sun, H.; Yang, H. Numerical investigation of the average wind speed of a single wind turbine and development of a novel three-dimensional multiple wind turbine wake model. *Renew. Energy* **2019**, *147*, 192–203. [CrossRef]
15. Ishihara, T.; Qian, G.-W. A new Gaussian-based analytical wake model for wind turbines considering ambient turbulence intensities and thrust coefficient effects. *J. Wind Eng. Ind. Aerodyn.* **2018**, *177*, 275–292. [CrossRef]
16. Qian, G.-W.; Ishihara, T. A New Analytical Wake Model for Yawed Wind Turbines. *Energies* **2018**, *11*, 665. [CrossRef]
17. Cheng, Y.; Zhang, M.; Zhang, Z.; Xu, J. A new analytical model for wind turbine wakes based on Monin-Obukhov similarity theory. *Appl. Energy* **2019**, *239*, 96–106. [CrossRef]
18. Lin, M.; Porté-Agel, F. Large-Eddy Simulation of Yawed Wind-Turbine Wakes: Comparisons with Wind Tunnel Measurements and Analytical Wake Models. *Energies* **2019**, *12*, 4574. [CrossRef]
19. Porté-Agel, F.; Bastankhah, M.; Shamsoddin, S. Wind-Turbine and Wind-Farm Flows: A Review. *Bound.-Layer Meteorol.* **2019**, *174*, 1–59. [CrossRef] [PubMed]
20. Fuertes, F.C.; Markfort, C.D.; Porté-Agel, F. Wind Turbine Wake Characterization with Nacelle-Mounted Wind Lidars for Analytical Wake Model Validation. *Remote Sens.* **2018**, *10*, 668. [CrossRef]
21. Archer, C.L.; Vasel-Be-Hagh, A.; Yan, C.; Wu, S.; Pan, Y.; Brodie, J.F.; Maguire, A.E. Review and evaluation of wake loss models for wind energy applications. *Appl. Energy* **2018**, *226*, 1187–1207. [CrossRef]
22. Foti, D.; Yang, X.; Guala, M.; Sotiropoulos, F. Wake meandering statistics of a model wind turbine: Insights gained by large eddy simulations. *Phys. Rev. Fluids* **2016**, *1*, 044407. [CrossRef]
23. Foti, D.; Yang, X.; Sotiropoulos, F. Similarity of wake meandering for different wind turbine designs for different scales. *J. Fluid Mech.* **2018**, *842*, 5–25. [CrossRef]
24. Foti, D.; Yang, X.; Campagnolo, F.; Maniaci, D.; Sotiropoulos, F. Wake meandering of a model wind turbine operating in two different regimes. *Phys. Rev. Fluids* **2018**, *3*, 054607. [CrossRef]
25. Xie, S.; Archer, C.L. A Numerical Study of Wind-Turbine Wakes for Three Atmospheric Stability Conditions. *Bound.-Layer Meteorol.* **2017**, *165*, 87–112. [CrossRef]
26. Yang, X.; Pakula, M.; Sotiropoulos, F. Large-eddy simulation of a utility-scale wind farm in complex terrain. *Appl. Energy* **2018**, *229*, 767–777. [CrossRef]
27. Liu, Z.; Lu, S.; Ishihara, T. Large eddy simulations of wind turbine wakes in typical complex topographies. *Wind Energy* **2021**, *24*, 857–886. [CrossRef]
28. Qian, G.-W.; Ishihara, T. Numerical study of wind turbine wakes over escarpments by a modified delayed detached eddy simulation. *J. Wind Eng. Ind. Aerodyn.* **2019**, *191*, 41–53. [CrossRef]
29. Foti, D.; Yang, X.; Shen, L.; Sotiropoulos, F. Effect of wind turbine nacelle on turbine wake dynamics in large wind farms. *J. Fluid Mech.* **2019**, *869*, 1–26. [CrossRef]
30. Santoni, C.; Carrasquillo, K.; Arenas-Navarro, I.; Leonardi, S. Effect of tower and nacelle on the flow past a wind turbine. *Wind Energy* **2017**, *20*, 1927–1939. [CrossRef]
31. Sedaghatizadeh, N.; Arjomandi, M.; Kelso, R.; Cazzolato, B.; Ghayesh, M.H. Modelling of wind turbine wake using large eddy simulation. *Renew. Energy* **2018**, *115*, 1166–1176. [CrossRef]
32. Japar, F.; Mathew, S.; Narayanaswamy, B.; Lim, C.M.; Hazra, J. Estimating the wake losses in large wind farms: A machine learning approach. In Proceedings of the Innovative Smart Grid Technologies Conference 2014, Washington, DC, USA, 19–22 February 2014; pp. 1–5. [CrossRef]
33. Sun, H.; Qiu, C.; Lu, L.; Gao, X.; Chen, J.; Yang, H. Wind turbine power modelling and optimization using artificial neural network with wind field experimental data. *Appl. Energy* **2020**, *280*, 115880. [CrossRef]
34. Wilson, B.; Wakes, S.; Mayo, M. Surrogate modeling a computational fluid dynamics-based wind turbine wake simulation using machine learning. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, USA, 27 November–1 December 2017; pp. 1–8. [CrossRef]
35. Ti, Z.; Deng, X.W.; Yang, H. Wake modeling of wind turbines using machine learning. *Appl. Energy* **2019**, *257*, 114025. [CrossRef]
36. Ti, Z.; Deng, X.W.; Zhang, M. Artificial Neural Networks based wake model for power prediction of wind farm. *Renew. Energy* **2021**, *172*, 618–631. [CrossRef]
37. Yang, X. Towards the development of a wake meandering model based on neural networks. *J. Phys. Conf. Ser.* **2020**, *1618*, 062026. [CrossRef]
38. Zhang, J.; Zhao, X. Machine-Learning-Based Surrogate Modeling of Aerodynamic Flow Around Distributed Structures. *AIAA J.* **2021**, *59*, 868–879. [CrossRef]
39. King, R.N.; Adcock, C.; Annoni, J.; Dykes, K. Data-Driven Machine Learning for Wind Plant Flow Modeling. *J. Phys. Conf. Ser.* **2018**, *1037*, 072004. [CrossRef]

40. Ali, N.; Calaf, M.; Cal, R.B. Clustering sparse sensor placement identification and deep learning based forecasting for wind turbine wakes. *J. Renew. Sustain. Energy* **2021**, *13*, 023307. [CrossRef]
41. Renganathan, S.A.; Maulik, R.; Letizia, S.; Iungo, G.V. Data-Driven Wind Turbine Wake Modeling via Probabilistic Machine Learning. 2021. Available online: http://arxiv.org/abs/2109.02411 (accessed on 9 October 2021).
42. Aird, J.A.; Quon, E.W.; Barthelmie, R.J.; Debnath, M.; Doubrawa, P.; Pryor, S.C. Region-Based Convolutional Neural Network for Wind Turbine Wake Characterization in Complex Terrain. *Remote Sens.* **2021**, *13*, 4438. [CrossRef]
43. Khosronejad, A.; Sotiropoulos, F. A short note on the simulation of turbulent stratified flow and mobile bed interaction using the continuum coupled flow and morphodynamics model. *Environ. Fluid Mech.* **2020**, *20*, 1511–1525. [CrossRef]
44. Germano, M.; Piomelli, U.; Moin, P.; Cabot, W.H. A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids A Fluid Dyn.* **1991**, *3*, 1760–1765. [CrossRef]
45. Gilmanov, A.; Sotiropoulos, F. A hybrid Cartesian/immersed boundary method for simulating flows with 3D, geometrically complex, moving bodies. *J. Comput. Phys.* **2005**, *207*, 457–492. [CrossRef]
46. Kang, S.; Lightbody, A.; Hill, C.; Sotiropoulos, F. High-resolution numerical simulation of turbulence in natural waterways. *Adv. Water Resour.* **2011**, *34*, 98–113. [CrossRef]
47. Yang, X.; Sotiropoulos, F. A new class of actuator surface models for wind turbines. *Wind Energy* **2018**, *21*, 285–302. [CrossRef]
48. Yang, X.; Zhang, X.; Li, Z.; He, G.-W. A smoothing technique for discrete delta functions with application to immersed boundary method in moving boundary simulations. *J. Comput. Phys.* **2009**, *228*, 7821–7836. [CrossRef]
49. Berg, J.; Bryant, J.; LeBlanc, B.; Maniaci, D.C.; Naughton, B.; Paquette, J.A.; Resor, B.R.; White, J.; Kroeker, D. Scaled Wind Farm Technology Facility Overview. In Proceedings of the 32nd ASME Wind Energy Symposium, National Harbor, MD, USA, 13–17 January 2014. [CrossRef]
50. Kelley, C.L.; Ennis, B.L. *SWiFT Site Atmospheric Characterization*; Technical Report; Sandia National Lab.: Albuquerque, NM, USA, 2016.
51. Herges, T.G.; Keyantuo, P. Robust Lidar Data Processing and Quality Control Methods Developed for the SWiFT Wake Steering Experiment. *J. Phys. Conf. Ser.* **2019**, *1256*, 012005. [CrossRef]
52. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
53. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2012**, *25*, 1097–1105. [CrossRef]
54. Guo, X.; Li, W.; Iorio, F. Convolutional Neural Networks for Steady Flow Approximation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 481–490. [CrossRef]
55. Sekar, V.; Jiang, Q.; Shu, C.; Khoo, B.C. Fast flow field prediction over airfoils using deep learning approach. *Phys. Fluids* **2019**, *31*, 057103. [CrossRef]
56. Morimoto, M.; Fukami, K.; Zhang, K.; Nair, A.G.; Fukagata, K. Convolutional Neural Networks for Fluid Flow Analysis: Toward Effective Metamodeling and Low-Dimensionalization. 2021. Available online: http://arxiv.org/abs/2101.02535 (accessed on 20 February 2021).
57. Nakamura, T.; Fukami, K.; Hasegawa, K.; Nabae, Y.; Fukagata, K. Convolutional neural network and long short-term memory based reduced order surrogate for minimal turbulent channel flow. *Phys. Fluids* **2021**, *33*, 025116. [CrossRef]
58. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
59. Hecht-Nielsen, R. Theory of the backpropagation neural network. In *Neural Networks for Perception*; Academic Press: Cambridge, MA, USA, 1992; pp. 65–93.
60. Khosronejad, A.; Flora, K.; Kang, S. Effect of Inlet Turbulent Boundary Conditions on Scour Predictions of Coupled LES and Morphodynamics in a Field-Scale River: Bankfull Flow Conditions. *J. Hydraul. Eng.* **2020**, *146*, 04020020. [CrossRef]
61. Voutsinas, S.; Rados, K.; Zervos, A. On the analysis of wake effects in wind parks. *Wind Eng.* **1990**, *14*, 204–219.
62. Ebtehaj, I.; Bonakdari, H. A comparative study of extreme learning machines and support vector machines in prediction of sediment transport in open channels. *Int. J. Eng. Trans. B Appl.* **2016**, *29*, 1499–1506.

electronics

MDPI

Article

A Frequency and Voltage Coordinated Control Strategy of Island Microgrid including Electric Vehicles

Peixiao Fan [1], Song Ke [1,*], Salah Kamel [2], Jun Yang [1], Yonghui Li [1], Jinxing Xiao [3], Bingyan Xu [3] and Ghamgeen Izat Rashed [1]

[1] School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China; whufpx0408@163.com (P.F.); JYang@whu.edu.cn (J.Y.); whusee2006@yahoo.com (Y.L.); ghamgeen@whu.edu.cn (G.I.R.)
[2] Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt; skamel@aswu.edu.eg
[3] Electric Power Company of Shanghai, State Grid, Shanghai 200122, China; xjx1122@163.com (J.X.); gh197493@yahoo.com (B.X.)
* Correspondence: kesong1997@whu.edu.cn; Tel.: +18-771-043566

Abstract: Frequency and voltage deviation are important standards for measuring energy indicators. It is important for microgrids to maintain the stability of voltage and frequency (VF). Aiming at the VF regulation of microgrid caused by wind disturbance and load fluctuation, a comprehensive VF control strategy for an islanded microgrid with electric vehicles (EVs) based on Deep Deterministic Policy Gradient (DDPG) is proposed in this paper. First of all, the *SOC* constraints of EVs are added to construct a cluster-EV charging model, by considering the randomness of users' travel demand and charging behavior. In addition, a four-quadrant two-way charger capacity model is introduced to build a microgrid VF control model including load, micro gas turbine (MT), EVs, and their random power increment constraints. Secondly, according to the two control goals of microgrid frequency and voltage, the structure of DDPG controller is designed. Then, the definition of space, the design of global and local reward functions, and the selection of optimal hyperparameters are completed. Finally, different scenarios are set up in an islanded microgrid with EVs, and the simulation results are compared with traditional PI control and $R(\lambda)$ control. The simulation results show that the proposed DDPG controller can quickly and efficiently suppress the VF fluctuations caused by wind disturbance and load fluctuations at the same time.

Keywords: islanded microgrid; electric vehicles; charger capacity model; VF control; DDPG

Citation: Fan, P.; Ke, S.; Kamel, S.; Yang, J.; Li, Y.; Xiao, J.; Xu, B.; Rashed, G.I. A Frequency and Voltage Coordinated Control Strategy of Island Microgrid including Electric Vehicles. *Electronics* **2022**, *11*, 17. https://doi.org/10.3390/electronics11010017

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 5 December 2021
Accepted: 20 December 2021
Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Microgrid refers to a small power generation and distribution system that is composed of distributed power sources, energy storage devices, energy conversion devices, related loads, monitoring, and protection devices. It is an autonomous system that can realize self-control, protection, and management. In addition, the microgrid can operate in grid-connected mode and islanded mode. In islanded mode, the power quality of the microgrid is usually maintained by the micro sources and flexible loads [1]. At the same time, with the development of vehicle-to-grid (V2G) technology, the research of EVs in the areas of grid peak and valley filling, suppression of power fluctuations, and microgrid stability control has also been deepened [2,3], which brings opportunities and challenges to the VF regulation of microgrids.

Due to the limited capacity of the islanded microgrid, ensuring the stability of frequency and voltage is the key for the operation safety of microgrid. In [4], a VF strategy of an islanded microgrid based on fuzzy logic controller is proposed, which can control active and reactive powers and decrease power losses of the microgrid, thus the effectiveness and robustness of the proposed controller over the conventional proportional- integral controller. In [5], a decoupled VF controller for DGs is proposed, which is able to keep the

grid VF magnitude constant, so as to enhance the resilience and increase the penetration of renewable energy to the stand-alone microgrid. In [6], an optimized solution is proposed for minimizing both frequency and voltage deviations. The simultaneous control of VF is achieved with proper load sharing among the DG units. However, the system parameter setting of the traditional control strategy mentioned above is complicated, and the control performance needs to be further improved when faced with complex working conditions such as wind disturbance and load fluctuation.

Therefore, various intelligent algorithms are gradually being widely used in the control of microgrids. In [7], a new scheme for the online minimization of harmonic distortion of an islanded microgrid based on a population-based optimization method is proposed, presenting a new central controller to optimize network voltage harmonics according to particle swarm optimization (PSO) algorithm, while active power is shared between distributed generation units. In [8], a coordinated load shedding control scheme based on Double-*Q* learning for an islanded microgrid is proposed to solve the problem of how to determine the appropriate load shedding amount and objects when frequency is disturbed by considering the relationship between the active power and frequency deviation of each distributed energy resource. However, the intelligent controllers mentioned above can only regulate frequency or voltage, that is, they cannot take both of frequency recovery and voltage adjustment.

Meanwhile, in the construction of the microgrid model, the access of EVs is not considered, and the boundary of the output power of each unit is ignored. Thus, there is room for further optimization in the microgrid model and control strategy. EVs have become a new type of distributed energy storage unit with its energy saving, environmental protection, and flexibility [9,10], which can provide power support for the islanded microgrid and improve its operational flexibility through V2G technology. In [11], an islanded microgrid LFC model including loads, distributed power sources, MT, EVs, and their constraints is established. However, the output power boundary of the EVs charging station model in this paper is a fixed value, which does not match the actual situation. In [12], a microgrid including micro gas turbine (MT), EVs, distributed power, and loads is established, and an improved robust model predictive frequency control strategy of microgrids with EVs is proposed, which can better suppress the frequency fluctuation with a faster response speed than other methods, but the random output power boundary is not refined from the perspective of cluster EVs. In addition, none of the above references considers the reactive power regulation effect of EVs on voltage stability. In fact, the power boundary of the charging station can be affected by user travel demand, charging behavior, and the characteristics of EV clusters. Thus, the active power P and reactive power Q output by the EVs charging station can be adjusted according to the control command and the power factor angle of the charger, so as to complete the stability control of VF.

In summary, the randomness of users can affect the charging behavior of EVs stations. In addition, there is no suitable intelligent control algorithm that can use EVs to realize the coordinated control of the VF of the islanded microgrid. Thus, a VF coordinated control strategy based on Deep Deterministic Policy Gradient (DDPG) is proposed in this paper, which is applied to the VF control of an islanded microgrid with EVs. The main contributions are as follows: (1) In order to solve the problem of randomness in the charging boundary of EVs caused by the users' randomness, the VF control model of EVs is established. The *SOC* constraint condition of the EVs is established, and a four-quadrant two-way charger capacity model is introduced. Thus, a microgrid VF control model including load, MT, EVs, and their random power boundary is built; and (2) the voltage and frequency fluctuations can be caused by wind disturbance and load fluctuations. Thus, the DDPG controller with online learning and experience playback capabilities is selected. The convergence characteristics of DDPG are great, so it can coordinate the frequency recovery and voltage regulation of the islanded microgrid greatly. (3) In order to achieve effective regulation of voltage and frequency at the same time, the structure of DDPG controller is designed according to the two control goals of microgrid frequency

and voltage. In addition, then the definition of space, the design of global and local reward functions, and the selection of optimal hyperparameters are completed. Thereby, it can simultaneously meet the VF control requirements.

2. Microgrid Control Model with EVs

The VF control in microgrid can be realized by distributed power supply, energy storage device, etc. In addition, EVs can also participate in microgrid VF regulation. For the VF control of microgrid, the microgrid control system of distributed power supply, load, MT, and EVs is established in this section.

2.1. Electric Vehicle Control Model

As a flexible energy storage device in microgrid control, EVs can regulate the charge and discharge power of the battery according to the instructions of the controller, thereby to control the interaction of active power with the grid [13]. At the same time, charger scheduling is applied to realize the regulation of voltage or reactive power. The two-way charger can realize four-quadrant operation [14], and the power factor cannot determine the transmission direction of reactive power, so the operating quadrant of the charger cannot be determined. Taking the power factor angle as the control variable can determine the transmission direction and magnitude of active and reactive power together, which is more conducive to the two-way transmission control of active and reactive power between the grid and the EV.

The function of EVs in microgrid control is similar to that of energy storage devices. In terms of active power, the charging and discharging power ranges of EV are limited within $\pm\lambda_e$, due to the limits of inverter capacity. The E_{max} is the maximum capacity of EVs station. In addition, the recommended maximum capacity $E_{rmax} = 0.9E_{max}$ and the recommended minimum capacity $E_{rmin} = 0.1E_{max}$ are set to ensure the safe and stable operation of EV station. When the current capacity E of the EVs station is higher than the E_{rmax}, the EV stations can discharge to the microgrid, and the discharge power range is 0–λ_e. Similarly, if the current capacity of the EVs station is lower than the E_{rmin}, the EV station can be charged from the microgrid within the charging power range is $-\lambda_e$–0. In addition, the EV control model can be affected by users' uncertain factors such as the randomness of travelling demands and charging behavior of users.

Firstly, the randomness of user travel demand affects the capacity and limitation of the charging station to be random. Therefore, it is necessary to establish the constraints of *SOC* to ensure that the user's normal travel is still satisfied under the interaction between EVs and the grid. In addition, the initial *SOC* of the battery in this paper is set as a random number [15] obeying Gaussian distribution, and its probability density function is expressed as Equation (1):

$$f(s) = \frac{1}{\sigma_s\sqrt{2\pi}} e^{\frac{-(s-\mu_s)^2}{2\sigma_s^2}} \tag{1}$$

where μ_s represents the average value of *SOC*, and σ_s represents the standard deviation.

According to the 2017 National Household Travel Survey (NHTS) of the US Department of Transportation [16], it can be obtained that the daily mileage L obeys lognormal distribution, and its probability density function is as follows:

$$f(L) = \frac{1}{L\sigma_L\sqrt{2\pi}} e^{\frac{-(lnL-\mu_L)^2}{2\sigma_L^2}} \tag{2}$$

where μ_L represents the average value of the daily mileage L, and σ_L represents the standard deviation.

According to the daily driving mileage, the charging time T_c is calculated:

$$T_c = \frac{LQ_{100}}{100P_c} \tag{3}$$

where P_c is the charging power, and Q_{100} is the power consumption per 100 km.

For the leaving time T_{leave}, it is required that $T_{\text{leave}} \geq T_c$. Thus, T_{leave} is set as follows:

$$T_{\text{leave}} = (1 + \sigma_T)T_c \tag{4}$$

where σ_T is a positive random number.

Based on the above parameters, the demanded SOC for future travel named SOC_m can be calculated [17]:

$$SOC_m = S_0 + \frac{L}{L_{\max}} \tag{5}$$

where S_0 is the initial SOC for EVs.

Therefore, for EVs in the station, the SOC can be maintained within the range of [SOC_{rmin}, SOC_{rmax}]. SOC_{rmax} and SOC_{rmin} are the recommended maximum and minimum value of SOC, which can ensure the life of the battery. To satisfy the sufficient SOC_m to make sure the follow-up driving when EVs leave, the constraint conditions are added to the SOC of EVs, as shown in Figure 1. The blue dotted line represents the charge boundary, which means that the EV can no longer charge when the SOC reaches SOC_{rmax}. The red dotted line represents the discharge boundary, which means that the EV can no longer discharge when the SOC reaches SOC_{rmin}. The solid green line represents the boundary of forced charging, which means that the EV is forced to charge to ensure the SOC_m when leaving the charging station.

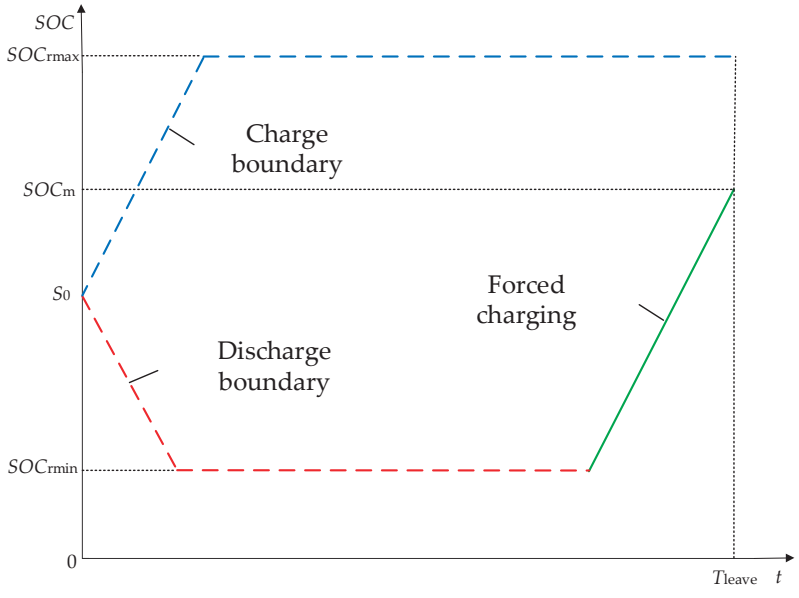


Figure 1. Boundary of charging and discharging constraints for EVs.

Furthermore, in terms of active power, the rated charging power of a single EV can be set to $P_{EV,i,}^{ch}$ and the rated discharging power to $P_{EV,i}^{dis}$ The relationship between the charging power of a single EV and the charging and discharging state can be obtained as follows: When $SOC_i \geq SOC_{\text{rmax}}$, the single EV can discharge positive power increment $0 < \Delta P_{EV,I} < P_{EV,i}^{dis}$, which can ensure that SOC_i is controlled below SOC_{rmax}. When $SOC_i \leq SOC_{\text{min}}$, the single EV can only be charged, that is, only the negative power increment can be discharged $-P_{EV,i}^{ch} < \Delta P_{EV,i} < 0$, which can ensure that SOC_i is controlled above SOC_{rmin}. When $SOC_{\text{rmin}} < SOC_i < SOC_{\text{rmax}}$, the single EV can be charged and discharged. Thus, the power increment satisfies $-P_{EV,i}^{ch} < \Delta P_{EV,i} < P_{EV,i}^{dis}$. In summary, the

instruction distribution of the EVs station through the controller is shown in Figure 2. In addition, the charging and discharging constraint boundary of a single EV can be obtained as follows:

$$P^{+}_{EV,i}(t) = \begin{cases} 0 & ,t \geq T_{leave,i}, 0 \leq t < T_{leave,i} \text{ and} SOC_{EV,i}(t) \leq SOC^{-}_{EV,i}(t) \\ P^{dis}_{EV,i}(t) & , 0 \leq t < T_{leave,i} \text{ and } SOC_{EV,i}(t) > SOC^{-}_{EV,i}(t) \end{cases} \tag{6}$$

$$P^{-}_{EV,i}(t) = \begin{cases} 0 & ,t \geq T_{leave,i}, 0 \leq t < T_{leave,i} \text{ and } SOC_{EV,i}(t) \geq SOC^{+}_{EV,i}(t) \\ P^{ch}_{EV,i}(t) & , 0 \leq t < T_{leave,i} \text{ and } SOC_{EV,i}(t) < SOC^{+}_{EV,i}(t) \end{cases} \tag{7}$$

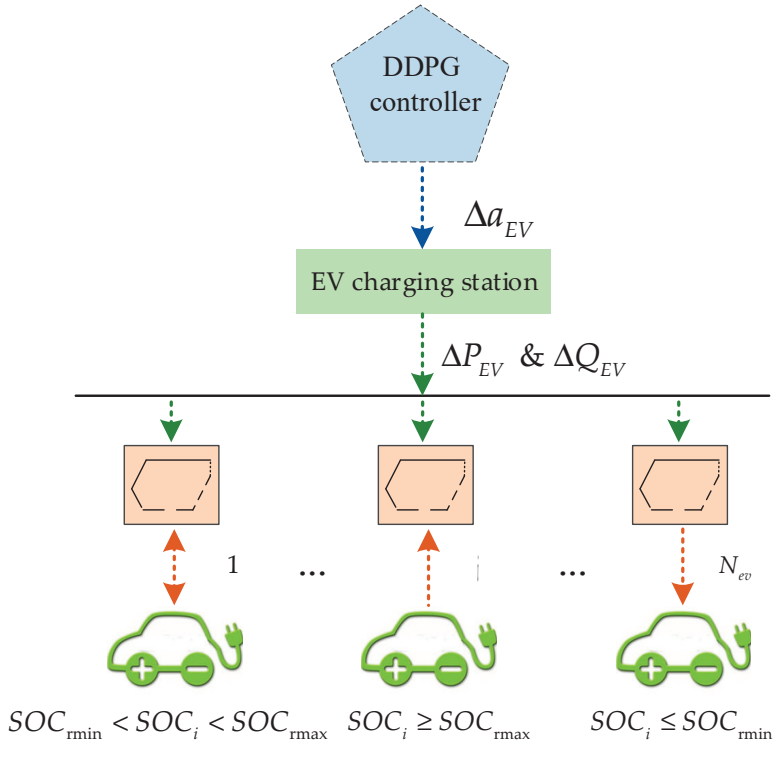


Figure 2. The distribution of controller commands in the charging station.

The charging and discharging constraint boundary of the cluster EVs' P_{EV} can be obtained from the boundary of a single EV as follows:

$$\begin{cases} P^{-}_{EV}(t) < \Delta P_{EV}(t) < P^{+}_{EV}(t) \\ P^{+}_{EV}(t) = \sum_{i=1}^{n_{EV}} P^{+}_{EV,i}(t) \\ P^{-}_{EV}(t) = \sum_{i=1}^{n_{EV}} P^{-}_{EV,i}(t) \end{cases} \tag{8}$$

where n_{EV} is the number of EV.

In addition, the active power capacity calculation is related to the number and the *SOC* state of EV:

$$E_{ct} = \sum_{i=1}^{N_{ev}} (SOC_i \times E_i)/E_{all} \tag{9}$$

where E_i represents the active power capacity of a single EV, E_{all} represents the total active power capacity of EVs, and E_{ct} represents the real time active power capacity of the EVs station.

From this, it can be obtained that the output power ΔP_{EV} of the EV charging station during the charging and discharging process should meet the following constraints:

$$\begin{cases} 0 < \Delta P_{EV} < \lambda_e , & E_{ct} > E_{\mathrm{rmax}} \\ -\lambda_e < \Delta P_{EV} < \lambda_e , & E_{\mathrm{rmin}} < E_{ct} < E_{\mathrm{rmax}} \\ -\lambda_e < \Delta P_{EV} < 0 , & E_{ct} < E_{\mathrm{rmin}} \end{cases} \tag{10}$$

when $E_{ct} > E_{\mathrm{rmax}}$, the real time active power capacity E_{ct} of the EV station is higher than the recommended maximum capacity E_{rmax}, due to the rapid increase in the number of EVs in the charging station. When $E_{ct} < E_{\mathrm{rmin}}$, the number of EVs in the charging station is too small, or the EVs in the charging station are all in a low battery state. When $E_{\mathrm{rmin}} < E_{ct} < E_{\mathrm{rmax}}$, the EV station can either discharge to the microgrid or charge from the microgrid.

Furthermore, the capacity state E of the EVs station is related to the EVs existing in the EVs station in different *SOC* states. Therefore, by combining Equations (8) and (10), it can obtain the constraint of active output power ΔP_{EV} considering the travel demand of users, the number of electric vehicles, and the real-time *SOC* of electric vehicles as:

$$\begin{cases} 0 < \Delta P_{EV} \le P_{EV}^{+}(t) , & E_{ct} > E_{\mathrm{rmax}} \\ P_{EV}^{-}(t) \le \Delta P_{EV} \le P_{EV}^{+}(t) , & E_{\mathrm{rmin}} < E_{ct} < E_{\mathrm{rmax}} \\ P_{EV}^{-}(t) \le \Delta P_{EV} < 0 , & E_{ct} < E_{\mathrm{rmin}} \end{cases} \tag{11}$$

After obtaining the boundary of the active discharge power ΔP_{EV} of the EVs, the reactive power boundary can be obtained through the power factor angle of the charger, and the circuit topology of the four-quadrant bidirectional charger mostly uses a double-buck AC–DC half-bridge conversion circuit, a traditional AC–DC half-bridge conversion circuit, and an AC–DC full-bridge conversion circuit. The capacity curve of the charger is shown in Figure 3 [18].

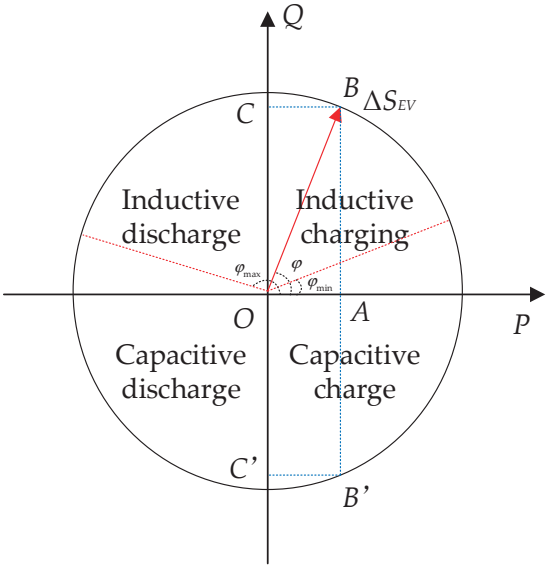


Figure 3. The capacity curve of the charger.

φ is the power factor angle when the apparent rated power is ΔS_{EV}. $\varphi_{\min}$ and $\varphi_{\max}$ are the minimum and maximum power factor angles of the charger. The positive axis of the P axis and Q axis represents the energy transferred from the grid to the EV charger. When the active power is OA, the adjustable range of reactive power is CC', and the length of OB is the apparent rated power ΔS. In addition, the relationship of the active and reactive power ΔP_{EV} and ΔQ_{EV} can be charged by Figure 3, as in the Formula (12):

$$\Delta Q_{EV} = \Delta P_{EV} \tan \varphi \tag{12}$$

$$\begin{cases} \varphi_{\min} < \varphi < \varphi_{\max} \\ -\varphi_{\min} < \varphi < -\varphi_{\max} \end{cases} \tag{13}$$

Thus, the power factor angle needs to meet the operating characteristics of the charger, and when $\Delta P_{EV} > 0$, the grid feeds active power to the EVs, when $\Delta Q_{EV} > 0$, the grid feeds reactive power to the EVs.

In summary, the boundary of the output power increment of the EV charging station is affected by the number of EV in the charging station N_{EV}, SOC state, electric vehicle charging station real time capacity E, and the angle of charging power factor.

2.2. VF Control Model of Microgrids with EVs

The output characteristics of distributed wind power and photovoltaic system are random, and load fluctuations simultaneously affect the output of active and reactive power. Therefore, in the process of microgrid VF control in this paper, the wind power and photovoltaic system are equivalent to disturbance sources [19]. In addition, the load response characteristics of wind power system and photovoltaic power system are similar, so only the microgrid load VF control under the wind power disturbance is considered, and it is applied using recorded historical data [20]. In addition, the MT is added to the microgrid system as a main control unit in this paper to ensure the flexibility and validity of microgrid regulation.

The structure of the microgrid is in Figure 4. The microgrid includes a MT, EVs, distributed wind power, and load.

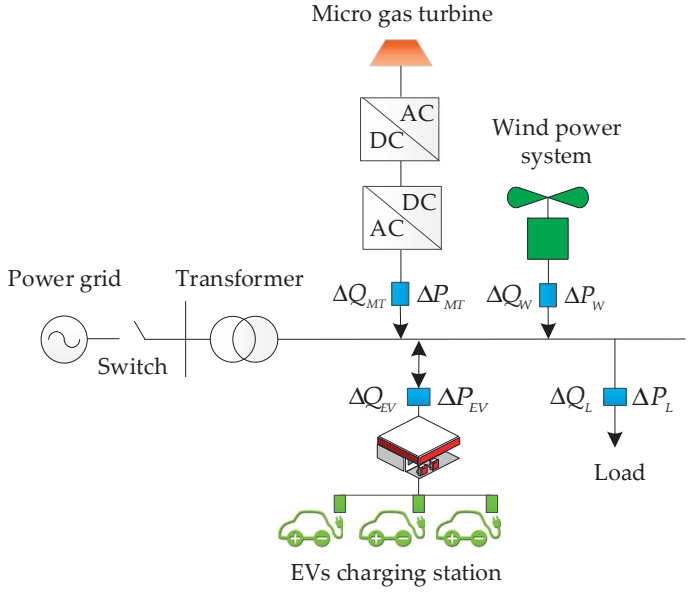


Figure 4. Structure of the islanded microgrid.

ΔP_L and ΔQ_L are the load disturbance power, ΔP_W and ΔQ_W are the wind disturbance power, ΔP_{MT} and ΔQ_{MT} are the power variation of *MT*, and ΔP_{EV} and ΔQ_{EV} are the power variation of EVs.

3. The Design of Microgrid VF Controller Based on DDPG

In the islanded microgrid, it is important to maintain the stability of VF, but there are some control problems such as various uncertainties and nonlinearities caused by DGs and EVs, which can inevitably cause the VF fluctuation and make it deviate from the reference value.

In addition, the Deep Reinforcement Learning (DRL) with online learning, experience playback capabilities and other advantages, is suitable for nonlinear systems [21]. Therefore, in this paper, a VF controller based on DDPG for islanded microgrid with EVs is designed. The frequency and voltage deviation is fed back to the DDPG controller, which adjusts the power output of each unit to ensure the stability of the frequency and voltage of the system.

3.1. Theoretical Analysis of DDPG

Q-learning and Deep Q-learning (DQN) are typical value-based reinforcement learning algorithms that use value functions to learn the optimal strategy during the interaction with the environment [22]. However, since the *Q*-learning cannot process continuous signals, it is necessary to discretize the action space. Therefore, it is difficult to realize the precise control of MT, EVs and chargers, which is not suitable for the design of this paper.

In addition, the learning of the DDPG can be carried out in a continuous action space [23]. The DDPG contains four networks, namely actor current network, actor target network, critic current network, critic target network. At t, the actor current network parameter is θ, and the actor target network parameter is θ', the critic current network parameter is ω, the critic target network parameter is ω'.

In the above four networks, the actor current network can generate action a_t according to the current status s_t. The actor target network can generate the action a_{t+1} at the $t+1$ time according to the subsequent state of the environment. The critic current network can calculate the value R_t corresponding to the status s_t and action a_t. The Critic target network can generate the value of ${Q_{value}}'(s_{t+1}, a_{t+1} \mid \omega')$ based on subsequent state s_{t+1} and action a_{t+1}, which is used to calculate the target value y, as shown in the Formula (14):

$$y = r_t + \gamma {Q_{value}}'(s_{t+1}, a_{t+1}, \omega') \tag{14}$$

where γ is a discount factor and $0 < \gamma < 1$, ${Q_{value}}'(s_{t+1}, a_{t+1} \mid \omega')$ is the value generated by subsequent state s_{t+1} and action a_{t+1}, which is used to calculate the target value y.

Meanwhile, the critic current network parameter ω is updated by the gratial direction of the neural network using a mean square difference loss functional Formula (15). In addition, the parameter of the actor current network θ is updated through the gradient of the neural network, as shown in Formula (16):

$$L = \frac{1}{m}\sum_{j=1}^{m}\left(y_j - Q(s_j, A_j, \omega)\right)^2 \tag{15}$$

$$\nabla J(\theta) = \frac{1}{m}\sum_{j=1}^{m}\left[\nabla_a Q(s, a, \omega)\Big|_{s=s_j, a=\pi_\theta(s)} \nabla_\theta \pi_\theta(s)\Big|_{s=s_j}\right] \tag{16}$$

where m is the number of samples, y_j is the target value of the j sample, $Q(s_j, a_j, \omega)$ is the output value of the critic current network for the j sample, and $\pi_\theta(\cdot)$ is the output value of the actor current network.

Furthermore, it is necessary to update the critic target network and actor target network parameters by Equation (17):

$$\begin{aligned} \omega' &\leftarrow \tau\omega + (1-\tau)\omega' \\ \theta' &\leftarrow \tau\theta + (1-\tau)\theta' \end{aligned} \tag{17}$$

where τ is an update coefficient, which is generally small.

In addition, the E is a termination function, which is to determine whether the Agent enters the termination. If the Agent enters the termination state, the iterative process stops and a new round of state sequence starts. If the Agent enters the non-termination state, the iterative process of the wheel can be continued.

In summary, status information, reward value, action information, and termination status information $\{s, a, R, s', E\}$ are formed into a sample unit and stored in the empirical playback set D. Then, m sample units of set D are taken to be trained by Formulas (14)–(17). A total of T rounds is trained, and the training step length of each round is T_{m}. The specific training process is shown in Figure 5.

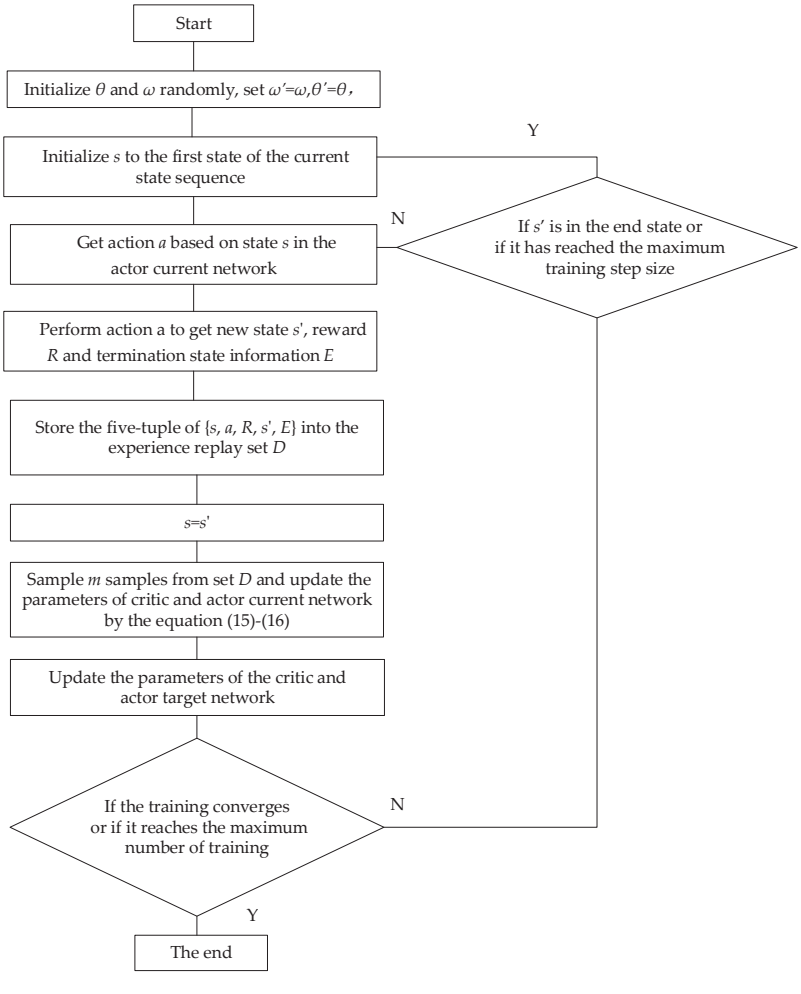


Figure 5. Structure of the islandeded microgrid.

3.2. Design of DDPG VF Controller Structure

Considering MT and EV output power increment limiting constraints, a VF controller structure based on DDPG is proposed, as shown in Figure 6. The controller is composed of two layers: coordinate layer and control layer. The coordinate layer provides real-time regulation signal ΔA to the control layer according to the frequency deviation Δf, voltage deviation ΔU, and the real-time boundary of output power of EV charging station, and then controls the output power of MT and EV to quickly suppress the frequency and voltage deviation.

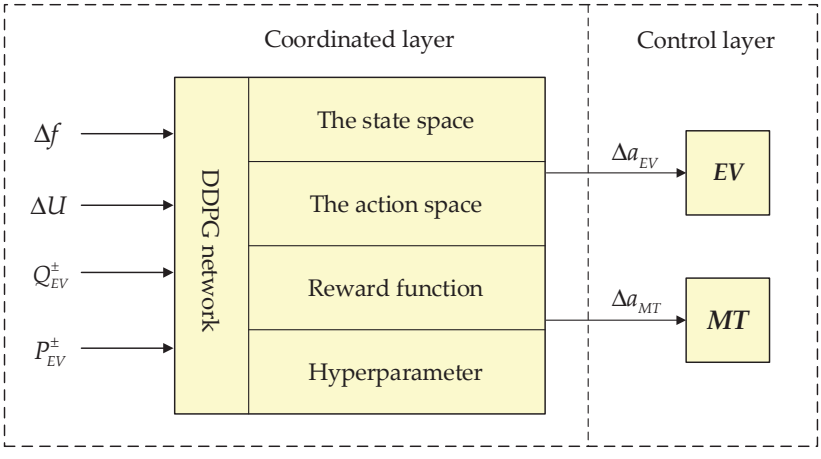


Figure 6. Microgrid LFC controller structure based on DDPG.

3.3. Definition of Space and Reward Function

As mentioned above, the state set of the control system is frequency deviation $\Delta F(t)$, voltage deviation $\Delta U(t)$, and the real-time boundary of output power of EV charging station $P_{EV}^{\pm}(t)$ and $Q_{EV}^{\pm}(t)$, so the state space S can be defined as follows:

$$S = [\Delta F(t), \Delta U(t),\ P_{EV}^{+}(t),\ P_{EV}^{-}(t), Q_{EV}^{+}(t), Q_{EV}^{-}(t)] \tag{18}$$

In addition, the joint action set A of the DDPG controller, namely the output of the controller, should be a real-time set of dispatch instruction of the active and reactive power output of MT, the output active power of EVs, and the power factor angle of the charger. Thus, the action space A can be defined as follows:

$$A = [\Delta A_{P,\mathrm{MT}}(t), \Delta A_{P,EV}(t), \Delta A_{Q,MT}(t), \Delta A_{\varphi,EV}(t)] \tag{19}$$

In addition, then, China's power safety work principle stipulates that the frequency of the power system during normal operation should be within the range of 50 ± 0.2 Hz, and the voltage deviation should within 5%. Thus, on this basis, a certain adjustment dead zone is considered, the discrete set of real-time frequency deviation ΔF(t) can be set as $(-\infty, -0.2)$, $[-0.2, -0.15)$, $[-0.15, -0.10)$, $[-0.10, -0.03)$, $[-0.03, 0.03]$, $(0.03, 0.10]$, $(0.10, 0.15]$, $(0.15, 0.2]$, $(0.2, +\infty)$, unit of Hz, and the discrete set of real-time voltage deviation $\Delta U(t)$ can be set as $(-1,-0.05)$, $[-0.05, -0.03)$, $[-0.03, -0.02)$, $[-0.02, -0.01)$, $[-0.01, 0.01]$, $(0.01, 0.02]$, $(0.02, 0.03]$, $(0.03, 0.05]$, $(0.05,1)$, unit of p.u.

Meanwhile, the control objectives in this paper are: ①Restore the frequency to the rated value; ②Regulate and control the voltage to restore to the best state. As a result,

a comprehensive reward function including two local reward functions can be set up to coordinate frequency recovery and voltage adjustment:

$$R = r_f + r_u \tag{20}$$

$$r_f = \begin{cases} 0 & |\Delta f| < 0.03 \\ -\mu_1|\Delta f| & 0.03 \le |\Delta f| < 0.10 \\ -\mu_2|\Delta f| & 0.10 \le |\Delta f| < 0.15 \\ -\mu_3|\Delta f| & 0.15 \le |\Delta f| < 0.2 \\ -\mu_4|\Delta f| & 0.2 \le |\Delta f| \end{cases} \tag{21}$$

$$r_u = \begin{cases} 0 & |\Delta U| < 0.01 \\ -\delta_1|\Delta U| & 0.01 \le |\Delta U| < 0.02 \\ -\delta_2|\Delta U| & 0.02 \le |\Delta U| < 0.03 \\ -\delta_3|\Delta U| & 0.03 \le |\Delta U| < 0.05 \\ -\delta_4|\Delta U| & 0.05 \le |\Delta U| < 1 \end{cases} \tag{22}$$

where R is the global reward, r_f is the frequency reward, r_u is the voltage reward, μ_1, μ_2, μ_3 and μ_4 are the weights corresponding to the reward function of each control region in the frequency penalty item r_f, and δ_1, δ_2, δ_3 and δ_4 and are the weights corresponding to the voltage control regions.

The control process needs to control the frequency through r_f, when $|\Delta f|$ is in adjusting dead zone $[-0.05, 0.05]$ Hz, and the frequency meets the minimum error requirement of normal operation, so the maximum reward value given to the DDPG controller at this time is 0. When $|\Delta f|$ is respectively in normal control (0.05, 0.10) and (0.10, 0.15) Hz, auxiliary control area (0.15, 0.2) Hz, emergency control area (0.2, $+\infty$) Hz, the controller can get the corresponding negative incentives, namely the penalty value. Meanwhile, when voltage control is performed, the voltage needs to be regulated by r_u, when $|\Delta U|$ is in adjusting dead zone $[-0.01, 0.01]$, the maximum reward value given to the DDPG controller at this time is 0, and when $|\Delta U|$ is respectively in normal control (0.01, 0.02) and (0.02, 0.03), auxiliary control area (0.03, 0.05), emergency control area (0.05, 1), the controller can get the corresponding penalty value.

When determining the values of the above parameters, it should be noted that the size of the reward value can affect the convergence effect and the learning speed. Therefore, it is necessary to perform simulation tests based on actual calculation examples, and the specific process will be discussed later.

In summary, the state space and reward function designed in this paper can realize the simultaneous adjustment of voltage and frequency. When the frequency is restored, it can consider whether the voltage exceeds the limit, and, when adjusting the voltage, it can also consider whether the frequency deviates from the rated value, which significantly improves the overall stability of the microgrid.

3.4. The Selection of Hyperparameter

In DRL, it is necessary to provide the agent with a set of optimal hyperparameters to improve the performance and effect of learning [24].

First of all, the larger the discount factor γ, the more the agent attaches importance to past experience and can give up current interests and pursue overall interests. However, if γ is too large, it will also cause the training of agent to fail to converge. The greater the learning rate α, the faster the agent converges, but the worse the stability; the smaller the α, the better the stability, but the slower the agent converges. Therefore, the convergence speed should be improved on the premise when the agent training can converge. In addition, the design of network structure can be discussed from two aspects: network type and network depth. The choice of network type depends largely on the state space, and the state space of the control system in this paper is frequency and voltage deviation, which belong to

one-dimensional vector, so the full connection layer can better meet the requirements of the storage strategy set. In addition, the network depth determines the generalization ability of the neural network, which includes the number of layers of the neural network h and the neurons in each layer u.

In addition, the specific values of γ, α, h and u need to be selected according to the calculation example.

3.5. Summary of Control Strategy

In summary, the control strategy of this paper is carried out in the following steps:

1. Firstly, definite the state set of the control system as $\Delta F(t)$, $\Delta U(t)$, $P_{EV}^{\pm}(t)$ and $Q_{EV}^{\pm}(t)$. In addition, the action space can be defined as $\Delta A_{P,MT}$ (t), $\Delta A_{P,EV}$ (t), $\Delta A_{Q,MT}$ (t), $\Delta A_{\varphi,EV}$ (t).
2. Secondly, the parameters are adjusted according to the actual calculation example, and the values of the reward function coefficients and hyperparameters are obtained.
3. Thirdly, perform agent training according to the process in Figure 5, and obtain the optimal value function Q network $Q_{\varphi(s,a)}$.
4. Finally, in different cases, input disturbances to the islanded microgrid system, and the agent can generate corresponding actions based on the disturbances to adjust the output of each unit, so as to ensure the frequency and voltage balance of the islanded microgrid system.

4. Simulation Results

In order to evaluate the control effect of the above strategy, the coupled islanded microgrid system is built as shown in Figure 7. In addition, the specific settings of equipment parameters are shown in Table 1. The verification of the calculation examples in this paper is carried out through simulation experiments. The computing platform is a PC with i7-1165G7@2.80GHz CPU and 16 GB RAM, and the software environment is Windows 10 Professional and MATLAB R2021a.

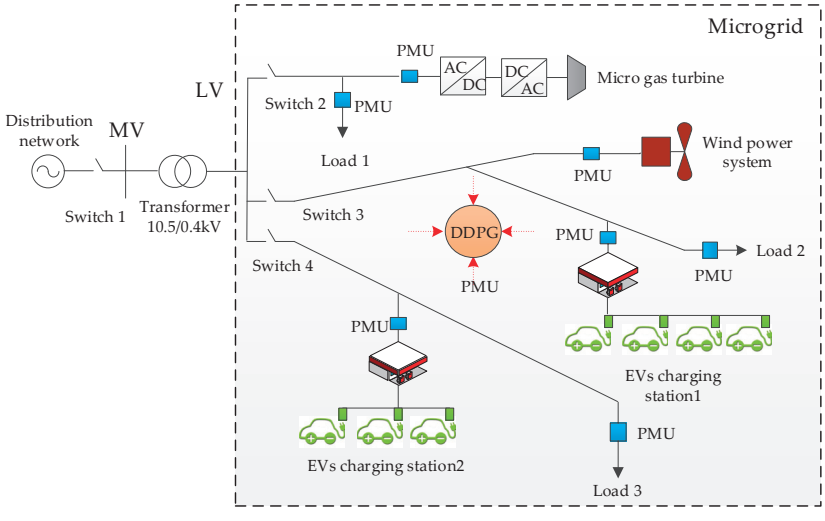


Figure 7. Microgrid LFC controller structure based on DDPG.

In the microgrid, there is a MT with capacity of 40 kW, a WT with capacity of 20 kW, an EV station1 with capacity of 16 kW, an EV station2 with capacity of 14 kW, and 60 kW ordinary loads. In addition, this paper assumes that the initial state of the microgrid is stable. Thus, when there is no external disturbance, the power output of MT, EV stations,

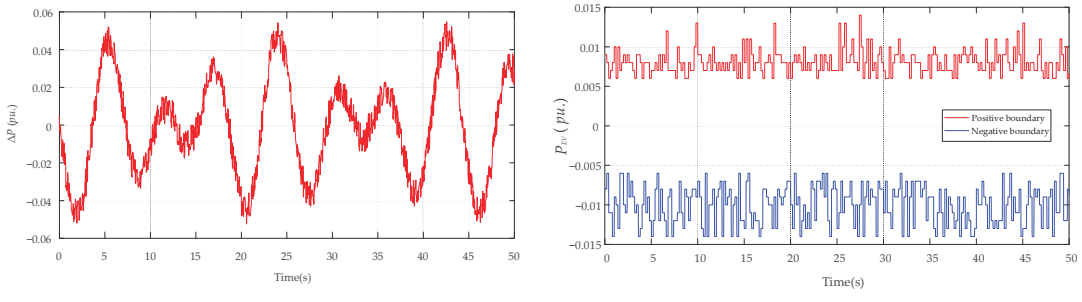
WT, and conventional loads are always in balance. Therefore, in the following calculation examples, only the per-unit value of the power fluctuations of MT, EVs stations, WT, and load need to be considered.

Table 1. Parameters of equipment in microgrids.

Unit	Parameter	Meaning	Value
MT	T_f	time constant of governor	10 s
	T_t	time constant of generator	0.1 s
	R_f	speed regulation factor	0.005 Hz/p.u.
	λ_{Pmtd}	lower limit of active power variation	−0.025 p.u.
	λ_{Pmtp}	upper limit of active power variation	0.025 p.u.
	λ_{Qmtd}	lower limit of reactive power variation	−0.025 p.u.
	λ_{Qmtp}	upper limit of reactive power variation	0.03 p.u.
EV1	T_{e1}	time constant of EV	1 s
	λ_{Ped1}	lower limit of active power variation	−0.016 p.u.
	λ_{Pep1}	upper limit of active power variation	0.016 p.u.
	λ_{Qed1}	lower limit of reactive power variation	−0.015 p.u.
	λ_{Qep1}	upper limit of reactive power variation	0.015 p.u.
	n_{EV1}	Initial number of EVs in station1	40
EV2	T_{e2}	time constant of EV	1.5 s
	λ_{Ped2}	lower limit of active power variation	−0.014 p.u.
	λ_{Pep2}	upper limit of active power variation	0.014 p.u.
	λ_{Qed2}	lower limit of reactive power variation	−0.0135 p.u.
	λ_{Qep2}	upper limit of reactive power variation	0.0135 p.u.
	n_{EV2}	Initial number of EVs in station2	35

4.1. Pre-Learning Stage

Before the controller is used, it needs to undergo a random trial and error learning process, which is called the pre-learning stage. In the initial stage of pre-learning, the controller has not accumulated any experience and has no intelligent control ability [25]. Only after accepting various state actions can the optimal value function Q network $Q_{\varphi(s,a)}$. Therefore, the wind and load disturbances superimposed by various different amplitudes and different types of functions are set up for repeated training of the controller. Meanwhile, according to the output capacity change data of the electric vehicle charging station, a boundary function of the output power increment that changes randomly over time is set. Take active power disturbance and the output boundary of the active power of EVs as examples. The random disturbance of a certain training process is shown in Figure 8.



(**a**) (**b**)

Figure 8. Random perturbation function in the pre-learning phase: (**a**) Random Function of Active Power Disturbance; (**b**) Random Function of EV Output Power Boundary.

Meanwhile, through a large number of simulation studies, μ_1, μ_2, μ_3, and μ_4 are referred as 1, 5, 10, and 20, respectively, δ_1, δ_2, δ_3 and δ_4 are referred as 5, 20, 50, and

100 respectively, and α and γ are referred as 0.01, 0.09. Meanwhile, the number of learning iterations of the DDPG controller is set to 500, each with 500 steps, and the step length is 0.1 s. Therefore, six groups of parameters (h, u) are set for the convergence test, and the learning results are shown in Table 2. It can be seen that the reward value of the system at convergence is the highest when $h = 5$ and $u = 50$.

Table 2. Convergence test results under different parameters.

SN	Parameter Settings	Average Reward	Final Award
1	$h = 3$, $u = 50$	−34.037	−1.83622
2	$h = 3$, $u = 200$	−26.673	−1.20923
3	$h = 5$, $u = 50$	−21.096	−0.65307
4	$h = 5$, $u = 200$	−27.075	−1.35723
5	$h = 10$, $u = 50$	−34.572	−2.54778
6	$h = 10$, $u = 200$	−40.922	−3.04643

Thus, when $h = 5$ and $u = 50$, the pre-learning process of the agent is shown in Figure 9.

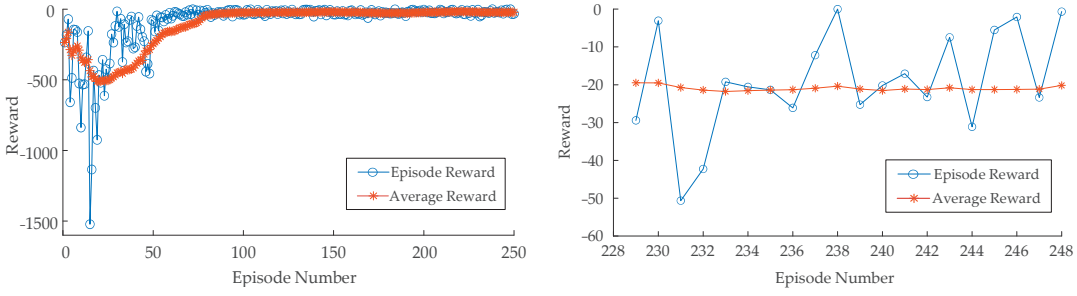


Figure 9. The complete trend graph of the reward function: (**a**) The complete trend graph of the reward function; (**b**) the trend graph of the reward function for the last 50 iterations.

It can be seen that the agent basically converges after 80 iterations, and the system judges that the learning process has been completed and stops the training after 248 iterations. In this case, the average reward is −21.096 and the final award is 0.65307, which shows that the controller can complete the subsequent simulation at this time.

4.2. The Implementation of Constraint Conditions in the EV Model

In order to verify the implementation of constraint conditions in the EV model, this paper selects several typical monomer EV *SOC* simulation situations as examples, as shown in Figures 10 and 11. In addition, to ensure the life of battery, the initial *SOC* is set between $SOC_{\mathrm{rmin}} = 0.2$ and $SOC_{\mathrm{rmax}} = 0.8$.

The first situation in Figure 10 shows that, when $SOC < SOC_{\mathrm{rmin}}$, the EV will be forced to enter the charging state. Only when $SOC > SOC_{\mathrm{rmin}}$ can the EV participate in system regulation. The second situation in Figure 10 shows that, when the EV is close to the leaving time and $SOC < SOC_m$, it will turn to the forced charging state to ensure that the *SOC* reaches the expected SOC_m when leaving the charging station. In general, the changes in the *SOC* of EVs participating in the regulation of the microgrid are shown in Figure 11. The *SOC* of EVs will change in the constraint range.

4.3. Case Study

After completing the pre-learning phase and the verification of the EV *SOC* constraints, the example can be simulated under different operation scenarios. Meanwhile, in order to

evaluate the effect of DDPG controller proposed in this paper, traditional PID controller and $R(\lambda)$ controller are used in the same scene respectively, and the corresponding controller parameters are shown in Table 3.

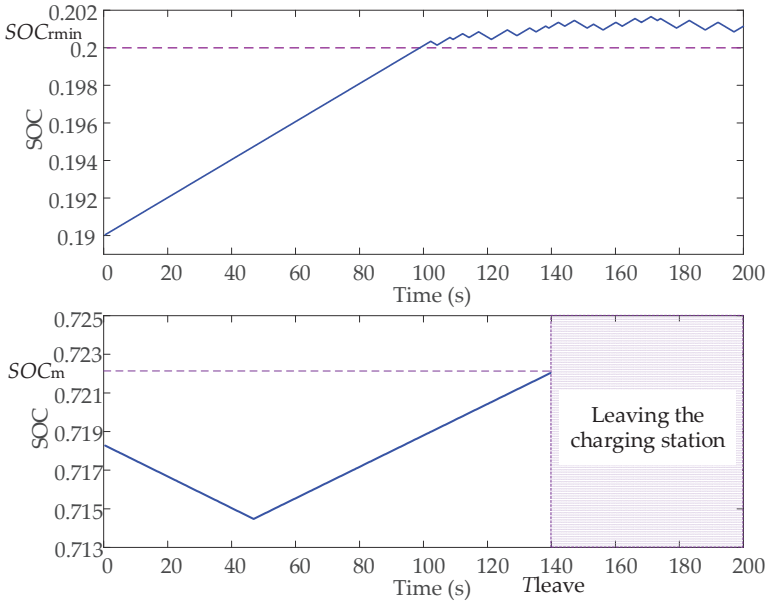


Figure 10. Changes of *SOC* of EVs at critical value.

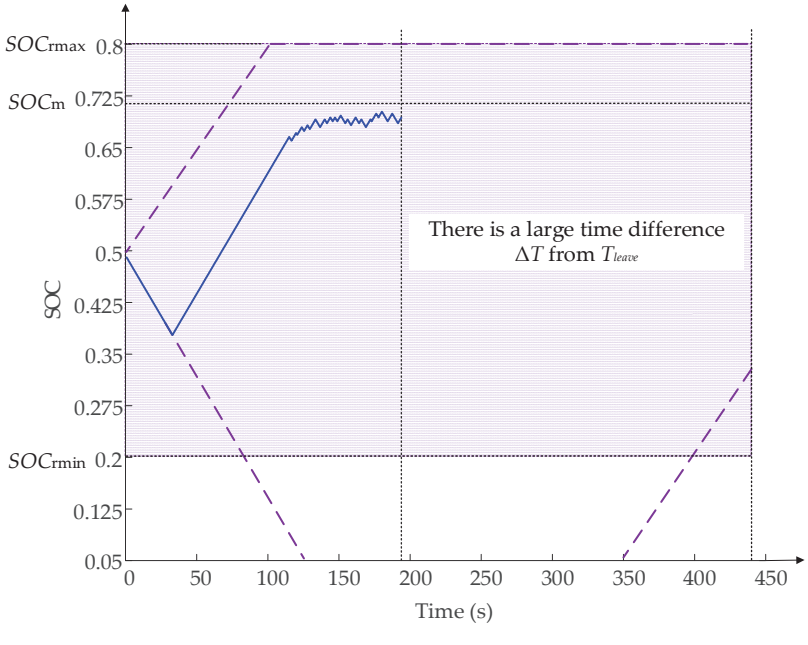


Figure 11. Changes of *SOC* of EVs in the normal range.

Table 3. PID and R(λ) controller system parameters.

Controller	Describe	Parameter	Value
PID	Proportional gain	K_P	4
	Integral gain	K_i	1.18
	Differential gain	K_D	0.5
R(λ)	learning rate	α	0.01
	discount factor	γ	0.9
	network depth	(h, u)	(3, 10)

4.3.1. Case 1: The Response of Wind Power Disturbance

First of all, wind power disturbance is added to the islanded microgrid system, and wind mainly provides active power disturbances to the grid. In order to compare the adjusting speed of each controller, the wind power disturbance ends after 43 s. The disturbance setting is shown in Figure 12.

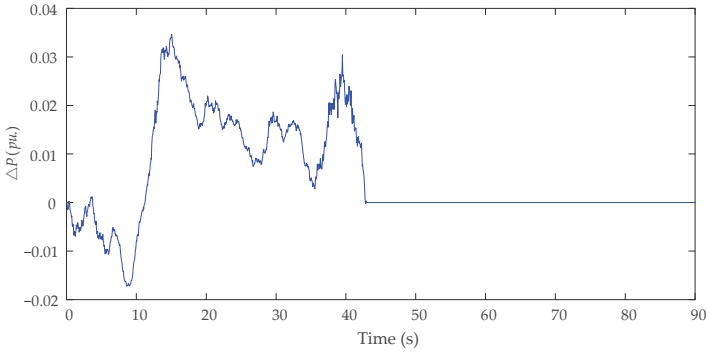


Figure 12. Wind power disturbance.

There is not the fluctuation of reactive power in this case, so the impact of voltage fluctuation is not considered here. The variation of frequency deviation under wind power disturbance is shown in Figure 13. Meanwhile, according to the simulation results, this paper takes the absolute value of $|\Delta f|$ as the evaluation object, and sets the threshold of the frequency deviation excellence rate to 2×10^{-4} Hz, and defines T_{recover} as the time which is taken for $|\Delta f|$ to recover to 5×10^{-5} Hz after the wind power disturbance ends. The results of the control test under wind disturbance are shown in Table 4.

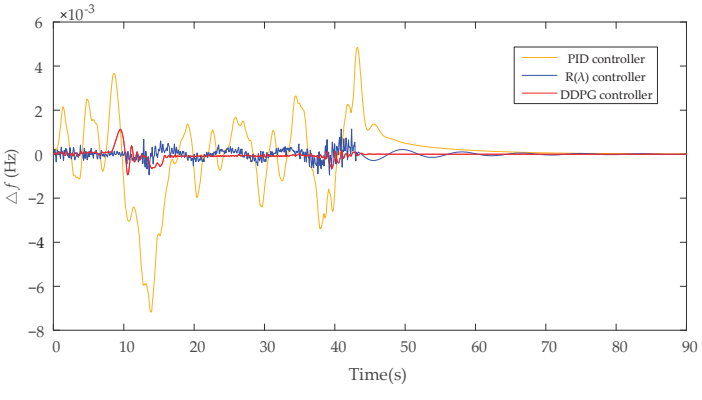


Figure 13. Performance of frequency control under wind power disturbance.

Table 4. Frequency simulation results under wind disturbance.

Indicators	PID	R(λ)	DDPG
Average (Hz)	0.9215×10^{-3}	0.1159×10^{-3}	0.0267×10^{-3}
Maximum (Hz)	7.174×10^{-3}	1.148×10^{-3}	1.133×10^{-3}
Proficiency (%)	30.56%	92.07%	98.35%
T_{recover} (s)	34.25	23.95	0.75

It can be seen from Figure 13 and Table 4 that, compared with the PID controller, the DDPG and $R(\lambda)$ controller with the ability of online learning and experience playback can more effectively deal with the highly random disturbance. Under the wind disturbance, the frequency fluctuation of the islanded microgrid under the DDPG controller can be limited in 2×10^{-4} Hz, and the excellent rate can reach 98%, which is significantly better than the traditional controller. In addition, if only analyzed from the perspective of frequency control, the control strategy of DDPG and $R(\lambda)$ controller in this paper possesses virtues of great control effect, smaller amplitude of frequency fluctuation, and faster regulation speed than a traditional controller. Furthermore, the regulation speed of DDPG controller is much faster than a R(λ) controller.

Furthermore, the power variations of each equipment in islanded microgrid under the DDPG controller are shown in Figure 14. It can be seen that, when the system suffers disturbance, the MT undertakes the main work of frequency regulation, and the output power of EV charging station is also significant. In addition, when the limit is reached, the power variations of different charging stations are different.

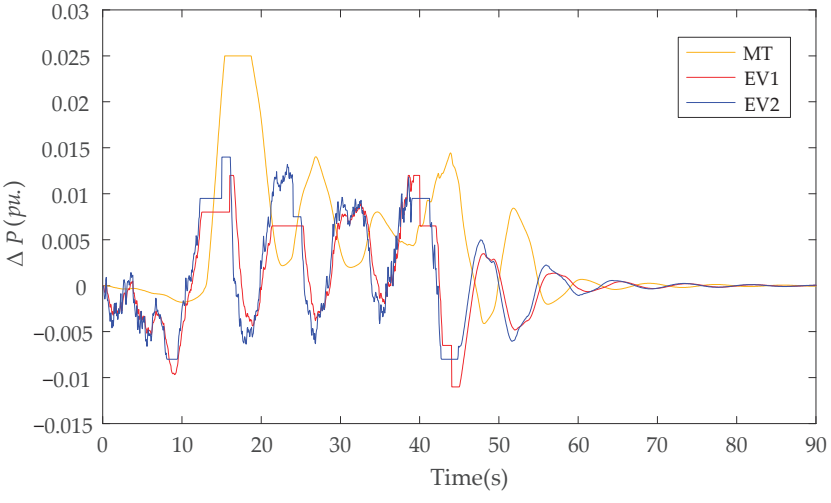


Figure 14. Power variations of each equipment under wind power disturbance.

4.3.2. Case 2: The Response to Load Power Disturbance

The fluctuation of load power is gentler than that of wind power, but the load change is abrupt and can cause fluctuations in active and reactive power at the same time. In this case, load power variations are set as $\Delta P_L = -0.025$ p.u during 10–40 s, $\Delta P_L = 0.005$ p.u during 40–55 s, $\Delta P_L = -0.0025$ p.u during 55–70 s, $\Delta P_L = 0.015$ p.u during 70–150 s, $\Delta Q_L = -0.04$ p.u during 7.5–34 s, $\Delta Q_L = -0.015$ p.u during 34–66 s, $\Delta Q_L = 0.005$ p.u during 66–96 s, $\Delta Q_L = -0.0075$ p.u during 96–150 s. The specific setting of load disturbance is shown in Figure 15.

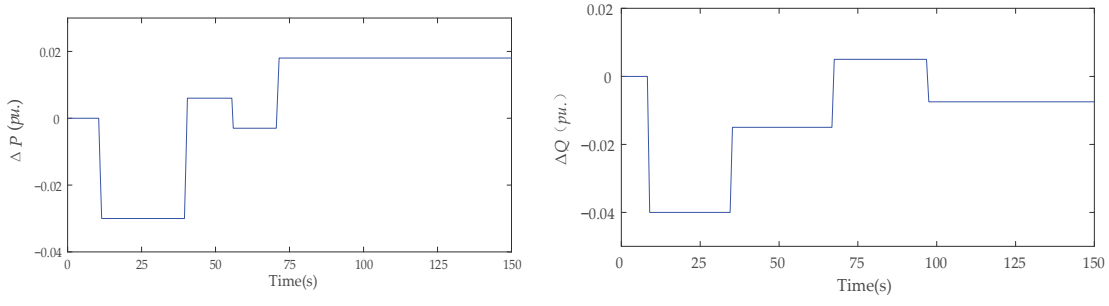


Figure 15. Load power disturbance: (**a**) load active power disturbance; (**b**) load reactive power disturbance.

The DDPG controller is compared with traditional PID and $R(\lambda)$ controller, and the frequency and voltage fluctuation are shown in Figures 16 and 17. The same as the case 1, this part takes $|\Delta f|$ and $|\Delta U|$ as the evaluation object, and sets the threshold of the $|\Delta f|$ excellence rate to 2×10^{-4} Hz, the $|\Delta U|$ excellence rate to 0.01 p.u. Meanwhile, T_{recover} is defined as the time which is taken for $|\Delta f|$ to recover to 5×10^{-5} Hz and $|\Delta U|$ to recover to 0.002 p.u after the load power disturbance no longer changes. Thus, the statistical results of the control test under load disturbance are shown in Tables 5 and 6.

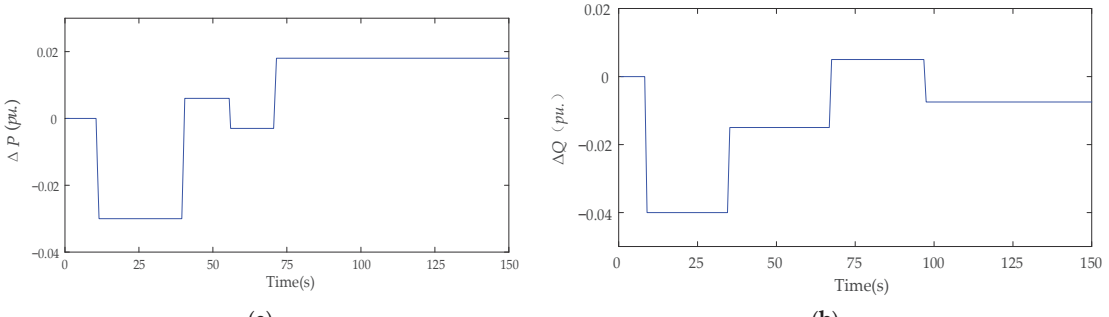


Figure 16. Performance of frequency control under load power disturbance.

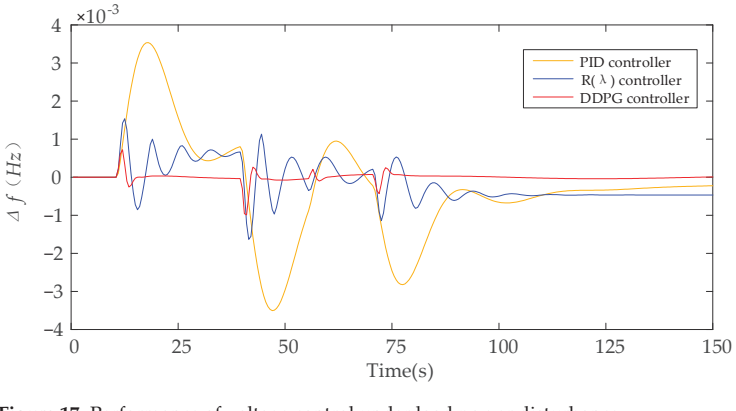


Figure 17. Performance of voltage control under load power disturbance.

Table 5. Frequency simulation results under load disturbance.

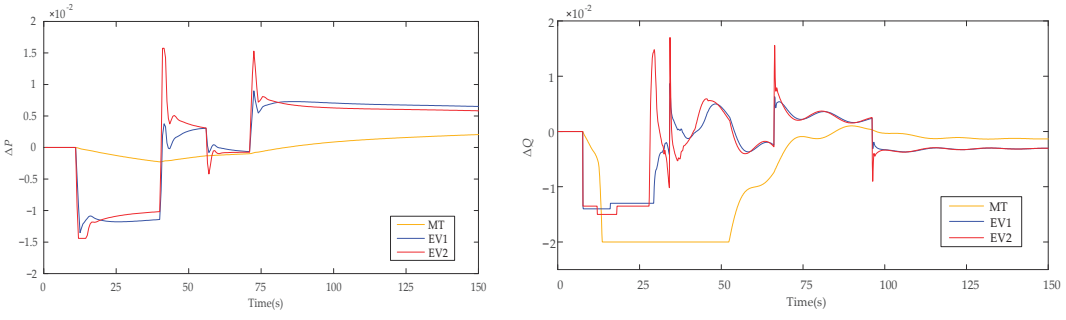
Indicators	PID	R(λ)	DDPG
Average (Hz)	0.995×10^{-3}	0.437×10^{-3}	0.0521×10^{-3}
Maximum (Hz)	3.54×10^{-3}	1.63×10^{-3}	1.09×10^{-3}
Proficiency (%)	9.6%	16.7%	98%
T_{recover} (s)	/	/	0.13

Table 6. Voltage simulation results under load disturbance.

Indicators	PID	R(λ)	DDPG
Average (p.u)	0.0093	0.0021	0.00047
Maximum (p.u)	0.0569	0.0112	0.0023
Proficiency (%)	30.3%	83.07%	100%
T_{recover} (s)	/	/	9.2

It can be seen from Figures 16 and 17 and Tables 5 and 6 that, when the load changes, compared with the PI controller and R(λ) controller, the DDPG controller can ensure that the frequency deviation of the microgrid is maintained within $\pm 1 \times 10^{-3}$ Hz Hz, and the voltage deviation is also close to 0, which is much smaller than the control index of the power quality of the power grid. In addition, compared with the $R(\lambda)$ controller, the DDPG controller can coordinate the frequency recovery and voltage adjustment of the islanded microgrid, so as to meet the VF control requirements at the same time, which has superior dynamic control characteristics.

Furthermore, the power variations of each equipment are shown in Figure 18. The MT in the micro grid is used as the main source to maintain the stability of the VF amplitude of the microgrid, while the EV_1 and EV_2 as the slave sources are mainly responsible for the regulation of the active power of the microgrid and also participate in the regulation of the reactive power. In addition, due to the randomness of users, the output power boundary of EV charging stations is random, showing obvious jagged shapes.



(**a**) (**b**)

Figure 18. Power variations of each equipment under load power disturbance: (**a**) active power increment; (**b**) reactive power increment.

5. Conclusions

To solve the problem in which the stability of island microgrid is greatly affected by random power sources, and it is difficult to control frequency and voltage together, a VF control strategy of islanded microgrids with EVs is proposed in this paper. The randomness of charging behavior is considered, and an islanded microgrid system including MT, WT,

EVs stations, and loads is established. Thus, a VF synergistic control strategy based on DDPG is proposed. The simulation results show that:

1. Compared with PID controller, the DDPG controller with the ability of online learning and experience playback can more effectively deal with the highly random disturbance. Under the wind disturbance, the frequency fluctuation of the islanded microgrid under the DDPG controller can be limited in 2×10^{-4} Hz, and the excellent rate can reach 98%, which is significantly better than the traditional controller.
2. Compared with the $R(\lambda)$ controller, the DDPG controller in this paper can coordinate the frequency recovery and voltage adjustment of the island microgrid, so as to meet the VF control requirements at the same time, which is more suitable for the stable control of the microgrid. When the load changes, the DDPG controller can ensure that the frequency deviation of the microgrid is maintained within $\pm 1 \times 10^{-3}$ Hz, and the voltage deviation is also close to 0.
3. The EV charging station has the characteristics of small inertia and fast regulation speed in the microgrid control, which can play an important role in VF regulation;
4. The realization effect of the constraint conditions in the EV model is great. The single EV can judge whether it participates in the adjustment of the microgrid system according to the *SOC* situation.

For microgrid systems with more complex structures and larger volumes, it is necessary to consider the multi-microgrid interconnection technology. In addition, multi-agent algorithms such as MA-DDPG, COMA, CommNet, etc. will also be applied to the control of multi-microgrid. The follow-up work will focus on in-depth analysis and research in these directions, and add corresponding hardware circuit experiments or semi-physical simulation experiments.

Author Contributions: P.F., S.K. (Song Ke) and S.K. (Salah Kamel) conceptualized the idea of this research, P.F.; performed the experiments and data analysis, P.F. and S.K. (Song Ke) wrote the paper; J.Y., Y.L., J.X., B.X. and G.I.R. provided supervision and reviewed the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the science and technology project of the State Grid Corporation of China Research and the application of flexible control technology for a distribution system with large-scale distributed generation and a multi microgrid (No. 52093220000H).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Lam, Q.L.; Bratcu, A.I.; Riu, D.; Boudinet, C.; Labonne, A.; Thomas, M. Primary frequency H_∞ control in stand-alone mi-crogrids with storage units: A robustness analysis confirmed by real-time experiments. *Int. J. Electr. Power Energy Syst.* **2020**, *115*, 105507.1–105507.13. [CrossRef]
2. Chae, S.; Kim, G.; Choi, Y.-J.; Kim, E.-H. Design of Isolated Microgrid System Considering Controllable EV Charging Demand. *Sustainability* **2020**, *12*, 9746. [CrossRef]
3. Ziras, C.; Prostejovsky, A.M.; Bindner, H.W.; Marinelli, M. Decentralized and discretized control for storage systems offering primary frequency control. *Electr. Power Syst. Res.* **2019**, *177*, 106000.1–106000.10. [CrossRef]
4. Roudbari, E.S.; Beheshti, M.T.H.; Rakhtala, M. Voltage and frequency regulation in islandeded microgrid with PEM fuel cell based on a fuzzy logic voltage control and adaptive droop control. *IET Power Electron.* **2019**, *13*, 78–85. [CrossRef]
5. Joung, K.W.; Kim, T.; Park, J.W. Decoupled frequency and voltage control for stand-alone microgrid with high renewable penetration. In Proceedings of the IEEE/IAS Industrial & Commercial Power Systems Technical Conference IEEE, Niagara Falls, ON, Canada, 7–10 May 2018.
6. Dhua, R.; Goswami, S.K.; Chatterjee, D. An Optimized Frequency and Voltage Control Scheme for Distributed Generation Units of an Islandeded Microgrid. In Proceedings of the 2021 Innovations in Energy Management and Renewable Resources (IEMRE), Kolkata, India, 5–7 February 2021.
7. Keypour, R.; Adineh, B.; Khooban, M.H.; Blaabjerg, F. A New Population-Based Optimization Method for Online Minimiza-tion of Voltage Harmonics in Islandeded Microgrids. *IEEE Trans. Circuits Syst. II: Express Briefs* **2019**, *67*, 1084–1088. [CrossRef]

8. Wang, C.; Mei, S.; Dong, Q.; Chen, R.; Zhu, B. Coordinated Load Shedding Control Scheme for Recovering Frequency in Is-landeded Microgrids. *IEEE Access* **2020**, *8*, 215388–215398. [CrossRef]
9. Iqbal, S.; Xin, A.; Jan, M.U.; Salman, S.; Abdelbaky, M.A. V2G Strategy for Primary Frequency Control of an Industrial Mi-crogrid Considering the Charging Station Operator. *Electronics* **2020**, *9*, 549. [CrossRef]
10. Fan, H.; Lin, J.; Zhang, C.K.; Mao, C. Frequency regulation of multi-area power systems with plug-in electric vehicles con-sidering communication delays. *IET Gener. Transm. Distrib.* **2016**, *10*, 3481–3491. [CrossRef]
11. Yang, J.; Zeng, Z.; Tang, Y.; He, H.; Wu, Y. Load Frequency Control in Isolated Micro-Grids with Electrical Vehicles Based on Multivariable Generalized Predictive Theory. *Energies* **2015**, *8*, 2145–2164. [CrossRef]
12. Rao, Y.; Yang, J.; Xiao, J.; Xu, B.; Liu, W.; Li, Y. A frequency control strategy for multi-microgrids with V2G based on the im-proved robust model predictive control. *Energy* **2021**, *222*, 119963.1–119963.13. [CrossRef]
13. Li, P.; Hu, W.; Xu, X.; Huang, Q.; Liu, Z.; Chen, Z. A frequency control strategy of electric vehicles in microgrid using virtual synchronous generator control. *Energy* **2019**, *189*, 116389. [CrossRef]
14. Kisacikoglu, M.C.; Ozpineci, B.; Tolbert, L. EV/PHEV Bidirectional Charger Assessment for V2G Reactive Power Operation. *IEEE Trans. Power Electron.* **2013**, *28*, 5717–5727. [CrossRef]
15. Cao, Y.; Tang, S.; Li, C.; Zhang, P.; Tan, Y.; Zhang, Z.; Li, J. An Optimized EV Charging Model Considering TOU Price and SOC Curve. *IEEE Trans. Smart Grid* **2011**, *3*, 388–393. [CrossRef]
16. U.S. U.S. Department of Transportation. Federal Highway Administration, 2017 National Household Travel Survey. Available online: http://nhts.ornl.gov (accessed on 20 November 2021).
17. Gjelaj, M.; Hashemi, S.; Andersen, P.B.; Traeholt, C. Optimal infrastructure planning for EV fast-charging stations based on prediction of user behaviour. *IET Electr. Syst. Transp.* **2020**, *10*, 1–12. [CrossRef]
18. Mojdehi, M.N.; Ghosh, P. An On-Demand Compensation Function for an EV as a Reactive Power Service Provider. *IEEE Trans. Veh. Technol.* **2015**, *65*, 4572–4583. [CrossRef]
19. Vachirasricirikul, S.; Ngamroo, I. Robust LFC in a Smart Grid with Wind Power Penetration by Coordinated V2G Control and Frequency Controller. *IEEE Trans. Smart Grid* **2014**, *5*, 371–380. [CrossRef]
20. Zhu, X.; Xia, M.; Chiang, H.-D. Coordinated sectional droop charging control for EV aggregator enhancing frequency stability of microgrid with high penetration of renewable energy sources. *Appl. Energy* **2018**, *210*, 936–943. [CrossRef]
21. Zhang, J.; Zhang, C.; Chien, W.C. Overview of Deep Reinforcement Learning Improvements and Applications. *J. Internet Technol.* **2021**, *22*, 239–255.
22. Giannopoulos, A.; Spantideas, S.; Kapsalis, N.; Karkazis, P.; Trakadas, P. Deep Reinforcement Learning for Energy-Efficient Multi-Channel Transmissions in 5G Cognitive HetNets: Centralized, Decentralized and Transfer Learning Based Solutions. *IEEE Access* **2021**, *9*, 129358–129374. [CrossRef]
23. Yang, Q.; Zhu, Y.; Zhang, J.; Qiao, S.; Liu, J. UAV Air Combat Autonomous Maneuver Decision Based on DDPG Algorithm. In Proceedings of the 2019 IEEE 15th International Conference on Control and Automation (ICCA) IEEE, Edinburgh, Scotland, 16–19 July 2019.
24. Gulde, R.; Tuscher, M.; Csiszar, A.; Riedel, O.; Verl, A. Deep reinforcement learning using cyclical learning rates. In Proceedings of the 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), Irvine, CA, USA, 21–23 September 2020.
25. Yu, T.; Zhou, B.; Chan, K.W.; Yuan, Y.; Yang, B.; Wu, Q.H. R(λ) imitation learning for automatic generation control of in-terconnected power grids. *Automatica* **2012**, *48*, 2130–2136. [CrossRef]

sustainability

MDPI

Article

Demand Management for Resilience Enhancement of Integrated Energy Distribution System against Natural Disasters

Yuting Xu *, Songsong Chen, Shiming Tian and Feixiang Gong

China Electric Power Research Institute Co., Ltd., Beijing 100192, China; 15901168062@163.com (S.C.); oldtian@sina.com.cn (S.T.); gongfeixiangouc@126.com (F.G.)
* Correspondence: yutingxu163@163.com

Abstract: For energy sustainability, the integrated energy distribution system (IEDS) is an efficient and clean energy system, which is based on the coordinated operation of a power distribution network, a gas distribution network and a district heating system. In this paper, considering the damage of natural disasters to IEDS, a demand management strategy is proposed to improve resilience of IEDS and ensure stable operation, which is divided into three stages. In the first stage, the electricity, natural gas and thermal energy are co-optimized in the simulating fault state to develop the importance ranking of transmission lines and gas pipelines. In the second stage, the natural disasters are classified as surface natural disasters and geological natural disasters. According to the types of natural disasters, the demand management strategy includes semi-emergency demand management scheme and full-emergency demand management scheme in the electrical resilience mode and the integrated resilience mode, respectively. In the third stage, the non-sequential Monte-Carlo simulation and scenario reduction algorithm are applied to describe potential natural disaster scenarios. According to the importance ranking of transmission lines and gas pipelines, a demand management strategy is formulated. Finally, the proposed strategy is applied on an IEEE 33-bus power system and a 19-node natural gas system. Its effectiveness is verified by numerical case studies.

Keywords: demand management; integrated energy distribution system; resilience; co-optimization; non-sequential Monte-Carlo simulation; scenario reduction algorithm

Citation: Xu, Y.; Chen, S.; Tian, S.; Gong, F. Demand Management for Resilience Enhancement of Integrated Energy Distribution System against Natural Disasters. *Sustainability* **2022**, *14*, 5. https://doi.org/10.3390/su14010005

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 24 October 2021
Accepted: 30 November 2021
Published: 21 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the development of society and economy, predatory energy consumption has caused environmental pollution [1] and energy crisis [2]. The integrated energy distribution system (IEDS) [3,4] takes full account of electricity, natural gas, heat and other forms of energy coupling. It can achieve the effect of energy mutual benefit according to the energy consumption characteristics of electricity, natural gas and heat. Hence, IEDS is an efficient and clean energy system.

On the other hand, frequent natural disasters have severely affected the energy system. In 2011, the Great East Japan Earthquake caused power outages in 8.71 million homes in the affected area [5]. In 2012, approximately 7.5 million customers suffered power outages in the Hurricane Sandy in New York and the disaster caused an economic loss of 65 billion US dollars [6]. Many typhoons from the Pacific will land in China every year. For example, Jiangsu Province was hit by a typhoon in 2016, which caused two 500-kV transmission lines, four 220-kV transmission lines, and eight 110-kV transmission lines to trip and left many customers without power [7]. In this regard, it is necessary and exigent to enhance the IEDS resilience. The IEDS resilience is derived from the extension of power system resilience, which can be defined as the ability to anticipate, resist, absorb and recover from disruptions caused by extreme natural disasters such as earthquakes and hurricanes [8].

At present, the previous studies mainly focused on power system resilience [9–15]. Two types of strategies can be adopted to enhance resilience: operational measures [9–11] and hardening measures [12–15]. The operational measures include scheduling flexible back-up resources [9], using decentralized control strategies [10] and altering network topology [11]. The study in [9] proposes dispatching mobile emergency generators to restore critical loads and improve power distribution system resilience. In [10], the networked micro-grids (MGs) are scheduled by a decentralized control strategy, which can improve power quality by supporting and interchanging electricity among the networked MGs. In [11], power grid reconfiguration is adopted in an active distribution system to reduce load demand response and improve power grid resilience. On the other hand, the hardening measures mean physically enhancing the infrastructure to reduce its susceptibility to disruptions [12], In [13], a tri-level defender-attacker-defender (DAD) model is proposed and its result can provide the system hardening decisions.

Due to the increasing interaction among the electricity distribution system, natural gas distribution system and district heating system, it is not suitable to consider power system resilience solely and it is necessary to consider integrated energy system resilience. However, the literature contains little research on IEDS resilience [14,15]. The study in [14] considers that the overhead power grid can be hardened by replacing fragile overhead transmission lines with underground natural gas pipelines and proposes a two-stage robust model to formulate this issue. Combined with the natural gas system, the DAD model is further expanded in [15] to accommodate electricity and gas storage facilities. The existing enhanced resilience methods in the aforementioned research are listed in Table 1.

Table 1. The existing enhanced resilience methods.

System Types	References	Enhanced Resilience Methods
Power system	[9]	Scheduling flexible back-up resources
	[10]	Decentralized control strategy for MGs
	[11]	Grid reconfiguration
	[12]	Enhancing infrastructure
	[13]	
Integrated energy system	[14]	Enhancing infrastructure
	[15]	
	This paper	Demand management

Although the IEDS resilience has not been fully studied yet, there is much research focusing on the operation of integrated energy system [16–19]. In [16], an integrated framework based on the Newton-Raphson technique is proposed to solve the steady-state energy flow among electrical, natural gas, and district heating networks. Furthermore, the study in [17] proposes a coordinated optimal operation method of the regional energy internet, considering the combined cooling, heating and power (CCHP) units. The study in [18] proposes a robust day-ahead scheduling model for electricity and natural gas system, which minimizes the total cost including fuel cost, spinning reserve cost and cost of operational risk while ensuring the feasibility for all scenarios within the uncertainty set. In particular, the study in [19] proposes a novel linear method for Weymouth equation of natural gas system, which develops a more robust, flexible and tractable formulation of the integrated power and gas network.

The above research on the integrated energy system operation lay the foundation for IEDS resilience. Therefore, motivated by the aforementioned facts, this paper proposes a novel demand management strategy for improving the resilience of the power distribution network and gas distribution network in different types of natural disasters. The demand management strategy is divided into three stages. In the first stage, the critical transmission lines and gas pipelines are identified by co-optimizing electricity, natural gas and thermal

energy under the assumptive fault. In other word, the importance ranking of transmission lines and gas pipelines are developed by simulating fault state. In the second stage, the natural disasters are classified as surface natural disasters (SNDs) and geological natural disasters (GNDs). The demand management strategy includes two modes. In the case of SNDs, the demand management strategy adopts the electrical resilience mode, in which the overhead transmission lines in power distribution system will be damaged, while the underground gas pipelines in gas distribution system still work without fault. In the case of GNDs, the demand management strategy adopts the integrated resilience mode, in which both overhead transmission lines and underground gas pipelines will be damaged. In the third stage, the non-sequential Monte-Carlo simulation and scenario reduction algorithm are applied to describe potential natural disaster scenarios. According to the importance ranking of transmission lines and gas pipelines, the demand management strategy is formulated. The advantages of the proposed strategy are demonstrated by numerical case studies.

The objective of the study is to propose a demand management strategy for resilience enhancement of IEDS against natural disasters and prove the effectiveness and practicality of this strategy. The major contributions of this paper are summarized as follows:

(1) Aiming at improving the IEDS resilience, this paper proposes a demand management strategy, which innovatively expands the traditional demand management in the power system.

(2) The demand management strategy includes semi-emergency demand management scheme and full-emergency demand management scheme in the electrical resilience mode and the integrated resilience mode, respectively. In the electrical resilience mode, gas distribution system can operate normally and play a role of energy storage to help power distribution system reduce demand response load. In the integrated resilience mode, the whole integrated energy system suffers damage and demand management is required.

(3) The demand management strategy is formulated in accordance with the importance of transmission lines and gas pipelines, taking into account the impact of the transmission grid and gas transmission network structure on demand management, which can increase the resilience of IEDS and reduce the economic subsidy cost of demand management.

The rest of this paper is organized as follows. The type of natural disasters, the stochastic approaches, and the framework of demand management strategy are discussed in Section 2. Then, Section 3 presents the scheduling model for the IEDS. In Section 4, numerical results are provided to validate the proposed strategy. Finally, Section 5 draws main conclusions in this paper.

2. The Demand Management Strategy for the IEDS under Natural Disasters

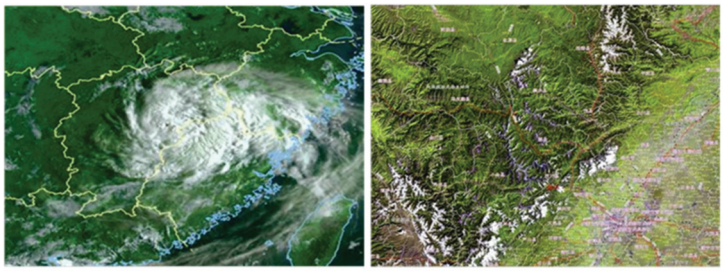
2.1. The Type of Natural Disasters

In this paper, natural disasters are divided into SNDs and GNDs according to whether the underground gas pipeline is destroyed or not. The SNDs include hurricane, typhoon and blizzard, while the GNDs include earthquake, landslide and mud-rock flow. The typical surface natural disaster and geological natural disaster are shown in Figure 1.

Because the overhead transmission lines are exposed to the environment, they are vulnerable to SNDs, such as typhoons and hurricanes. However, the underground gas pipelines have a certain resistibility to SNDs, so the gas distribution system can operate normally in the case of SNDs. Hence, the gas distribution system plays a role of energy storage to helps power distribution system improve resilience. In summary, the mode in which the overhead transmission lines are destroyed and the underground gas pipelines are not destroyed is called the electrical resilience mode.

On the other hand, both overhead transmission lines and underground gas pipelines may be damaged by the GNDs. In this case, the underground gas pipelines will be destroyed, and the gas distribution system also needs to be restored. Therefore, the mode in which the overhead transmission lines and the underground gas pipelines are destroyed is called the integrated resilience mode.

In this paper, the proposed demand management strategy can be applied to two types of natural disasters. The main difference between the two modes is whether the gas loads need to be reduced in the recovery process.



(**a**) (**b**)

Figure 1. The typical surface natural disaster and geological natural disaster. (**a**) Satellite cloud picture of "Sangmei" typhoon; (**b**) Satellite telemetry picture of Wenchuan Earthquake.

2.2. The Fragility Curves and Stochastic Approaches

The non-sequential Monte-Carlo simulation method generates many disaster scenarios with the random parameters of typhoon or earthquake. In order to get a typical disaster scenario, the scenario reduction algorithm is applied to reduce the number of samples. Then, as a calculation result of the scenario reduction algorithm, several main scenarios and their weights are derived. A detailed description of this algorithm can be found in [20].

The failure probabilities of the overhead transmission lines and underground gas pipelines in each simulated scenario are provided through their fragility curves, which express their failure probability as a function of the disaster parameter. The generic fragility curves of the overhead transmission lines and underground gas pipelines are shown in Figure 2 [8].

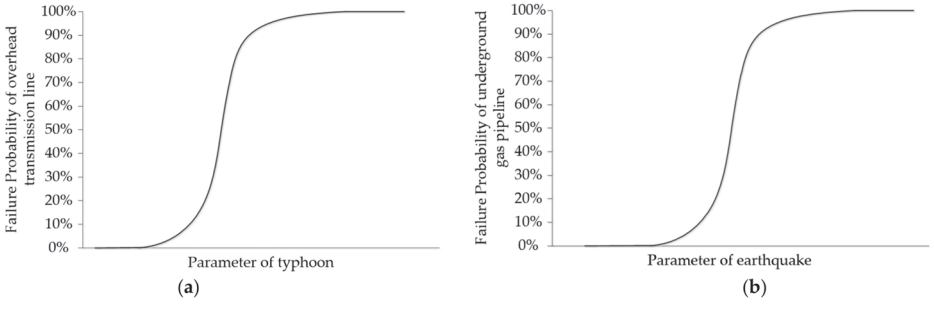


Figure 2. The generic fragility curves. (**a**) The fragility curve of overhead transmission line under typhoon; (**b**) The fragility curve of underground gas pipeline under earthquake.

2.3. The Framework of the Demand Management Strategy

As noted earlier, the demand management strategy is divided into three stages and its framework is depicted in Figure 3.

The first stage is the pre-disaster stage, in which the power distribution system, gas distribution system and district heating system co-optimize under the assumptive fault to identify critical transmission lines and gas pipelines, which is formulated in Section 3. Then, the second stage is mainly to distinguish the modes of demand management strategy

corresponding to different types of natural disasters. On this basis, the third stage generates the simulated scenarios and develops the demand management scheme.

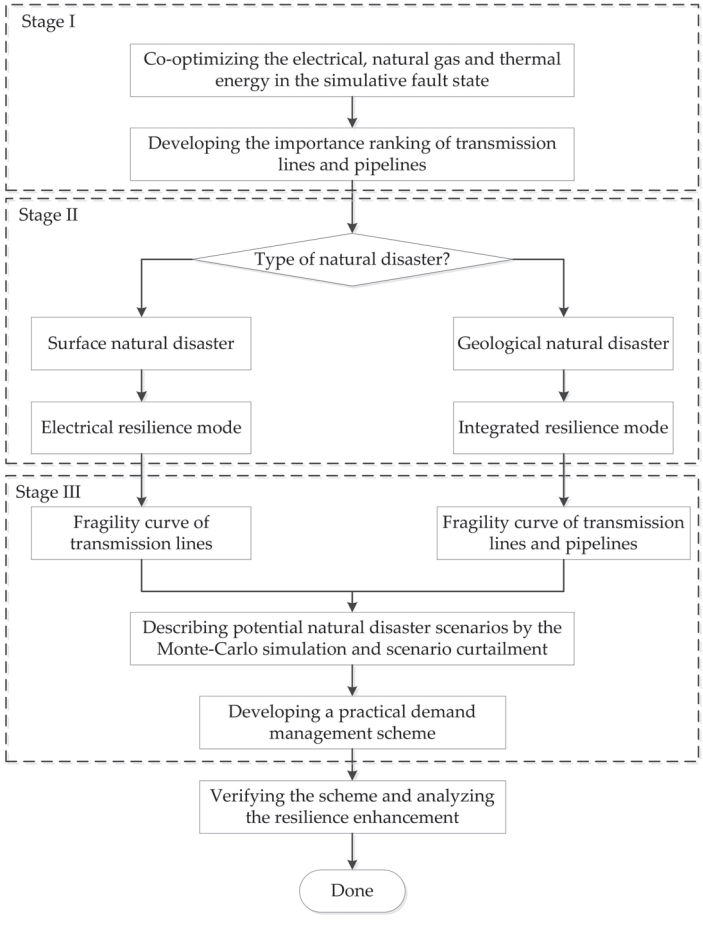


Figure 3. The framework of the demand management strategy.

3. Model Formulation

3.1. Objective Function

The goal of scheduling is to minimize the total operation cost of IEDS as shown in (1). Please refer to the Abbreviations for the explanation of the variables, indices, sets and parameters.

$$\begin{aligned}\min\{\sum_{s=1}^{N_s}\lambda_s(\sum_{t=1}^{T}(&\underbrace{\sum_{i^{pe}\in\delta_{pe}(b)}c_t^p p_{i^{pe},t,s}^{pur}}_{i}+\underbrace{\sum_{i^{pe}\in\delta_{pe}(b)}c_t^q q_{i^{pe},t,s}^{pur}}_{ii}+\underbrace{\sum_{i^{pg}\in\delta_{pg}(n)}c_t^f f_{i^{pg},t,s}^{pur}}_{iii}+\underbrace{\sum_{i^{CHP}\in\delta_{CHP}(b)}c_{i^{CHP}}^q q_{i^{CHP},t,s}^{CHP}}_{iv}\\&+\underbrace{\sum_{b=1}^{N_{bus}}\pi^p p_{b,t,s}^{demand}}_{v}+\underbrace{\sum_{b=1}^{N_{bus}}\pi^q q_{b,t,s}^{demand}}_{vi}+\underbrace{\sum_{n=1}^{N_{node}}\pi^f f_{n,t,s}^{demand}}_{vii}+\underbrace{\pi^h h_{t,s}^{demand}}_{viii}+\underbrace{\sum_{i^w\in\delta_w(b)}\kappa_{i^w}^w p_{i^w,t,s}^{w,cut}}_{ix}))\}\end{aligned}\tag{1}$$

The goal of the optimization is to minimize operation cost. The operation cost includes 9 parts, which can be divided into 3 categories. The first category is purchasing cost,

including active power purchased (i) from electricity transmission system, reactive power purchased (ii) from electricity transmission system, natural gas purchased (iii) from gas transmission system, reactive power purchased (iv) from combined heating and power (CHP) units. The second category is demand response cost, including active power load demand response (v), reactive power load demand response (vi), gas load demand response (vii), and heat load demand response (viii). The third category is penalty cost, including penalty cost of wind power curtailment (ix).

3.2. The Power Distribution System Constraints

The branch flow model is widely applied to compute the alternative current power flow of a power distribution system, but it is not suitable for optimization problems, because this model includes the power flow constraints, which is non-convex. For this difficulty, the second-order cone (SOC) relaxation is employed to relax the non-convex constraints in the branch flow model [21]. In detail, following the SOC relaxation method, the power flow equality constraints are replaced by inequality constraints and this modification still ensures the exactness, so the non-convex branch flow model is converted to the DistFlow model. Furthermore, the DistFlow model can be linearized to formulate microgrid operation constraints. In fact, the linear DistFlow model has been extensively used and justified in power distribution systems.

The linear DistFlow model gives constraints (2)–(13).

$$\begin{aligned}&\sum_{i^{pe}\in\delta_{pe}(b)} p^{pur}_{i^{pe},t,s} + \sum_{\substack{i^{le}\in\delta_{le}(b)\\ j^{ls}\in l(i^{le})}} p^{l}_{i^{le}j^{ls},t,s} + \sum_{i^{CHP}\in\delta_{CHP}(b)} p^{CHP}_{i^{CHP},t,s} + \sum_{i^{w}\in\delta_{w}(b)} p^{w}_{i^{w},t,s}\\ &= \sum_{\substack{i^{ls}\in\delta_{ls}(b)\\ j^{le}\in l(i^{ls})}} p^{l}_{i^{ls}j^{le},t,s} + \sum_{i^{EHP}\in\delta_{EHP}(b)} p^{EHP}_{i^{EHP},t,s} + p^{load}_{b,t,s} - p^{demand}_{b,t,s} \qquad \forall b,\forall t,\forall s\end{aligned} \tag{2}$$

$$\begin{aligned}&\sum_{i^{pe}\in\delta_{pe}(b)} q^{pur}_{i^{pe},t,s} + \sum_{\substack{i^{le}\in\delta_{le}(b)\\ j^{ls}\in l(i^{le})}} q^{l}_{i^{le}j^{ls},t,s} + \sum_{i^{CHP}\in\delta_{CHP}(b)} q^{CHP}_{i^{CHP},t,s} + \sum_{i^{w}\in\delta_{w}(b)} q^{w}_{i^{w},t,s}\\ &= \sum_{\substack{i^{ls}\in\delta_{ls}(b)\\ j^{le}\in l(i^{ls})}} q^{l}_{i^{ls}j^{le},t,s} + q^{load}_{b,t,s} - q^{demand}_{b,t,s} \qquad \forall b,\forall t,\forall s\end{aligned} \tag{3}$$

$$V_{i^{ls},t,s} - V_{j^{le},t,s} = \frac{p^{l}_{i^{ls}j^{le},t,s} r_{i^{ls}j^{le}} + q^{l}_{i^{ls}j^{le},t,s} x_{i^{ls}j^{le}}}{V_{base}} \qquad i^{ls}\in\delta_{ls}(b), j^{le}\in l(i^{ls}),\forall t,\forall s \tag{4}$$

$$-\overline{p}^{l}_{i^{ls}j^{le}} \le p^{l}_{i^{ls}j^{le},t,s} \le \overline{p}^{l}_{i^{ls}j^{le}} \qquad i^{ls}\in\delta_{ls}(b), j^{le}\in l(i^{ls}),\forall t,\forall s \tag{5}$$

$$-\overline{q}^{l}_{i^{ls}j^{le}} \le q^{l}_{i^{ls}j^{le},t,s} \le \overline{q}^{l}_{i^{ls}j^{le}} \qquad i^{ls}\in\delta_{ls}(b), j^{le}\in l(i^{ls}),\forall t,\forall s \tag{6}$$

$$0 \le p^{pur}_{i^{pe},t,s} \le \overline{p}^{pur} \qquad i^{pe}\in\delta_{pe}(b),\forall t,\forall s \tag{7}$$

$$0 \le q^{pur}_{i^{pe},t,s} \le \overline{q}^{pur} \qquad i^{pe}\in\delta_{pe}(b),\forall t,\forall s \tag{8}$$

$$\underline{V}_{b} \le V_{b,t,s} \le \overline{V}_{b} \qquad \forall b,\forall t,\forall s \tag{9}$$

$$0 \le p^{demand}_{b,t,s} \le p^{load}_{b,t,s} \qquad \forall b,\forall t,\forall s \tag{10}$$

$$0 \le q^{demand}_{b,t,s} \le q^{load}_{b,t,s} \qquad \forall b,\forall t,\forall s \tag{11}$$

$$p^{w}_{i^{w},t,s} = \overline{p}^{w}_{i^{w},t,s} - p^{w,cut}_{i^{w},t,s} \qquad i^{w}\in\delta_{w}(b),\forall t,\forall s \tag{12}$$

$$q^{w}_{i^{w},t,s} = \tan(\arccos(\phi_{i^{w}}))\, p^{w}_{i^{w},t,s} \qquad i^{w}\in\delta_{w}(b),\forall t,\forall s \tag{13}$$

Specifically, (2) and (3) imply active power balance and reactive power balance, which means that the active (or reactive) power injection amount is equivalent to outflow amount

at each bus. (4) is the DistFlow equation, which relates the active power and reactive power flows of a transmission line to the voltage magnitude of its two terminal buses. (5) and (6) denote active power limit and reactive power limit on overhead transmission lines, respectively. (7) and (8) refer to the active (or reactive) power purchasing limits. (9) gives voltage limit at each bus. (10) and (11) imply the load demand response of active power and reactive power, respectively. (12) and (13) describe power outputs of wind generators.

3.3. The Gas Distribution System Constraints

The operation constraints of the gas distribution system are shown in (14)–(19).

$$\begin{aligned}\sum_{i^{pg}\in\delta_{pg}(n)} f^{pur}_{i^{pg},t,s} + \sum_{\substack{i^{ge}\in\delta_{ge}(n)\\ j^{gs}\in g(i^{ge})}} f^{g}_{i^{ge}j^{gs},t,s} = \sum_{\substack{i^{gs}\in\delta_{gs}(n)\\ j^{ge}\in g(i^{gs})}} f^{g}_{i^{gs}j^{ge},t,s}\\ + \sum_{i^{CHP}\in\delta_{CHP}(n)} f^{CHP}_{i^{CHP},t,s} + \sum_{i^{boi}\in\delta_{boi}(n)} f^{boi}_{i^{boi},t,s} + f^{load}_{n,t,s} - f^{demand}_{n,t,s} \quad \forall n,\forall t,\forall s\end{aligned} \tag{14}$$

$$-\overline{f}^{g}_{i^{gs}j^{ge}} \le f^{g}_{i^{gs}j^{ge},t,s} \le \overline{f}^{g}_{i^{gs}j^{ge}} \quad i^{gs}\in\delta_{gs}(n), j^{ge}\in l(i^{gs}),\forall t,\forall s \tag{15}$$

$$0 \le f^{pur}_{i^{pg},t,s} \le \overline{f}^{pur} \quad i^{pg}\in\delta_{pg}(n),\forall t,\forall s \tag{16}$$

$$\underline{z}_{n} \le z_{n,t,s} \le \overline{z}_{n} \quad \forall n,\forall t,\forall s \tag{17}$$

$$0 \le f^{demand}_{n,t,s} \le f^{load}_{n,t,s} \quad \forall n,\forall t,\forall s \tag{18}$$

$$f^{g}_{i^{gs}j^{ge},t,s} = K_{i^{gs}j^{ge}}\sqrt{\beta(z_{i^{gs},t,s},z_{j^{ge},t,s})\left(z^{2}_{i^{gs},t,s}-z^{2}_{j^{ge},t,s}\right)} \quad i^{gs}\in\delta_{gs}(n), j^{ge}\in l(i^{gs}),\forall t,\forall s \tag{19}$$

(14) denotes natural gas balance, which means that the natural gas flow injection amount is equivalent to amount at each node. (15) refers to gas flow limit in gas pipelines. (16) refers to the active natural gas purchasing limits. In order to avoid damage or malfunction of gas pipeline caused by too low or too high gas pressure, (17) gives gas pressure limit at each node. (18) implies load demand response of natural gas. (19) is the Weymouth equation, which relates gas flow in gas pipeline and pressure of its two terminal nodes. Specifically, the direction coefficient $\beta(z_{i^{gs},t,s},z_{j^{ge},t,s})$ describes the relationship between gas flow direction and gas pressure value at both ends of gas pipeline, which is denoted as follows:

$$\beta(z_{i^{gs},t,s},z_{j^{ge},t,s}) = \begin{cases} 1 z_{i^{gs},t,s} \ge z_{j^{ge},t,s} \\ -1 z_{i^{gs},t,s} < z_{j^{ge},t,s} \end{cases} \quad i^{gs}\in\delta_{gs}(n), j^{ge}\in l(i^{gs}),\forall t,\forall s \tag{20}$$

(20) denotes the fact that gas flow runs from higher pressure to lower pressure. Moreover, the Weymouth equation (19) can be rewritten as (21).

$$\left(f^{g}_{i^{gs}j^{ge},t,s}\right)^{2} = \beta(z_{i^{gs},t,s},z_{j^{ge},t,s})\left(K_{i^{gs}j^{ge}}\right)^{2}\left(z^{2}_{i^{gs},t,s}-z^{2}_{j^{ge},t,s}\right) \quad i^{gs}\in\delta_{gs}(n), j^{ge}\in l(i^{gs}),\forall t,\forall s \tag{21}$$

Apparently, since gas flow is squared, its direction cannot connect to gas pressure value. In other words, the direction coefficient is useless in (21). For the flow direction problem, (21) is equivalently reformulated as (22)–(25) by the Big-M theory with additional binary variables [22].

$$-M\left(1-\delta_{i^{gs}j^{ge},t,s}\right) \le \left(f^{g}_{i^{gs}j^{ge},t,s}\right)^{2} - \left(K_{i^{gs}j^{ge}}\right)^{2}\left(z^{2}_{i^{gs},t,s}-z^{2}_{j^{ge},t,s}\right) \le M\left(1-\delta_{i^{gs}j^{ge},t,s}\right) i^{gs}\in\delta_{gs}(n), j^{ge}\in l(i^{gs}),\forall t,\forall s \tag{22}$$

$$-M\delta_{i^{gs}j^{ge},t,s} \le \left(f^{g}_{i^{gs}j^{ge},t,s}\right)^{2} - \left(K_{i^{gs}j^{ge}}\right)^{2}\left(z^{2}_{i^{gs},t,s}-z^{2}_{j^{ge},t,s}\right) \le M\delta_{i^{gs}j^{ge},t,s} \quad i^{gs}\in\delta_{gs}(n), j^{ge}\in l(i^{gs}),\forall t,\forall s \tag{23}$$

$$-M\left(1-\delta_{i^{gs}j^{ge},t,s}\right) \le z_{i^{gs},t,s} - z_{j^{ge},t,s} \le M\delta_{i^{gs}j^{ge},t,s} \quad i^{gs}\in\delta_{gs}(n), j^{ge}\in l(i^{gs}),\forall t,\forall s \tag{24}$$

$$-M\left(1-\delta_{i^{gs}j^{ge},t,s}\right) \le f^{g}_{i^{gs}j^{ge},t,s} \le M\delta_{i^{gs}j^{ge},t,s} \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}), \forall t, \forall s \tag{25}$$

It is worth mentioning that the quadratic terms in (22) and (23) are nonlinear, which makes the optimization model difficult to solve, so the quadratic terms are linearized by the piecewise linear approximation method. The gas pressure square magnitude (i.e.,$z^{2}_{i^{gs},t,s}$) in (19) can be directly replaced by a positive continuous variable $Z_{i^{gs},t,s}$. However, the same approach cannot be applied to tackle gas flow square magnitude, because the linear terms of gas flow are expressed in natural gas balance constraint (14) and gas pipelines limit constraint (15). Therefore, as an efficient piecewise linear approximation method, the incremental linearization method [19] is employed to deal with the nonlinearity of gas flow square magnitude, which is appropriate for mixed-integer linear programming.

The domain of gas flow in pipeline is $f^{g}_{i^{gs}j^{ge},t,s} \in \left[-\overline{f}^{g}_{i^{gs}j^{ge}}, \overline{f}^{g}_{i^{gs}j^{ge}}\right]$, which is denoted by (15). The domain $\left[-\overline{f}^{g}_{i^{gs}j^{ge}}, \overline{f}^{g}_{i^{gs}j^{ge}}\right]$ is divided into KP intervals by breakpoints $x_{i^{gs}j^{ge},k}$ as shown in (26):

$$-\overline{f}^{g}_{i^{gs}j^{ge}} = x_{i^{gs}j^{ge},1} \le x_{i^{gs}j^{ge},2} \le \cdots \le x_{i^{gs}j^{ge},k} \le \cdots \le x_{i^{gs}j^{ge},KP} \le x_{i^{gs}j^{ge},KP+1} = \overline{f}^{g}_{i^{gs}j^{ge}} \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}) \tag{26}$$

$$\varepsilon_{i^{gs}j^{ge}} = x_{i^{gs}j^{ge},k+1} - x_{i^{gs}j^{ge},k} \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}), k = 1, 2, \cdots, KP-1 \tag{27}$$

$$KP = \left[2\overline{f}^{g}_{i^{gs}j^{ge}} / \varepsilon_{i^{gs}j^{ge}}\right] \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}) \tag{28}$$

The size of interval is calculated by (27) and the number of intervals is calculated by (28). Corresponding to each $x_{i^{gs}j^{ge},k}$, the ordinate value is calculated by $y_{i^{gs}j^{ge},k} = x^{2}_{i^{gs}j^{ge},k}$. Figure 4 shows the piecewise linearization of nonlinear function. The quadratic function (solid red line) is approximated by piecewise linear function (solid blue line).

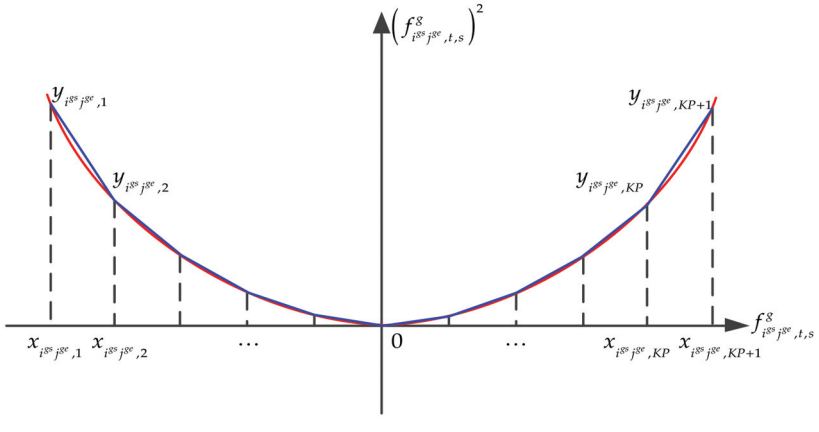


Figure 4. The piecewise linearization of nonlinear function.

Then, the linear terms of (22) and (23) are formulated by

$$\left(K_{i^{gs}j^{ge}}\right)^2\left(Z_{i^{gs},t,s} - Z_{j^{ge},t,s}\right) = y_{i^{gs}j^{ge},1} + \sum_{k=1}^{KP}\left(y_{i^{gs}j^{ge},k+1} - y_{i^{gs}j^{ge},k}\right)\mu_{i^{gs}j^{ge},t,s,k} \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}), \forall t, \forall s, k = 1, 2, \cdots, KP-1 \tag{29}$$

$$f^{g}_{i^{gs}j^{ge},t,s} = x_{i^{gs}j^{ge},1} + \sum_{k=1}^{KP}\left(x_{i^{gs}j^{ge},k+1} - x_{i^{gs}j^{ge},k}\right)\mu_{i^{gs}j^{ge},t,s,k} \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}), \forall t, \forall s, k = 1, 2, \cdots, KP-1 \tag{30}$$

$$0 \le \mu_{i^{gs}j^{ge},t,s,k} \le 1 \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}), \forall t, \forall s, \forall k \tag{31}$$

$$\mu_{i^{gs}j^{ge},t,s,k+1} \le \xi_{i^{gs}j^{ge},t,s,k} \le \mu_{i^{gs}j^{ge},t,s,k} \qquad i^{gs} \in \delta_{gs}(n), j^{ge} \in l(i^{gs}), \forall t, \forall s, k = 1, 2, \cdots, KP-1 \tag{32}$$

where $\mu_{i^{gs}j^{ge},t,s,k}$ and $\xi_{i^{gs}j^{ge},t,s,k}$ are auxiliary variables at kth interval of piecewise linear function. Constraints (32) ensures that there is at most one index k of an interval with $0 < \mu_{i^{gs}j^{ge},t,s,k} < 1$ and other indexes are equal to 0 or 1. Specifically, $\mu_{i^{gs}j^{ge},t,s,k+1} = 0$, if $\xi_{i^{gs}j^{ge},t,s,k} = 0$; $\mu_{i^{gs}j^{ge},t,s,k+1} = 1$, if $\xi_{i^{gs}j^{ge},t,s,k} = 1$.

From the above, the operation constraints of the gas distribution system consists of (14)–(19) and the nonlinear constraint (19) is reformulated by (22)–(25) and (29)–(32).

3.4. The District Heating System Constraints

The district heating system constraints consist of heat balance and operation of heat generating equipment, which includes CHP units, electrical heat pumps (EHPs) and boilers.

The heat balance constraint:

$$\sum_{i^{CHP} \in \delta_{CHP}(n)} h^{CHP}_{i^{CHP},t,s} + \sum_{i^{EHP} \in \delta_{EHP}(b)} h^{EHP}_{i^{EHP},t,s} + \sum_{i^{boi} \in \delta_{boi}(n)} h^{boi}_{i^{boi},t,s} = h^{load}_{t,s} - h^{demand}_{t,s} \qquad \forall t, \forall s \tag{33}$$

$$0 \le h^{demand}_{t,s} \le h^{load}_{t,s} \qquad \forall t, \forall s \tag{34}$$

The constraints (33) means that the heat generating amount is equivalent to heat demand amount in the district heating system. (34) implies load demand response of heat.

The CHP units constraints are as follows:

$$\underline{p}^{CHP}_{i^{CHP}} \le p^{CHP}_{i^{CHP},t,s} \le \overline{p}^{CHP}_{i^{CHP}} \qquad i^{CHP} \in \delta_{CHP}(b), \forall t, \forall s \tag{35}$$

$$\underline{q}^{CHP}_{i^{CHP}} \le q^{CHP}_{i^{CHP},t,s} \le \overline{q}^{CHP}_{i^{CHP}} \qquad i^{CHP} \in \delta_{CHP}(b), \forall t, \forall s \tag{36}$$

$$p^{CHP}_{i^{CHP},t,s} = \eta^{CHP,ele}_{i^{CHP}} f^{CHP}_{i^{CHP},t,s} \qquad i^{CHP} \in \delta_{CHP}(b), \forall t, \forall s \tag{37}$$

$$h^{CHP}_{i^{CHP},t,s} = \eta^{CHP,heat}_{i^{CHP}} f^{CHP}_{i^{CHP},t,s} \qquad i^{CHP} \in \delta_{CHP}(b), \forall t, \forall s \tag{38}$$

(35) and (36) denote output limit of CHP units, which means that active power and reactive power must be between the minimum output and maximum output. (37) and (38) imply power generation efficiency and heat production efficiency of CHP units, respectively.

The EHPs constraints are as follows:

$$0 \le p^{EHP}_{i^{EHP},t,s} \le \overline{p}^{EHP}_{i^{EHP}} \qquad i^{EHP} \in \delta_{EHP}(b), \forall t, \forall s \tag{39}$$

$$h^{EHP}_{i^{EHP},t,s} = \eta^{EHP}_{i^{EHP}} p^{EHP}_{i^{EHP},t,s} \qquad i^{EHP} \in \delta_{EHP}(b), \forall t, \forall s \tag{40}$$

(39) denotes power consumption limit of EHP. (40) refers to heat production efficiency of EHP.

The boilers constraints are as follows:

$$0 \le f^{boi}_{i^{boi},t,s} \le \overline{f}^{boi}_{i^{boi}} \qquad i^{boi} \in \delta_{boi}(n), \forall t, \forall s \tag{41}$$

$$h^{boi}_{i^{boi},t,s} = \eta^{boi}_{i^{boi}} f^{boi}_{i^{boi},t,s} \qquad i^{boi} \in \delta_{boi}(n), \forall t, \forall s \tag{42}$$

(41) denotes natural gas consumption limit of boiler. (42) refers to heat production efficiency of boiler.

The proposed scheduling model of IEDS is a mixed integer linear programming (MILP), which can be solved with the solvers, such as CPLEX, GUROBI.

4. Numerical Results

4.1. The Operation of IEDS

The proposed hardening strategy is examined on a test system, which consists of IEEE 33-bus power distribution system, 19-node gas distribution system and district heating system, shown in Figure 5. The numbers without brackets in Figure 5 represent the node number, and the numbers with brackets represent the line number.

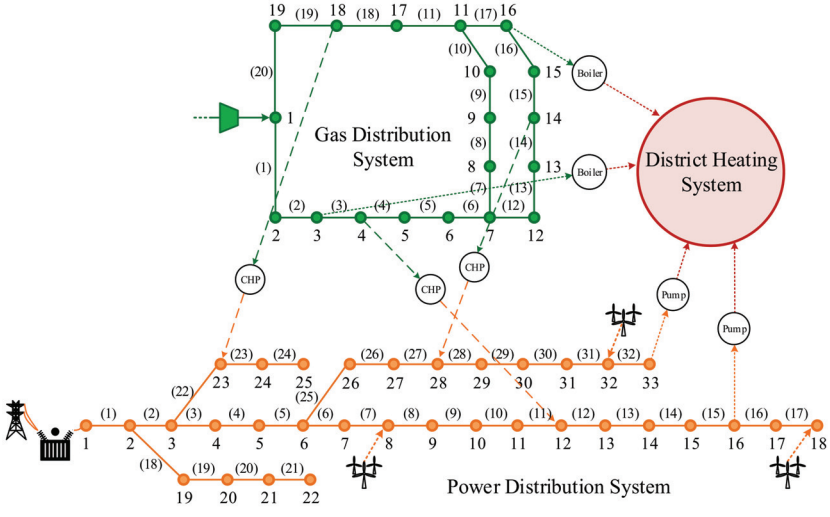


Figure 5. The integrated energy distribution system.

As shown in Figure 5, there are three CHP units, two EHPs, two boilers and three wind generators. In fact, the three energy systems are geographically overlapping. For convenience, Figure 5 is a schematic diagram of the wiring of three systems.

First of all, in order to facilitate the explanation of the test system, the operation of IEDS is introduced in normal state, which is the basis of fault operation. The supply of active power load, gas load and heat load are shown in Figure 6.

As shown in Figure 6, the supply and demand of IEDS are balanced at each time slot in normal operation. However, if the overhead transmission lines or underground gas pipelines are broken, the energy supply will be restricted and load demand will not be met, so an effective demand management strategy is urgently needed to alleviate the tight supply of IEDS.

4.2. The Importance Ranking of Transmission Lines and Pipelines

After breaking transmission lines and pipelines, the amount of load demand response is calculated at each simulating fault state. This section calculates the economic subsidy caused by load demand response. The corresponding economic subsidy caused by the damaged transmission lines and pipelines is shown in Tables 2 and 3.

Then, according to the economic subsidy in each fault state, the importance of transmission lines and pipelines is ranked, which is divided into three levels. In the A-level and B-level, economic subsidy is greater than 120 K and 60 K, respectively. Economic subsidy less than 60 K is the C-level.

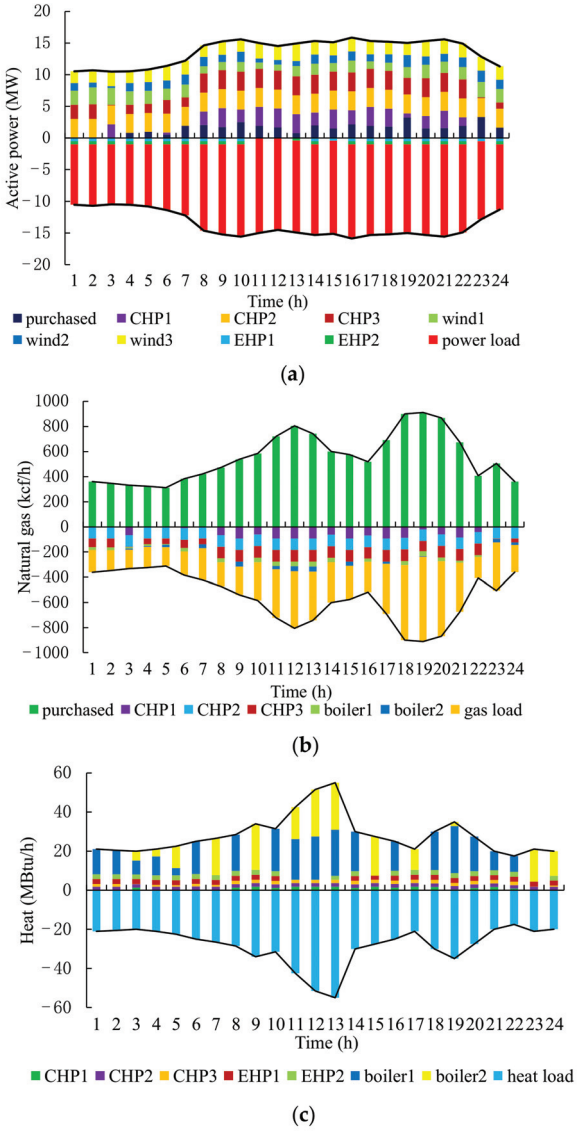


Figure 6. The supply of load in normal state. (**a**) The supply of active power load; (**b**) The supply of gas load; (**c**) The supply of heat load.

4.3. The Electrical Resilience Mode and Integrated Resilience Mode

In the electrical resilience mode, the surface natural disasters are simulated. In total, 100 fault scenarios are generated by non-sequential Monte-Carlo simulation, and then 5 main scenarios are obtained by scenario reduction algorithm, which are used to calculate the fault operation conditions caused by the surface natural disasters. The main scenario with the largest probability is 0.34. Transmission lines No. 6, 14, 17, 19, 29, and 30 have failed. After the failure, the power load cannot be fully supplied. Comparing the normal operation state, a demand response is required, as shown in Figure 7.

Table 2. The economic subsidy of the damaged transmission lines.

A	232.1 K	213.7 K	199.5 K	187.7 K	187.5 K	183.2 K	162.0 K	156.4 K	143.0 K	142.5 K	133.7 K	131.0 K
Line Number	18	23	12	30	19	13	28	14	20	31	24	15
B	119.4 K	109.9 K		106.4 K	98.5 K		88.9 K	84.2 K	71.4 K	63.5 K		62.8 K
Line Number	1	16		29	21		17	32	6	7		3
C	58.4 K	58.1 K	56.9 K	56.5 K	56.2 K		56.2 K	55.9 K	55.7 K	54.4 K	54.3 K	54.3 K
Line Number	5	4	11	10	22		9	8	2	25	26	27

Table 3. The economic subsidy of the damaged pipelines.

A	155.0 K		146.4 K		136.9 K		135.8 K		131.4 K	126.0 K	
Pipeline Number	20		19		11		18		7	1	
B	117.3 K	113.4 K	106.8 K	101.2 K	96.2 K	91.1 K	86.7 K	83.0 K	83.0 K	77.2 K	64.8 K
Pipeline Number	6	2	12	13	8	3	4	17	5	10	9
C	54.2 K				51.1 K				8.2 K		
Pipeline Number	14				16				15		

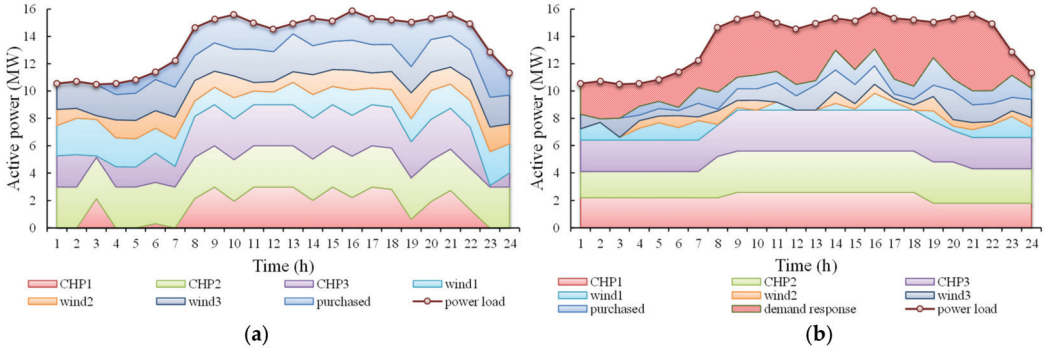


Figure 7. The power load supply situation. (**a**) The power load supply situation in normal operation state; (**b**) The power load supply situation in failure operation state.

As shown in Figure 7, destruction of transmission lines results in an amount of power loads that cannot obtain electricity through outsourcing. The wind turbines are damaged due to natural disasters and the output decreases. At this time, the distribution network forms many small island grids, and CHP provides power supply to the island grids.

However, affected by the supply capacity of CHP, a certain amount of power load is still needed as a demand response resource to participate in demand management. This demand management scheme that considers the support of the natural gas system to the distribution network is called a semi-emergency scheme in this paper, and it is not necessary to use all demand response capabilities.

In the integrated resilience mode, the geological natural disasters are simulated. Similarly, 100 failure scenarios are generated and reduced to 5 main scenarios to analyze the operation of IEDS after geological natural disaster. The main scenario with the largest probability is 0.29. Transmission lines No. 7, 9, 14, 16, 20, 26, and 29 have failed. Gas pipelines No. 4, 8, 11, and 15 have failed. After the failure, not only can the power load not be fully supplied, but the natural gas is also unable to supply all the gas load due to the failure of some pipelines. The power load and gas load supply conditions are shown in Figure 8.

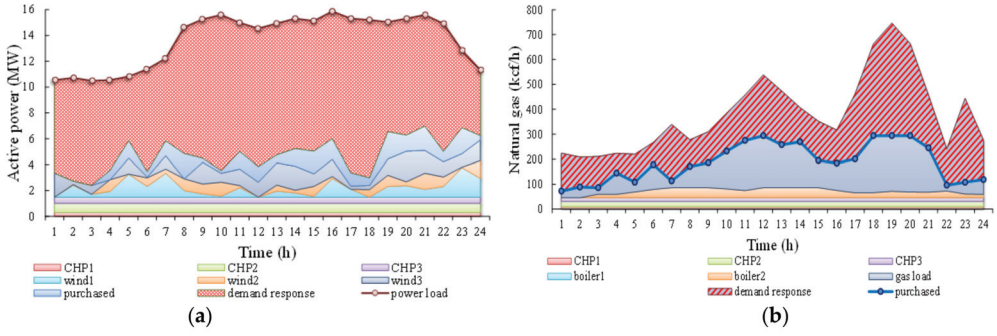


Figure 8. The power load and gas load supply situation. (**a**) The power load supply situation in failure operation state; (**b**) The gas load supply situation in failure operation state.

As shown in Figure 8, after the natural gas distribution network was destroyed by geological natural disasters, some gas pipeline failures caused a gap in natural gas supply. The red shaded part in Figure 8b is the amount of natural gas involved in gas load demand response, which is equal to the total gas load minus the largest purchased natural gas in each period.

In addition, due to the difficulty of gas supply, CHP's ability to support power supply is extremely weak, so they all operate at a very low level of power generation. At this time, since the power load cannot get assistance from the natural gas system, almost all loads have to participate in demand management. This demand management scheme that does not consider natural gas support is called full-emergency scheme in this paper, and it is necessary to use all demand response capabilities.

It is worth mentioning that regarding the supply of heat load, in the electrical resilience mode, although pumps cannot provide heat due to insufficient power supply, gas can be used to provide sufficient heat through the boilers without causing heat load loss. In the integrated resilience mode, due to the failure of No. 8, 11, and 15 pipelines, No. 1 boiler cannot obtain gas for heating, and No. 2 boiler is used for heating, supplementally. Although part of the heat load is involved in demand response, it is not the focus of this paper and will not be analyzed in detail.

4.4. The Demand Management Strategy

From the above, the demand management strategy includes a semi-emergency demand management scheme and a full-emergency demand management scheme, according to the type of natural disaster. Moreover, the demand management strategy is based on the importance of lines and pipelines. As shown in Tables 1 and 2, the failure of each transmission line and gas pipeline corresponds to a certain economic subsidy to the load participating in the demand response. According to the semi-emergency scheme and the full-emergency scheme, the upper and lower cost limits of the corresponding demand management for each failure transmission line and gas pipeline can be obtained, as shown in Figures 9 and 10.

As shown in Figures 9 and 10, according to the faults caused by natural disasters generated by non-sequential Monte-Carlo simulation, the economic subsidies for demand management caused by the failure of various transmission lines and gas pipelines fluctuate relatively widely. Therefore, the demand management strategy proposed in this paper can obtain the optimal demand management cost according to the failure situation, which can ensure that IEDS can not only improve the resilience, but also have a lower operating cost.

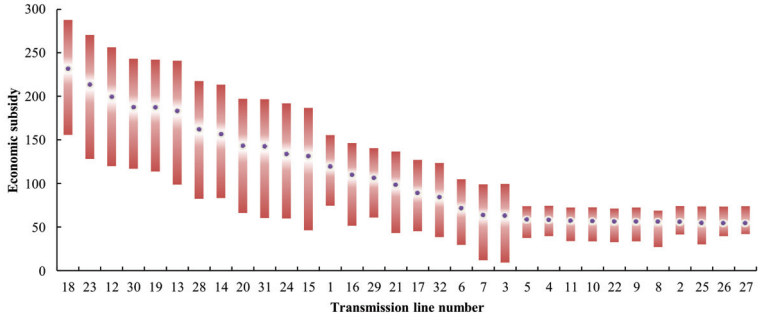


Figure 9. The economic subsidy of power demand response.

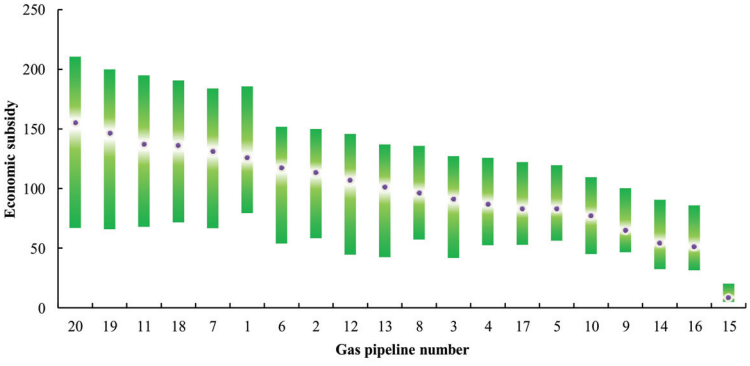


Figure 10. The economic subsidy of gas demand response.

5. Conclusions

IEDS is an efficient, clean and sustainable energy system. However, increasingly frequent natural disasters pose a severe threat to the security of IDES. Hence, this paper studies demand management for IEDS to improve resilience against extreme events. The demand management strategy includes a semi-emergency demand management scheme and a full-emergency demand management scheme, according to the resilience mode. The strategy generation process is divided into three stages. In short, combined with component generic fragility curves, through the non-sequential Monte-Carlo simulation method and the scene reduction algorithm, the main scene of the failure caused by the natural disaster is simulated, and then according to the importance ranking of transmission line and gas pipeline the demand management economic subsidy costs under different failure situations are obtained. The main conclusions are summarized as follows:

(1) The proposed scheduling model of IEDS co-optimizes the electricity, natural gas and thermal energy in normal and failure operation state, which promotes efficient and sustainable energy consumption.

(2) This paper proposes a demand management strategy, which innovatively expands the traditional demand management in a power system. The demand management of the IEDS includes two parts: power demand management and gas demand management. This paper analyzes the economic subsidies for demand response in these two parts, respectively.

(3) The demand management strategy proposed in this paper can increase the resilience of IEDS and reduce the economic subsidy cost of demand management.

Author Contributions: Conceptualization, Y.X. and S.C.; methodology, S.T. and F.G.; software, Y.X.; validation, Y.X., S.C., S.T. and F.G.; formal analysis, Y.X.; investigation, Y.X.; resources, S.C.; data curation, S.T.; writing—original draft preparation, Y.X.; writing—review and editing, Y.X. and F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Science and Technology Project of State Grid "Research on Large-scale Interactive Response Technology of Demand-Side Flexible Resources with High Penetration of Renewable Energy" (No. 5100-202114296A-0-0-00).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Indices and Sets

t	Index of hours running from 1 to T.
s	Index of scenarios running from 1 to N_s.
b	Index of buses running from 1 to N_{bus}.
n	Index of nodes running from 1 to N_{node}.
i^{pe}/i^{pg}	Index of power/natural gas purchase points.
i^{ls}/i^{le}	Index of overhead transmission line start/end points.
j^{le}/j^{ls}	Index of overhead transmission line end/start points corresponding to i^{ls}/i^{le}.
i^{gs}/i^{ge}	Index of gas pipeline start/end points.
j^{ge}/j^{gs}	Index of gas pipeline end/start points corresponding to i^{gs}/i^{ge}.
i^{w}	Index of wind generators.
$i^{CHP}/i^{EHP}/i^{boi}$	Index of CHP units/electrical heat pumps/boilers.
$\delta_{pe}(b)/\delta_{pg}(n)$	Set of power/natural gas purchase points.
$\delta_{ls}(b)/\delta_{le}(b)$	Set of overhead transmission line start/end points.
$l(i^{le})/l(i^{ls})i^{w}$	Set of adjacent buses.
$\delta_{gs}(n)/\delta_{ge}(n)$	Set of gas pipeline start/end points.
$g(i^{ge})/g(i^{gs})$	Set of adjacent nodes.
$\delta_{CHP}(b)/\delta_{CHP}(n)$	Set of access points of CHP units in power/gas distribution system.
$\delta_{EHP}(b)/\delta_{boi}(n)/\delta_{w}(b)$	Set of access points of electrical heat pumps/boilers/wind generators.

Parameters

T	Duration of scheduling horizon.
N_s	Number of scenarios.
N_{bus}	Number of buses
N_{node}	Number of nodes
λ_s	Probability of scenarios s.
$c_t^p/c_t^q/c_t^f$	Active power/reactive power/natural gas price at hour t.
$c_{i^{CHP}}^q$	Cost of reactive power from CHP units i^{CHP}.
$\pi^p/\pi^q/\pi^f/\pi^h$	Subsidy cost of active power/reactive power/natural gas/heat load demand response.
$\kappa_{i^w}^w$	Penalty cost of wind generator i^w power curtailment.
$p_{b,t,s}^{load}/\ q_{b,t,s}^{load}$	Active power load/reactive power load at bus b in Scenarios s at hour t.

$f_{n,t,s}^{load}$	Natural gas load at node n in Scenarios s at hour t.
$h_{t,s}^{load}$	Heat load in Scenarios s at hour t.
$r_{i^{ls}j^{le}} / x_{i^{ls}j^{le}}$	Resistance/reactance of overhead transmission line ij.
V_{base}	Base value of voltage.
$\overline{p}_{i^{ls}j^{le}}^{l} / \overline{q}_{i^{ls}j^{le}}^{l}$	Maximum active/reactive power of overhead transmission line ij.
$\overline{p}^{pur} / \overline{q}^{pur}$	Maximum active/reactive power purchased from IEDS.
$\overline{V}_b / \underline{V}_b$	Maximum/minimum voltage at bus b.
$\overline{p}_{i^w,t,s}^{w}$	Forecasted output of wind generator i^w in Scenarios s at hour t.
ϕ_{i^w}	Power factor of wind generator i^w.
$K_{i^{gs}j^{ge}}$	Gas flow constant of gas pipeline ij.
$\beta(z_{i^{gs},t,s}, z_{j^{ge},t,s})$	Direction coefficient of gas flow in gas pipeline ij.
$\overline{f}_{i^{gs}j^{ge}}^{g}$	Maximum natural gas flow of gas pipeline ij.
$\overline{f}^{pur}$	Maximum natural gas purchased from IEDS.
$\underline{z}_n / \overline{z}_n$	Maximum/minimum pressure at node n.
M	Sufficiently big number
$x_{i^{gs}j^{ge},k}$	Abscissa breakpoints of domain
$y_{i^{gs}j^{ge},k}$	Ordinate value corresponding to $x_{i^{gs}j^{ge},k}$
$\varepsilon_{i^{gs}j^{ge}} / KP$	Size/number of interval.
$\overline{P}_{i^{CHP}}^{CHP} / \underline{P}_{i^{CHP}}^{CHP}$	Maximum/minimum active power output of CHP units i^{CHP}.
$\overline{q}_{i^{CHP}}^{CHP} / \underline{q}_{i^{CHP}}^{CHP}$	Maximum/minimum reactive power output of CHP units i^{CHP}.
$\eta_{i^{CHP}}^{CHP,ele} / \eta_{i^{CHP}}^{CHP,heat}$	Electrical/heat efficiency of CHP units i^{CHP}.
$\overline{P}_{i^{EHP}}^{EHP} / \overline{f}_{i^{boi}}^{boi}$	Maximum power/gas consumption of EHP i^{EHP}/boiler i^{boi}.
$\eta_{i^{EHP}}^{EHP} / \eta_{i^{boi}}^{boi}$	Heat efficiency of EHPi^{EHP}/boiler i^{boi}.

Variables

$p_{i^{pe},t,s}^{pur} / q_{i^{pe},t,s}^{pur} / f_{i^{pg},t,s}^{pur}$	Active power/reactive power/natural gas purchased from IEDS in scenarios s at hour t.
$p_{b,t,s}^{demand} / q_{b,t,s}^{demand} / f_{n,t,s}^{demand} / h_{t,s}^{demand}$	Active power/reactive power/natural gas/heat load demand response at bus b/at node n in Scenarios s at hour t.
$p_{i^w,t,s}^{w,cut}$	Power curtailment of wind generator i^w in Scenarios s at hour t.
$p_{i^w,t,s}^{w} / q_{i^w,t,s}^{w}$	Scheduled active/reactive power output of wind generatori^w in Scenarios s at hour t.
$q_{i^w,t,s}^{w}$	Scheduled reactive power output of wind generator i^w in Scenarios s at hour t.
$p_{i^{CHP},t,s}^{CHP} / q_{i^{CHP},t,s}^{CHP}$	Active power/reactive power of CHP unit i^{CHP} in Scenarios s at hour t.
$p_{i^{EHP},t,s}^{EHP}$	Power consumption of electrical heat pump i^{EHP} in Scenarios s at hour t.
$p_{i^{ls}j^{le},t,s}^{l} / q_{i^{ls}j^{le},t,s}^{l}$	Active power/reactive power of overhead transmission line ij in Scenarios s at hour t.
$V_{i^{ls},t,s} / V_{j^{le},t,s}$	Voltage at overhead transmission line start/end point in Scenarios s at hour t.
$V_{b,t,s}$	Voltage at bus b in Scenarios s at hour t.
$f_{i^{CHP},t,s}^{CHP}$	Gas consumption of CHP unit i^{CHP} in Scenarios s at hour t.

$f_{i^{boi},t,s}^{boi}$	Gas consumption of boiler i^{boi} in Scenarios s at hour t.
$f_{i^{ge}j^{ge},t,s}^{g}$	Natural gas flow of gas pipeline ij in Scenarios s at hour t.
$z_{i^{gs},t,s}/z_{j^{ge},t,s}$	Gas pressure at gas pipeline start/end point in Scenarios s at hour t.
$z_{n,t,s}$	Gas pressure at node n in Scenarios s at hour t.
$\delta_{i^{gs}j^{ge},t,s}$	Binary variable of gas flow direction of gas pipeline ij in Scenarios s at hour t.
$Z_{i^{gs},t,s}/Z_{j^{ge},t,s}$	Gas pressure square at gas pipeline start/end point in Scenarios s at hour t.
$\mu_{i^{gs}j^{ge},t,s,k}/\xi_{i^{gs}j^{ge},t,s,k}$	Auxiliary continuous/binary variable at kth interval of gas pipeline ij in Scenarios s at hour t.
$h_{i^{CHP},t,s}^{CHP}$	heat output of CHP unit i^{CHP} in Scenarios s at hour t.
$h_{i^{EHP},t,s}^{EHP}$	Heat output of electrical heat pump i^{EHP} in Scenarios s at hour t.
$h_{i^{boi},t,s}^{boi}$	Heat output of boiler i^{boi} in Scenarios s at hour t.

References

1. Wang, S.; Li, Y.; Haque, M. Evidence on the Impact of Winter Heating Policy on Air Pollution and Its Dynamic Changes in North China. *Sustainability* **2019**, *11*, 2728. [CrossRef]
2. Rashid, L.; Lin, L. Foreign Direct Investment in the Power and Energy Sector, Energy Consumption, and Economic Growth: Empirical Evidence from Pakistan. *Sustainability* **2019**, *11*, 192.
3. Ye, J.; Yuan, R. Integrated Natural Gas, Heat, and Power Dispatch Considering Wind Power and Power-to-Gas. *Sustainability* **2018**, *9*, 602. [CrossRef]
4. Wu, J.; Yan, J.; Jia, H. Integrated Energy Systems. *Applied Energy* **2016**, *167*, 155–157. [CrossRef]
5. Marnay, C.; Aki, H.; Hirose, K.; Kwasinski, A.; Ogura, S.; Shinji, T. Japan's Pivot to Resilience: How Two Microgrids Fared after the 2011 Earthquake. *IEEE Power Energy Mag.* **2015**, *13*, 44–57. [CrossRef]
6. Che, L.; Khodayar, M.; Shahidehpour, M. Only Connect: Microgrids for Distribution System Restoration. *IEEE Power Energy Mag.* **2014**, *12*, 70–81.
7. Bie, Z.; Lin, Y.; Li, G. Battling the Extreme: A Study on the Power System Resilience. *Proc. of the IEEE* **2017**, *105*, 1253–1266. [CrossRef]
8. Mathaios, P.; Pierluigi, M. Modeling and Evaluating the Resilience of Critical Electrical Power Infrastructure to Extreme Weather Events. *IEEE Systems Journal* **2017**, *11*, 1733–1742.
9. Lei, S.; Wang, J.; Chen, C.; Hou, Y. Mobile Emergency Generator Pre-Positioning and Real-Time Allocation for Resilient Response to Natural Disasters. *IEEE Trans. Smart Grid* **2018**, *9*, 2030–2041. [CrossRef]
10. Wang, Z.; Chen, B.; Wang, J.; Chen, C. Networked Microgrids for Self-Healing Power Systems. *IEEE Trans. Smart Grid* **2016**, *7*, 310–319. [CrossRef]
11. Sergio, F.; Desta, Z.; Marco, R.; Carlos, M. Impacts of Optimal Energy Storage Deployment and Network Reconfiguration on Renewable Integration Level in Distribution Systems. *Appl. Energy* **2017**, *185*, 44–55.
12. Abdullahi, M.; Li, Y.; Stewart, M. Evaluating System Reliability and Targeted Hardening Strategies of Power Distribution Systems Subjected to Hurricanes. *Reliab. Eng. Syst. Saf.* **2015**, *144*, 319–333.
13. Lin, Y.; Bie, Z. Tri-level Optimal Hardening Plan for a Resilient Distribution System Considering Reconfiguration and DG Islanding. *Applied Energy* **2018**, *210*, 1266–1279. [CrossRef]
14. Shao, C.; Shahidepour, M.; Wang, X.; Wang, X.; Wang, B. Integrated Planning of Electricity and Natural Gas Transportation Systems for Enhancing the Power Grid Resilience. *IEEE Trans. Power Syst.* **2017**, *32*, 4418–4429. [CrossRef]
15. Cong, H.; He, Y.; Wang, X.; Jiang, C. Robust Optimization for Improving Resilience of Integrated Energy Systems with Electricity and Natural Gas Infrastructures. *J. Mod. Power Syst. Clean Energy* **2018**, *6*, 1066–1078. [CrossRef]
16. Amin, S.; Ali, R. An Integrated Steady-State Operation Assessment of Electrical. Natural Gas, and District Heating Networks. *IEEE Trans. Power Syst.* **2016**, *31*, 3636–3647.
17. Long, R.; Liu, J.; Shi, J.; Zhang, J. Coordinated Optimal Operation Method of the Regional Energy Internet. *Sustainability* **2017**, *9*, 848. [CrossRef]
18. Yao, L.; Wang, X.; Qian, T.; Qi, S.; Zhu, C. Robust Day-Ahead Scheduling of Electricity and Natural Gas Systems via a Risk-Averse Adjustable Uncertainty Set Approach. *Sustainability* **2018**, *11*, 3848. [CrossRef]
19. Carlos, M.; Pedro, S. Integrated Power and Natural Gas Model for Energy Adequacy in Short-Term Operation. *IEEE Trans. Power Syst.* **2015**, *30*, 3347–3355.

20. Dupacová, J.; GroWe-Kuska, N.; RoMisch, W. Scenario Reduction in Stochastic Programming an Approach Using Probability Metrics. *Mathe. Program.* **2003**, *3*, 493–511. [CrossRef]
21. Ding, T.; Lin, Y.; Li, G.; Bie, Z. A New Model for Resilient Distribution Systems by Microgrids Formation. *IEEE Trans. Power Syst.* **2017**, *32*, 4145–4147. [CrossRef]
22. He, C.; Wu, L.; Liu, T.; Shahidehpour, M. Robust Co-Optimization Scheduling of Electricity and Natural Gas Systems via ADMM. *IEEE Trans. Sustain. Energy* **2017**, *8*, 658–670. [CrossRef]

sustainability

MDPI

Review

Key Issues and Technical Applications in the Study of Power Markets as the System Adapts to the New Power System in China

Jun Dong [1], Dongran Liu [1], Xihao Dou [1], Bo Li [2], Shiyao Lv [1], Yuzheng Jiang [1] and Tongtao Ma [1,*]

[1] Beijing Key Laboratory of New Energy and Low-Carbon Development, School of Economics and Management, North China Electric Power University, Beijing 102206, China; dongjun@ncepu.edu.cn (J.D.); 1182106008@ncepu.edu.cn (D.L.); xijicc@126.com (X.D.); l33189sy@163.com (S.L.); 120202206107@ncepu.edu.cn (Y.J.)

[2] The State Key Laboratory of Power Transmission Equipment & System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing 400044, China; boli9301@163.com

* Correspondence: mtt@ncepu.edu.cn

Abstract: To reach the "30·60" decarbonization target (where carbon emissions start declining in 2030 and reach net zero in 2060), China is restructuring its power system to a new energy-based one. Given this new situation, this paper reviews previous studies on the power market and highlights key issues for future research as we seek to adapt to the new power system (NPS). Based on a systematic literature review, papers on the operational efficiency of the power market, participants' bidding strategies and market supervision were identified. In a further step, papers with high relevance were analyzed in more detail. Then, key studies that focused on market trading under China's new power system were picked out for further discussion. New studies were searched for that pertained to new energy mechanisms and bidding, the transition from coal-fired power, flexible resources and the technical applications of simulations. The quantitative analysis supports the construction of a basic paradigm for the study of power markets that is suitable for the new power system. Finally, the theoretical basis and application suggestions for power market simulations are introduced. This study summarized the existing research on the power market and further explored the key issues relating to the power market as it adapts to the NPS, hoping to inspire better research into China's power sector, and promote safe, low-carbon, and sustainable development in China's power industry.

Keywords: new power system (NPS); power market; new energy; flexible resources; experimental economics; agent-based computational economics

Citation: Dong, J.; Liu, D.; Dou, X.; Li, B.; Lv, S.; Jiang, Y.; Ma, T. Key Issues and Technical Applications in the Study of Power Markets as the System Adapts to the New Power System in China. *Sustainability* **2021**, *13*, 13409. https://doi.org/10.3390/su132313409

Academic Editor: Attila Bai

Received: 19 October 2021
Accepted: 1 December 2021
Published: 3 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The reform of a power system is a major measure and an important part of China's energy restructuring. Since the second round of the electricity reform that was launched in 2015, China's power market has been further liberalized [1]. Meanwhile, the Chinese government has pledged to halt further CO_2 emissions by 2030 and become carbon neutral by 2060 [2]. This requires higher levels of sustainable development in the power system. In recent years, China has been accelerating the replacement of fossil fuels. The proportion of new energy in primary energy consumption continues to rise. Renewable energy power output reached 2.2 trillion kilowatt-hours in 2020, accounting for 29.5% of national electricity consumption. China has fulfilled its commitment that non-fossil energy consumption must account for at least 15% of the primary energy consumption by 2020 as scheduled [3]. In March 2021, China's Central Financial and Economic Affairs Commission called for the building of a new power system (NPS) with new energy as the main source. This government directive is vital to the power industry's direction of development. Against this background, it is particularly important to study the new power system and the power market trading systems that feature a high proportion of renewable energy.

The new power system, compared with traditional power systems, functions in new industrial models. Every part of the power system, including power generation, grid operation, and power load control and energy storage, is changing. On the power supply side, renewable energy will gradually become the main source of electricity generation. On the consumption side, many prosumers are emerging. With regard to grid operation, large-sized grids still dominate, while multiple forms of grids will coexist. From the perspective of the whole system, the operating mechanism and balancing mode will undergo profound changes. The new power system is a clean and low-carbon, safe and controllable, flexible and efficient, intelligent and friendly, open and interactive system. The change of the power market will follow these trends.

At present, China's power market is based on the provincial power market, and power trading is based on medium- and long-term contracts. A spot trading pilot for power is being promoted in the provincial market. New energy power generation enterprises have priority clearance rights in some provinces. That means that they are allowed to sell electricity directly at the volume that they bid, or that they can even sell power without applying for trading volume and prices before the transaction. Their power output will be bought at market prices without any quoting before the trade. Meanwhile, researchers are accelerating the search for strategies to promote regional power markets, which can prompt new energy companies to participate further in the market.

In terms of thermal power, the utilization hours of traditional coal power units are declining as the proportion of new energy is rising [4]. Government requirements for carbon emission reduction are also squeezing the living space of thermal power, as burning clean coal will surely make it more expensive. At present, China's coal-electric installation accounts for about 50% of the power system. Its survival is not only related to the development of the power system, but also related to the social economy and people's livelihoods. We need to pay enough attention to the survival of thermal power. Problems also exist for new energy and the flexible resources that are participating in the power market. The value of flexible resources has not been clarified, and there is no corresponding market mechanism to ensure that large quantities of electricity generated from renewable energy resources can be consumed at market prices. The market mechanisms of auxiliary services also needs to be improved urgently.

Due to the two-track characteristics of China's power market, with both government planning and active markets, a series of issues concerning market operation and convergence are emerging, including the disposal of imbalanced funds [5,6], power market supervision/market power testing [7–9], demand response [10], renewable energy participating in the electricity market [11], the synergy between the electricity market and the carbon market, and the construction of future financial markets for power and market mechanisms for standby capacity. There is no mature international experience that can be drawn on that would help solve these problems.

The goal of this systematic review is to find solutions to develop the Chinese power market, combining China's actual situation with successful foreign experience in power market construction and operation. Therefore, a basic paradigm for the study of power markets is identified by analyzing primary studies that are related to issues on market entities and operational mechanisms. Then, key issues pertaining to power market construction as it relates to China's new power system (NPS) are discussed. To achieve these objectives, we conducted the research guided by the following questions:

1 What are the hot topics in primary studies regarding the construction and operation of power markets, and how do these topics interconnect with each other?
2 What are the key issues in the study of power markets as the system seeks to adapt to the NPS, which will feature a high proportion of new energy?
3 What is the basic paradigm for the study of power markets that would be suitable for the NPS?
4 Which theoretical basis can help simulate the power market to further research into power market mechanisms under the NPS?

5 From which dimensions should simulation modeling be established and how should researchers set related parameters?

The review is structured as follows. Section 2 describes the methodology. Section 3 introduces the analysis of general issues that concern researchers in the study of power markets according to primary research. Sections 4 and 5 discuss key points related to the NPS power market from both theoretical and practical perspectives. In Section 4, key issues and basic paradigm for studying the power market under the NPS are presented, while specific suggestions for the applications of power market simulations are provided in Section 5. Section 6 gives the conclusions and implications.

2. Methodology

The systematic literature review (SLR) framework [12] is adopted in this work to identify, evaluate, interpret and describe the existing body of knowledge on power markets. We focus especially on the operation and supervision of the power market and the bidding strategies adopted by market participants that have caught the most attention. This study follows PRISMA guidance [13] for articles so that we can formulate robust and reproducible research. Figure 1 illustrates the design and logical framework of the respective literature review. The research questions have been outlined in Section 1.

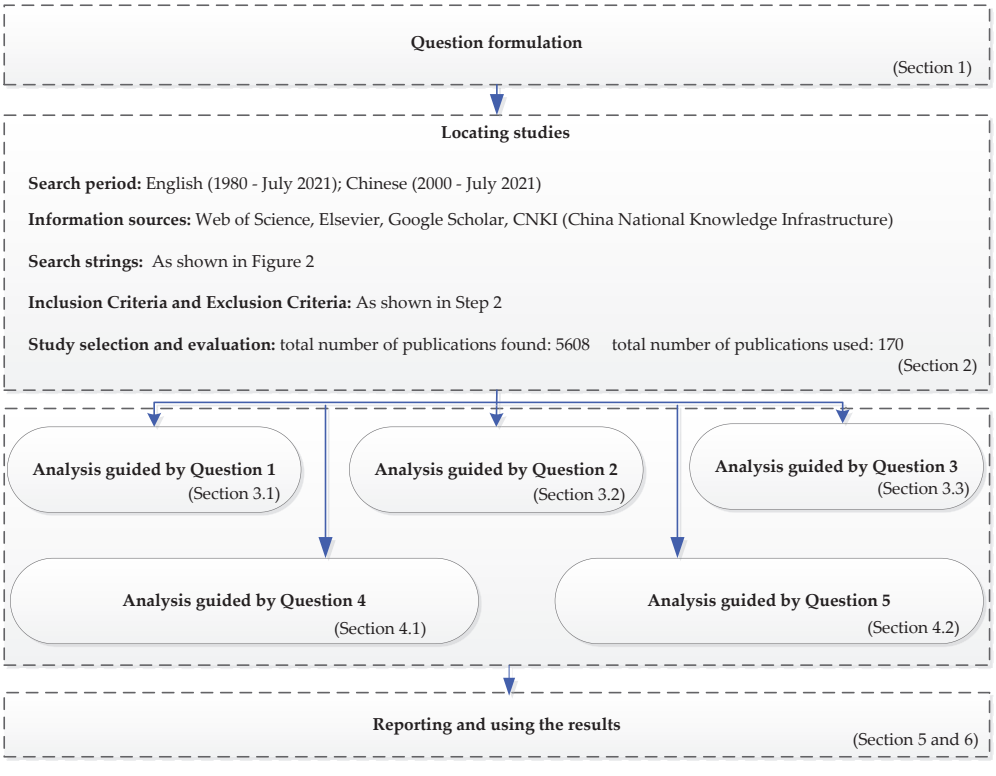


Figure 1. Review Design for the Study of Power Markets as Researchers Adapt to the NPS in China.

We carried out the search process in five steps, including identifying the information sources, defining the eligibility criteria, composing the search strategy and selection processes, which aims to present unbiased, rigorous, and auditable results.

Step 1: Classify the search strings groups and identify the information sources.

This review has searched for both previous studies of the power market and new studies that show how research is adapting to the NPS, focusing on the research objectives and main research questions. Older studies on the power market were searched for with three groups of strings, including "power marker operation", "bidding strategy" and "power market supervision". New studies were searched for with four groups of strings with terms that include "high renewable energy", "coal-fired power units", "flexible resources" and "technical application". The search criteria groups are shown in Figure 2.

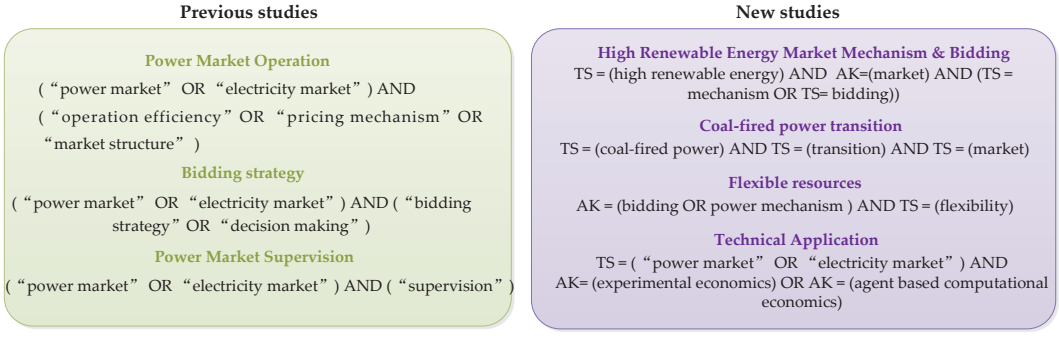


Figure 2. SLR search criteria.

Meanwhile, relevant databases that could be searched were also determined in this stage. We collected literature primarily from Web of Science, Elsevier, Google Scholar and CNKI (China National Knowledge Infrastructure). Other websites of authoritative research institutions that we used included the National Renewable Energy Laboratory and the Lawrence Berkeley National Laboratory. Collectively, these databases covered the main published papers/reports, unpublished manuscripts and conference papers (full-text access) that focused on the power markets of China and foreign countries.

Step 2: Set eligibility criteria for inclusion and exclusion of identified papers.

Eligibility criteria were determined in this phase to exclude ineligible studies from this review. There were several standards for studies to be included or excluded:

- Papers should be in social sciences, business and economics, energy, or relative fields.
- The title or subject must be strictly matched with the search strings. For example, papers that focused on storage efficiency in the power market [14] could be found when we searched for "power market operation efficiency". Alternatively, articles that used experimental methods with neural networks [15] could also be found when we searched for "experimental economics". These kinds of articles should be excluded.
- Papers or reports must be written in English or Chinese. Literature in other languages was excluded in this review.
- English articles must be published between 1980 and 2021 (access in July 2021) and Chinese articles must be published between 2000 and 2021.
- Articles must have at least one keyword in the title or abstract, or cover the topic in the full article.

Step 3: Search and export the studies on power markets.

Based on the search criteria groups in Figure 2, suitable combinations of Boolean variables, such as "AND", "OR", and "NOT", were employed to find the suitable literature. Because there are a range of topics that pertain to the power market, which has been studied extensively, "precise search" was used during the search process. The rankings of the literature were based on their relevance to the strings, which is helpful for further selection. Then, the identified studies were exported for deduplication in Step 4. The detailed search processes are as follows:

Firstly, papers published in refereed journals were searched for using the "Power Market Operation" criteria, which resulted in 1023 primary papers being selected from the business and economics fields from the Web of Science. With the same search strings, 411 papers were found in Elsevier, and 307 papers were found in the CNKI.

Secondly, the term "Bidding strategy", searched for in Web of Science, resulted in 896 papers that were indexed in the fields of business and economics. A further 1460 papers were found in the field of economics and energy in Elsevier, while 177 papers were identified in the CNKI, which was much lower than the number of that in the English databases.

Thirdly, studies on "power market supervision" were searched for using the strings. In this step, 180 papers were found in Web of Science, and, of these, 176 papers satisfied the inclusion criteria. Only three of 72 papers were selected from Elsevier, however. We found 121 articles that were published in the Chinese database.

Then, new studies regarding renewable energy, coal-fired power units, and flexible resources and "technical applications" were searched for using the same method. As a result, 187, 54, 132 and 51 of them were identified in Web of Science, respectively, while 13, 2, 2 and 48 of them were selected in Elsevier, and 18, 5, 5 and 66 of them were identified in CNKI. All eligible papers were exported for further deduplication.

Step 4: Remove duplicates of searched papers.

Because the search process was carried out using different groups of strings, Citespace [16] was used to remove the duplicates in this review. We removed 189 previous studies in this phase according to the names of the exported .txt files of all of the suitable literature.

Step 5: Choosing eligible and relevant studies.

With the assistance of the relevance ranking function in the databases, the last pages with less relevant articles were excluded from further selection. After selecting the most relevant articles, we screened their titles, abstracts, and keywords. The illegible papers were excluded based on the inclusion and exclusion criteria. Then, we further read the conclusion and implications to identify the most important pieces of literature. Finally, through careful reading of the full text, we used the "snowballing" method to find articles that were not identified in the search. This meant that we found a wide range of literature. We removed 194 articles that were duplicates. The screening led to the exclusion of 4256 publications that did not meet the eligibility criteria. There were 893 old studies and 232 new studies that were further analyzed. Finally, a total of 170 studies were included in the review. To express the search process more clearly, the search strategy flow is shown in Figure 3.

After reviewing research hotspots and solutions of power markets around the world, general issues in the study of power markets were summarized. Then key issues in the study of power markets as it pertains NPS were put forward. This involves mechanisms and strategies for renewable energy, the survival and transformation of coal-fired power plants, and the value realization for flexible resources. After theoretical analysis, technical applications of experimental economics and agent-based computational economics in the study of NPS as it relates to the power market are presented. Further, simulation and parameters in terms of power system, power market and market participants are implemented.

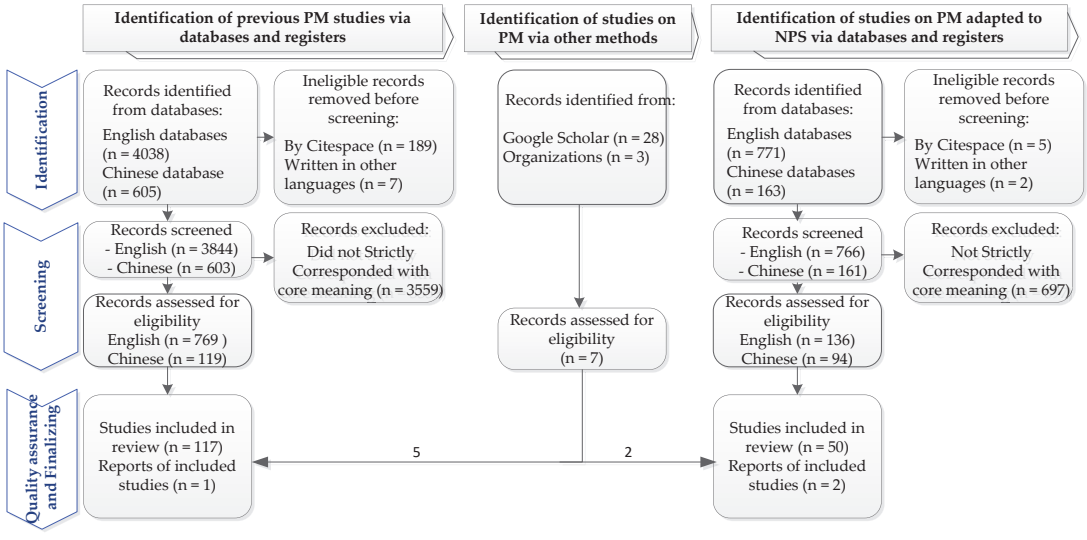


Figure 3. Search process.

3. Key Issues and the Basic Paradigm of Research into New Power Systems in the Power Market

After merging and splitting the spot sales and future trading markets [17], the Nordic power market finally grew into a situation where financial markets for power and spot trading complemented each other. Following the NETA [18] (New Electricity Trading Arrangements) and the BETTA [18] (British Electricity Trading and Transmission Arrangements), the UK has gradually liberalized the market for large and small industrial users and achieved efficient cross-border trading of electricity. With a high proportion of nuclear power and relatively low standard electricity prices, France has relevant experience in regulating the market pricing mechanisms that are adapted to a high proportion of nuclear power. In addition, issues related to electricity markets in Australia [19], California [20], Brazil [21], New Zealand [22] and Germany [23], have also been hotly debated in academia. It can be seen that the construction of electric power markets varies from country to country, due to the unique characteristics of power systems that results from different national conditions.

At present, China's power market is constructed around the provincial electricity energy market, with power transactions relying on medium- and long-term contracts. Research on China's power market is focused on the spot trading of electricity, and mainly revolves around the design of spot market price mechanisms, the bidding games of market subjects, and the interface mechanism between spot trading, medium- and long-term contracts, and auxiliary services. The further development of the power selling side, with energy storage equipment operators and other emerging entities participating in the electric sector and auxiliary services market, is also attracting the attention of scholars. Some other emerging hot research issues include demand response prices or incentives, power financial markets, electricity retail market prices and package design, as well as market risk and market assessment.

3.1. General Issues in the Study of Power Markets

A healthy power market depends on sound market mechanisms, competitive market subjects and strong market supervision. A large amount of research and practice has therefore been carried out on the operational efficiency of the power market, market subject strategies and market supervision and evaluation. Detailed research issues are shown in Figure 4.

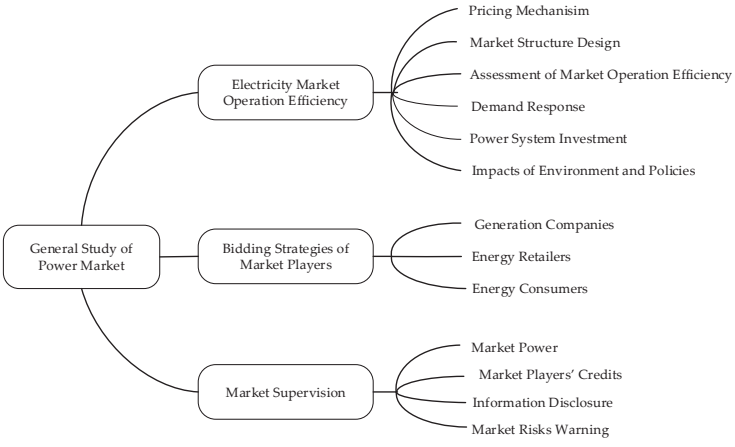


Figure 4. General Issues in the Worldwide Study of Power Markets.

3.1.1. Operational Efficiency of the Power Market

This part reviewed literature which studied market mechanisms or operational efficiency. Therefore, such papers that studied energy efficiency [24] or the operational efficiency of power plants [25] in a market were excluded. After the selection and evaluation of search results with "power market operation efficiency" strings, only 70 papers in the databases corresponded tightly with this topic. In addition, five reports were added into the pool for further analysis.

Around the time of the first power system reform in China, the operational efficiency of the power market was mainly analyzed theoretically and empirically from the perspective of market power [26]. Some scholars also studied the key problems facing the operational efficiency of the power market from a market structure perspective [27]. Other scholars put forward systemic theories and evaluation methods for the operational efficiency of the power market [28].

Apart from evaluating market efficiency [19] based on actual market operation data, foreign scholars have also studied the factors affecting the efficiency of market operations, such as price mechanism design, market structure design, market power, demand response, power system investment and environmental policies, as shown in Table 1. These studies provide a powerful reference for designing new market systems or behaviors.

Table 1. The Study of Power Markets: General Issues and References.

Research on the Operation of the Power Market		References
Price mechanism design	Regional marginal pricing	[29]
	Locational marginal pricing	[30]
	Capacity pricing mechanism	[31–34]
	Ancillary service pricing mechanism	[35]
	Combined clearing of power energy and ancillary service	[36]
	Distributed trading mechanism	[37]
Market structure design		[38]
Market power	Evaluating metrics	[19,39–46]
	Evaluating methods	[47–50]
Demand response		[51–54]
Market efficiency assessment		[19]
Investment in power systems		[55,56]
Environmental policy		[57–59]

Pricing mechanisms are the core of market operation design. A lot of practice and research on the applicability and effect of pricing mechanisms has been carried out by the industrial sector and in academia. For example, the original pricing mechanism of the Texas power market in the US was based on regional marginal pricing in the early stage of market construction, but it changed to a node marginal pricing mechanism [30] in 2010 as the market operating efficiency decreased with the deepening of power system congestion in the region. Focusing on the redispatching market in Europe, some scholars have quantitatively assessed the impact of regional pricing and node tariff mechanisms on the operational efficiency of the power system in a market environment where there is incomplete competition [60]. Such studies provide useful references for the design of electricity pricing mechanisms. According to other investigations, pricing mechanisms in China vary from province to province. Gansu and Jiangsu implement the regional marginal price [29], and the clearing price in the province is the same. Shanxi, Zhejiang and Guangdong, on the other hand, implement the locational marginal price, meaning that the clearing prices in different nodes fluctuate. Based on a full development of the electric energy market, PJM, California and other power markets in the Americas began to explore the key role [31–34] that capacity pricing mechanisms play in balancing between strengthening the reliability of power systems and increasing income for power producers. In addition, because distributed power supply has developed rapidly, reaching a large scale in recent years, its trading mechanisms [37] have also aroused wide concern.

Demand response is the process whereby end users adjust their own electricity consumption behavior in response to time-based rates. End users are provided with compensation [61] for reducing power usage when wholesale market prices are high or the reliability of power systems is threatened. At present, foreign research on demand response is focused on resident demand response strategies [62,63], business user demand response models and optimization methods [54], as well as on microgrids. Some other emerging flexible resources are also seen as key elements [64] of demand response, including flexible loads [65–67], as well as resource aggregators (RA) [54] that combine distributed power supplies, energy storage systems, and controlled loads.

The optimization of investment planning in power systems is the subject of power trading research at the macro level. Investment is also an important means of building a new power system in stages. However, most research on investment planning in power systems is focused on the physical characteristics of the electricity system, while influences from market factors are ignored by most studies. Typical simulation software for investment planning for power systems include Energy Plan [68], SWITCH [69–71], NEPLAN [72] and PLEXOS [73]. Only PLEXOS takes into account market factors (including recent and real-time market price movements, changes in demand, market risks, etc.), ancillary services and system flexibility. It can also simulate carbon trading management, showing the near-real re-emergence of multi-market coupling and system-to-market coupling. However, other investment planning simulation software are based on the technical and economic characteristics of the power system. They take into account environmental or policy impact, rather than the impact of inner-market factors on system investment planning. Such software specializes in long-term planning and optimization of the energy structure of the system.

3.1.2. Bidding Strategy of Market Participants

The quotation strategy is where the feedback from market participants corresponds to the power market mechanism, and it is also an effective test of the market mechanism. In early research, most scholars focused on power producers. They judged their winning probability by predicting the market clearance price [74], or their competitor's strategy [75], and drew their own bidding strategy based on the probability theory method. The game theory revenue matrix and incomplete competition game models were also the tools that they used to form their own bidding strategies.

During the first round of power system reform in China, a power generation side competition model was formed because the state-owned power grid company is the only buyer. Around 2002, many scholars studied the bidding strategies of power generation enterprises that were based on coal and other fossil fuels, and obtained rich research results. The quotation strategies that were used in those studies were mainly based on cost analysis [76], electricity price forecasts [77], optimal methods [78] and game theory [79]. After that, some scholars used power market simulation experiments to study the bidding strategies and bidding trading systems of power producers based on EWA algorithms [80], Repast algorithms [81], Agents [82,83], and MAS [84]. However, those strategies only served traditional thermal power plants, and the target market was limited to provincial electricity energy markets. The market is changing as China pushes ahead with a new round of power system reform. The power selling side of the market is opening. The types of market participant are increasing, and trading variety is growing. The focus of current research has moved to the portfolio strategies of market participants (represented by power producers, electricity sellers and large users) to cope with medium- and long-term contracts (physical and financial), spot trading, ancillary services and carbon markets. With the progress of technology, research on new market subjects has gradually increased. The bidding strategies of virtual power plants or hybrid power plants [85], microgrids, load aggregators [86], prosumers, energy storage and electric vehicles are becoming hot issues. Some studies have identified trading strategies for a certain type of market, and some researchers have proposed joint optimization strategies for multiple types of markets. Common research methods include two-stage robust optimization, which takes into account risks, and the Markov decision process [87]. Typical references are as shown in Table 2.

Table 2. Typical Literature on Market Players' Bidding Strategies.

Market Players	Strategy Target	References
Traditional power producers	Medium- and long-term market	[88]
	Day-ahead market	[89]
	Real-time market	[90]
	Day-ahead market and real-time market combined	[91]
Electricity retailer	Day-ahead market	[92,93]
Virtual power plant	Day-ahead market and ancillary services	[94,95]
Microgrids	Day-ahead market	[96]
	Day-ahead market and real-time market combined	[91]
	Electric energy market and ancillary services	[97,98]
Prosumer	Day-ahead market	[86]
	Day-ahead market and ancillary services	[99]
Energy storage	Day-ahead market	[100]
	Real-time market and ancillary services	[101]
Electric vehicles	Day-ahead market	[102]
	Day-ahead market and ancillary services	[103]

As can be seen from the table above, most of this research is focused on the electric energy market. The joint strategy for real-time markets and ancillary services is an especially hot research issue. There is little research that focuses on a single secondary service and capacity market strategy. It is worth noting that the market participation strategy for new thermal power units with flexible regulation capabilities has not yet been studied. However, the right strategy is crucial for thermal power units to survive during market reforms. Under new situations, when developing trading strategies, thermal power units should consider not only the income from selling electricity, but also potential benefits from providing ancillary services or spare capacity through flexible resources. Risk aversion through financial contracts [104–106], arbitrage and linkage strategies with the carbon market should also be taken into account. For instance, Bjorgan analyzed the impacts of resources constraints when utilizing a portfolio of contracts to manage adverse markets risks [104]. Furthermore, T. S. Chung introduced a new electricity forward contract that

included a bilateral financial contract that allows both the seller and buyer to cope with market price fluctuations [105]. Focusing on the forward risk premia, Spodniak studied the differences between the trading price of electricity in forward contracts and the spot prices to hedge transmission risks [106]. In addition, the feasibility of trading strategies that combine new energy, thermal power, gas and energy storage and other sources is also worth exploring.

3.1.3. Power Market Regulation

A sound electricity market requires regulation. Power market regulators around the world perform nearly the same functions, including detecting potential risks to the healthy operation of the market (usually to detect whether market participants behave in accordance with market rules, standards and processes), judging whether market participants use market power appropriately, looking for defects in market rules and detecting how regional transmission organizations (RTO) influence market operations [107]. Market regulators usually exist in four major forms: (1) a single regulation department that is embedded in the system scheduling operational structure, (2) a single regulatory body that is independent of the system scheduling agency, (3) a single regulatory body that is controlled by the government, and (4) a dual regulatory body, one that belongs to the system scheduling agency, and the other which is independent. Typical regulators include the Office of Gas and Electricity Markets (Ofgem) in the UK, the European Securities and Markets Authority (ESMA), the Australian Energy Market Commission (AEMC) and the Australian Energy Regulator (AER), as well as the Federal Energy Regulatory Commission (FERC) in the US.

A variety of operational risks within the power market have been studied by academics, including supply-side market structure risks, mechanism design risks, the credit risks of market participants, power system security risks and external environmental risks. As an important index of market supervision, market power has attracted much attention from researchers.

In research outside of China, S. Prabhakar Karthikeyan [108] and his team comprehensively expounded a range of market power assessment indicators, common simulation tools, as well as common theories and algorithms. After analyzing typical cases in different countries, they proposed measures to control market power. Common indicators of market power assessment include the Herfindahl–Hirschman index (HHI) [39], the Lerner index (LI) [40,41], the must-run ratio (MRR) [42] and the residual supply index [43]. In addition, the system interchange capacity (SIC) [44], the location privilege (LP) surplus deviation index [45], the contribution congestion factor matrix [46] and other indicators have been put forward and applied by some scholars. Some other scholars claim that HHI evaluation has shortcomings, and have thus put forward a method to evaluate market power in the electricity sector based on oil market simulation [109]. In terms of regions, electricity markets in California of US [110,111], Wales (UK) [112], Germany [113], India [114], and Iran [114] has been surveyed to evaluate their market power.

The research on market power by Chinese scholars can be divided into three major types. For the first type, researchers presented a review of market power regulation and mitigation measures in foreign electricity markets, such as North America and northern Europe. For the second type, which is based on market power indicators, researchers carried out market efficiency assessments and proposed risk warnings for power producers. This research was conducted before the second round of power reform. Some were based on principal component analysis [47], and some were based on game theory [48,49]. Only Zhang Fuqiang [50] and his team considered the strategic behavior of power producers and used intelligent simulations to evaluate market power in different settlement modes. A third type of research has emerged since China deregulated its retail power market in the second round of power reform. This type focuses on the market power of various participants [115], including electricity producers [116], retailers and users. Some of the recent studies help electricity producers discover spot markets using support

vector machines (SVM). Some other research has been conducted on the breakdown of medium- and long-term contracts, taking into account the market power of the parties to the contracts [117].

In order to help market participants avoid risks, market designers provide financial instruments based on the operating boundaries of power systems when designing price mechanisms, such as contracts for differences (CFD), transmission congestion contracts (TCCs), and financial transmission rights (FTRs). In addition, some scholars have proposed the concept of an operational health degree for power markets [118], which can be used to measure the electricity market so as to help government departments, system operators (e.g., independent system operators, ISOs) and market management committees monitor market conditions in a timely manner and take appropriate measures if necessary.

3.2. Key Issues in the Study of Power Markets as the System Adapts to the New Power System

As a major measure for reaching China's carbon reduction targets, constructing a new power system based on renewable energy is a complex systematic project that involves various sides in the electricity sector, including electricity producers, grids and load control sources. There will be new requirements for every part of the power system. Electricity generation units should be more flexible and capable of coordinating with the grid. The grid should be optimized for the configuration of clean energy. End users are expected to interact flexibly with the grid to help ensure power security. The market mechanism is the basic guarantee when constructing a new power system. Furthermore, the operation of new energy sources and thermal power units with regulatory capabilities and flexible loads [119], needs to be reinterpreted under the new situation. Therefore, research on how the power market needs to adapt to the new power system should be focused on building market mechanisms to promote the participation of new energy in the market, helping thermal power units survive during periods of market reform and discovering flexible resources.

3.2.1. Mechanisms and Strategies for Renewable Energy Producers Participating in the Power Market

Governments in different countries have introduced price mechanisms [120–122] and incentive policies [123] to encourage renewable energy producers to produce more. The fixed electricity prices and tradable green certification mechanisms [124] that were introduced by the EU have been successful at stimulating renewable power generation. Germany has introduced a fixed feed-in tariff mechanism, a market premium mechanism and a bidding system for the different stages of renewable energy development. The German government can adjust its fiscal incentives flexibly considering the downward trend of renewable energy costs. With these policies, Germany is seeing a rapid growth in renewable energy production. The United States and Australia promote new energy through renewable energy quota systems and renewable energy certification mechanisms. They guarantee the trading volume of wind and photovoltaic power through power purchase agreements (PPAs), and encourage renewable energy companies to directly participate in the wholesale electricity market based on contracts for difference (CFDs).

Problems appear when renewable energy producers participate in the electricity market. Firstly, the wholesale price in the power market has fallen as the scale of new energy generation has rapidly expanded. Falling prices have suppressed investment from conventional power suppliers, and this may harm the safety and stability of the whole power system. Secondly, the unstable output from new energy generation makes the markets' clearing prices more volatile. The gap between clearing prices in day-ahead markets and real-time markets widens, adding to the uncertainty regarding the returns gained by conventional thermal power plants. Thirdly, the climbing market share of renewable energy poses challenges to the design of market mechanisms.

Research into the power market is currently focusing on how to innovate the spot market price mechanism, auxiliary service price mechanism and capacity price mechanism to adapt to the new market, which now features a high proportion of new energy. Many

scholars are proposing new market mechanisms for auxiliary services that are adapted to the new situation [125]. Godoy and his team are among these researchers. They designed a cost-based ancillary services market mechanism and verified its technical and economic feasibility through a simulation of the Chilean [126] electricity market. Some scholars studied the joint clearance of the electricity spot market and the auxiliary services market [127–129], while some others have proposed mechanisms to prompt flexible resource providers, including energy storage [130] and distributed power supply [131], to participate in the secondary services market.

The optimization of the power system and microgrid structure in the green power market [132] is also attracting the attention of researchers. S. Arango Aramburo [133] has studied the impact of the access to new energy on the investment cycle of power system capacity. A. Hasankhani [134] compared the differences between the scale of renewable energy in smart microgrids before and after participating in the power market, and optimized the energy management of microgrids. The issue of how demand response contributes to promoting new energy electricity connections to the grid [135] has also been explored, and researchers weighed the relationship between the scale of demand response and social welfare.

3.2.2. The Survival and Transformation of Coal-Fired Power Units

The rapid expansion of new energy power generation is squeezing out traditional coal-fired power plants. Firstly, coal-fired power generators have to keep their overall utilization hours at low levels because large quantities of electricity generated from renewable energy sources must be consumed before thermal power. Meanwhile, coal power plants have to adjust their own output to make up for new energy generation and maintain the stable operation of the grid. When electricity generated by renewable resources is sufficient, thermal power units need to reduce their output. When renewable power plants are unable to support power consumption at peak times, thermal electric plants need to increase their generation. Thus, the operational cost of coal-fired power plants is rising. Secondly, the wholesale price of electricity is falling as a result of the rapid expansion of renewable power generation. Coal-fired power plants are losing competitiveness due to falling electricity prices. Some thermal power plants even report severe losses in regions with high prices of coal. Thirdly, the environmental cost of coal-fired power plants is also rising because China is promoting green and low-carbon development through the carbon trading market, which further cuts the competitiveness of thermal power plants. Fourthly, the complexity and cost of decision making is surging as coal-fired power plants adapt to the power and carbon market linkage system. It is a new challenge for traditional power suppliers which have just become accustomed to the market environment in the past few years.

However, thermal power accounts for more than 60% [136] of China's total installed power capacity, with nearly 50% generated by coal-fired power plants. Therefore, the thermal power sector plays an important role in China's power security. Meanwhile, China is limited to just relying on energy storage and demand-side resources to provide auxiliary services for the new power system. Flexible thermal power units' capacity for adjusting electricity consuming load and frequency is worth exploring, especially with the target of transforming 200 million KW of coal-fired power units into flexible resources during China's 14th five-year plan period [137]. In addition, the survival and development of coal-fired power units is also worth an in-depth discussion from the perspective of people's livelihood. Therefore, whether coal-fired power plants can be successfully transformed through technological improvements and reasonable market mechanisms is a key issue for the construction of the new power system.

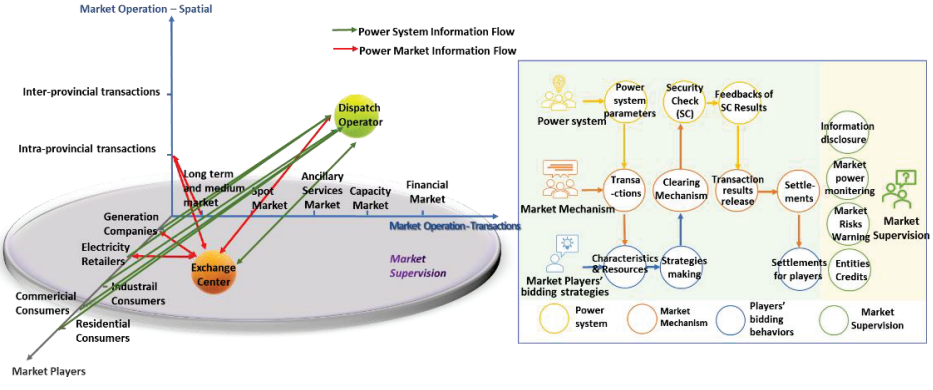
3.2.3. Value Realization and Supporting Mechanism for Flexible Resources

The value of flexible resources in the new power system is expected to be realized using market mechanisms. More and more attention is being paid to the value of flexible resources, including energy storage, electric vehicles, demand-side resources and thermal

power plants with flexible regulation capabilities. However, current research on flexible resources is generally limited to optimization planning or resource allocation within power systems, discussing, for example, the system optimization of flexible resources for power supply, grid operation and load management. Another limited research issue is studying how to optimize the configuration of flexible resources for adequate demand response [138]. Only a few studies have explored the value [139,140] of flexible resources in the electricity market, which lack systematic quantitative analysis. Significantly, there are no studies on value modeling, market system design, and simulation deduction and decision analysis when it comes to flexible resources that will operate under market conditions within the new power system. To fill in this gap, we need to recognize the value of flexible resources and design a market mechanism that can clearly reflect price signals, so as to help renewable energy resources participate in the electricity market.

3.3. Basic Paradigm for the Study of Power Markets under the New Situation

A power market is a complex economic system, with branches forming where the system couples with others. To simplify this study, we built a framework with three dimensions, including market operations, market player strategies and market supervision, to outline a basic paradigm for power market research under the new situation. As shown in Figure 5a, the power market operates through the interaction of different sub-markets. Depending on its different functions, transaction times and spaces, the power market can also be divided into different submarkets. Market players make strategic decisions in different scenarios based on the market mechanisms, which leads to diverse market results. The interaction between market participants and the market operation mechanisms affects the health of the power market. Connecting all kinds of market players, the power trading center is responsible for publishing market information, organizing transactions of different types, managing the credit of market participants and assessing market risk. The dispatch center checks the safety of the power system according to the clearing information of the trading center, and forms the final clearing and dispatch orders. The market supervision department supervises the credit of market players, information disclosure, market force detection and risk warning. As shown in Figure 5b, the power system, market mechanisms and strategies of market participants are key factors that help us understand market systems. These three aspects will be further discussed in the following part of this paper.



(**a**) (**b**)

Figure 5. Basic paradigm for the study of power markets: (**a**) research dimensions of power market; (**b**) general process of power transactions.

Typical types of power markets include the electric energy market, the auxiliary services market, the capacity market, the power financial market and the transmission

rights market. The electric energy market can be further divided into the wholesale market and the retail market according to the trading volume. In the wholesale market, participants can trade through the trading center, or they are allowed to make direct transactions through bilateral negotiations.

From the perspective of the time dimension, the electric energy market can be divided into the medium- and long-term market, and the spot trading market. The medium- and long-term power market is designed to avoid market risks and maintain the safety of power systems. Transactions in the medium- and long-term market can be made either physically or in a financial manner. From the perspective of the spatial dimension, power markets can be classified as provincial markets, regional markets, national markets and transnational markets, according to the different structures of the power grids. Different countries choose different trading modes according to the unique structures of their grids.

Auxiliary services are helpful for maintaining the reliable operations of the grid and ensuring the quality of electricity transmission. The ancillary services market is an important complement to the spot market. As China advances the reform of the power market, a key issue is how it can determine the supply boundaries [141] and supply costs of auxiliary services within the auxiliary services market. Another hot issues is how the supply costs would change [142] with different operational strategies.

One study [141] is a review of the international research results from studies of auxiliary service standby mechanisms that are based on the actual operations of regional power markets, covering the types of auxiliary service standby mechanisms, the coordination and optimization mechanisms of the standby power market, and methods for assessing backup demand. If the auxiliary services market is regarded as the safety mechanism to ensure the stability of power system in real time, then the capacity market is the mechanism that ensures the stable supply of electricity over the medium and long term. The key issue when studying the capacity market is how to encourage market participants to build effective power generation units, improve investment efficiency and ensure the stable supply of electricity [143] under the premise of reducing investment costs. Financial derivatives are important tools for risk management in the power market. In the power markets of typical countries, market participants use contracts for difference (CFDs), power futures and power options products and other financial derivatives [144,145] to suppress the risks from fluctuating spot prices.

With the urgent need to build a new power system and market trading system, it is especially necessary to explore the market mechanisms, capacity mechanisms or capacity markets [146], and market interface mechanisms that can be adapted to China's own conditions. Under China's new power system, based as it is on renewable energy, the important value of flexible resources, an emerging market resource, needs not only to be quantified, but also to be realized under the market mechanisms. Therefore, whether to design a new market mechanism for renewable energy and flexible resources, how to design and connect the mechanisms, what effect the new mechanism could bring, and other questions are issues worth exploring. Further, the research on the portfolio and bidding strategies of market participants, including traditional thermal power capacities and emerging renewable energy and flexible resources, is of great significance to help market players integrate into the new market conditions and to promote the healthy development of the market.

4. The Theoretical Basis and Technical Applications of Power Market Simulations to Help Markets Adapt to the New Power System

The effects of designing market mechanisms and conducting research into market participants' strategies need to be verified under specific market circumstances. Power market simulation technology is an effective means of avoiding the loss of benefits that the strategy's design may cause to the actual system. The power market is an open and dynamic, complex and adaptive system with parallel interactions and nonlinear dynamics. Each market participant is independent but can interact with any other. Traditional simulation tools that are based on unified decision making are not suitable for complex power

systems with decentralized decision making as the basic feature [147]. The simulation of power markets based on experimental economics and computational economics not only considers the rational and irrational behavior that may occur in the real environment, but can also simulate people's limited rational decision making and realize complex computations using intelligent agent technology. Therefore, it is an effective tool for studying market mechanisms and the strategies of market participants.

4.1. The Theoretical Basis of Power Market Simulation

4.1.1. Principles and Applications of Experimental Economics

Experimental economics (EE) is a branch of economics that uses data from experiments to address economic questions. Experimental economics was developed in the second half of the 20th century and matured during the 1960s and 1980s. Drawing on the experimental methods that are used in the natural sciences, EE simulates the actual decision making that goes into market transactions according to the strategy of selected subjects. Experimental economics is used to prove the theoretical hypotheses of economics and test the effects of specific rules in the economic system through controlled experiments. The research issues of experimental economics have developed from simple theoretical tests to summarizing regularities and forming new theories using systematic experiments. The design process of economic experiments has also changed from simply copying the scenes in classic models to designing "tailor-made" experiments [148] for new research issues.

With the development of computer technology, experimental methods have become diversified, which makes large-scale experiments and remote interactive experiments feasible. The scope of research has been extended from microscopic and purely theoretical topics to macro [149], financial [150–153] and social networks [154]. Many government agencies use the results from experimental economics research to guide practice [155] in auctions and central bank policy making. Through systematic reviewing-related research, some scholars divide experimental economics into seven branches [148]: individual preference and decision making, game theory, industrial organization, labor economics, public economics, finance, and macroeconomics. Power market simulations that are based on experimental economics can help verify the theoretical value of the above theories when studying multi-subject interactions in markets and market dynamics, providing a key basis for understanding the competitive behavior and theoretical modeling of market participants. Through power market simulation, we can also research the influence of different trading mechanisms that consider the rational and irrational behavior of real market participants on the efficiency of resource allocation.

4.1.2. Principles and Applications of Agent-Based Computational Economics

Only a limited number of market players can be considered to be the 'rational economic men' of classical economics. Market participants' behavior is not universal, and it is therefore difficult to interpret using traditional mathematical models. With the convergence of technologies in multiple fields and the progress of research and development technologies, agent-based computational economics (ACE) has become an effective tool for overcoming the shortcoming of traditional economics.

Through artificial intelligence and computer modeling and simulation technology, ACE can be used to study complex phenomena in economic systems. Based on the system evolution model, which constitutes a large number of independent market players, the decision-making methods and intelligence levels of different market participants can be calculated using a series of calculations. The simulation system based on agent-based computational economics (ACE) can not only verify some existing economic theories, but also address problems that cannot be well explained through other branches of economics. For example, in the real market environment, even if there is no top-down government/management influence, why do some market participants still follow behavior patterns that are beneficial to the overall development of the market system? Such a problem can be explained using agent-based computational economics (ACE).

In addition, the ACE model can help researchers understand the impact of different systems on the social efficiency and individual welfare of economic systems, and solve the problems of long-term games and individual psychological behavior, which are weak points in existing economic research. In power market simulations, each agent is given different capabilities in their initial state. During the trial cycle, individual agents predict potential changes in the market and give feedback based on their own characteristics. All agents in the system compete with each other and influence each other, together constituting a complex, dynamic, evolutionary system.

Major application scenarios of the existing research based on ACE are as follows. The first type of research is the simulation of individual learning mechanisms, of which the complex network evolution game, based on individual learning abilities, is a cutting-edge problem [156]. This simulation method is mainly based on genetic algorithm and neural network technology.

The second type is to depict behavior patterns. Typical research in earlier periods concentrated on the impact of cultural communication on economic globalization. After introducing the concept of open innovation, Chinese scholars combined computational economics with open innovation problems, and gave market players different equity preference attributes, so as to build a multi-intelligence simulation model. Then, market subjects' preferred strategies, based on their different preference types in different innovation modes, are deduced. Such research provides a useful point of reference [157] for studying open innovation alliances and enterprise innovation development from the micro perspective.

Thirdly, the ACE model is also applied in the study of trading networks in economic systems, focusing mainly on the formation and evolution of networks, as well as the optimization of network organization. Social problems, such as the seasonal migration of shepherds, are typical issues at the heart of this research.

The fourth is to study the market model based on ACE. The evolutionary process and institutional influences on various kinds of markets attract the most researchers. The simulation of the securities market conducted by SFI(Santa Fe Institute) [158] is a typical scenario where ACE is applied.

4.2. Simulation Applications and Parameter Settings for the Power Market as It Adapts to the NPS

The power market simulation system is a comprehensive system that uses specific rules and design processes to simulate the real power market based on the characteristics of real power systems and real markets. The core modules of power market simulation cover power system modeling, power market rules and the characteristics of market participants. Based on the core modules, new scenarios can be designed by adding, deleting and changing parameters, so as to study key issues and test hypotheses pertaining to market constructions under the new power system.

4.2.1. Parameter Settings for the Modeling and Simulation of Power Systems

The modeling and simulation of power systems is based on the physical properties [159–161] of power systems, such as the structure of regional power supply, load, grid nodes and the limit of power transmission lines. As China expects to embrace a new power system with a large amount of renewable energy, the value of all kinds of electricity generation units (especially flexible resources) needs to be verified for the new market situation [162]. In order to simulate the market to study specific issues, a variable scenario for modeling power systems can be constructed by adding the types and proportions of electricity generation units. The types of units covered by the simulation include new energy sources, energy storage (usually in form of electrochemical, hydrogen storage, and flywheel), distributed power supplies and thermal power units with flexible regulation capabilities. Common parameters of power system simulation that are based on electric power system theory and prediction theory are shown in Table 3. In addition, under the construction of the new power system, the market share occupied by new energy, the rate of new energy penetration, and other indicators can be added. By adding such indicators,

other issues [163], such as testing paths towards the dual carbon emission target, discovering the value of diverse resources and improving the strategies of market participants, can be researched under the simulation with extended parameters.

Table 3. Power System Simulation Parameters.

Parameters	Description
Structure of the power supply	Type, location, installed capacity and other properties.
Forecast of electricity output (accuracy) generated from renewable energy of the system	The reliability of the system can be verified by setting a proper coefficient of power generation using renewable energy units. Based on weather forecast data, the dispatch center estimates potential output from renewable resources and provides the power generation coefficient of new energy to market participants for reference.
Standby rate	The standby rate is the ratio of the gap between the system's total available capacity and the peak load. It is an important index to measure the reliability of the power system [164].
Number of nodes	The complexity of the power system can also be reflected by the number of nodes. The closer the location and number of nodes is set to the actual grid topology, the more effective the power system simulation will be.
Power generation schedule of units	Factors influencing the power generation schedule of the units include system operation factors and non-system operation factors.
Power flow	Excited by the potential of the power supply, electric current flows from the power supply to every load in the power system through distribution components, and thus power is distributed throughout the whole grid.
The constraint of power flow in N-1 transmission section	When grid topology operates normally according to the given constraint, if disconnecting any line in the section, power flow in the rest of the lines is not overloaded. The constraint of power flow in a section is the maximum allowable power current when the section is running normally [165]. If the power flow surpasses the limitation, a nodal price would be formed.
Constraint of active power balance	Active power balance means that active power on the generation side and the consumption side is equal. The frequency of the power system is directly related to the balance of active power, and the auxiliary services of frequency regulation can be referenced.
Constraint of reactive power balance	Reactive power balance is the condition that reactive power on the generation side and the consumption side is equal. When the reactive power is insufficient, the reactive compensation is required.
Plans for maintenance of power generation, transmission and transformation equipment	Plans for maintenance covers two types of equipment, including power generation equipment, as well as power transmission and transformation equipment. The maintenance plans for power distribution and transformation equipment affect the structure of power transmission for a short time.
System load forecasting (accuracy)	The dispatch center predicts the total load of the power system for several days based on data obtained about the system. The accuracy of system load forecasting directly affects the analysis of market supply and the demand of market participants, and then affects market quotation behavior.

4.2.2. Parameter Settings to Simulate the Rules and Conditions of the Power Market

The simulation of market rules and conditions is focused on market mechanisms and related policies. Through setting different parameters, diverse market scenarios can be

simulated. Market rules are reflected by a range of parameters, such as the entry and exit of market players, symbols of trade, and rules of clearance and settlement. Market conditions are simulated by the ratio of supply and demand, market forces and policy-related indicators. Common parameter settings that are based on power market trading, theory and industrial economics theory are shown in Table 4.

Table 4. Power Market Rules and Environment Simulation Parameters.

Parameters	Description
Trading methods	Major trading methods include bilateral consultations, centralized bidding and listing transactions. The spot trading market is divided into the day-ahead market, the intra-day marker and the real-time market. A variety of trading methods can be combined according to the transaction cycle, thus new symbols can be built in to increase the diversity of market symbols.
Trading transactions	Depending on different types of markets, typical symbols include medium- and long-term (yearly, monthly, weekly) transactions, contract power transfer, power generation rights trading, spot trading and ancillary services trading. New trading instruments for renewable energy and flexible resources will be added in order to meet the needs of further construction in the power market.
Limitations on the declared price	Based on the relationship between demand and supply, in order to avoid market manipulation and vicious competition, minimum and maximum declared prices must be set for the direct trading of electricity in different transaction modes. Different ranges for price limitations can be set through power market simulation to verify the market's capacity.
Rules for clearance	Major rules of clearance include high–low matching and unified clearance. High–low matching is the priority match between the highest spread on the demand side and the lowest spread on the supply side. Whether the spread pair matches can be judged by the following formula: $L_n - S_m \geq 0$, match $L_n - S_m < 0$, not match where L_n is the spread pair on the demand side, while S_m represents the spread pair on the supply side. Unified clearance refers to the selection of the last pair of spread pairs that match successfully according to the principles of high–low matching. The arithmetic average of the selected pair will be designated as the closing spread of all participants. When studying the development degree and diversification of the power market, different rules of clearance and trading methods can be combined to conduct diversified trading. Thus, the market affordability can be tested.
Settlement mechanisms	Settlement is reached either at the marginal electricity price or at the actual declared price by power generations units: $\min F_m,\ F_m = \sum_{i=1}^{I} C_{om} \cdot P_i$ where F_m represents the cost of power purchases paid by the grid, P_i represents the power generation volume bid by the power producer (or electricity generation unit) i, and i refers to power producer. I is the total number of power producers, and C_{om} represents the marginal electricity price of the system.

Table 4. *Cont.*

Parameters	Description
Changing rate of the balancing funds account	As the price fluctuations of electricity bought from the power market by the grid cannot be transmitted to power retail prices in time, a balancing funds account is built to link power sales prices to power purchase prices: $V_i = \frac{F_i - F_{i-1}}{F_{i-1}}$ where V_i is the changing rate of the funds in the balancing account, F_i represents the amount of balancing funds this year, while F_{i-1} refers to the amount of balancing funds last year.
Market turnover rate	Market turnover rate is the ratio representing the number of market participants declaring during the statistical period (which can also be a transaction for one time) divided by the number of market subjects that finally complete the transaction. The turnover rate is used to analyze whether market competition is insufficient or judge collusion by market participants: $V_i = \frac{F_i}{G_i}$ where Vi is the turnover rate, Fi refers to the number of market subjects which finally complete the transaction, and G_i is number of market participants declaring during the statistical period.
HHI	The Herfindahl–Hirschman index (HHI) is a composite index that measures industrial concentration. It reflects the changes in market share, i.e., the dispersion of the size of manufacturers in the market: $I_{HHI} = \sum_{i=1}^{N} (100 S_i)^2$ where S_i is the market share occupied by market supplier i per trade sequence during the evaluation cycle, while N represents the number of suppliers in the market during the evaluation cycle.
Renewable energy subsidies	According to the effectiveness of the market mechanism and the development of new energy, the amount of renewable energy subsidy can be changed to verify the risks brought about by the marketization of renewable energy resources.
Deviation parameters	The parameters reflecting the deviation of power generation (more or less) on the generation side include the exemption range and the appraisal price. The parameters measuring the deviation of electricity consumption (more or less) also include the exemption range and the appraisal price.

To study the key issues relating to the market trading system under the new power system, the access of emerging market participants can be modified through the module of participant characteristics. When designing market mechanisms to stimulate flexible resources, it is necessary to make explicit the value and price conduction mechanism of the flexible resources. Simultaneously, some parameters, including the price declaration mode, the upper and lower limits of the price, the clearing method and the settlement method, can be modified so as to check market feedback.

In addition, whether the symbols are sufficient to meet market demand can be verified through combining multiple symbols and trading methods, with the actual situation in the simulation area taken into consideration. To verify the settlement rules, the model, content and the method of settlement that is used by different users can be combined and simulated through power market simulation. Thus, the settlement rules, with their regional characteristics, such as deviation assessment and double rule assessment, can be included in the study.

4.2.3. Market Participant Characteristic Parameters

As the most basic units of the power market, market participants affect the construction of the trading mechanism to a certain extent. The market subject module is focused on the

characteristics of transaction behaviors and the strategies of market participants. In the simulation system, the subject that is making decisions can be intelligent or real people. Therefore, parameters are designed to simulate the characteristics of human-machine decision making, including the type, quantity, distribution, feedback behavior and learning ability of market participants. The parameters that are commonly used to simulate market participants characteristic, which are based on power market theory, behavioral economics theory and forecasting techniques, are shown in Table 5.

Table 5. Market Participants Characteristic Parameters.

Parameters	Description
Power producer, asset type and status constraints	This parameter covers the type of units and operating parameters of the power producer. When the user of the simulation system is the power producer, the power generation enterprise can develop its own strategies by simulating the trading strategy of its competitors based on market conditions.
Cost composition	Cost is an important basis for making transaction decisions. Power generation costs primarily cover start-up costs, empty operating costs, fuel costs, environmental costs, and other economic costs. Under China's dual carbon reduction goal, environmental costs have become an important constraint on the development of thermal power enterprises. Power sales companies need to consider retail contract prices, wholesale prices and other related parameters.
Marginal revenue of the power generation unit	This parameter refers to the income earned through increasing or decreasing the power generation of the unit, and its relationship with the power generation and market price is: $MR_i = \frac{d(PG_i)}{dG_i} - P$ where MR_i is the marginal revenue of the power generation unit i, G_i refers to the output of unit i, and P is the electricity price in the power market.
Forecasts for market prices and power generation	A power plant forecasts its own power generation output based on its installed capacity. Combining this with market price forecasts, the power plant declares reasonable prices and a power generation plan. At the same time, the market price and power generation coefficient can also be set to verify the anti-jamming of the quoting strategy.
Power generation plans	Power generation enterprises make production plans based on existing contracts. Power market simulation provides users with a chance to choose the optimal solution from different generation plans so as to avoid market shocks.
Intelligent agent simulation	Through simulating the trading behaviors of competitors, intelligent agent simulation technology helps power producers to formulate their own trading strategies. By simulating the transactions of multiple participants, it can also help verify market results. Intelligent agents can be divided into rational and irrational, or radical and conservative. Commonly used parameters include the number of intelligent agents, learning mechanisms and feedback mechanisms.

To better study the key issues relating to the market trading system under the new power system, on the one hand, we should expand the research to cover a diversity of market subjects. A range of participants, such as different renewable energy resources, energy storage, distributed power supplies, flexible thermal power units, and flexible loads that can respond to power consumption peaks, should be included. On the other hand, stronger decision-making abilities should be developed for market subjects. Training the learning, prediction, portfolio and risk management abilities of the intelligent agents is helpful. By enriching the parameters for the simulation of market transaction behaviors, the effect of different trading mechanisms can be tested, providing beneficial points of reference for market participants for better decision making.

5. Discussion

This paper provided a rigorous systematic review of the study of power markets with clear framework conditions using pertinent literature databases and full-text access for query articles. The disregard of non-English and non-Chinese language publications in the databases and the disregard of conference articles without full-text access could be regarded as limitations of this review. Moreover, the exclusion of the last pages papers in the relevance rankings could also be a limitation.

This review covered existing research on the study of power markets, including market operational efficiency, the bidding strategies of market participants and power market supervision. Based on the search results, it can be concluded that English papers focused more on bidding strategies than on the operational efficiency of markets. However, Chinese papers paid more attention to the operational efficiency of markets and supervision than to bidding strategy.

Previous studies on the operational efficiency of power markets mainly focused on the structure design or evaluation. As illustrated in Section 3.1.1, such studies have three major flaws. Firstly, such empirical analyses were limited to the unilateral bidding market scenario that was formed during the first electric power reforms in China. Secondly, previous researchers ignored the constraints and influences of environmental factors. Thirdly, those studies have not considered the systematic impact of a high proportion of new energy connecting to the power system and power market. Given the different roles of market players, studies on bidding strategies can be sorted according to three aspects, namely, generation companies' strategies, energy retailers' strategies, and energy consumers' strategies. With regard to the development of the power market, studies on the strategies of prosumers [86], virtual power plants [94,95], microgrids [91] and energy storage [91] were given more attention. Furthermore, most of them focused on the spot market, especially the day-ahead market. Moreover, previous studies in the bidding strategy field were implemented using cost analysis, electricity price forecasts, optimization algorithms or game theory. Little research was conducted using a power market trading simulation system, which is based on experimental economics and agent-based computational economics. To some extent, these can provide points of reference for bidding strategy studies that focus on the adaptation to the NPS. However, most of them studied participants as price takers, and only a few publications [166] studied the market strategies of flexible resources in joint energy and flexible product markets. This work emphasized the significance of power market simulation systems for the study of power markets, and proposed parameter setting suggestions for power system simulations, power market rules simulation and market participants characteristics.

This literature review introduced several key issues in Section 3.2 that relate to the adaptation to the NPS. In Section 3.2.1, relatively new studies were discussed, with 42 papers found that were published in Web of Science and 121 papers in Elsevier. However, only two of these papers were published after the NPS was proposed. Only one of them focused on the financial implications for investors in renewable energy [167]. The other one introduced a day-ahead clearing model for the simultaneous energy and ancillary services market with a high penetration of renewable energy sources [127]. Although some papers also talked about market issues with renewable energy resources, most of them focused more on power system operations [168,169] than the power market operations. In conclusion, few studies discussed the mechanisms and strategies for an investor in a market with a high amount of renewable energy.

Only five papers were identified as highly relevant publications based on in the "Coal-fired power transition" search term. Some studies discussed phase-out issues [170,171], rather than the transition strategy for coal-fired plants to adapt to the energy transition. Although one paper [172] proposed that coal-fired power plant could be retired in Australia in the future, it did not discuss the policy implications for coal-fired power plants during this transition. Under the construction of the NPS, one should consider not only the income

from selling electricity, but also the potential benefits from providing flexibility using coal-fired units.

As for the study of flexible resources in power markets, Muñoz [173] discussed the current limitations of Chile's electricity market going forward, which aimed to provide incentives for the efficient resources to give the system more flexibility. Moreover, Das, P. [174] argued that technology, policy and new modeling strategies are needed to expand large-scale renewable energy, which supports the standpoint of this review.

6. Conclusions and Implications

Research on electric power markets are now discussing how the markets can adapt to the new power system, which will be based on renewable energy. Market mechanisms, market participants' strategies, and market supervision are three key elements for the healthy operation of the power market. This paper provided a rigorous systematic review of the study of power markets given this new situation. The article also provided a comprehensive overview of the key issues in the new studies on power markets and its applications and techniques. After the analysis, a basic paradigm for the study of power markets that is suitable for the new situation was put forward to provide a clear review of the study of power markets and inspire better research on market trading mechanisms under the new power system. The main contributions of our research are as follows:

1. The general issues of power market research were summarized in terms of market operation, bidding strategies of market players, and market supervision. The quantitative analysis shows that China's previous power market studies paid more attention to the market mechanism design and operation supervision. However, bidding strategies were discussed more than power market supervision in other countries.
2. Key issues related to research on the power market were picked out for further study as the market adapts to the new power system, which features a high proportion of new energy. These issues include market mechanisms to promote the participation of new energy, flexible resource value discovery and supporting mechanisms, as well as the survival of thermal power units. The systematic review indicates that the mechanism and strategies for a new energy-dominated power market were not fully considered in previous studies. Moreover, for the survival of coal-fired power plants in this new environment, most studies focused on the phase out issues rather than the market mechanisms and strategies relating to their transformation. In addition, research on flexible market products is still rare, although more attention is being paid to this issue [174].
3. A basic paradigm for the study of power markets that is suitable for the new power system was established, which provides the basic direction for power market research under the new situation.
4. The theoretical basis of power market simulation was presented in our paper. Meanwhile, the applications of experimental economics and agent-based computational economics in the study of power markets were reviewed. The quantitative analysis shows that although the current studies can provide a point of reference for the study of power markets, there are few pieces of literature on the new power market with the application of EE and ACE. More research is needed in the future to adapt to the construction of the NPS.
5. Specific parameter settings were recommended for simulating power market transactions, which can serve as a theoretical basis and practical guidance for other researchers to better study the power market.

The power market is a complex economic system. To better construct a new power market that is dominated by new energies, more studies on the key issues are needed. Studies on the new power market can be developed using EE and ACE techniques., which can help to reduce the costs of power market construction, promote the low-carbon transformation of the electricity sector, and improve the market trading mechanisms of the new power system.

Author Contributions: Conceptualization, J.D. and D.L.; methodology, D.L. and T.M.; validation, D.L., T.M. and X.D.; formal analysis, D.L. and X.D.; investigation, S.L., Y.J. and B.L.; resources, D.L. and B.L.; data curation, D.L. and X.D.; writing—original draft preparation, D.L. and X.D.; writing—review and editing, D.L., B.L., T.M. and X.D; visualization, X.D.; supervision, J.D. and T.M.; project administration, J.D. and T.M.; funding acquisition, J.D. and T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Higher Education Discipline Innovation and Talent Plan—China Green Power Development Research Discipline Innovation and Talent Base (B18021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This study was supported by the Energy Market Research Institute of North China Electric Power University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, D.; Zhang, Z.; Yang, X.; Zhang, Y.; Li, Y.; Zhao, Y. Multi-Scenario Simulation on the Impact of China's Electricity Bidding Policy Based on Complex Networks Model. *Energy Policy* **2021**, *158*, 112573. [CrossRef]
2. Huo, T.; Xu, L.; Feng, W.; Cai, W.; Liu, B. Dynamic Scenario Simulations of Carbon Emission Peak in China's City-Scale Urban Residential Building Sector through 2050. *Energy Policy* **2021**, *159*, 112612. [CrossRef]
3. China News. China's Renewable Energy Power Generation Capacity Reached 2.2 Trillion KWh in 2020. Available online: https://www.chinanewsweb.com/index.php/2021/03/30/chinas-renewable-energy-generation-will-reach-2-2-trillion-kilowatt-hours-by-2020 (accessed on 15 August 2021).
4. Ma, L.; Fan, M.; Qu, H.; Li, J.; Zhao, Z.; Wu, C.; Chen, K. Construction path and key problems of China's power market. *Electric Power* **2020**, *53*, 1–9.
5. Chen, H. Causes and Countermeasures of "unbalanced funds" in power market. *China Electr. Power Enterp. Manag.* **2020**, *25*, 27–30.
6. Zhao, Z.; Gu, W.; Chen, Y.; Li, X.; Jiang, Y.; Gao, D.; Duan, R.; Wu, Y. Analysis of unbalanced capital in power market under dual track system. *Electr. Power* **2020**, *53*, 47–54.
7. Chen, Q.; Yang, J.; Huang, Y.; Lu, E.; Wang, Y. Review on Market Power Monitoring and Mitigation Mechanisms in Foreign Electricity Markets. *South. Power Syst. Technol.* **2018**, *12*, 9–15+63.
8. Deng, S.; Xu, K.; Liu, L. Review on the Reasons for Formation and Inhibition Mechanism of Market Power in the Electricity Market. *Telecom Power Technol.* **2018**, *35*, 129–131.
9. Wu, T.; Ding, Y.; Shang, N.; Bao, M.; Song, Y. Joint decomposition algorithm of medium and long-term contract electricity of multi class units considering market power test. *Autom. Electr. Power Syst.* **2021**, *45*, 72–81.
10. Zhang, Q.; Wang, X.; Wang, J.; Feng, C.; Liu, L. Review on demand response research in power market. *Autom. Electr. Power Syst.* **2008**, *3*, 97–106.
11. Xiao, Y.; Wang, X.; Wang, X.; Bie, Z. Review on Electricity Market Towards High Proportion of Renewable Energy. *Proc. CSEE* **2018**, *38*, 663–674.
12. Keele, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; EBSE Technical Report: Durham, UK, 2007.
13. PRISMA-P Group; Moher, D.; Shamseer, L.; Clarke, M.; Ghersi, D.; Liberati, A.; Petticrew, M.; Shekelle, P.; Stewart, L.A. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 Statement. *Syst. Rev.* **2015**, *4*, 1–9. [CrossRef]
14. Shrestha, T.K.; Karki, R. Impact of Market-Driven Energy Storage System Operation on the Operational Adequacy of Wind Integrated Power Systems. *J. Energy Storage* **2020**, *32*, 101792. [CrossRef]
15. Bigdeli, N.; Afshar, K.; Fotuhi-Firuzabad, M. Bidding Strategy in Pay-as-Bid Markets Based on Supplier-Market Interaction Analysis. *Energy Convers. Manag.* **2010**, *51*, 2419–2430. [CrossRef]
16. Chen, C. Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5303–5310. [CrossRef] [PubMed]
17. Chen, Y.; Qin, Z. Review of Research on Electricity Market Reform. *Electrotech. Electr.* **2019**, *4*, 1–6+12.
18. Newbery, D. Electricity Liberalization in Britain and the Evolution of Market Design. In *Electricity Market Reform*; Elsevier: Amsterdam, The Netherlands, 2006; pp. 109–143.
19. Marshall, L.; Bruce, A.; MacGill, I. Assessing Wholesale Competition in the Australian National Electricity Market. *Energy Policy* **2021**, *149*, 112066. [CrossRef]
20. Borenstein, S.; Bushnell, J.B.; Wolak, F.A. Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market. *Am. Econ. Rev.* **2002**, *92*, 1376–1405. [CrossRef]

21. Daglish, T.; de Bragança, G.G.F.; Owen, S.; Romano, T. Pricing Effects of the Electricity Market Reform in Brazil. *Energy Econ.* **2021**, 105197. [CrossRef]
22. Bertram, G. Weak Regulation, Rising Margins, and Asset Revaluations. In *Evolution of Global Electricity Markets*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 645–677.
23. Keles, D.; Bublitz, A.; Zimmermann, F.; Genoese, M.; Fichtner, W. Analysis of Design Options for the Electricity Market: The German Case. *Appl. Energy* **2016**, *183*, 884–901. [CrossRef]
24. Nilsson, M. Red Light for Green Paper: The EU Policy on Energy Efficiency. *Energy Policy* **2007**, *35*, 540–547. [CrossRef]
25. Pérez-Díaz, J.I.; Wilhelmi, J.R.; Arévalo, L.A. Optimal Short-Term Operation Schedule of a Hydropower Plant in a Competitive Electricity Market. *Energy Convers. Manag.* **2010**, *51*, 2955–2966. [CrossRef]
26. Xu, D. *Research on Operation Efficiency of Power Market*; Zhejiang University: Hangzhou, China, 2003.
27. Luan, F. *Research on Theory and Application of Economic Efficiency Evaluation of Power Market*; North China Electric Power University: Beijing, China, 2007.
28. Wang, W.; Shao, S.; Li, L.; Tang, H.; Zhao, Y.; Xia, Q. Power market efficiency theory and its evaluation method. *Power Grid Technol.* **2009**, *33*, 66–71.
29. Ren, Y.; Liu, S.; Bie, Z. Regional Marginal Price Mechanism in the China Northwest Power Grid. In Proceedings of the 2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP), Xi'an, China, 21 October 2019; pp. 1115–1119.
30. Daneshi, H.; Srivastava, A.K. ERCOT Electricity Market: Transition from Zonal to Nodal Market Operation. In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 24–28 July 2011; pp. 1–7.
31. Höschle, H.; De Jonghe, C.; Le Cadre, H.; Belmans, R. Electricity Markets for Energy, Flexibility and Availability—Impact of Capacity Mechanisms on the Remuneration of Generation Technologies. *Energy Econ.* **2017**, *66*, 372–383. [CrossRef]
32. BOWRING, J. Capacity Markets in PJM. *Econ. Energy Environ. Policy* **2013**, *2*, 47–64. [CrossRef]
33. Bowring, J.E. The evolution of the PJM capacity market: Does it address the revenue sufficiency problem? In *Evolution of Global Electricity Markets*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 227–264.
34. Hobbs, B.F.; Hu, M.-C.; Inon, J.G.; Stoft, S.E.; Bhavaraju, M.P. A Dynamic Analysis of a Demand Curve-Based Capacity Market Proposal: The PJM Reliability Pricing Model. *IEEE Trans. Power Syst.* **2007**, *22*, 3–14. [CrossRef]
35. Song, X.; Zhai, X.; Chen, W.; Xue, J.; Wang, P. Study on Three-Part Pricing Method of Pumped Storage Power Station in China Considering Peak Load Regulation Auxiliary Service. *IOP Publ.* **2021**, *675*, 012110. [CrossRef]
36. Wang, J.; Zhong, H.; Yang, Z.; Lai, X.; Xia, Q.; Kang, C. Incentive Mechanism for Clearing Energy and Reserve Markets in Multi-Area Power Systems. *IEEE Trans. Sustain. Energy* **2020**, *11*, 2470–2482. [CrossRef]
37. Wang, J.; Zhong, H.; Wu, C.; Du, E.; Xia, Q.; Kang, C. Incentivizing Distributed Energy Resource Aggregation in Energy and Capacity Markets: An Energy Sharing Scheme and Mechanism Design. *Appl. Energy* **2019**, *252*, 113471. [CrossRef]
38. Wolak, F.; Patrick, R. *The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market*; National Bureau of Economic Research: Cambridge, MA, USA, 2001; p. w8248.
39. Tirole, J. *The Theory of Industrial Organization*; MIT Press: Cambridge, MA, USA, 1988.
40. Landes, W.M.; Posner, R.A. Market Power in Antitrust Cases. *Harv. Law Rev.* **1981**, *94*, 937. [CrossRef]
41. Visudhiphan, P.; Ilic, M.D. Dependence of Generation Market Power on the Demand/Supply Ratio: Analysis and Modeling. In Proceedings of the 2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.00CH37077), Singapore, 23–27 January 2000; Volume 2, pp. 1115–1122.
42. Gan, D.; Bourcier, D.V. Locational Market Power Screening and Congestion Management: Experience and Suggestions. *IEEE Trans. Power Syst.* **2002**, *17*, 180–185. [CrossRef]
43. Rahimi, A.F.; Sheffrin, A.Y. Effective Market Monitoring in Deregulated Electricity Markets. *IEEE Trans. Power Syst.* **2003**, *18*, 486–493. [CrossRef]
44. Goncalves, M.J.D.; Vale, Z.A. Evaluation of Transmission Congestion Impact in Market Power. In Proceedings of the 2003 IEEE Bologna Power Tech Conference Proceedings, Bologna, Italy, 23–26 June 2003; Volume 4, pp. 438–443.
45. Bompard, E.; Ma, Y.C.; Napoli, R.; Jiang, C.W. Assessing the Market Power Due to the Network Constraints in Competitive Electricity Markets. *Electr. Power Syst. Res.* **2006**, *76*, 953–961. [CrossRef]
46. Moreira e Silva, R.; Terra, L.D.B. Market Power under Transmission Congestion Constraints. In Proceedings of the PowerTech Budapest 99. Abstract Records. (Cat. No.99EX376), Budapest, Hungary, 29 August–2 September 1999; p. 82.
47. Li, H.; Wang, B.; Guo, S. Evaluation on Market Power of Generation Companies in Regional Electricity Market Based on Principal Component Analysis. *Mod. Electr. Power* **2011**, *28*, 85–89.
48. Yang, T.; Fu, S.; Wang, B. Analysis of market power in power market based on Game Theory. *J. Shandong Electr. Power Coll.* **2011**, *14*, 13–18.
49. Zhang, Z.; Guo, X.; Wang, F. Analysis of financial transmission rights and market power based on oligopoly competition model. *Autom. Electr. Power Syst.* **2011**, *35*, 30–33+59.
50. Zhang, F.; Wen, F.; Yan, H.; Yu, Z.; Zhong, Z.; Hunag, J. Analysis of reactive power market power based on agent simulation. *Autom. Electr. Power Syst.* **2009**, *33*, 18–24.
51. Ambrosius, M.; Grimm, V.; Sölch, C.; Zöttl, G. Investment Incentives for Flexible Demand Options under Different Market Designs. *Energy Policy* **2018**, *118*, 372–389. [CrossRef]

52. Richstein, J.C.; Hosseinioun, S.S. Industrial Demand Response: How Network Tariffs and Regulation (Do Not) Impact Flexibility Provision in Electricity Markets and Reserves. *Appl. Energy* **2020**, *278*, 115431. [CrossRef]
53. Spees, K.; Lave, L.B. Demand Response and Electricity Market Efficiency. *Electr. J.* **2007**, *20*, 69–85. [CrossRef]
54. Wang, J.; Zhong, H.; Ma, Z.; Xia, Q.; Kang, C. Review and Prospect of Integrated Demand Response in the Multi-Energy System. *Appl. Energy* **2017**, *202*, 772–782. [CrossRef]
55. Grimm, V.; Rückel, B.; Sölch, C.; Zöttl, G. The Impact of Market Design on Transmission and Generation Investment in Electricity Markets. *Energy Econ.* **2021**, *93*, 104934. [CrossRef]
56. Tómasson, E.; Hesamzadeh, M.R.; Söder, L.; Biggar, D.R. An Incentive Mechanism for Generation Capacity Investment in a Price-Capped Wholesale Power Market. *Electr. Power Syst. Res.* **2020**, *189*, 106708. [CrossRef]
57. Barazza, E.; Strachan, N. The Co-Evolution of Climate Policy and Investments in Electricity Markets: Simulating Agent Dynamics in UK, German and Italian Electricity Sectors. *Energy Res. Soc. Sci.* **2020**, *65*, 101458. [CrossRef]
58. Feng, T.; Li, R.; Zhang, H.; Gong, X.; Yang, Y. Induction Mechanism and Optimization of Tradable Green Certificates and Carbon Emission Trading Acting on Electricity Market in China. *Resour. Conserv. Recycl.* **2021**, *169*, 105487. [CrossRef]
59. Kraan, O.; Kramer, G.J.; Nikolic, I.; Chappin, E.; Koning, V. Why Fully Liberalised Electricity Markets Will Fail to Meet Deep Decarbonisation Targets Even with Strong Carbon Pricing. *Energy Policy* **2019**, *131*, 99–110. [CrossRef]
60. Sarfati, M.; Hesamzadeh, M.R.; Holmberg, P. Production Efficiency of Nodal and Zonal Pricing in Imperfectly Competitive Electricity Markets. *Energy Strategy Rev.* **2019**, *24*, 193–206. [CrossRef]
61. U.S. Department of Energy. *Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them*; U.S. Department of Energy: Washington, DC, USA, 2006.
62. Davarzani, S.; Pisica, I.; Taylor, G.A.; Munisami, K.J. Residential Demand Response Strategies and Applications in Active Distribution Network Management. *Renew. Sustain. Energy Rev.* **2021**, *138*, 110567. [CrossRef]
63. Pallonetto, F.; De Rosa, M.; D'Ettorre, F.; Finn, D.P. On the Assessment and Control Optimisation of Demand Response Programs in Residential Buildings. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109861. [CrossRef]
64. Lu, X.; Li, K.; Xu, H.; Wang, F.; Zhou, Z.; Zhang, Y. Fundamentals and Business Model for Resource Aggregator of Demand Response in Electricity Markets. *Energy* **2020**, *204*, 117885. [CrossRef]
65. Yang, S.; Liu, J.; Yao, J.; Ding, H.; Wang, K.; Li, Y. Multi time scale coordinated flexible load interactive response scheduling model and strategy. *Proc. CSEE* **2014**, *34*, 3664–3673.
66. Yuan, B.; Yang, Q.; Yan, W. Demand response under real-time price for domestic energy system. *J. Mech. Electr. Eng.* **2015**, *32*, 857–862.
67. Jiang, T.; Li, Y.; Ju, P.; Yang, Y.; Zhao, J. Overview of Modeling Method for Flexible Load and its Control. *Smart Power* **2020**, *48*, 1–8.
68. Lund, H.; Münster, E.; Tambjerg, L.H. *EnergyPlan: Computer Model for Energy System Analysis: Version 6*; Technology, Environment and Society, Department of Development and Planning, Aalborg University: Aalborg, Denmark, 2004.
69. He, G.; Avrin, A.-P.; Nelson, J.H.; Johnston, J.; Mileva, A.; Tian, J.; Kammen, D.M. SWITCH-China: A Systems Approach to Decarbonizing China's Power System. *Environ. Sci. Technol.* **2016**, *50*, 5467–5473. [CrossRef] [PubMed]
70. He, G.; Lin, J.; Sifuentes, F.; Liu, X.; Abhyankar, N.; Phadke, A. Rapid Cost Decrease of Renewables and Storage Accelerates the Decarbonization of China's Power System. *Nat. Commun.* **2020**, *11*, 2486. [CrossRef] [PubMed]
71. Li, B.; Ma, Z.; Hidalgo-Gonzalez, P.; Lathem, A.; Fedorova, N.; He, G.; Zhong, H.; Chen, M.; Kammen, D.M. Modeling the Impact of EVs in the Chinese Power System: Pathways for Implementing Emissions Reduction Commitments in the Power and Transportation Sectors. *Energy Policy* **2021**, *149*, 111962. [CrossRef]
72. NEPLAN. Available online: http://ieeexplore.ieee.org/document/1687803/ (accessed on 24 July 2021).
73. Hungerford, Z.; Bruce, A.; MacGill, I. The Value of Flexible Load in Power Systems with High Renewable Energy Penetration. *Energy* **2019**, *188*, 115960. [CrossRef]
74. Ren, Y.; Zou, X.; Zhang, X. Bidding Model of Power Plant Company with Incomplete Information. *Autom. Electr. Power Syst.* **2003**, *9*, 11–14.
75. Lei, B.; Wang, X.; Gao, Y.; Wang, X. Analysis on bidding strategy of independent power producer in days-ahead market. *Autom. Electr. Power Syst.* **2002**, *26*, 8–14.
76. Wang, X.; Wang, X.; Chen, H. *Fundamentals of Electricity Market*; Xi'an Jiaotong University Press: Xi'an, China, 2003.
77. Conejo, A.J.; Nogales, F.J.; Arroyo, J.M. Price-Taker Bidding Strategy under Price Uncertainty. *IEEE Power Eng. Rev.* **2002**, *22*, 57. [CrossRef]
78. Wen, F.; David, A. Optimal Bidding Strategies and Modeling of Imperfect Information among Competitive Generators. *IEEE Trans. Power Syst.* **2002**, *16*, 15–21.
79. Kang, D.-J.; Kim, B.H.; Hur, D. Supplier Bidding Strategy Based on Non-Cooperative Game Theory Concepts in Single Auction Power Pools. *Electr. Power Syst. Res.* **2007**, *77*, 630–636. [CrossRef]
80. Jing, Z.; Yang, Y. Application of the EWA Algorithm in Electricity Market Simulation. *Autom. Electr. Power Syst.* **2010**, *34*, 46–50.
81. Zhu, J. *Study on Bidding Strategy of Generation Companies with Evolution Game Theory Based on Repast Platform*; North China Electric Power University: Beijing, China, 2010.
82. Zhu, J. *Research on Bidding Strategy of Generator Based on Agent*; Beijing Jiaotong University: Beijing, China, 2010.
83. Ren, X. *A Study on Bidding Behavior of Power Producers Based on Multi-Agent Simulation*; North China Electric Power University: Beijing, China, 2009.

84. Liang, Z. *Simulation of Generation Side Power Market Bidding Trading System Based on MAS*; North China Electric Power University: Beijing, China, 2012.
85. Ghavidel, S.; Ghadi, M.J.; Azizivahed, A.; Aghaei, J.; Li, L.; Zhang, J. Risk-Constrained Bidding Strategy for a Joint Operation of Wind Power and CAES Aggregators. *IEEE Trans. Sustain. Energy* **2020**, *11*, 457–466. [CrossRef]
86. Iria, J.; Soares, F.; Matos, M. Optimal Supply and Demand Bidding Strategy for an Aggregator of Small Prosumers. *Appl. Energy* **2018**, *213*, 658–669. [CrossRef]
87. Dong, Y.; Dong, Z.; Zhao, T.; Ding, Z. A Strategic Day-Ahead Bidding Strategy and Operation for Battery Energy Storage System by Reinforcement Learning. *Electr. Power Syst. Res.* **2021**, *196*, 107229. [CrossRef]
88. Cheng, L.; Liu, G.; Huang, H.; Wang, X.; Chen, Y.; Zhang, J.; Meng, A.; Yang, R.; Yu, T. Equilibrium Analysis of General N-Population Multi-Strategy Games for Generation-Side Long-Term Bidding: An Evolutionary Game Perspective. *J. Clean. Prod.* **2020**, *276*, 124123. [CrossRef]
89. Wang, Y.; Wang, J.; Sun, W.; Zhao, M. Optimal Day-Ahead Bidding Strategy for Electricity Retailer with Inner-Outer 2-Layer Model System Based on Stochastic Mixed-Integer Optimization. *Math. Probl. Eng.* **2019**, *2019*, 1–14. [CrossRef]
90. Yang, M.; Ai, X.; Tang, L.; Guo, S.; Luo, G. Optimal Trading Strategy in Balancing Market for Electricity Retailer Considering Risk Aversion. *Power Syst. Technol.* **2016**, *40*, 3300–3309.
91. Mirzaei, M.A.; Hemmati, M.; Zare, K.; Abapour, M.; Mohammadi-Ivatloo, B.; Marzband, M.; Anvari-Moghaddam, A. A Novel Hybrid Two-Stage Framework for Flexible Bidding Strategy of Reconfigurable Micro-Grid in Day-Ahead and Real-Time Markets. *Int. J. Electr. Power Energy Syst.* **2020**, *123*, 106293. [CrossRef]
92. Jia, C.; Du, X. Optimization of power purchase strategy of power selling companies under medium and long-term trading mechanism. *Electric Power* **2019**, *52*, 140–147.
93. Sun, B.; Wang, F.; Xie, J.; Sun, X. Electricity Retailer Trading Portfolio Optimization Considering Risk Assessment in Chinese Electricity Market. *Electr. Power Syst. Res.* **2021**, *190*, 106833. [CrossRef]
94. Mashhour, E.; Moghaddas-Tafreshi, S.M. Bidding Strategy of Virtual Power Plant for Participating in Energy and Spinning Reserve Markets—Part I: Problem Formulation. *IEEE Trans. Power Syst.* **2011**, *26*, 949–956. [CrossRef]
95. Sadeghi, S.; Jahangir, H.; Vatandoust, B.; Golkar, M.A.; Ahmadian, A.; Elkamel, A. Optimal Bidding Strategy of a Virtual Power Plant in Day-Ahead Energy and Frequency Regulation Markets: A Deep Learning-Based Approach. *Int. J. Electr. Power Energy Syst.* **2021**, *127*, 106646. [CrossRef]
96. Nguyen, D.T.; Le, L.B. Optimal Bidding Strategy for Microgrids Considering Renewable Energy and Building Thermal Dynamics. *IEEE Trans. Smart Grid* **2014**, *5*, 1608–1620. [CrossRef]
97. Wang, J.; Zhong, H.; Tang, W.; Rajagopal, R.; Xia, Q.; Kang, C.; Wang, Y. Optimal Bidding Strategy for Microgrids in Joint Energy and Ancillary Service Markets Considering Flexible Ramping Products. *Appl. Energy* **2017**, *205*, 294–303. [CrossRef]
98. Fang, Y.; Zhao, S. Look-Ahead Bidding Strategy for Concentrating Solar Power Plants with Wind Farms. *Energy* **2020**, *203*, 117895. [CrossRef]
99. Iria, J.; Soares, F.; Matos, M. Optimal Bidding Strategy for an Aggregator of Prosumers in Energy and Secondary Reserve Markets. *Appl. Energy* **2019**, *238*, 1361–1372. [CrossRef]
100. Wang, Y.; Dvorkin, Y.; Fernandez-Blanco, R.; Xu, B.; Qiu, T.; Kirschen, D.S. Look-Ahead Bidding Strategy for Energy Storage. *IEEE Trans. Sustain. Energy* **2017**, *8*, 1106–1117. [CrossRef]
101. Xie, Y.; Guo, W.; Wu, Q.; Wang, K. Robust MPC-Based Bidding Strategy for Wind Storage Systems in Real-Time Energy and Regulation Markets. *Int. J. Electr. Power Energy Syst.* **2021**, *124*, 106361. [CrossRef]
102. Zheng, Y.; Yu, H.; Shao, Z.; Jian, L. Day-Ahead Bidding Strategy for Electric Vehicle Aggregator Enabling Multiple Agent Modes in Uncertain Electricity Markets. *Appl. Energy* **2020**, *280*, 115977. [CrossRef]
103. Vagropoulos, S.I.; Bakirtzis, A.G. Optimal Bidding Strategy for Electric Vehicle Aggregators in Electricity Markets. *IEEE Trans. Power Syst.* **2013**, *28*, 4031–4041. [CrossRef]
104. Bjorgan, R.; Liu, C.-C.; Lawarree, J. Financial Risk Management in a Competitive Electricity Market. *IEEE Trans. Power Syst.* **1999**, *14*, 1285–1291. [CrossRef]
105. Chung, T.S.; Zhang, S.H.; Yu, C.W.; Wong, K.P. Electricity Market Risk Management Using Forward Contracts with Bilateral Options. *IEE Proc. Gener. Transm. Distrib.* **2003**, *150*, 588. [CrossRef]
106. Spodniak, P.; Collan, M. Forward Risk Premia in Long-Term Transmission Rights: The Case of Electricity Price Area Differentials (EPAD) in the Nordic Electricity Market. *Util. Policy* **2018**, *50*, 194–206. [CrossRef]
107. Zhong, J.; He, Y.; Wang, D.; Sun, Y.; He, T.; Peng, Y.; Yuan, T.; Luo, X. Review on market power regulation and mitigation measures in power market. *Proc. CSEE* **2018**, *9*. Available online: https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CPFD&dbname=CPFDLAST2019&filename=JDSD201810001025&v= (accessed on 2 July 2021).
108. Prabhakar Karthikeyan, S.; Jacob Raglend, I.; Kothari, D.P. A Review on Market Power in Deregulated Electricity Market. *Int. J. Electr. Power Energy Syst.* **2013**, *48*, 139–147. [CrossRef]
109. Borenstein, S.; Bushnell, J.; Knittel, C.R. Market Power in Electricity Markets: Beyond Concentration Measures. *Energy J.* **1999**, *20*, 65–88. [CrossRef]
110. Borenstein, S.; Bushnell, J.; Kahn, E.; Stoft, S. Market Power in California Electricity Markets. *Util. Policy* **1995**, *5*, 219–236. [CrossRef]

111. Wolak, F.A. Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market, 1998–2000. *Am. Econ. Rev.* **2003**, *93*, 425–443. [CrossRef]
112. Sweeting, A. Market Power in the England and Wales Wholesale Electricity Market 1995–2000. *Econ. J.* **2007**, *117*, 654–685. [CrossRef]
113. Möst, D.; Genoese, M. Market Power in the German Wholesale Electricity Market. *J. Energy Mark.* **2009**, *2*, 47. [CrossRef]
114. Shukla, U.K.; Thampy, A. Analysis of Competition and Market Power in the Wholesale Electricity Market in India. *Energy Policy* **2011**, *39*, 2699–2710. [CrossRef]
115. Shang, N. *Market Power Risk Assessment of Multi-Type Markets Participants Considering Reliability in Electricity Markets*; Zhejiang University: Hangzhou, China, 2019.
116. Dong, L.; Wang, S.; Huang, H.; Guo, H. Identification of Market Power Abuse in Spot Market of Chinese Electric Market. *Proc. CSEE* **2021**, 1–11. [CrossRef]
117. Xu, H.; Cheng, Z.; Zhang, H.; Dong, L.; Hua, H. Market Power Abuse Identification of Power Generation Enterprises Based on Improved Support Vector Machine. *J. North China Electr. Power Univ.* **2020**, *47*, 86–95.
118. Dong; Wang; Liu; Ainiwaer; Nie Operation Health Assessment of Power Market Based on Improved Matter-Element Extension Cloud Model. *Sustainability* **2019**, *11*, 5470. [CrossRef]
119. Wang, Z.; Zhang, Y.; Huang, K.; Wang, C. Robust Optimal Scheduling Model of Virtual Power Plant Combined Heat and Power Considering Multiple Flexible Loads. *Electr. Power Constr.* **2021**, *42*, 1–10.
120. Abban, A.R.; Hasan, M.Z. Solar Energy Penetration and Volatility Transmission to Electricity Markets—An Australian Perspective. *Econ. Anal. Policy* **2021**, *69*, 434–449. [CrossRef]
121. Hu, X.; Jaraitė, J.; Kažukauskas, A. The Effects of Wind Power on Electricity Markets: A Case Study of the Swedish Intraday Market. *Energy Econ.* **2021**, *96*, 105159. [CrossRef]
122. Spodniak, P.; Ollikka, K.; Honkapuro, S. The Impact of Wind Power and Electricity Demand on the Relevance of Different Short-Term Electricity Markets: The Nordic Case. *Appl. Energy* **2021**, *283*, 116063. [CrossRef]
123. Mays, J. Missing Incentives for Flexibility in Wholesale Electricity Markets. *Energy Policy* **2021**, *149*, 112010. [CrossRef]
124. The European Parliament; The Council of the European Union. *Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the Promotion of the Use of Energy from Renewable Sources and Amending and Subsequently Repealing Directives 2001/77/EC and 2003/30/EC*; The European Parliament: Brussels, Belgium; The Council of the European Union: Brussels, Belgium, 2009.
125. Banshwar, A.; Sharma, N.K.; Sood, Y.R.; Shrivastava, R. Renewable Energy Sources as a New Participant in Ancillary Service Markets. *Energy Strategy Rev.* **2017**, *18*, 106–120. [CrossRef]
126. Godoy-González, D.; Gil, E.; Gutiérrez-Alcaraz, G. Ramping Ancillary Service for Cost-Based Electricity Markets with High Penetration of Variable Renewable Energy. *Energy Econ.* **2020**, *85*, 10455. [CrossRef]
127. Goudarzi, H.; Rayati, M.; Sheikhi, A.; Ranjbar, A.M. A Clearing Mechanism for Joint Energy and Ancillary Services in Non-Convex Markets Considering High Penetration of Renewable Energy Sources. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106817. [CrossRef]
128. Zarnikau, J.; Tsai, C.H.; Woo, C.K. Determinants of the Wholesale Prices of Energy and Ancillary Services in the U.S. Midcontinent Electricity Market. *Energy* **2020**, *195*, 117051. [CrossRef]
129. Banshwar, A.; Sharma, N.K.; Sood, Y.R.; Shrivastava, R. Market Based Procurement of Energy and Ancillary Services from Renewable Energy Sources in Deregulated Environment. *Renew. Energy* **2017**, *101*, 1390–1400. [CrossRef]
130. Glass, E.; Glass, V. Enabling Supercapacitors to Compete for Ancillary Services: An Important Step towards 100 % Renewable Energy. *Electr. J.* **2020**, *33*, 106763. [CrossRef]
131. Hu, Q.; Zhu, Z.; Bu, S.; Wing Chan, K.; Li, F. A Multi-Market Nanogrid P2P Energy and Ancillary Service Trading Paradigm: Mechanisms and Implementations. *Appl. Energy* **2021**, *293*, 116938. [CrossRef]
132. Stürmer, B.; Theuretzbacher, F.; Saracevic, E. Opportunities for the Integration of Existing Biogas Plants into the Austrian Electricity Market. *Renew. Sustain. Energy Rev.* **2021**, *138*, 110548. [CrossRef]
133. Arango-Aramburo, S.; Bernal-García, S.; Larsen, E.R. Renewable Energy Sources and the Cycles in Deregulated Electricity Markets. *Energy* **2021**, *223*, 120058. [CrossRef]
134. Hasankhani, A.; Hakimi, S.M. Stochastic Energy Management of Smart Microgrid with Intermittent Renewable Energy Resources in Electricity Market. *Energy* **2021**, *219*, 119668. [CrossRef]
135. Ambec, S.; Crampes, C. Real-Time Electricity Pricing to Balance Green Energy Intermittency. *Energy Econ.* **2021**, *94*, 105074. [CrossRef]
136. Lianhe Ratings. Research Report and Prospect of Thermal Power Industry in 2020. Available online: https://pdf.dfcfw.com/pdf/H3_AP202101131450227605_1.pdf?1610555525000.pdf (accessed on 7 June 2021).
137. National Development and Reform Commission (NDRC). Circular of the National Development and Reform Commission and the National Energy Administration on Carrying out the Transmission and Upgrading of Coal-Fired Power Units Nationwide. Available online: www.gov.cn/zhengce/zhengceku/2021-11/03/content_5648562.htm (accessed on 7 November 2021).
138. Jian, Q.; Liu, X.; Yang, J.; Liu, C.; Wang, X.; Liu, D. Optimal Allocation of Power System Flexible Resources Considering Demand Response. *Mod. Electr. Power* **2021**, *38*, 286–296.

139. Kazempour, J.; Hobbs, B.F. Value of Flexible Resources, Virtual Bidding, and Self-Scheduling in Two-Settlement Electricity Markets with Wind Generation—Part I: Principles and Competitive Model. *IEEE Trans. Power Syst.* **2018**, *33*, 749–775.
140. Huang, B.; Hu, J.; Jiang, L.; Li, Q.; Feng, K.; Yuan, B. Application Value Assessment of Grid Side Energy Storage Under Typical Scenarios in China. *Electr. Power* **2021**, *54*, 158–165.
141. Shi, J.; Guo, Y.; Sun, H.; Wu, C. Review of Research and Practice on Reserve Market. *Proc. CSEE* **2021**, *41*, 123–134+403.
142. Willis, L.; Finney, J.; Ramon, G. Computing the Cost of Unbundled Services [Power Transmission]. *IEEE Comput. Appl. Power* **1996**, *9*, 16–21. [CrossRef]
143. Sun, S.; Chi, D.; Yu, B.; Zhou, M. Building a new power market system and electricity price mechanism. *Macroecon. Manag.* **2021**, *3*, 71–77.
144. Leng, Y.; Gu, W. Operating Mechanism of Australian Electric Financial Derivatives Market and Its Implications for Electricity Market Construction in China. *Electr. Power* **2021**, *54*, 36–43.
145. van Koten, S. The Forward Premium in Electricity Markets: An Experimental Study. *Energy Econ.* **2021**, *94*, 105059. [CrossRef]
146. Fang, X.; Hu, Q.; Bo, R.; Li, F. Redesigning Capacity Market to Include Flexibility via Ramp Constraints in High-Renewable Penetrated System. *Int. J. Electr. Power Energy Syst.* **2021**, *128*, 106677. [CrossRef]
147. Liu, M.; Yang, L.; Gan, D. A survey on agent based electricity market simulation. *Power Syst. Technol.* **2005**, *29*, 76–80.
148. Bao, T.; Wang, G.; Dai, Y. Future oriented experimental economics: Literature review and Prospect. *Manag. World* **2020**, *36*, 218–237.
149. Duffy, J. Macroeconomics: A Survey of Laboratory Research. *Handb. Exp. Econ.* **2016**, *2*, 1–90.
150. Binmore, K.; Klemperer, P. The Biggest Auction Ever: The Sale of the British 3G Telecom Licences. *Econ. J.* **2002**, *112*, C74–C96. [CrossRef]
151. Arifovic, J.; Duffy, J.M.; Jiang, J.H. *Adoption of a New Payment Method: Theory and Experimental Evidence*; Bank of Canada Staff Working Paper; Bank of Canada: Ottawa, ON, Canada, 2017.
152. Blinder, A.S.; Morgan, J. Do Monetary Policy Committees Need Leaders? A Report on an Experiment. *Am. Econ. Rev.* **2008**, *98*, 224–229. [CrossRef]
153. Hommes, C.; Massaro, D.; Weber, M. Monetary Policy under Behavioral Expectations: Theory and Experiment. *Eur. Econ. Rev.* **2019**, *118*, 193–212. [CrossRef]
154. Davis, D.D.; Holt, C.A. *Experimental Economics*; Princeton University Press: Princeton, NJ, USA, 1993.
155. List, J.A.; Lucking-Reiley, D. Demand Reduction in Multiunit Auctions: Evidence from a Sportscard Field Experiment. *Am. Econ. Rev.* **2000**, *90*, 961–972. [CrossRef]
156. Fan, R.; Ye, Q.; Du, J. Frontier Development of Agent—Based Computational Economics: A Survey. *Econ. Rev.* **2013**, *2*, 145–150.
157. Mi, J.; Lin, R. How equity preference affects open innovation: A study based on Computational Economics. *Chin. J. Manag. Sci.* **2015**, *23*, 157–166.
158. Arthur, W.B.; Holland, J.H.; Lebaron, B.; Palmer, R.; Taylor, P. *Asset Pricing under Endogenous Expectation in an Artificial Stock Market*; Working Papers; Santa Fe Institute: Santa Fe, NM, USA, 1996.
159. Barrows, C.; Preston, E.; Staid, A.; Stephen, G.; Watson, J.-P.; Bloom, A.; Ehlen, A.; Ikaheimo, J.; Jorgenson, J.; Krishnamurthy, D.; et al. The IEEE Reliability Test System: A Proposed 2019 Update. *IEEE Trans. Power Syst.* **2020**, *35*, 119–127. [CrossRef]
160. Gacitua, L.; Gallegos, P.; Henriquez-Auba, R.; Lorca, Á.; Negrete-Pincetic, M.; Olivares, D.; Valenzuela, A.; Wenzel, G. A Comprehensive Review on Expansion Planning: Models and Tools for Energy Policy Analysis. *Renew. Sustain. Energy Rev.* **2018**, *98*, 346–360. [CrossRef]
161. Xu, Y.; Myhrvold, N.; Sivam, D.; Mueller, K.; Olsen, D.J.; Xia, B.; Livengood, D.; Hunt, V.; d'Orfeuil, B.R.; Muldrew, D.; et al. U.S. Test System with High Spatial and Temporal Resolution for Renewable Integration Studies. In Proceedings of the 2020 IEEE Power & Energy Society General Meeting (PESGM), Montreal, QC, Canada, 2–6 August 2020; pp. 1–5.
162. Denholm, P.; Hand, M. Grid Flexibility and Storage Required to Achieve Very High Penetration of Variable Renewable Electricity. *Energy Policy* **2011**, *39*, 1817–1830. [CrossRef]
163. Li, Z.; Chen, S.; Dong, W.; Liu, P.; Du, E.; Ma, L.; He, J. Low Carbon Transition Pathway of Power Sector Under Carbon Emission Constraints. *Proc. CSEE* **2021**, *41*, 3987–4001.
164. Wang, B.; Xia, Y.; Xia, Q.; Zhang, H.; Han, H. Model and Methods of Generation and transmission scheduling of Inter-regional Power Grid via HVDC Tie-line. *Autom. Electr. Power Syst.* **2016**, *40*, 8–13+26.
165. Li, X.; Zhang, G.; Guo, Z. N-1 principle steady state security restriction on power flow of transmission tie line group. *Electr. Power Autom. Equip.* **2004**, *11*, 10–13+17.
166. Khoshjahan, M.; Moeini-Aghtaie, M.; Fotuhi-Firuzabad, M.; Dehghanian, P.; Mazaheri, H. Advanced Bidding Strategy for Participation of Energy Storage Systems in Joint Energy and Flexible Ramping Product Market. *IET Gener. Transm. Distrib.* **2020**, *14*, 5202–5210. [CrossRef]
167. Castillo-Ramírez, A.; Mejía-Giraldo, D. Measuring Financial Impacts of the Renewable Energy Based Fiscal Policy in Colombia under Electricity Price Uncertainty. *Sustainability* **2021**, *13*, 2010. [CrossRef]
168. Guo, W.; Liu, P.; Shu, X. Optimal Dispatching of Electric-Thermal Interconnected Virtual Power Plant Considering Market Trading Mechanism. *J. Clean. Prod.* **2021**, *279*, 123446. [CrossRef]
169. Lasemi, M.A.; Arabkoohsar, A. Optimal Operating Strategy of High-Temperature Heat and Power Storage System Coupled with a Wind Farm in Energy Market. *Energy* **2020**, *210*, 118545. [CrossRef]

170. Rentier, G.; Lelieveldt, H.; Kramer, G.J. Varieties of Coal-Fired Power Phase-out across Europe. *Energy Policy* **2019**, *132*, 620–632. [CrossRef]
171. Trencher, G.; Healy, N.; Hasegawa, K.; Asuka, J. Discursive Resistance to Phasing out Coal-Fired Electricity: Narratives in Japan's Coal Regime. *Energy Policy* **2019**, *132*, 782–796. [CrossRef]
172. Webb, J.; de Silva, H.N.; Wilson, C. The Future of Coal and Renewable Power Generation in Australia: A Review of Market Trends. *Econ. Anal. Policy* **2020**, *68*, 363–378. [CrossRef]
173. Muñoz, F.D.; Suazo-Martínez, C.; Pereira, E.; Moreno, R. Electricity Market Design for Low-Carbon and Flexible Systems: Room for Improvement in Chile. *Energy Policy* **2021**, *148*, 111997. [CrossRef]
174. Das, P.; Mathuria, P.; Bhakar, R.; Mathur, J.; Kanudia, A.; Singh, A. Flexibility Requirement for Large-Scale Renewable Energy Integration in Indian Power System: Technology, Policy and Modeling Options. *Energy Strategy Rev.* **2020**, *29*, 100482. [CrossRef]

energies

MDPI

Article

Improving Artificial Intelligence Forecasting Models Performance with Data Preprocessing: European Union Allowance Prices Case Study

Miguel A. Jaramillo-Morán [1,*], Daniel Fernández-Martínez [1], Agustín García-García [2] and Diego Carmona-Fernández [1]

[1] Department of Electrical Engineering, Electronics and Automation, School of Industrial Engineering, University of Extremadura, Avda. Elvas s/n, 06006 Badajoz, Spain; danielfm@unex.es (D.F.-M.); dcarmona@unex.es (D.C.-F.)
[2] Department of Economics, Faculty of Economics and Business Sciences, University of Extremadura, Avda. Elvas s/n, 06006 Badajoz, Spain; agarcia@unex.es
* Correspondence: miguel@unex.es; Tel.: +34-924-289-928

Abstract: European Union Allowances (EUAs) are rights to emit CO_2 that may be sold or bought by enterprises. They were originally created to try to reduce greenhouse gas emissions, although they have become assets that may be used by financial intermediaries to seek for new business opportunities. Therefore, forecasting the time evolution of their price is very important for agents involved in their selling or buying. Neural Networks, an artificial intelligence paradigm, have been proved to be accurate and reliable tools for time series forecasting, and have been widely used to predict economic and energetic variables; two of them are used in this work, the Multilayer Preceptron (MLP) and the Long Short-Term Memories (LSTM), along with another artificial intelligence algorithm (XGBoost). They are combined with two preprocessing tools, decomposition of the time series into its trend and fluctuation and decomposition into Intrinsic Mode Functions (IMF) by the Empirical Mode Decomposition (EMD). The price prediction is obtained by adding those from each subseries. These two tools are combined with the three forecasting tools to provide 20 future predictions of EUA prices. The best results are provided by MLP-EMD, which is able to achieve a Mean Absolute Percentage Error (MAPE) of 2.91% for the first predicted datum and 5.65% for the twentieth, with a mean value of 4.44%.

Keywords: European Union allowances; CO_2 price prediction; emission allowances; neural networks; forecasting

Citation: Jaramillo-Morán, M.A.; Fernández-Martínez, D.; García-García, A.; Carmona-Fernández, D. Improving Artificial Intelligence Forecasting Models Performance with Data Preprocessing: European Union Allowance Prices Case Study. *Energies* **2021**, *14*, 7845. https://doi.org/10.3390/en14237845

Academic Editor: Nuno Carlos Leitão

Received: 18 October 2021
Accepted: 20 November 2021
Published: 23 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Since the European Union (EU) created the Emission Trading System (EU ETS) in 2005 to combat climate change, it has become one of the cornerstones of the European environmental policy, with strong implications for industrial activities and repercussions that reach all economic and social sectors. Its main goal is to reduce greenhouse gas emission. It is supposed that companies producing carbon emissions must effectively manage associated costs by buying or selling rights to emit CO_2, the so-called European Union Allowances (EUAs). The EU ETS is a cap-and-trade system, which includes only large stationary sources of emissions belonging to the most pollutant industrial sectors of the European economy (power plants, oil refineries, ferrous metallurgy, cement clinker or lime, glass—including glass fiber—ceramic products by firing, and pulp, paper and board).

Companies involved can either use EUAs to compensate their emissions or sell them to others that need them [1]; they are allowed to trade emission allowances freely within the EU, so the system seeks to ensure that overall emissions are reduced, but also that cuts are made by those companies that can achieve the most efficient abatement costs [2,3].

Although the main EUA market goal was to give firms an incentive to move towards a less fossil fuel-intensive production, it also provides a new asset and new business development opportunities for financial intermediaries. Thus, current allowance prices, as well as the predicted EUA prices, are critical for companies, brokers, traders, and investors, and they can also affect decarbonization investment decisions [4–7]. Therefore, it may be stated that, although the market is designed to encourage decarbonization investments in the industrial sectors subject to the system, the consideration of emission allowances as financial assets introduces a new component that may complicate the market's effectiveness in environmental terms.

Regarding the consideration of EUA as financial assets, it would be helpful to obtain short-term reliable forecasts, since agents involved in this market need to make quick buying and selling decisions in order to obtain maximum profitability [8].

However, in terms of the environmental component underlying the market, agents involved in the decisions need to use a longer time horizon. For the system to work properly, the EUA market must provide information that incentivizes decarbonization decisions, even though investments in technological improvements or fuel substitution can take long payback periods. Therefore, incentives for decarbonization decisions must be credible and long-lasting, so that, rather than the information contained in the day-to-day fluctuation of prices, it is the trend of EUA prices that is of interest. In order to favor market stability and fulfillment of the objective of incentivizing decarbonization decisions by maintaining EUA prices, in 2019 the EU created a mechanism—the Market Stability Reserve (MSR)—with the aim of removing the excess of emission allowances that had generated the crisis since 2008. The MSR was designed to absorb excess EUAs in the short-term and to match the supply of EUAs in case of severe shortages in the long-term. However, EUA prices have risen sharply in the last years and this increase can hardly be explained by the purchasing needs of the companies included in the system, but rather by the arrival of other investors in the market, outside the polluting sectors, which are governed by objectives other than those initially set out in the EU ETS. In this case, we are talking about the behavior of EUAs as financial assets, whose price has shown not only rapid growth but also high volatility in the short-term.

The evolution of EUA prices has complicated the current economic situation, as their sharp rise has affected the costs of various sectors, causing significant increases in electricity prices. For example, the wholesale electricity market prices in Spain in September 2021 are three times higher than the year before. Although the cost of EUAs is not the only factor responsible for this price increase, according to some estimates [9], in the case of Spain, around 20% of this increase would be related to the rise of CO_2 prices in the EU ETS. Other European markets have experienced an evolution of wholesale electricity prices very similar to the Spanish one. In this way, the energy price increase in Europe has become macroeconomically significant [10]. Several factors, in addition to the CO_2 rising trend, are responsible for the rise of electricity prices: an increase of natural gas demand forced by higher demand of electric energy along with a decrease in renewable electricity production and a significant increase in coal prices. The pricing system in European (and other) electricity markets assumes that the price of electricity reflects the marginal production cost of the most expensive technology involved in generation. Therefore, fossil fuel power producers incorporate the price of EUAs into the marginal cost, passing on CO_2 prices into electricity prices and, where appropriate, incentivizing investment in renewable sources. The maintenance of high EUAs prices, although may be compatible with the environmental objective of the system and reinforce the incentives for decarbonization, can also lead to problems derived from the increase in costs in all sectors, including the loss of commercial competitiveness in Europe. In addition, the effects of higher electricity prices on consumers can be very significant, affecting different social classes in different ways. Therefore, prediction of the time evolution of EUA prices has become a fundamental tool for enterprises dealing with them, both for the short-term, to manage their day-to-

day evolution—linked to its behavior as a financial asset—and the long-term, related to investments and decisions aimed at reducing the emission of greenhouse gases.

Prediction of EAU prices can be carried out by organizing them as times series so that tools usually used to carry out predictions in this field could be applied to obtain those future price predictions. It is generally accepted that economic variables follow nonlinear processes [11]; non-linearity represents a major difficulty when modeling the dynamics of time series describing those economic (and financial) variables' evolution [12,13]. To deal with those kinds of complex problems, classical linear forecasting tools such as ARIMA are used along with other forecasting tools in [14] a Fourier Series Expansion optimized with Particle Swarm Optimization (PSO) was used to refine the predictions provided by a seasonal ARIMA to forecast electricity consumption, while in [15] ARIMA was combined with Autoregressive Conditional Heteroscedasticity (ARCH) to forecast CO_2 emissions in Europe. The hybrid models clearly outperformed the basic ARIMA. Nevertheless, other forecasting tools which could provide more accurate predictions when dealing with a nonlinear behavior have been also used. The ARCH and Generalized ARCH (GARCH) models have proved very useful for financial time series analysis. Thus, they have been also used to forecast CO_2 allowance prices, sometimes without any other tool [16], where a modification of its basic structure (fractionally integrated asymmetric power GARCH) is used, integrated into another forecasting model such as Markov chains [4,13] or by forming a hybrid model with other forecasting tool such as ARIMA [15].

Despite the good results obtained by those models with some time series, the development of new forecasting tools based on artificial intelligence have driven many researchers to use them to forecast time series, as they are especially well suited to deal with the nonlinear behavior of complex series [17]. In this work several forecasting tools were tested and Artificial Intelligence models clearly outperformed statistical ones such as ARIMA or GARCH in electricity price forecasting. Between them, Neural Networks (NN) have been widely used to forecast variables related to economy or energy, as they have been able to provide very accurate predictions. One of the most popular ones is the Multilayer Perceptron (MLP); it is one of the first neural models developed, and despite its simplicity, it has been widely used, as it is able to provide very accurate predictions of complex nonlinear time series. There are several fields where it has been used, such as forecasting of electric power transactions [18], natural gas demand [19], electric energy consumption [20], stock market variables [21,22] or electricity prices [23]. Despite its simplicity, MLP has been able to provide accurate and reliable predictions of different variables such as those mentioned above. In fact, it is able to provide predictions that are as good as those obtained with other more elaborated neural models and, indeed, to outperform some of them [18]. They have been also used to forecast CO_2 emission allowance prices [24,25]. In [24] it provided direct predictions of EUA prices while in [25] it was combined with a mixed data sampling regression (MIDAS) to forecast carbon prices in a Chinese market with the help of several energy, weather and environmental variables. Nevertheless, MLP is not the only neural model usually used for time series forecasting. The development of new complex neural structures known as deep learning neural networks, so-called because of the high number of processing elements, has driven many researchers to use some of those structures to forecast time series [17]. Long Short-Term Memory (LSTM) is one of such structures, as it has provided very accurate and reliable results when applied to carry out very complex data processing such as speech or text recognition, and they are able to analyze the time and contextual dependencies present in those problems. This is why they are supposed to be able to provide accurate predictions in time series forecasting; indeed, they have been used to predict electric energy load [26–28] or electricity prices [29]. In [29] LSTM clearly outperformed ARIMA. In [27] LSTM made up a hybrid model with VMD and a Genetic Algorithm, while in [28] LSTM combined with Empirical Mode Decomposition (EMD) and information related to day similarity was able to provided better predictions than ARIMA, MLP and Support Vector Regression (SVR). Other Artificial Intelligence tools have been also used for time series forecasting such as Random Forest (RF), Gradient

Boosting (GB) and Extreme Gradient Boosting (XGBoost) [30] or SVR [31], although neural networks are nowadays the preferred option because they are better suited for the time series forecasting problem. In fact, when comparing performances, neural networks usually outperform other forecasting tools, especially those known as statistical methods such as ARIMA [20,25,27–29].

Although the forecasting tools described are able to provide accurate and reliable predictions when forecasting time series, a lot of work has been carried out in order to improve their performance by preprocessing available data. The aim is to modify the time series to provide several ones which could be more efficiently predicted or to extract information about the series time evolution that could help the forecasting tools to improve performance. The Empirical Mode Decomposition (EMD) decomposes a time series into a set of subseries, each one with a proper oscillatory behavior which is easier to predict; they are separately forecasted and then added to obtain the original series forecasting. This is a heuristic algorithm that suffers from a lack of mathematical theory supporting it, and to overcome this problem, the Variational Mode Decomposition has been developed. They both have been used with different forecasting tools such as LSTM [27,28], SVR [31] or spiking neurons [32]. A simplified version of those decomposition processes could be obtained by splitting the time series only into its trend and fluctuations. In this way two series are obtained: one describing the global trend of data and the other their seasonal and cyclic oscillations. It has provided good results when applied to electric consumption forecasting [20] and also to EUA prices prediction [33]. Another approximation to this decomposition is a regression algorithm that samples a dataset at different frequencies (MIDAS), which have been developed to deal with econometric series, and have been also applied to carbon prices forecasting [25]. All these algorithms have provided good results, and it is not possible to select one as the best option, since they all have their pros and cons and selecting one or another depends on the problem at hand and the researcher's own experience. In any case, preprocessing has become a fundamental step in the forecasting process, as many works have proved that it has improved the performance of forecasting tools when properly selected and applied to the time series to be predicted.

There are not many works devoted to forecast CO_2 prices [24,25,32,33]. Nevertheless, the increasing interest in environmental preservation and the influence that free auctioning of EUA has on the final price of electric energy have increased the interest of researchers in this field. Several works have appeared in which CO_2 allowance prices are predicted not only in Europe but also in other countries [25,31]. Most of them use neural networks tools along with some kind of preprocessing to carry out this task.

The aim of this work is to test several forecasting tools along with a proper preprocessing of data to provide accurate and robust predictions of 20 days ahead of carbon prices; usually, a one-day-ahead prediction is provided in most works. Nevertheless, multistep predictions could be potentially more interesting than those of one only data ahead because a more complete time evolution of the forecasted variable is provided. Despite this, there are few works that provide multistep predictions [25,33,34]. So, in this work 20 future prices are provided each time a prediction is carried out. In this way, both short and long-term predictions are provided at once, so that this information could be valuable for both traders considering EUA as financial assets (who carry out shot-term buying and selling operations) and agents involved in decision-making related to decarbonization polices (who would prefer a long-term prediction of the price evolution).Two neural networks (MLP and LSTM) that have proved to be very accurate forecasting tools have been used. Another machine learning tool (XGBoost) has also been tested because, although it has been little used in time series forecasting, it has provided very good results in classification problems. Two preprocessing strategies have been tested to improve the prediction accuracy: the decomposition of the original time series into its trend and fluctuations components and the Empirical Mode Decomposition. The results obtained were analyzed to find out the structure providing the best performance. They showed that a proper preprocessing of data before being predicted by the forecasting tools clearly improves the prediction accuracy.

The paper is organized as follows: Section 2 describes the models used to forecast future CO_2 prices along with the preprocessing algorithms, while Section 3 describes the figures of merit used to measure the forecasting performance and the data preprocessing applied to the original time series to improve the forecasting models accuracy. Then, the simulation results obtained with the three models tested and the two preprocessing algorithms are described. In Section 4, those results are compared, and the best performance identified. Finally, Section 5 presents the conclusions.

2. Materials and Methods

2.1. Artificial Intelligence Tools for Time Series Forecasting

Neural Networks are a set of artificial intelligence algorithms which simulate the structure of brains to try to mimic some of their abilities. They have been widely used to forecast time series because of their capability to learn the dynamic behavior of complex systems. They have provided very accurate predictions when dealing with nonlinear systems, where other tools fail to provide them. There are several neural models, although only some of them are used to carry out time series forecasting. One of the most widely used in these tasks is Multilayer Perceptron (MLP) [35]. It is a very simple classical model which is organized in a multilayer structure, with an input layer, several hidden ones and an output layer. Information is processed while it flows from input to output, which is why they are known as Feedforward Neural Networks (FFNN). There are several neural models which also process information following this data flow, although MLP is the most popular one; however, they are not able to deal with data strongly dependent on past information, such as that found in speech processing. Thus, in order to address these kinds of problems, new neural models have been developed in which feedback has been added to a FFNN to provide the network the ability to retain past information to be processed with present data. Thus, information may flow back from one layer to another preceding it, or among neurons in the same layer, providing the network with a kind of "memory", as those data may be seen as past states of the network which can be processed along with new data presented to the network. These types of networks are known as Recurrent Neural Networks (RNN), and some of their models have been used to predict time series, as they are supposed to perform well in forecasting tasks because of their ability to process "past" information along with the present data. Since their structure is more complex than that of FFNNs, MLP will be first described and then, based on its structure, that of the RNN used in this work will be studied.

2.1.1. Multilayer Perceptron

The succeed of MLPs to provide accurate predictions comes from the fact that they have proved to be universal approximations [36,37], as they can approximate any continuous function with one hidden layer, provided that this one has enough neurons. Its simplicity and simple programming, along with this property, have made them one of the most popular neural models for time series forecasting. In MLPs, the first layer is actually the set of input data to the neural network. In the hidden layer (or layers if several of them are considered), the information provided to the network is processed and then passed to the output layer, which provides the network response. Each neuron in a layer processes the information it receives from all the neurons in the previous one:

$$y_j = \sigma\left(\sum w_{ji}\, x_i + b_j\right), \tag{1}$$

where x_i represents the ith input of the jth neuron, w_{ji} the strength (weight) of the connections between this neuron and all those in the previous layer, y_j the neuron output and b_j a bias constant.

$\sigma(\cdot)$ is an activation function which provides the network the nonlinear characteristic that allows the identification of the nonlinear behavior inherent to complex dynamics. In the hidden layers it is usually the hyperbolic tangent or the logistic function:

$$y_j = \frac{1}{1 + e^{-(\sum w_{ji}\ x_i + b_j)}} \tag{2}$$

In the output layer, this function is usually a linear one, because it is usually assumed that this layer only provides an adaptation of the neural network response to the data structure.

The neural network's ability to approximate any system is provided by its learning capability. Thus, it must be trained to learn the behavior of the system it tries to reproduce. To do this, it must be trained with a specific set of data which must be arranged in pairs of network inputs (patterns) and desired outputs, so that each time one of those input patterns is presented to the network it provides an output response, which must be compared with the desired one to obtain an error measurement. The errors obtained for all the patterns will be summed to obtain a global error whose value must be minimized by properly adapting each neuron's weights in order to guarantee that the network has learned all the patterns. The algorithm performing this process is the well-known "Backpropagation" [35].

A dataset different from that used for training must be processed to validate the network to guarantee that it is able to provide a proper response to patterns different from those previously learned. When dealing with time series, this means that the network will be able to provide an accurate prediction of future values when past ones are provided as network inputs.

2.1.2. Long Short-Term Memories

Long Short-Term Memory (LSTM) [38] have a multilayer structure similar to that of MLP, but now the neural outputs of a layer are fed back to all the neurons in that layer. In addition, a sort of "memory" is stored in each cell, recording past information received by the neuron. However, all this information—new data, feedback and "memory"—is not processed by neurons directly; in fact, several activation gates decide which of them will be used whether or not a neuron output will be provided. To carry out this control process, the neural model of the LSTM has three activation gates which process data from the previous layer along with those from neurons in the same one providing signals that control inputs, "memory" update and output:

$$i_j = \sigma\left(\boldsymbol{W_i}\cdot\left[\boldsymbol{x}^t,\ \boldsymbol{h}^{t-1}\right] + b_i\right), \tag{3}$$

$$f_j = \sigma\left(\boldsymbol{W_f}\cdot\left[\boldsymbol{x}^t,\ \boldsymbol{h}^{t-1}\right] + b_f\right), \tag{4}$$

$$o_j = \sigma\left(\boldsymbol{W_o}\cdot\left[\boldsymbol{x}^t,\ \boldsymbol{h}^{t-1}\right] + b_o\right). \tag{5}$$

In these formulas, $\boldsymbol{W_i}$, $\boldsymbol{W_f}$ and $\boldsymbol{W_o}$ represent weight matrices while $[\boldsymbol{x}^t, \boldsymbol{h}^{t-1}]$ represents an input vector made up with data from the previous layer, $\boldsymbol{x}^t$, and feedbacks (one time step delayed) from neurons in the same layer, $\boldsymbol{h}^{t-1}$. b_o, b_f and b_o are bias weights. σ is an activation function which may be the logistic one or the hyperbolic tangent, although the first one is preferred for the activation gates.

The cell input is:

$$z_j = \sigma\left(\boldsymbol{W_z}\cdot\left[\boldsymbol{x}^t,\ \boldsymbol{h}^{t-1}\right] + b_z\right). \tag{6}$$

where $\boldsymbol{W_z}$ and b_z represent a weight matrix and a bias, respectively.

The cell "memory" c_j^t is updated by taking into account both the cell input z_j and its past value c_j^{t-1} according to the expression:

$$c_j^t = i_j \cdot z_j + f_j \cdot c_j^{t-1} \tag{7}$$

where i_j decides whether or not new information is added to the "memory" and f_j controls whether old information should be retained or forgotten. Thus, it is usually known as the "forget" gate.

Finally, the cell output is:

$$y_j = o_j \cdot \sigma\left(c_j^t\right). \tag{8}$$

As this cellular structure is rather more complex than that of other neural models, it is usually known as "cellular block", which, in addition, may be made up of one or several neurons. In this last case, all the neurons in a block share the same control gates. Thus, a layer will have several blocks, each one with one or several neurons. When used to carry out very complex tasks, such as text or speech recognition, a high number of layers are used. This is why this neural model (along with others with a high number of layers and neurons in each one) is known as "deep learning" models.

LSTMs are trained with a variation of the well-known "Backpropagation" algorithm, which is adapted to deal with the recurrent structure of this neural model. Two variants of the basic algorithm are used: that known as "truncated Backpropagation Through Time" (BPTT) for adjusting weights of cell outputs and output gates and "Real-Time Recurrent Learning (RTRL)," used to adapt weights of cell inputs, input gates and forget gates [39].

2.1.3. XGBoost

XGBoost (Extreme Gradient Boosting) [40] is a machine learning algorithm for decision trees boosting. It is an open-source library provided for most programming environments used nowadays, and has become a very popular tool for machine learning, since it was able to win many of the challenges proposed in the 2015 Kaggle and KDDcup competitions. As previously stated, XGBoost is a machine learning system for tree boosting, that is to say, XGBoost provides a procedure to define an ensemble of decision trees which carries out classification or regression of the data presented as input to the model (this is why these trees are usually known as CART: Classification and Regression Trees). A decision tree provides an answer which may be binary (the data presented belongs or not to a certain class) or numerical, which may be represented by a function. In tree boosting, this last one is used, so that the output of the tree ensemble has the form:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i), \tag{9}$$

where $\boldsymbol{x}_i$ represents an input (a vector defining a pattern to be classified), $f_k(\boldsymbol{x}_i)$ the function describing the answer of each decision tree, K the number of trees and $\hat{y}_i$ the answer of the whole ensemble.

The whole tree ensemble is to be trained with a set of input-output pairs $(\boldsymbol{x}_i, y_i)$, where $\boldsymbol{x}_i$ represents the pattern to be classified and $\boldsymbol{x}_i$ its desired output, by adjusting the parameters defining the tree structure by minimizing a cost function in a supervised process. Nevertheless, this is a harder problem than the learning strategies of other machine learning models, such as the descend gradient usually used with neural networks, since training all the trees at once may become too computationally intensive. Thus, a simplified iterative strategy, known as "boosting", is used to train one tree at each step.

The training process starts by fixing to zero the value of the first prediction:

$$\hat{y}_i^0 = 0. \tag{10}$$

In this expression, the superscript refers to the time step of the process. Now, a first tree, defined by its representative function, is added to the tree ensemble, whose output is now:

$$\hat{y}_i^1 = \hat{y}_i^0 + f_1(\boldsymbol{x}_i) = f_1(\boldsymbol{x}_i). \tag{11}$$

This new tree is trained with a subset of the training dataset and then predictions for the whole dataset are obtained. As a number of these predictions are probably different from their expected values, a new tree is added and then trained with the set of misclassified patterns. So, the ensemble prediction function now becomes:

$$\hat{y}_i^2 = \hat{y}_i^1 + f_2(\boldsymbol{x}_i) = f_1(\boldsymbol{x}_i) + f_2(\boldsymbol{x}_i). \tag{12}$$

The process is repeated until a certain accuracy is achieved, or the number of trees reaches a certain previously fixed value. The prediction function of the tree ensemble will be:

$$\hat{y}_i^t = \hat{y}_i^{t-1} + f_t(\boldsymbol{x}_i) = \sum_{k=1}^{t} f_k(\boldsymbol{x}_i). \tag{13}$$

The cost function to be minimized when training the trees is:

$$\mathcal{L} = \sum_i l(\hat{y}_i,\ y_i) + \sum_k \Omega(f_k) \tag{14}$$

where i stands for the number of patterns used for training and k for the number of trees. $l(\hat{y}_i,\ y_i)$ is a measure of the errors obtained in each prediction. It is usually the Mean Squared Error. $\Omega(f_k)$ is a regularization term that measures the simplicity of the tree structures. It helps to obtain a structure as simple as possible.

2.2. Data Preprocessing

Artificial Intelligence tools usually provide good performance when forecasting non-linear time series. Even so, those results may be improved when input data are adequately preprocessed in order to obtain a new dataset which could be more easily predicted by the tool. Several works have proved the success of this strategy when applied to neural models [20,27,31,32,41]. Many times, preprocessing has a very sophisticated structure which is more complex than that of the forecasting tool, therefore the question arises about whether it is the forecasting tools which provide the prediction with success or the preprocessing that is applied. To overcome this issue, two simple preprocessing strategies will be tested in this work in order to prove that it is not necessary to use such complex structures to improve the forecasting tool accuracy. The first one provides a simple decomposition of the original series into its trend and superimposed oscillations while the second is a more elaborated one, the Empirical Mode Decomposition (EMD), which decomposes the time series into several simpler ones by means of an iterative procedure.

2.2.1. Trend and Fluctuations Decomposition

Many time series show a combination of different behaviors: long-term ones, which define a certain trend of data, and short and medium-term variations superimposed on it. Hence, it is usual in time series forecasting to decompose a time series into three kinds of components: trend, seasonal, and cyclic. The first one represents, as pointed out above, a long time rising or decreasing evolution, the second, an oscillatory evolution associated to seasonal effects such as day of the week, month, season, weather, while the third oscillation is caused by economic or social influences on data. Sometimes, a fourth term related to noise may be also included, and in many time series, seasonal and cyclic factors are difficult to identify as two isolated components, although the series shows a clear oscillatory behavior. In those cases, the decomposition can be simplified if the series is only split into trend and fluctuations, which comprise both seasonal and cyclic components

around it (errors could be also assumed as integrated in this oscillatory component). This decomposition will be used in this work because of its simplicity and easy programming.

To carry it out the trend component will be first extracted by means of a softening process of the time series. There are a number of methods that can be used to do this (splines, low-pass filters, moving average, etc.) although in this work, the moving average with constant weights will be used because of its simplicity and because it has proved to provide accurate predictions [20,33]. This algorithm replaces each element of the time series by the mean value of the set made up of that element and $(n-1)$ ones preceding it:

$$x^t(t) = \frac{1}{n}(x(t) + x(t-1) + \ldots + x(t-(n-1))). \quad (15)$$

The fluctuations component will be obtained by subtracting the trend series from the original one. Then, they are forecasted separately, and their predictions added to obtain the prediction of the original series.

2.2.2. Empirical Mode Decomposition

As time series often have an oscillatory behavior, a good strategy to carry out the preprocessing process could be to identify and extract periodical components which could be more easily forecasted. This strategy has the drawback that it demands those components to be associated with frequencies, which should be clearly identified, and this is not usually the case. Instead, a lot of frequencies define the spectral profile of most of time series. To overcome this problem, an empirical tool has been developed which decomposes a time series with oscillatory behavior into a set of new series with an oscillatory behavior closely related to a certain frequency. This tool is known as Empirical Mode Decomposition (EMD) [42], and each one of the new series it provides is known as Intrinsic Mode Function (IMF); it is a numerical method that requires adjustment until proper IMFs are obtained. It is worth noting that an IMF is not a function but a time series which accomplish with two properties: the number of local minima and maxima must be equal or differ only by one and its mean value must be zero. The first condition may be also defined as: only one extreme point can be between two consecutive zero-crossing points. The second one means that the time series is stationary, a fact that makes its prediction easier.

In this way, an oscillatory time series can be decomposed into the sum of IMFs and a residue:

$$x(t) = \sum_n x_n(t) + r(t). \quad (16)$$

Taking into account these conditions, the EMD algorithm works as follows.

Take all maxima and minima points in the original time series and build two new series by interpolating each set of points with cubic splines. Thus, two envelops will be obtained, one for maxima and another for minima.

Obtain the mean series $m_n(t)$ of both envelops. Then, subtract this new one from the original series. It is a candidate to be an IMF:

$$c_n(t) = x(t) - m_n(t). \quad (17)$$

Verify whether this last series accomplishes the two properties of an IMF. If not, this series will be considered as a new "original" series ($x^s(t) = c_n(t)$) and the processes of maxima and minima extraction, mean calculation, subtraction and verification will be repeated (s = 1, 2, ... , S) until the series obtained accomplishes IMF's conditions or a stop criterion is achieved. This process is usually known as "sifting".

Once a new IMF is obtained ($x_n(t) = c_n(t)$) it will be subtracted from the series from which it was derived and the result will be considered as a new "original" one, which will undergo the process described above:

$$x^n(t) = x^{n-1}(t) - c_n(t). \quad (18)$$

The process will be repeated until a stop condition is achieved. Then, the last IMF will be subtracted from its "original" series and the result will be considered as a residue.

This procedure demands two stop conditions: one for "sifting" and the other for the whole process. The fists one should be accomplished when an IMF candidate fits the two required conditions. Nevertheless, this could lead the algorithm to over-sifting, providing a lot of meaningless IMFs. To avoid this effect, an early stopping criterion must be defined. Usually, a low threshold for the IMF candidate variance is fixed, so that once it is reached, the process will stop. The second stop condition will be reached when the residue, the series obtained after a new IMF is subtracted from its "original" series, accomplished with one of the following conditions: it is constant, has a constant slope or contains one only extreme.

3. Results

The time series used in this work is the daily spot price of a ton of CO_2 quoted on the European Energy Exchange (EEX) in Leipzig, Germany. It ranges from 14 October 2009 to 1 January 2021 with a total of 2890 data, as seen in Figure 1. This plot shows two different behaviors of prices: a more or less soft evolution with medium and low values until 2018, and a clear rising trend with high values and steep variations after this year. They have been arranged into a time series to train and then validate the performance of the forecasting tools proposed in this work. Only past data of prices have been used to forecast future values.

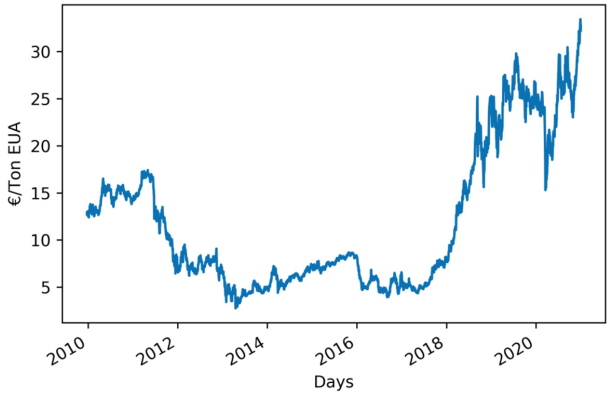


Figure 1. Daily spot prices of CO_2.

It is usual in time series forecasting to use the first data (60–80% of them) to train the model and the remaining (40–20%) to validate its performance. Two of such possible divisions have been tested in this work: 60–40% (training-validation) and 80–20%. Training data will be used to learn the times series behavior by adjusting the model's inner parameters. Once the forecasting tool (neural networks or XGBoost) is trained, the model so obtained is used to forecast with data from the validation dataset (the predictions obtained in this manner will be compared with the actual values of this dataset to obtain a measurement of the forecasting model accuracy).

Conversely, it should be taken into account that validation data with values significantly higher than those used for training and with steep variations (as those at the end of the time series) could jeopardize the accuracy of the forecasting models, as they have a behavior different from those used form training. To find out whether or not this behavior of the data worsens the performance of the forecasting models they will be also tested with a simpler dataset with a less steep evolution: the time series made up with data from 2009 to 2016, where validation data are similar to those used for training. In other words, the last data of the time series, those with higher values and steep variations, have been removed.

The different forecasting structures defined in this work will be tested with both datasets and their corresponding performances compared.

The different behavior of both datasets may be better understood when statistical information describing their data distribution is provided. They may be seen in Tables 1 and 2. The total number of data along their mean values, standard deviation, minimum and maximum values and the values defining each quartile are presented. From these data it may be concluded that the time series from 2009 to 2016 presents a more bonded behavior with lower fluctuations. Information regarding the scaled versions of both datasets (see below) is also provided.

Table 1. Statistics of the original dataset.

	Original	MinMax	Standard
Data	2890	2890	2890
Mean	11.895	0.297	0.000
Std	7.44	0.224	1.000
Min	2.750	0.000	−1.229
25%	5.920	0.103	−0.803
50%	8.145	0.175	−0.504
75%	15.875	0.427	0.534
Max	33.440	1.000	2.896

Table 2. Statistics of the reduced dataset.

	Original	MinMax	Standard
Data	1834	1834	1834
Mean	8.543	0.414	0.000
Std	3.848	0.271	1.000
Min	2.680	0.000	−1.523
25%	5.602	0.206	−0.764
50%	7.185	0.318	−0.352
75%	12.415	0.687	1.006
Max	16.840	1.000	2.156

The performance of predictions will be measured with two figures of merit: the Mean Absolute Percentage Error (MAPE) and the Root Mean Squared Error (RMSE):

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{A_i - F_i}{A_i} \right| \cdot 100, \tag{19}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (A_i - F_i)^2} \tag{20}$$

where A_i is an actual datum, F_i a forecasted one and N the total number of data predicted.

3.1. Data Scaling

The structure of the data shown in Figure 1 suggests that it may be difficult for the forecasting tool to provide accurate predictions because a lot of extreme values appear. To overcome this problem, very common in both regression and classification problems, data were scaled before using them, that is to say, they were transformed into a new bounded dataset. In the programming environment used in this work, Python, there are two algorithms which are mainly used to carry out this task: normalization and standardization. The first one transforms the original dataset into another in which values

are included in interval [1]. There are several ways to do this, although in this work, the Min-Max one has been selected because of its simplicity:

$$x_i^n = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{21}$$

where x_i^n represents the normalized datum, x_i the original one and x_{max} and x_{min} the maximum and minimum data in the original data set.

The second algorithm transforms the dataset into another one with zero mean and variations normalized to the standard deviation of data. This process is carried out with:

$$x_i^s = \frac{x_i - \overline{x}}{\sigma} \tag{22}$$

where x_i^s is the standardized datum, x_i^s the mean of the whole data set and σ their standard deviation.

Both algorithms will be used with all the forecasting tools used in this work to find out which one provides the best performance. It is worth noting that a process opposite to that of scaling must be applied to the forecasted data. The corresponding expressions will be obtained by reversing Equations (21) and (22).

3.2. Model Simulation

In this work, three artificial intelligence forecasting models have been tested: two Neural Networks (MLP and LSTM) and a popular machine learning tool widely used to solve data science problems, XGBoost. Each model will be simulated with three different preprocessing scenarios: no preprocessing, trend-fluctuations, decomposition, and EMD.

The first preprocessing method consists of splitting the CO_2 emission allowance price series into two subseries (Figure 2): its trend and fluctuations around it. They both will be independently forecasted, and their predictions added to obtain the predicted price.

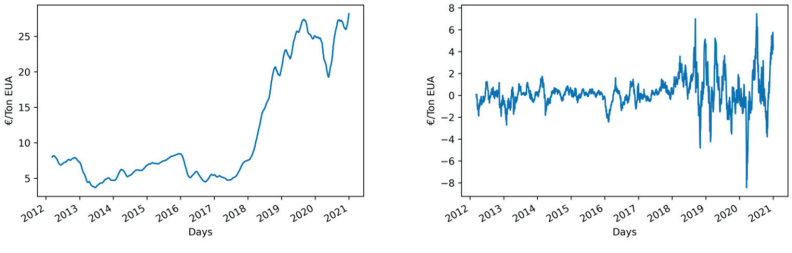


Figure 2. Trend-fluctuation decomposition of daily spot CO_2 prices series: (**a**) Trend series; (**b**) Fluctuations series.

In the second preprocessing model, the times series is split into eight stationary subseries (Figure 3), IMFs, which are independently forecasted and then added.

Before preprocessing the dataset, it has been scaled with both Min-Max normalization and standardization. The whole sequence of the different actions carried out to perform forecasting is described in the flowchart presented in Figure 4. As it may be seen, the process starts by scaling data and then splitting them into trend-fluctuations or IMF subseries, which are independently forecasted by each model. The predictions obtained are added to obtain the price predictions after rescaling the values provided by those summations. When data are not split, they are directly processed after scaled. This process is the same for both training and validation.

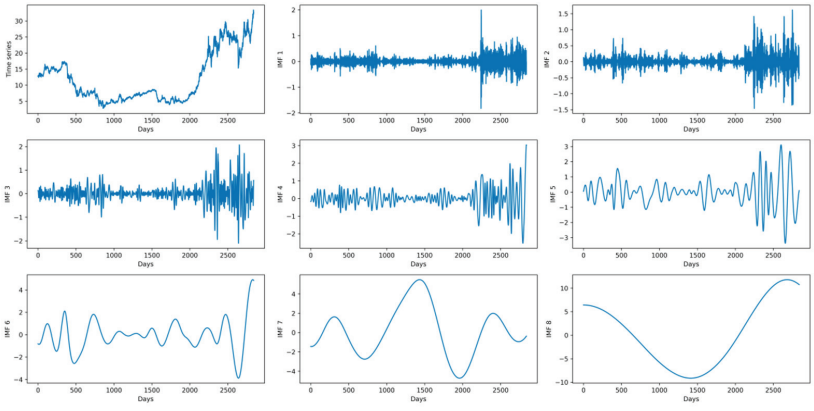


Figure 3. Original time series and the 8 IMFs of the EMD of the daily spot prices of CO_2.

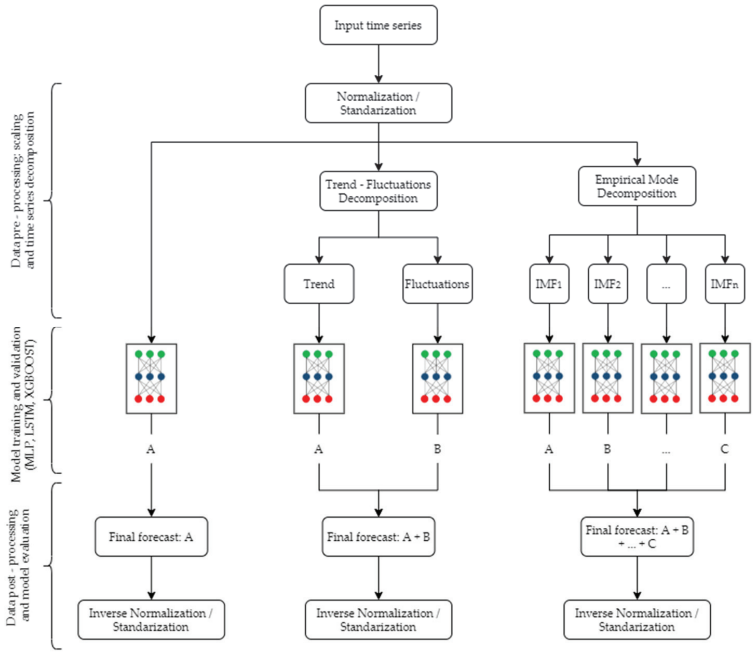


Figure 4. Flowchart of the different actions carried out in the forecasting process.

As pointed out above two datasets have been used: a simpler one, in which the data to be forecasted (validation subset) have a more or less stationary behavior, and the whole dataset, in which the data to be forecasted have values higher than those used for training with steep oscillations. The first defines a "simpler" problem without too extreme values and a "smooth" evolution. The second represents a harder problem with data with a different behavior from those used for training. The aim of defining two different scenarios is to check whether or not the forecasting models are able to provide good performances with both "easier" and "more difficult" problems of the same nature.

Regarding the two neural models, different numbers of layers and neurons in each one were tested. Nevertheless, structures with several hidden layers did not provide better performances than those with one only hidden layer. In fact, this last structure

outperformed those with several ones. This is a hardly surprising fact for a MLP because, as pointed out above, an MLP with one hidden layer with enough neurons behaves as a universal approximator. The case of LSTM is different, as it is usually used with a high number of layers and neurons, making up what is known as a "deep learning" neural network. However, this structure is applied to very complex problems, such as text or speech processing. Forecasting time series, no matter how nonlinear it is, is a much simpler problem to deal with. Thus, it looks reasonable to accept that a LSTM with one only layer will be enough to obtain accurate predictions. Therefore, only the results obtained with one hidden layer are presented in this work. The best performance was obtained with 100 neurons in the hidden layer for MLP and 100 memory blocks with one only cell in each one for LSTM. In this last network, the depth of the time delays in feedback was 3. Higher values were also tested, but the effect of gradient explosion appeared.

As these neural networks (and XGBoost) carry out a process of time series forecasting, past data of prices are used to forecast future ones. Those past values (several data preceding those to be forecasted) are the inputs to the neural networks (and XGBoost). The number of past data which provides the most accurate predictions should be determined by trial and error; therefore, several numbers of inputs between 1 and 100 were tested with the three models for the two datasets used. For MLP, the best performances were obtained with 60 inputs for the reduced dataset and 3 for the whole one. For LSTM, the best results were obtained with 3 inputs for the 2 datasets.

Several structures were also tested for the XGBoost algorithm. The best results were obtained with 1000 trees with a tree depth of 3 and 3 inputs for the two datasets.

The three forecasting models predicted 20 data at once, that is to say, they provided forecasted values of the daily spot price of CO_2 for the next 20 working days. So, the output layer of both MLP and LSTM has 20 neurons. This value was selected because it represents predictions for almost one month, four weeks, as the European Energy Exchange does not work at weekend days. The aim of providing 20 future values is to obtain both short-term and long-term predictions at once with one only forecasting model. The accuracy of the predictions (errors) will refer to that of 1 day ahead, 2 days ahead, and so on for the 20 predicted data.

All simulations have been programmed in Python with Tensor Flow and Keras packages. The XGBoost library for Python has been also used. The programs have been run in a personal computer with an Intel core i7-9700, 3.6 GHz with 32 Gbytes of RAM memory. Simulations have intensively used the GPU included in a RTX 2070 SUPER graphic card from Nvidia.

3.3. Prediction with MLP

The prediction errors (RMSE and MAPE) obtained with MLP are presented for both the simplified dataset and the whole one in Tables 3 and 4. Several divisions of data for training and validation were tested and the best results were obtained with 60–40% for the first dataset and 80–20% for the second. This is hardly surprising, because the first one takes into account data with values similar to those to be forecasted in the training subset, nevertheless for the whole dataset a division of 60–40% does not consider data with a steep rising trend for training, while in that of 80–20% a lot of data with that behavior are used.

The results in Table 3 show that the best performance was obtained when EMD was used to preprocess the reduced dataset. Nevertheless, although the best results were obtained with 60 inputs, a noticeable result was also obtained with a lower number of inputs (3): while the short-term horizon predictions are clearly improved those of the long-term ones got worse, providing a wider range of errors (as the higher standard deviation obtained shows). So, they both have been included in Table 3 as structures providing the best performance. It is difficult to decide which of them is the most accurate; in fact, it becomes a matter of preference; it depends on which prediction horizon the user is interested in.

Table 3. Predictions with MLP for the reduced data set (2009–2016). 60% of them were used for training and 40% for validation. Data have been scaled with standardization for "Without Prep." and "Trend-Fluc." and with Min-Max for EMD.

Days Ahead	Without Prep.		Trend-Fluc.		EMD (3 Inputs)		EMD (60 Inputs)	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	0.19	2.38	0.21	2.59	0.13	1.59	0.20	2.47
2	0.23	2.94	0.25	3.05	0.15	1.84	0.21	2.69
3	0.26	3.27	0.27	3.48	0.15	1.81	0.21	2.66
4	0.29	3.60	0.32	4.01	0.25	3.38	0.22	2.83
5	0.32	4.22	0.36	4.53	0.18	2.21	0.24	3.00
6	0.35	4.52	0.38	4.79	0.20	2.53	0.24	2.95
7	0.39	4.97	0.40	5.13	0.21	2.73	0.22	2.73
8	0.41	5.25	0.43	5.41	0.24	3.12	0.24	3.04
9	0.42	5.35	0.49	6.11	0.25	3.23	0.26	3.32
10	0.45	5.77	0.52	6.43	0.26	3.41	0.27	3.42
11	0.48	6.04	0.55	6.76	0.26	3.41	0.26	3.29
12	0.51	6.41	0.57	7.00	0.28	3.68	0.27	3.42
13	0.53	6.66	0.63	7.81	0.31	4.04	0.33	4.31
14	0.53	6.67	0.63	7.72	0.36	4.80	0.30	3.74
15	0.55	6.90	0.67	8.23	0.32	4.18	0.35	4.54
16	0.58	7.23	0.67	8.22	0.34	4.53	0.29	3.64
17	0.59	7.42	0.71	8.71	0.38	5.04	0.32	3.90
18	0.62	7.74	0.75	9.15	0.36	4.71	0.31	3.91
19	0.64	7.92	0.75	9.19	0.40	5.33	0.34	4.49
20	0.64	7.99	0.77	9.42	0.40	5.24	0.35	4.44
Mean	0.45	5.66	0.52	6.39	0.27	3.54	0.27	3.44
Std	0.14	1.69	0.18	2.11	0.08	1.15	0.05	0.64

Table 4. Predictions with MLP for the whole data set (2009–2020). 80% of them were used for training and 20% for validation. Data have been scaled with standardization for "Without Prep." and "Trend-Fluc." and with Min-Max for EMD.

Days Ahead	Without Prep.		Trend-Fluc.		EMD	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	0.72	2.19	0.72	2.24	0.92	2.91
2	0.97	2.99	0.96	2.99	0.97	3.14
3	1.18	3.66	1.16	3.66	1.01	3.26
4	1.35	4.18	1.31	4.13	1.08	3.48
5	1.51	4.60	1.49	4.67	1.16	3.68
6	1.64	5.05	1.63	5.11	1.19	3.75
7	1.74	5.42	1.75	5.47	1.22	3.91
8	1.86	5.77	1.82	5.76	1.28	4.05
9	1.94	6.19	1.91	6.01	1.35	4.30
10	2.04	6.47	2.00	6.23	1.45	4.69
11	2.13	6.72	2.08	6.43	1.42	4.52
12	2.20	6.98	2.15	6.71	1.48	4.78
13	2.28	7.25	2.23	6.91	1.52	4.85
14	2.38	7.52	2.31	7.15	1.55	4.98
15	2.45	7.73	2.39	7.41	1.60	5.03
16	2.53	7.89	2.46	7.64	1.64	5.22
17	2.64	8.29	2.53	7.86	1.69	5.43
18	2.70	8.49	2.62	8.15	1.73	5.51
19	2.78	8.64	2.65	8.35	1.76	5.68
20	2.86	9.00	2.74	8.68	1.77	5.65
Mean	2.00	6.25	1.95	6.08	1.39	4.44
Std	0.60	1.92	0.57	1.78	0.27	0.85

The results obtained when the whole dataset was used (Table 4) show that, again, the best predictions were obtained when EMD was used to preprocess the input data. It is worth noting that these results were obtained, in all cases, with only 3 inputs.

When comparing the results obtained with the two datasets, it may be seen that those obtained with the whole dataset are worse than those obtained with the reduced one. Nevertheless, these worse results are not too high, and it may be stated that reliable predictions were obtained. This fact shows the robustness of MLP as a forecasting tool, as its performance has suffered only a slight worsening when dealing with more complex data. In addition, it only needed three inputs to provide those results with the whole dataset, a structure much simpler than that with 60 inputs for the reduced one.

3.4. Prediction with LSTM

Tables 5 and 6 show the results obtained with the two datasets used. The best results were obtained with a distribution training-validation 60–40% for the reduced data set and 80–20% for the whole one, a distribution equal to that obtained with MLP. Results in Table 5 shows that now the best performance was obtained when no preprocessing was applied to the reduced dataset. The accuracy obtained when data were preprocessed by splitting them into trend and fluctuation was slightly better for the first two forecasted data than those without preprocessing, although they are clearly worse for the remaining ones. Data preprocessed with EMD provide clearly worse predictions than the option without preprocessing for short-term forecasting, although similar results were obtained for the long-term ones. When compared with the trend-fluctuations decomposition, it provides significantly worse results for short-term predictions, while providing better results for long-term ones.

Table 5. Predictions with LSTM for the reduced data set (2009–2016); 60% of them were used for training and 40% for validation. Data have been scaled with standardization for "Without Prep." and "Trend-Fluc." and with Min-Max for EMD.

Days Ahead	Without Prep.		Trend-Fluc.		EMD	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	0.18	2.14	0.16	1.91	0.48	5.61
2	0.23	2.77	0.22	2.72	0.48	5.64
3	0.26	3.29	0.27	3.32	0.50	5.95
4	0.30	3.74	0.31	3.88	0.50	5.90
5	0.33	4.18	0.35	4.39	0.48	5.81
6	0.36	4.57	0.39	4.92	0.51	6.13
7	0.39	4.90	0.43	5.41	0.52	6.39
8	0.41	5.27	0.46	5.89	0.50	6.16
9	0.44	5.59	0.50	6.34	0.49	6.14
10	0.46	5.88	0.53	6.75	0.52	6.42
11	0.49	6.14	0.56	7.14	0.60	7.51
12	0.51	6.40	0.59	7.54	0.51	6.42
13	0.53	6.66	0.62	7.91	0.61	7.69
14	0.56	6.91	0.65	8.32	0.55	6.98
15	0.58	7.17	0.68	8.71	0.56	7.08
16	0.60	7.42	0.70	9.08	0.62	7.81
17	0.62	7.66	0.73	9.46	0.57	7.30
18	0.63	7.89	0.75	9.80	0.61	7.75
19	0.65	8.10	0.78	10.12	0.59	7.51
20	0.67	8.32	0.80	10.36	0.68	8.82
Mean	0.46	5.75	0.52	6.70	0.54	6.75
Std	0.14	1.80	0.19	2.51	0.06	0.87

Table 6. Predictions with LSTM for the whole data set (2009–2020). 80% of them were used for training and 20% for validation. Data have been scaled with standardization for "Without Prep." and "Trend-Fluc." and with Min-Max for EMD.

Days Ahead	Without Prep.		Trend-Fluc.		EMD	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	0.95	2.94	0.76	2.43	4.36	17.32
2	1.13	3.57	0.98	3.16	4.40	17.38
3	1.30	4.11	1.17	3.76	4.39	17.28
4	1.45	4.56	1.32	4.24	4.41	17.30
5	1.60	5.01	1.49	4.75	4.32	16.83
6	1.72	5.42	1.62	5.15	4.49	17.47
7	1.82	5.80	1.72	5.48	4.41	17.04
8	1.91	6.12	1.80	5.74	4.42	17.00
9	2.00	6.47	1.88	5.90	4.41	16.84
10	2.09	6.83	1.95	6.08	4.38	16.63
11	2.18	7.13	2.03	6.28	4.47	16.89
12	2.25	7.33	2.11	6.53	4.54	17.11
13	2.34	7.65	2.18	6.77	4.41	16.49
14	2.44	7.96	2.27	7.01	4.47	16.63
15	2.51	8.18	2.35	7.26	4.48	16.59
16	2.57	8.29	2.43	7.50	4.19	15.22
17	2.65	8.50	2.51	7.76	4.21	15.16
18	2.72	8.73	2.59	8.04	4.39	15.78
19	2.78	8.97	2.64	8.28	4.30	15.30
20	2.85	9.18	2.71	8.51	4.47	15.87
Mean	2.06	6.64	1.93	6.03	4.40	16.61
Std	0.55	1.83	0.55	1.68	0.09	0.73

Performances when the whole dataset was used are shown in Table 6. Now the best accuracy was obtained with the trend-fluctuations preprocessing, although it was only slightly better than that obtained without preprocessing. It provides an accuracy only slightly worse than that obtained without preprocessing when the reduced dataset was used (the best option for that case), and clearly better for the last forecasted data when the same preprocessing process was applied to the reduced dataset. Nevertheless, the results obtained when the data were preprocessed with EMD are surprisingly poor, and what is more, long-term predictions are slightly better than short-term ones. They are much worse than those obtained with the reduced dataset. So, it may be stated that LSTM when EMD preprocessing was used has not been able to deal with the steep changes that appear at the end of the whole time series, while the structures without preprocessing and with trend-fluctuations decomposition were able to provide predictions that are only slightly worse that those obtained with the reduced dataset. This fact shows the robustness of LSTM with those two preprocessing models, as their performance suffers only a slightly worsen when a more complex time series is forecasted.

3.5. Prediction with XGBoost

The results obtained with XGBoost are shown in Tables 7 and 8. The first one presents the results obtained with the reduced dataset with a 60–40% division for training and validation and the second those with the whole one and an 80–20% division. The best performance obtained with the reduced dataset (Table 7) may be assumed as that provided by the model without preprocessing. Nevertheless, this statement demands a detailed explanation. The mean error of the twenty predictions is 8.54% for this model although that obtained with the EMD decomposition is 8.27%. However, the errors provided by this last model are almost constant for all predictions (as their very low standard deviation shows), while those obtained with the model without preprocessing are lower for the short-term predictions and higher for the long-term ones (higher standard deviation). This represents a more logical behavior of the forecasting tool, which provides a balanced evolution, since

predictions get worse as the time horizon increases. But if the user considers-long-term predictions as more valuable than short-term ones, the best model should be that with EMD decomposition.

Table 7. Predictions with XGBoost for the reduced data set (2009–2016). 60% of them were used for training and 40% for validation. Data have been scaled with standardization for "Without Prep." and "Trend-Fluc." and with Min-Max for EMD.

Days Ahead	Without Prep.		Trend-Fluc.		EMD	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	0.18	2.26	2.06	2.58	0.59	7.91
2	0.25	3.21	2.77	3.56	0.61	8.09
3	0.30	3.81	3.32	4.32	0.62	8.19
4	0.36	4.60	3.77	4.92	0.63	8.27
5	0.41	5.21	4.30	5.65	0.64	8.32
6	0.46	5.98	4.79	6.30	0.65	8.30
7	0.52	6.75	5.25	6.95	0.65	8.31
8	0.58	7.49	5.76	7.65	0.66	8.36
9	0.62	7.99	6.16	8.23	0.65	8.34
10	0.68	8.82	6.68	8.94	0.66	8.26
11	0.74	9.52	7.29	9.75	0.66	8.33
12	0.78	10.15	7.76	10.37	0.66	8.22
13	0.82	10.66	8.19	10.99	0.66	8.23
14	0.84	10.96	8.54	11.53	0.66	8.12
15	0.88	11.41	8.98	12.14	0.66	8.19
16	0.91	11.82	9.38	12.65	0.66	8.27
17	0.94	12.21	9.82	13.27	0.66	8.26
18	0.96	12.50	10.07	13.61	0.66	8.30
19	0.97	12.57	10.32	13.92	0.67	8.43
20	0.99	12.78	10.57	14.28	0.68	8.62
Mean	0.66	8.54	6.79	9.08	0.65	8.27
Std	0.26	3.35	2.63	3.62	0.02	0.14

Table 8. Predictions with XGBoost for the whole data set (2009–2020). 80% of them were used for training and 20% for validation. Data have been scaled with standardization for "Without Prep." and "Trend-Fluc." and with Min-Max for EMD.

Days Ahead	Without Prep.		Trend-Fluc.		EMD	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
1	3.40	9.72	2.60	8.66	30.49	123.22
2	3.39	10.14	2.75	9.29	30.48	123.14
3	3.73	11.50	2.84	9.67	30.52	123.20
4	3.40	10.78	2.93	10.07	30.50	123.05
5	3.55	11.51	3.06	10.47	30.41	122.65
6	3.66	11.98	3.06	10.56	30.40	122.50
7	3.44	11.37	3.14	10.83	30.39	122.35
8	3.63	12.06	3.19	11.04	30.39	122.30
9	3.68	12.38	3.23	11.13	30.39	122.19
10	2.93	9.61	3.29	11.33	30.37	121.96
11	2.88	9.33	3.36	11.61	30.36	121.84
12	2.84	9.07	3.39	11.69	30.31	121.49
13	3.00	9.60	3.49	12.02	30.34	121.53
14	3.28	10.63	3.63	12.58	30.33	121.44
15	3.32	10.64	3.73	12.85	30.34	121.39
16	3.62	11.71	3.78	12.91	30.31	121.10
17	3.69	11.87	3.85	13.04	30.30	120.95
18	3.97	12.89	3.93	13.23	30.31	120.90
19	3.95	12.77	4.03	13.48	30.27	120.66
20	4.25	14.00	4.08	13.74	30.30	120.70
Mean	3.48	11.18	3.37	11.51	30.38	121.93
Std	0.37	1.31	0.43	1.43	0.07	0.84

The predictions obtained with the whole dataset (Table 8) are clearly worse that those obtained with the reduced one. The best performances were obtained with both the

trend-fluctuation decomposition and without it, and it is surprising that the performance obtained with the EMD decomposition is especially bad. The errors are almost constant but with a so high value that it must be discarded as a forecasting model. When comparing the results obtained by the two datasets (only for no preprocessing and the trend-fluctuation decomposition) it may be seen that predictions get worse for the short-term remaining similar for the long-term. This means that XGBoost have problems to deal with a more complex dataset, at least in the short-term.

4. Discussion

When comparing the performance of the models simulated, it is clear that both MLP and LSTM outperform XGBoost in all cases. Only short-term predictions errors when no preprocessing and trend-fluctuations were used with the reduced dataset were similar to those of MLP or LSTM. For the whole dataset, errors are so high, especially in the case of EMD, that it may be stated that XGBoost is not a useful tool for forecasting this time series. This is not surprising if one bears in mind that XGBoots was designed to carry out classification tasks, so that it is not well suited for time series prediction. As it is proved in this work, XGBoost is not able to provide better results than other machine learning tools usually used in time series forecasting, such as the neural models used here.

Both MLP and LSTM were able to provide good predictions with the simplified dataset. In fact, very similar results were obtained when forecasting without preprocessing and when the trend-fluctuations were used: mean errors of 5.66% for MLP and 5.75% for LSTM without preprocessing and 6.39 and 6.70% for the trend-fluctuation preprocessing were obtained. This data also show that the accuracy of both models gets worse for the trend-fluctuations decomposition. Nevertheless, the performance of MLP is clearly improved when EMD is applied to the dataset (mean errors of 3.44 and 3.54%) while that of LSTM remains unchanged (mean error of 6.75%). As pointed out before, two structures have been tested with the MLP-EMD model because, although they provided almost equal mean errors, the time evolution of the prediction accuracies clearly differ: the structure with 3 inputs provides lower errors for short-term predictions, while that with 60 ones is better for long-term. From these results, it may be stated that the performance of MLP is clearly improved when EMD is applied, so that this structure provides the best performance of all the models tested with this dataset.

When the whole dataset is used, both MLP and LSTM provide similar results when no preprocessing was used (mean error of 6.25% for MLP and 6.64% for LSTM) and with the trend-fluctuations preprocessing (6.08 and 6.03%, respectively). However, their behavior clearly differs when preprocessed with EMD: while the mean error provided by MLP clearly decreases, providing a mean error of 4.44%, that obtained with LSTM undergoes a strong increase to 16.61%.

When comparing the results provided by MLP and LSTM with the two datasets, it may be seen that they are very similar, with a slight increase when no preprocessing was used (they pass from 5.66 and 5.75% for MLP and LSTM, respectively, with the reduced dataset to 6.25 and 6.64% with the whole one) and a slight decrease when trend-fluctuation was applied (6.39 and 6.70% to 6.08 and 6.03%).

These facts prove the robustness of MLP and LSTM when forecasting time series, as they are able to provide accurate prediction with both simpler and more complex time series providing that the training dataset is properly selected (the whole dataset was split into an 80–20% decomposition for the training-validation division instead of the 60–40% used for the reduced one).

Both MLP and LSTM are able to provide equally accurate predictions when they carry out forecasting directly, without preprocessing, achieving very similar errors. This common behavior changes when preprocessing is included. They both provide similar mean results when trend-fluctuations are included, although MLP behaves better for long-term predictions while LSTM does for short-term ones; but, for the reduced dataset, when EMD is included, MLP is able to improve its performance by decreasing its mean

error to 3.44 and 3.54%, whereas LSTM is only able to provide a mean error almost equal to that obtained with trend-fluctuations (although now with worse errors for short-time predictions). This behavior is more striking for the whole data set, since while MLP clearly improves accuracy when EMD was applied (its mean error falls to 4.44%) LSTM undergoes a strong increase to 16.16% and, what is more striking, with all values very close to that mean.

Therefore, it may be stated that both MLP and LSTM are able to provide accurate and robust prediction when no preprocessing is included in the forecasting model, but when it is included MLP clearly improves its performance when EMD is used, whereas LSTM provides much worse results, in fact the worst ones achieved with the three options tested.

On the other hand, it may be very interesting to analyze the evolution of predictions regarding their time horizon in order to identify how they behave, and their accuracies evolve. To do that, the one-day-ahead and the twenty-days-ahead predictions obtained in the validation set of the whole dataset provided by MLP-EMD are presented in Figures 5 and 6. It may be seen in Figure 5 that one-day-ahead predictions are able to follow fluctuations of CO_2 price, providing a reliable estimation of its daily evolution. Nevertheless, predictions with a time horizon of 20 days, as seen in Figure 6, are not able to follow daily fluctuations but, instead, they represent a sort of mean value of fluctuating prices, in other words, they provide a sort of trend of the time evolution of CO_2 prices. So, it could be stated that the model provides a prediction of the trend of the price evolution for the long-term instead of an estimation of actual prices on that time horizon. This behavior may be very useful for traders interested in the long-term evolution of prices because those predictions describe a sort of trend of how they will evolve.

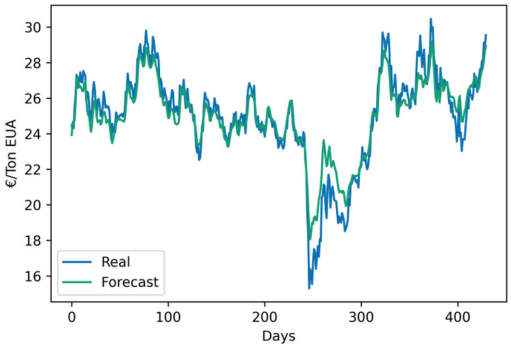


Figure 5. One-day-ahead prediction for the whole data set.

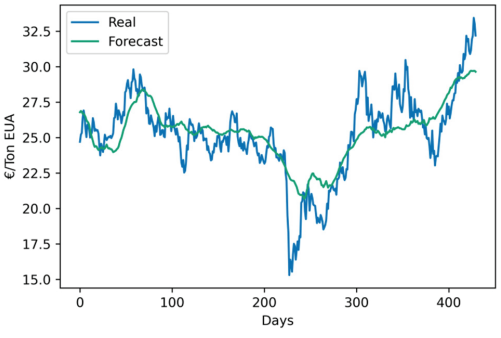


Figure 6. Twenty days ahead prediction for the whole data set.

This twofold behavior of the predicted data may be very useful, because it provides valuable information to agents involved in purchasing and selling EUAs. They can serve

as a reference for managers of companies included in the EU ETS, as they can manage their allowances portfolio with reliable information to help them make decisions about their costs in the short-term. In addition, long-term forecasts may be very useful for managers of polluting companies, as they provide fairly tight predictions of the price evolution trend. The proposed model gives reliable trend information, with a time horizon that is more suitable for making decisions on decarbonization in production processes.

5. Conclusions

Forecasting daily spot prices of CO_2 has become a key issue in recent years due to its upward trend that is affecting the final price of electricity. Several tools may be used to carry out this task, although Neural Networks seem to be the most reliable option, since they have proved to be one of the most accurate tools for time series forecasting. Nevertheless, not all models are able to provide accurate and reliable predictions and the best option for each particular time series should be identified by testing several ones. In this work, two popular neural models (MLP and LSTM) have been used to forecast the time series of daily spot prices of CO_2. Another popular artificial intelligence tool (XGBoost) has been also used for the sake of comparison. It provided poor performance compared with those obtained with the neural models. Several works have proved that the prediction accuracy of the forecasting models may be significantly improved when a suitable data preprocessing is applied prior to carry out the forecasting process. Thus, in this work, two techniques have been tested, trend-fluctuation decomposition and EMD, to provide data preprocessing. The best results were obtained with a hybrid MLP-EMD model, which provided a significant decrease in the forecasting errors. However, several of the combinations tested were not able to overcome the corresponding single forecasting tool, providing, in some cases, significantly higher errors. The robustness of the proposed models has been tested by using two datasets: a simplified one with a soft evolution and an enlarged one that included updated data with a rising trend with steep variations. The combination of MLP-EMD was able to provide accurate and reliable predictions with them both. Only a small increase of forecasting errors with the enlarged dataset was obtained, a fact that proves the model to be a robust forecasting tool. It is worth noting that MLP clearly outperforms LSTM as a forecasting tool despite the fact that this last seems, at first glance, to be better fitted for time series forecasting because of its recurrent behavior and inner "memory". In fact, MLP has proved to be a more accurate and robust forecasting tool. In addition, it has been able to take advantage of preprocessing techniques to improve accuracy while LSTM was not.

Therefore, an accurate and robust forecasting tool has been proposed to predict the time evolution of the daily spot price of CO_2. Predictions for 20 working days (four weeks) are provided at once with good accuracy, as means error of 2.91 % for the nearest prediction (1 day ahead) and 5.65% for the furthest one (20 days ahead) were provided. Thus, it may be a very useful tool for enterprises selling and purchasing emission allowances as well as for electric energy trading companies, as the forecasting model presented in this work is able to provide reliable predictions of the time evolution of daily spot prices of CO_2, what may help them to make decisions concerning their selling or purchasing activities.

On other hand, obtaining reliable allowance price predictions is important to market participants (such as affected companies, traders or brokers) who need accuracy price predictions to better manage their portfolios. Moreover, it is also crucial for the design of environmental policies, since CO_2 emission allowance prices provide information on the marginal abatement costs in the industry. Thus, based on the evolution of prices, the effectiveness of the environmental policies can be evaluated and the emission cap adjusted. Therefore, a more accurate carbon price forecasting is essential to establish a stronger and more efficient emission market.

Based on the results presented in this work, we aim at carrying out further research to test the efficiency of preprocessing strategies different from those used here. Refinements of EMD such as Ensemble Empirical Mode Decomposition (EEMD) or Variational Mode

Decomposition (VMD) could be applied and their performances compared with those obtained in this work. In addition, a study of different strategies to split data into training and validation sets based on cross-validation should also be carried out to try to define a model independent of the training-validation division dependence of accuracy pointed out in this work.

Author Contributions: Conceptualization, M.A.J.-M., D.F.-M., A.G.-G. and D.C.-F.; data curation, M.A.J.-M., D.F.-M. and A.G.-G.; formal analysis, M.A.J.-M., D.F.-M., A.G.-G. and D.C.-F.; funding acquisition, M.A.J.-M.; investigation, M.A.J.-M., D.F.-M., A.G.-G. and D.C.-F.; methodology, M.A.J.-M. and D.F.-M.; project administration, M.A.J.-M.; resources, D.F.-M. and A.G.-G.; software, M.A.J.-M. and D.F.-M.; supervision, M.A.J.-M., D.F.-M., A.G.-G. and D.C.-F.; validation, M.A.J.-M., D.F.-M., A.G.-G. and D.C.-F.; visualization, M.A.J.-M., D.F.-M., A.G.-G. and D.C.-F. writing—original draft, M.A.J.-M., D.F.-M. and A.G.-G.; writing—review and editing, M.A.J.-M., D.F.-M. and A.G.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Junta de Extremadura through the Grant GR18075 of its Research Groups Support Program (co-financed by FEDER funds).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Reinaud, J. *CO_2 Allowances and Electricity Price Interaction—Impact on Industry's Electricity Purchasing Strategies in Europe*; International Energy Agency OECD/IEA: Paris, France, 2007.
2. European Commission. *Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 Establishing a Scheme for Greenhouse Gas Emission Allowance Trading within the Community and Amending Council Directive 96/61/EC*; European Commission: Brussels, Belgium, 2003.
3. Ellerman, A.D. *Lessons for the United States from the European Union's CO_2 Emissions Trading Scheme. Cap-and-Trade: Contributions to the Design of a U.S. Greenhouse Gas Program*; MIT Center for Energy and Environmental Policy Research: Cambridge, MA, USA, 2008.
4. Benz, E.; Trück, S. Modeling the price dynamics of CO_2 emission allowances. *Energy Econ.* **2009**, *31*, 4–15. [CrossRef]
5. Fuss, S.; Johansson, D.; Szolgayová, J.; Obersteiner, M. Impact of Climate Policy Uncertainty on the Adoption of Electricity Generating Technologies. *Energy Policy* **2009**, *37*, 733–743. [CrossRef]
6. Fuss, S.; Szolgayová, J. Fuel price and technological uncertainty in a real options model for electricity planning. *Appl. Energy* **2010**, *87*, 2938–2944. [CrossRef]
7. Shahnazari, M.; McHugh, A.; Maybee, B.; Whale, J. Evaluation of power investment decisions under uncertain carbon policy: A case study for converting coal fired steam turbine to combined cycle gas turbine plants in Australia. *Appl. Energy* **2018**, *118*, 271–279. [CrossRef]
8. Barakat, M.R.; Elgazzar, S.H.; Hanafy, K.M. Impact of macroeconomic variables on stock markets: Evidence from emerging markets. *Int. J. Econ. Financ.* **2016**, *8*, 195–207. [CrossRef]
9. Pacce, M.; Sánchez-García, I.; Suárez-Varela, M. Recent Developments in Spanish Retail Electricity Prices: The Role Played by the Cost of CO_2 Emission Allowances and Higher Gas Prices. Banco de España Occasionals Paper No. 2020. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3903158 (accessed on 11 August 2021).
10. Tagliapietra, S.; Zachmann, G. Is Europe's Gas and Electricity Price Surge a One-Off? Bruegel Blog. 13 September 2021. Available online: https://www.bruegel.org/2021/09/is-europes-gas-and-electricity-price-surge-a-one-off/ (accessed on 15 September 2021).
11. Granger, C.; Teräsvirta, T. *Modelling Non-Linear Economic Relationships*; Oxford University Press: Oxford, UK, 1993.
12. Qi, M. Nonlinear Predictability of Stock Returns Using Financial and Economic Variables. *J. Bus. Econ. Stat.* **1999**, *17*, 419–429.
13. Lutz, B.J.; Pigorsch, U.; Rotfu, W. Nonlinearity in cap-and-trade systems: The EUA price and its fundamentals. *Energy Econ.* **2013**, *40*, 222–232. [CrossRef]
14. Wang, Y.; Wang, J.; Zhao, G.; Dong, Y. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy* **2012**, *48*, 284–294. [CrossRef]
15. Dritsaki, M.; Dritsaki, C. Forecasting European Union CO_2 Emissions Using Autoregressive Integrated Moving Average-autoregressive Conditional Heteroscedasticity Models. *Int. J. Energy Econ. Policy* **2020**, *10*, 411–423. [CrossRef]
16. Christian, C.; Rittler, D.; Rotfuß, W. Modeling and explaining the dynamics of European Union Allowance prices at high-frequency. *Energy Econ.* **2012**, *34*, 316–326. [CrossRef]
17. Lago, J.; de Ridder, F.; de Schutter, B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Appl. Energy* **2018**, *221*, 386–405. [CrossRef]

18. Bak, G.; Bae, Y. Predicting the Amount of Electric Power Transaction Using Deep Learning Methods. *Energies* **2020**, *13*, 6649. [CrossRef]
19. Szoplik, J. Forecasting of natural gas consumption with artificial neural networks. *Energy* **2015**, *85*, 208–220. [CrossRef]
20. González-Romera, E.; Jaramillo-Morán, M.A.; Carmona-Fernández, D. Monthly electric energy demand forecasting based on trend extraction. *IEEE Trans. Power Syst.* **2006**, *21*, 1946–1953. [CrossRef]
21. Moghaddam, A.H.; Moghaddam, M.H.; Esfandyari, M. Stock market index prediction using artificial neural network. *J. Econ. Financ. Adm. Sci.* **2016**, *21*, 89–93. [CrossRef]
22. Göçken, M.; Özçalıcı, M.; Boru, A.; Dosdogru, A.T. Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction. *Expert Syst. Appl.* **2016**, *44*, 320–331. [CrossRef]
23. Keles, D.; Scelle, J.; Paraschiv, F.; Fichtner, W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Appl. Energy* **2016**, *162*, 218–230. [CrossRef]
24. Fan, X.; Li, S.; Tian, L. Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. *Expert Syst. Appl.* **2015**, *42*, 3945–3952. [CrossRef]
25. Han, M.; Ding, L.; Zhao, X.; Kang, W. Forecasting carbon prices in the Shenzhen market, China: The role of mixed-frequency factors. *Energy* **2019**, *171*, 69–76. [CrossRef]
26. Ciechulski, T.; Osowski, S. High Precision LSTM Model for Short-Time Load Forecasting in Power Systems. *Energies* **2021**, *14*, 2983. [CrossRef]
27. Jin, Y.; Guo, H.; Wang, J.; Song, A. A Hybrid System Based on LSTM for Short-Term Power Load Forecasting. *Energies* **2020**, *13*, 6241. [CrossRef]
28. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [CrossRef]
29. Viviani, E.; di Persio, L.; Ehrhardt, M. Energy Markets Forecasting. From Inferential Statistics to Machine Learning: The German Case. *Energies* **2021**, *14*, 364. [CrossRef]
30. Lucas, A.; Pegios, K.; Kotsakis, E.; Clarke, D. Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression. *Energies* **2020**, *13*, 5420. [CrossRef]
31. Zhu, B.; Han, D.; Wang, P.; Wu, Z.; Zhang, T.; Wei, Y.-M. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Appl. Energy* **2017**, *191*, 521–530. [CrossRef]
32. Sun, G.; Chen, T.; Wei, Z.; Sun, Y.; Zang, H.; Chen, S. A Carbon Price Forecasting Model Based on Variational Mode Decomposition and Spiking Neural Networks. *Energies* **2016**, *9*, 54. [CrossRef]
33. Jaramillo-Morán, M.A.; García-García, A. Applying Artificial Neural Networks to Forecast European Union Allowance Prices: The Effect of Information from Pollutant-Related Sectors. *Energies* **2019**, *12*, 4439. [CrossRef]
34. Lamphiere, M.; Blackledge, J.; Kearney, D. Carbon Futures Trading and Short-Term Price Prediction: An Analysis Using the Fractal Market Hypothesis and Evolutionary Computing. *Mathematics* **2021**, *9*, 1005. [CrossRef]
35. Bishop, C.M. *Neural Networks for Pattern Recognition*; Oxford University Press: New York, NY, USA, 1995.
36. Hornik, K.; Stinchcombe, M.; While, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
37. Cybenco, G. Approximation by superpositions of a sigmoidal function. *Math. Control. Signals Syst.* **1989**, *2*, 303–314. [CrossRef]
38. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
39. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [CrossRef] [PubMed]
40. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
41. Liang, Y.; Niu, D.; Hong, W.-C. Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy* **2019**, *166*, 653–663. [CrossRef]
42. Zeiler, A.; Faltermeier, R.; Keck, I.R.; Tomé, A.M.; Puntonet, C.G.; Lang, E.W. Empirical Mode Decomposition—An introduction. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; IEEE: Barcelona, Spain, 2010; pp. 1–8. [CrossRef]

energies

MDPI

Article

Static Analysis and Optimization of Voltage and Reactive Power Regulation Systems in the HV/MV Substation with Electronic Transformer Tap-Changers

Jarosław Korpikiewicz * and Mostefa Mohamed-Seghir

Departamet of Ship Automation, Faculty of Electrical Engineering, Gdynia Maritime University, 81-225 Gdynia, Poland; m.mohamed-seghir@we.umg.edu.pl
* Correspondence: j.korpikiewicz@we.umg.edu.pl

Abstract: The quality of electricity is a very important indicator. The durability and reliable operation of all connected devices depend on the quality of the network voltage. Rapid changes in loads, changes in network connections and the presence of uncontrolled energy sources require the development of new voltage regulation systems. This requires voltage regulation systems capable of responding quickly to sudden voltage changes. In substations with control transformers, it is possible thanks to the use of semiconductor tap changers. Moreover, voltage regulation and reactive power compensation systems should be built as one system. This is due to the close dependence of voltage and reactive power in the network node. Therefore, it was proposed to use artificial intelligence methods to build a new voltage regulation and reactive power compensation system using all measurement voltages of network nodes. In the first stage of the research, active and reactive powers, as well as the voltage of the reference node, were selected for 6420 periods of the mains voltage. The simulation results were compared for the classic voltage regulation system with semiconductor tap changers and the evolution algorithm based on voltage measurements from the entire MV network. A significant improvement in the quality of voltage regulation with the use of an evolutionary algorithm was demonstrated. Then, a second set of input data with increased values of reactive power was generated. The results of the evolutionary algorithm after the application of the classic, independent reactive power compensation system and two-criteria optimization were compared. It has been shown that only the two-criteria optimization algorithm keeps both $|tg\varphi|$ within the acceptable range and the quality of voltage regulation is the best. The article compares different working algorithms for semiconductor tap changers.

Keywords: power system; voltage control; control tap-changer; evolution algorithm; multi-criteria optimization

Citation: Korpikiewicz, J.; Mohamed-Seghir, M. Static Analysis and Optimization of Voltage and Reactive Power Regulation Systems in the HV/MV Substation with Electronic Transformer Tap-Changers. *Energies* **2022**, *15*, 4773. https://doi.org/10.3390/en15134773

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 23 May 2022
Accepted: 24 June 2022
Published: 29 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The currently operated voltage regulation systems in HV/MV stations use only the transformer voltage on the MV side. The analysis of voltage regulation systems using electromechanical tap-changers of the transformer is presented in [1,2]. The design of a traditional tap-changer is shown in Figure 1a. The view of the power transformer with the on-load tap-changer is shown in Figure 1b.

Measurements of electrical quantities in MV networks (smart grids) are more often available. There are works on the use of semiconductor tap changers for voltage regulation in HV, MV and LV networks [3–12]. The differences between the electromechanical and semiconductor control algorithms are presented, among others, in [13,14]. There are applications of power semiconductors in high-voltage and high-power circuits, e.g., [15]. There is a need to develop optimally integrated voltage regulation and reactive power compensation [16–20]. It is indispensable to use artificial intelligence methods to design voltage regulation and reactive power compensation systems [21,22].

(**a**) (**b**)

Figure 1. Electromechanical tap-changer for power transformer from ABB [www.abb.com, accesed on 1 April 2020] (**a**), view of power transformer from ABB [www.abb.com, accesed on 1 April 2020] (**b**).

Due to the increase in the number of operating non-linear loads in the power grid, problems with the quality of electricity appear. For this reason, a new power electronic device was proposed to be installed in the substation [23,24]. The paper presents a novel strategy of predictive control for shunt active power filter (APF). The proposed control includes feedback from the supply current and combines the advantages of control in an open and closed loop—the transient response speed after changing the load current and a very high compensation efficiency. The high quality of the compensation current also results from the use of predictive algorithms in the control, as well as from the fact of connecting the converter to the network via the LCL circuit. The article presents the results of simulation tests of the proposed control algorithm. In [24] is presented an active filter connected in parallel to the power supply and electric energy receivers. It is made up of two sections of the coil sections L1 and L2, with a capacitor section connected in parallel between them. The other end of the filter is connected to six power transistors and a capacitor. The article demonstrates the effective filtration of harmonics up to the 50th. This type of device can be used in power stations to which industrial plants generating disturbances are connected, e.g., steel mills. These are examples of the use of automation and power electronics in power networks. Another example of the use of power electronics in the power industry is presented in [25].

1.1. Solid-State on-Load Tap Changer Technology

On-load tap changers have been used for a long time in HV substations. Currently, electromechanical tap changers are used. They have considerable disadvantages, including the formation of an electric arc on the contacts, limited switching frequency and limited total switching frequency, e.g., up to several times a day. Currently, there are more and more receivers and generators in the power grid with high dynamics of power changes. As a result, there is a need to build a voltage regulation system with high switching dynamics. The use of semiconductor tap changers enables quick voltage regulation. AC connectors should be used here. IGBT power transistors are currently the most popular in power electronics.

Currently, SiC-based power semiconductors made in the form of IGBT transistors have the highest switching frequencies. At the same time, the permissible operating temperature of SiC semiconductor elements is higher than the others. Thus, such elements can be used in the power industry. The regulating winding in HV/MV transformers is on the higher voltage side. The windings of these transformers on the 110 kV side, i.e., HV, are star-connected. One end of the regulating winding is connected to the neutral conductor, the other end to the working winding. The phase voltage of the entire winding is 63.5 kV.

Assuming that the control system regulation range should be from −20% to +20%, the required reverse voltage for semiconductor modules is 12.7 kV.

For publicly available high-voltage single semiconductor elements, it is at most 6.5 kV, e.g., 5SNA0400J650100 from ABB (collector current Ic = 400 A, turn-on delay time maximum is 700 ns, turn-off delay time maximum is 1700 ns); however, this semiconductor module is very expensive. It is possible to build modules for higher voltages, ensuring that the elements in series change the switching state practically simultaneously. In addition, during the construction of the module, a reserve of voltage resistance is additionally provided.

In [4] is presented a prototype of a tap changer controlled by a microcontroller. The prototype has five tap changers realized by means of pairs of thyristors. This applies to the low voltage 230/115 VAC system. For the correct commutation, the detection of current through zero was used (in regulating winding). This was realized by Zero Current Detection Card. The system was tested with a slight change in load or a slight change in input voltage but is working properly. The article does not present the implemented algorithm in the microcontroller or the use of voltage measurement on the primary side of the transformer

Any variation in the output voltage of the distribution transformer will be sensed by the microcontroller and compared with the reference value as per the program. This will produce the appropriate command to trigger the appropriate pair of anti-parallel thyristors for change in the suitable tapping of the transformer. The system stability is improved because of the quick response. Because of static devices, the maintenance cost is reduced due to the elimination of frequent sparking. The output voltage can be regulated in the range of ±5 V of nominal voltage [4].

In article [11] is presented the construction of a transformer semiconductor tap-changer regulator. Figure 2 shows the structure of the proposed voltage regulator controlling the semiconductor tap changers. The first block introduces a deadband that prevents oscillations when a voltage error changed the value on the border of two adjacent taps. The electronic tap changer operates fast and real-time measurement of the RMS value of the regulation bus is expected. One of the best substitutions for the RMS value of the voltage is the instantaneous RMS value of the voltage. Compensating block is used instead of a delay block, which is a special type of compensator. In [26], it has been shown that the use of an integrator in the compensating block of an electronic tap-changer seems interesting from a quality point of view. The integrator produces typically the delay proportional to DB, and it also has memory. The integrator gain influences the stability and also the speed of the system. The tap changer control is not a continuous control. For this reason, a quantization block is required. A discrete value of the tap number will be assigned to the continuous value of the voltage error. Due to the fact that the tap switching should take place when a current close to zero flows through the winding. Then, there are no overvoltages and disturbances. The S&H block remembers the selected tap number and, after receiving the permission to switch from the zero-crossing current detection block, performs the tap change-over. The loop-up table for the selected tap number displays the states of semiconductor switches similar to Table 1.

1.2. Volt Var Control

In many stations, reactive power compensation is required. Independent voltage regulation and reactive power compensation may cause deterioration of the operation of at least one of them. For example, when the voltage in the substation is close to the upper acceptable limit, i.e., 1.1 Un, and the reactive power compensation system additionally switches on capacitors, it may lead to the exceeding of this limit. This is due to the dependence of the node voltage and reactive power. For this reason, it is necessary to build integrated systems of voltage regulation and reactive power compensation called Volt Var Control or Volt/Var Management System. This is especially important when distributed generation or energy storage occurs in the distribution network.

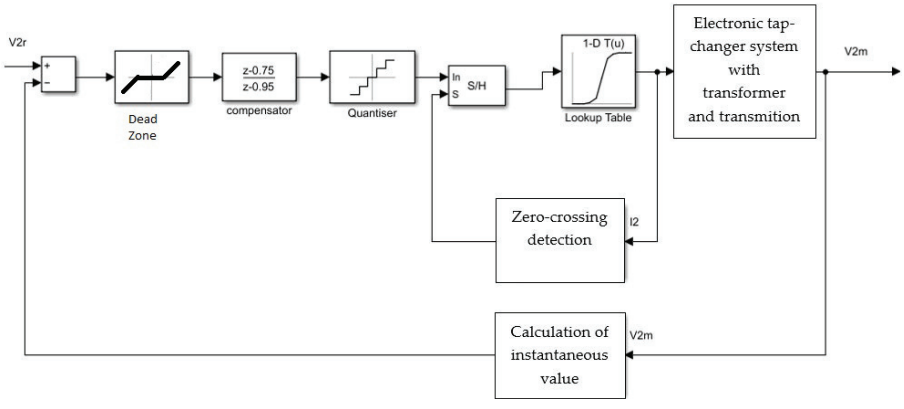


Figure 2. Schematic of the proposed electronic tap-changer [11].

Table 1. Switches states depending on the OLTC position.

OLTC Position	Percentage Change in Voltage on the MV Side	Ratio in p.u.	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14
1	−20.00%	0.80	1			1				1	1			1	1	
2	−19.00%	0.81	1			1				1	1	1			1	
3	−18.00%	0.82	1			1				1		1	1		1	
4	−17.00%	0.83	1			1			1		1			1	1	
5	−16.00%	0.84	1			1			1		1	1			1	
6	−15.00%	0.85	1			1			1			1	1		1	
7	−14.00%	0.86	1			1		1			1			1	1	
8	−13.00%	0.87	1			1		1			1	1			1	
9	−12.00%	0.88	1			1		1				1	1		1	
10	−11.00%	0.89	1			1	1				1			1	1	
11	−10.00%	0.90	1			1	1				1	1			1	
12	−9.00%	0.91	1		1					1	1	1			1	
13	−8.00%	0.92	1		1					1		1	1		1	
14	−7.00%	0.93	1		1				1		1			1	1	
15	−6.00%	0.94	1		1				1		1	1			1	
16	−5.00%	0.95	1		1				1			1	1		1	
17	−4.00%	0.96	1		1			1			1			1	1	
18	−3.00%	0.97	1		1			1			1	1			1	
19	−2.00%	0.98	1		1			1				1	1		1	
20	−1.00%	0.99	1		1		1				1			1	1	
21	0.00%	1.00	1	1												
22	1.00%	1.01		1	1		1				1			1		1
23	2.00%	1.02		1	1			1				1	1			1
24	3.00%	1.03		1	1			1			1	1				1
25	4.00%	1.04		1	1			1			1			1		1
26	5.00%	1.05		1	1				1		1	1	1			1
27	6.00%	1.06		1	1				1		1			1		1

Table 1. *Cont.*

OLTC Position	Percentage Change in Voltage on the MV Side	Ratio in p.u.	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14
28	7.00%	1.07		1	1				1		1			1		1
29	8.00%	1.08		1	1					1		1	1			1
30	9.00%	1.09		1	1					1	1	1				1
31	10.00%	1.10		1		1	1				1	1				1
32	11.00%	1.11		1		1	1				1			1		1
33	12.00%	1.12		1		1		1				1	1			1
34	13.00%	1.13		1		1		1			1	1				1
35	14.00%	1.14		1		1		1			1			1		1
36	15.00%	1.15		1		1			1			1	1			1
37	16.00%	1.16		1		1			1		1	1				1
38	17.00%	1.17		1		1			1		1			1		1
39	18.00%	1.18		1		1				1		1	1			1
40	19.00%	1.19		1		1				1	1	1				1
41	20.00%	1.20		1		1				1	1			1		1

2. Materials and Methods

2.1. Assumptions and Description

Input data, such as the voltage of the balancing (refencing) node and the active and reactive power of the load nodes, were randomized with certain assumptions. Then, the tested algorithm set the tap number and possibly turned on the appropriate number of batteries for reactive power compensation. In the calculation of the power flow, the voltages in all network nodes were determined. The input data sets were 6420 in size. On this basis, histograms were created, which allows to graphically present the range of changes and the frequency of occurrence of a given value of the voltage error, the coefficient $tg\varphi$.

The tested network consists of a referencing node number one in the depths of the network, with the network impedance calculated on the basis of the short-circuit power on the HV busbars of the 110/15 kV substation (equivalent to the rest of the power system—Thevenin's theorem). The structure of the network is presented in Figure 1 below. Nodes 4 to 15 are receiving nodes for which the value of active and reactive power is randomized, as in the actual network operation (the load powers of individual stations 15/0.4 change over time). The drawing of relative power values in individual load nodes was performed according to the following dependence (1) P,Q in p.u.:

$$\bigwedge_{i=4}^{15} P_i = random(0.3:1) \quad \text{where } random - \text{random number of uniform distribution} \tag{1}$$

The apparent power of a station in relative units is equal to 1: S_i = 1 p.u. Due to the apparent power of the station, the maximum reactive power is (2):

$$\bigwedge_{i=4}^{15} Q_{i\_\max} = \sqrt{1 - P_i^2} \tag{2}$$

Moreover, assuming that the maximum reactive power cannot exceed 60% of the value of the randomly selected active power in the node, we finally obtain the maximum reactive power (3):

$$\bigwedge_{i=4}^{15} Q_{\max\_i} = \min(Q_{i\_\max}; 0.6 * P_i) \tag{3}$$

The reactive power at the node i is (4):

$$\bigwedge_{i=4}^{15} Q_i = random(0.3 : Q_{\max\_i}) \tag{4}$$

Moreover, the voltage at node 1 is also variable in time and randomized (uniform distribution) in the range from 0.7 to 1.3 Un. The structure of the studied network is presented in Figure 3. Availability of online voltage measurements at all MV nodes and nodal powers was assumed. In order to perform the simulation, the Newton–Raphon method was used for 6420 samples (input data). The input data are the power consumed in the load nodes (P, Q) and the voltage of the referencing node no 1. The output data are voltages at 15 kV nodes and power flows.

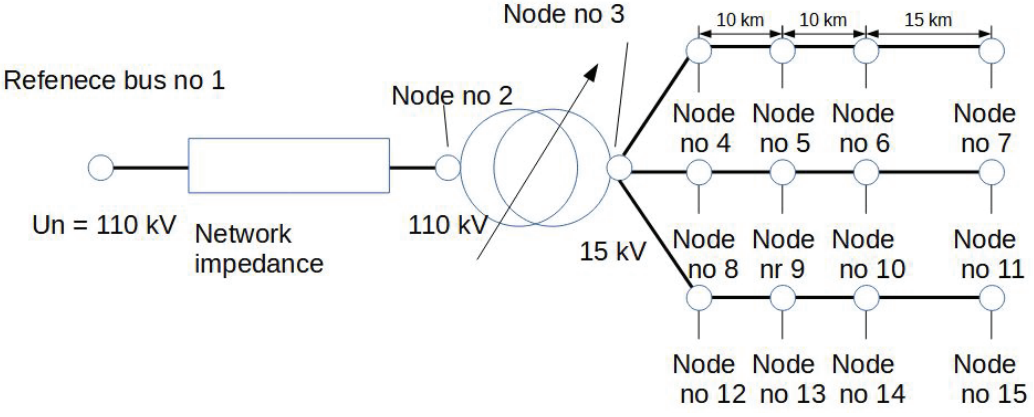


Figure 3. Structure of the tested power network.

The control winding was designed to be switched with semiconductor switches in such a way as to ensure smooth voltage regulation with a minimum number of taps and switches—Figure 4. Table 1 presents the switch states configuration depending on the required ratio.

In order to optimize the voltage regulation, the node voltage evaluation function was determined according to the Formula (5):

$$f(U_i) = \begin{cases} e_i = |1 - U_i| \\ 0, \ where \ e_i \le 0.05 \\ (e_i - 0.05 + 1)^6 - 1, where \ e_i > 0.05 \ and \ e_i \le 0.1 \\ (e_i - 0.1 + 1)^9 - 1 + 0.340095640625 \ where \ e_i > 0.1 \end{cases} \tag{5}$$

where U_i—The voltage module in the relative units of the node i, where i = 3.15.

The diagram of the node voltage evaluation function is shown in Figure 5. This function is continuous so that optimization is convergent. It is an internal function of the penalty for the voltage acceptable limit ±10% Un. The penalty function becomes non-zero after exceeding the absolute value of the error above 5% (see Figure 5). With an increase in the absolute value of the voltage error, the derivative of this function also increases. This provides a choice of optimization solutions with small voltage deviations in the nodes than solutions with large voltage deviation in at least one node.

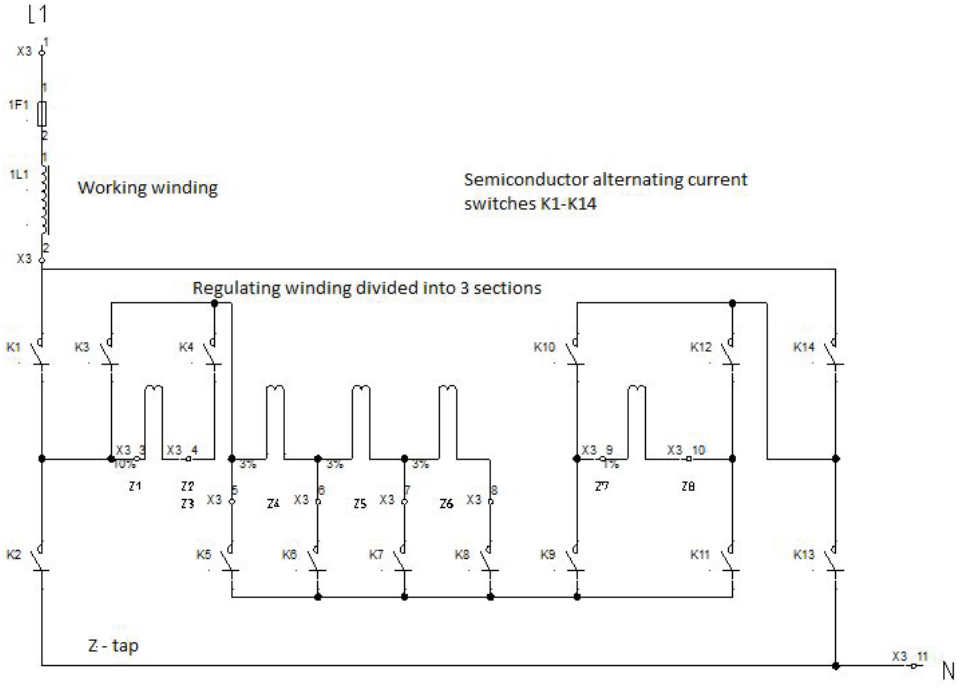


Figure 4. Structure of regulating winding for one phase.

For the entire network, the evaluate function is the sum of all MV node ratings:

$$J_m = \sum_{i=4}^{15} f(U_{i,m}\left(z; U_{ref,m}; P_{rec,m}; Q_{rec,m}\right)) \tag{6}$$

where: i—number node in 15 kV network, m—number sample of input data set, z_m—optimal number state regulation (see Table 1), $U_{ref,m}$—voltage of referencing node for m sample—input data, $P_{rec,m}$, $Q_{rec,m}$—vector active/reactive power for all receiving node in m sample—input data, $U_{i,m}$—the voltage at the node i for the input data set m—result of power flow analysis, $f(U_i)$—evaluation function for voltage node (5).

Optimization formula is (7):

$$\bigwedge_{m=1}^{6420} \min(\sum_{i=4}^{15} f\left(U_{i,m}\left(z_m; U_{ref,m}; P_{rec,m}; Q_{rec,m}\right)\right) \tag{7}$$

The evaluation of the operation of the control system was determined as (8):

$$J = \sum_{m=1}^{6420}\left(\sum_{i=4}^{15} f(U_{i,m}\left(z_m; U_{ref,m}; P_{rec,m}; Q_{rec,m}\right)\right) \tag{8}$$

This is the sum of the whole network scores for all input data samples.

2.2. Simulation Research Using Power Flow Calculations in Power Network

The simulations were carried out in several variants. In the first one without voltage regulation and reactive power compensation, the relative transformer ratio was equal to 1.

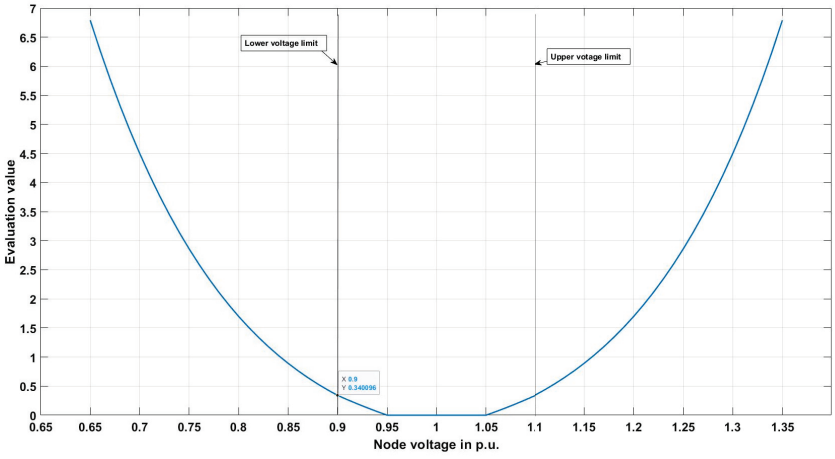


Figure 5. Evaluation function for voltage node.

The second variant is the classic regulation with semiconductor tap changers without reactive power compensation. The third variant is the use of evolutionary algorithms to determine the optimal tap using the voltage values in all load nodes. Results were presents in Section 3.1.

The fourth variant is the inclusion of the classic algorithm for connecting capacitor banks for the variant without voltage regulation. The fifth and sixth variant is also the launch of the classic algorithm for reactive power compensation for variants of the classic voltage regulation and optimization of the tap number using evolutionary algorithms.

The calculations were made in Matlab. MatPower was used to calculate voltages and power flow in the power grid.

The evolutionary algorithm is widely used to solve various optimization tasks or control systems in many fields of science, such as in [27–29] or other methods of artificial intelligence [30,31]. The eventual algorithm was started with the following Matlab commands ga.

It is an integer optimization with a limitation of the optimization variable value ranging from 1 to 41. The population size was set to 20 individuals. The maximum number of generations is 500.

2.3. Simulation Research Using Power Flow Calculations in Power Network with High Reactive Power Consumption

For high-reactive power data, the formula was applied regardless of the actual active power and the allowable apparent power. The remaining parameters of the simulation were left as in the previous one.

In that variant the reactive power at the node i is (9). Compared to Formula 4, the lower limit has been increased to 0.7 value of active power P_i and the upper limit to 0.8 p.u.:

$$\bigwedge_{i=4}^{15} Q_i = random(0.7 \cdot P_i : 0.8) \tag{9}$$

During the generation of new data, it was taken into account that not all the results obtained (reactive power in load nodes) will allow for the execution of power and voltage flow calculations. For this reason, after drawing the reactive power in load nodes, the input data was verified by means of power flow calculations. The remaining parameters, such as the referencing node voltages and active powers in nodes 4 to 15, remained the same as in the previous set of input data, i.e., 6420 periods.

Then, for the data thus obtained, an evolutionary algorithm was run in order to implement optimal voltage regulation. Moreover, the required value of reactive power for compensation and the number of capacitor banks with a capacity of 30 kVAr were determined. The next step was to update the power grid model, taking into account capacitor banks. For the obtained results, the calculations of voltages and power flow were made again for the determined degree of regulation control and connected batteries for reactive power compensation. Results were presented in Section 3.2.

2.4. Simulation Research Using Power Flow Calculations in Pareto Multi-Criteria Optimizing

Simultaneous and integrated voltage regulation and reactive power compensation are necessary to ensure correct operation of the substation. The problem of multi-criteria optimization arises. On the one hand, the system should ensure correct voltage values in the entire power network and at the same time compensate the reactive power to the required value—use of multi-criteria optimization—Pareto—simultaneous voltage and reactive power regulation. From the set of non-dominated solutions, a solution was selected that meets the voltage quality requirements with as much as possible reactive power compensation for each set of input data.

The first optimization criterion is minimizing the entire network, the evaluate function is the sum of all MV node ratings (6). The second criterion is the reactive power compensation assessment. According to the legal requirements, the reactive power should not exceed the value determined by the relationship (10):

$$\left| tg\phi = \frac{Q}{P} \right| \leq 0.4 \tag{10}$$

The reactive power compensation evaluation function should have a value of 0 when the reactive power does not exceed the value of 40% of active power. The evaluation function used is as follows (11):

$$\begin{gathered} tg\phi_{T,m} = \begin{cases} \frac{Q_{T,m}}{P_{T,m}}, where\ Q_{T,m} \geq 0\ and\ P_{T,m} > 0 \\ 0, other\ case \end{cases} \\ Q_{2compens} = \begin{cases} 0,\ where\ tg\phi_{T,m} \leq 0.4 \\ Q_{T,m} - 0.4 \cdot P_{T,m;} \end{cases} \\ J_{Q,m} = f_{compensation}(P_{T,m;} Q_{T,m;} Q_{bat}) = \begin{cases} 0, where\ Q_{2compens} \leq 0 \\ \frac{Q_{2compens}}{Q_{bat}}, other\ case \end{cases} \end{gathered} \tag{11}$$

where: $P_{T,m}$, $Q_{T,m}$—active and reactive power flowing through the transformer for m number sample of input data set, Q_{bat}—reactive power of capacitor bank, $Q_{2compens}$—required reactive power value to be compensated.

The graph of the Evaluation Function for the reactive power compensation is shown in Figure 6.

The problem of two-criteria optimization can be written as follows:

$$\begin{gathered} F(PPZ, numQ) = \min \begin{bmatrix} J_{Q,m} \\ J_m \end{bmatrix} \\ PPZ \epsilon 1 \ldots 41, integer\ number \\ numQ \epsilon 0 \ldots Q_{\max}, integer\ number \end{gathered} \tag{12}$$

where $J_{Q,m}$—reactive power compensation evaluation function for m number sample of input data set, J_m—voltage evaluate function for the entire network (Formula (6)) for m number sample of input data set, PPZ—tap-changer position (see Table 1), $numQ$—number of connected capacitor banks, $Q_{\max}$—maximum number of capacitor banks required in the entire simulation.

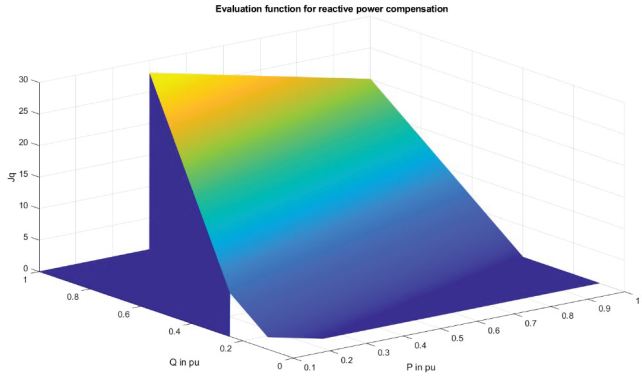


Figure 6. Evaluation function for reactive power compensation.

PPZ and *numQ* are integer optimization variables.

The multi-criteria optimization settings were implemented using the Matlab command gamutilobj, with population size of 200 and ParetoFraction factor of 0.7, with limitations.

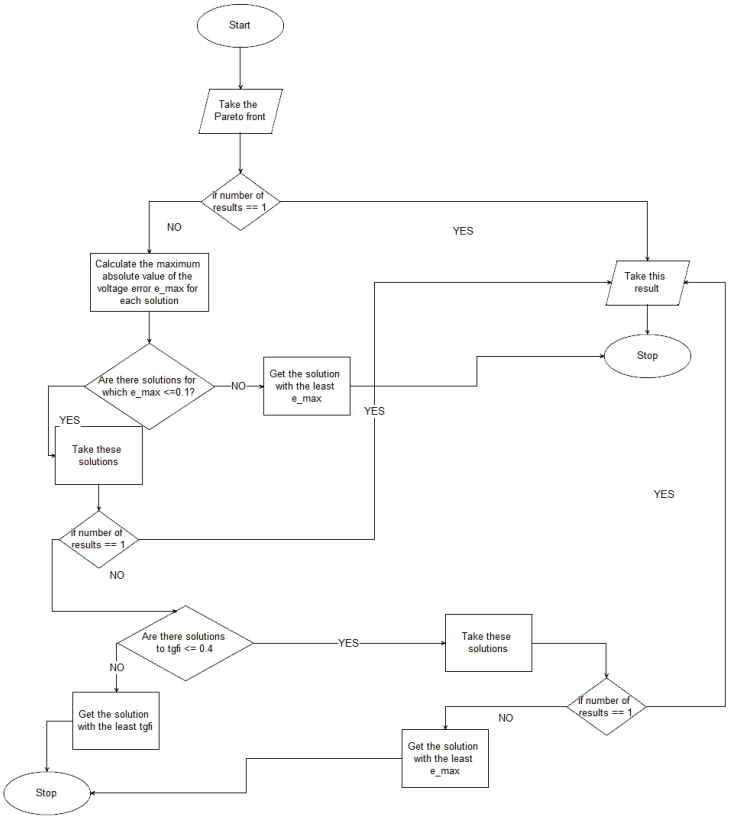


Figure 7. Algorithm for selecting one solution from the Pareto front.

Multi-criteria optimization consists in determining the Pareto front for each input data. After determining the Pareto front, one solution should be chosen. The solution selection algorithm is shown in Figure 7.

Results were presented in Section 3.3.

3. Results

3.1. Results of Simulation Research Using Power Flow

The histograms of the whole network evaluation function values for all input data are shown below. The range for the evaluation value in the histograms is five (width column—*X*-axis. On the *Y*-axis, we have a normalized number of results for a given interval of the evaluation function value. The lower the value of the evaluation function, the smaller the voltage error.

Figure 8 shows the results when the voltage regulation and reactive power compensation system are turned off. The ratio transformer is 110/15. 34% of the results fall within the first range of the evaluation function value. However, there are results with values above 200. Figure 9 shows the simulation results for the classic tap semiconductor control algorithm using only the voltage measurement on the HV/MV transformer. You can see a significant improvement in the quality of the voltage. Most of the simulation results fall within the first four columns of the histogram. Figure 10 shows the optimization result performed with the evolutionary algorithm. This algorithm used voltage measurements at all 15 kV nodes. You can see that almost all the results fall within the first range of the evaluation function value.

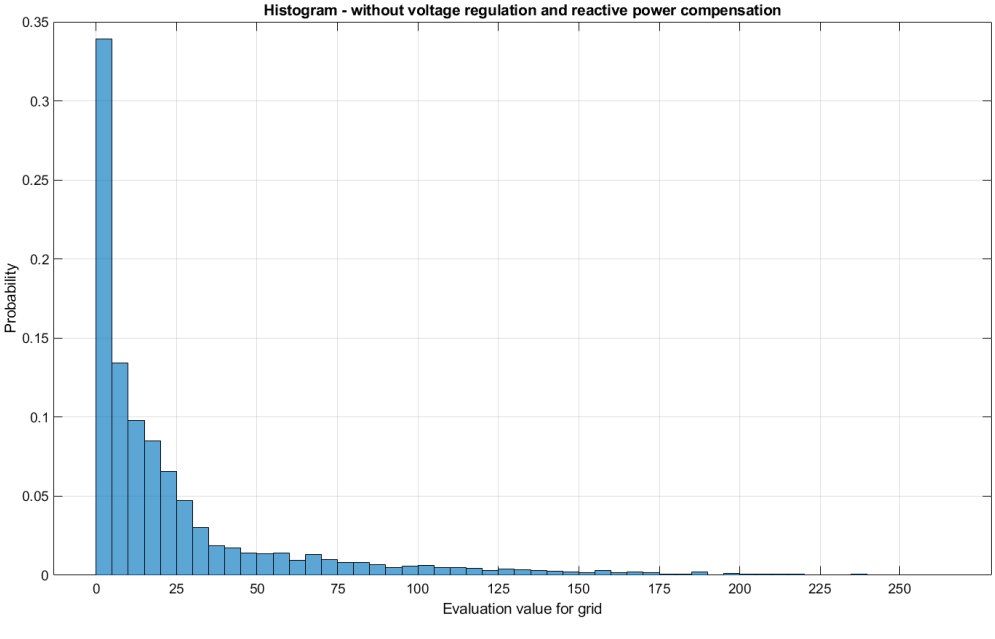


Figure 8. Histogram of evaluation value—without voltage regulation and reactive power compensation.

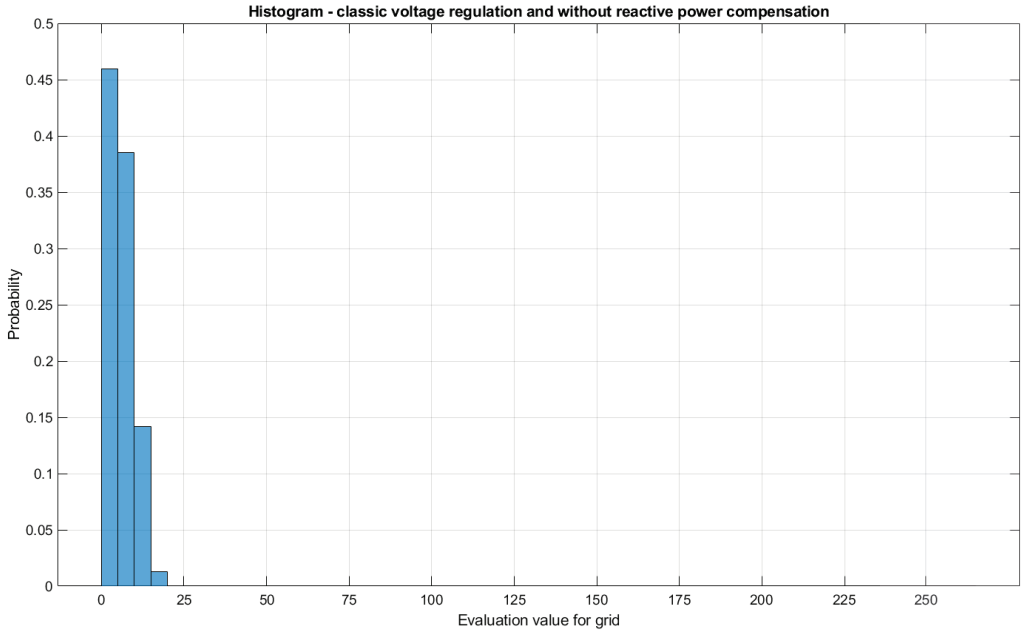


Figure 9. Histogram of evaluation value—classic voltage regulation and without reactive power compensation.

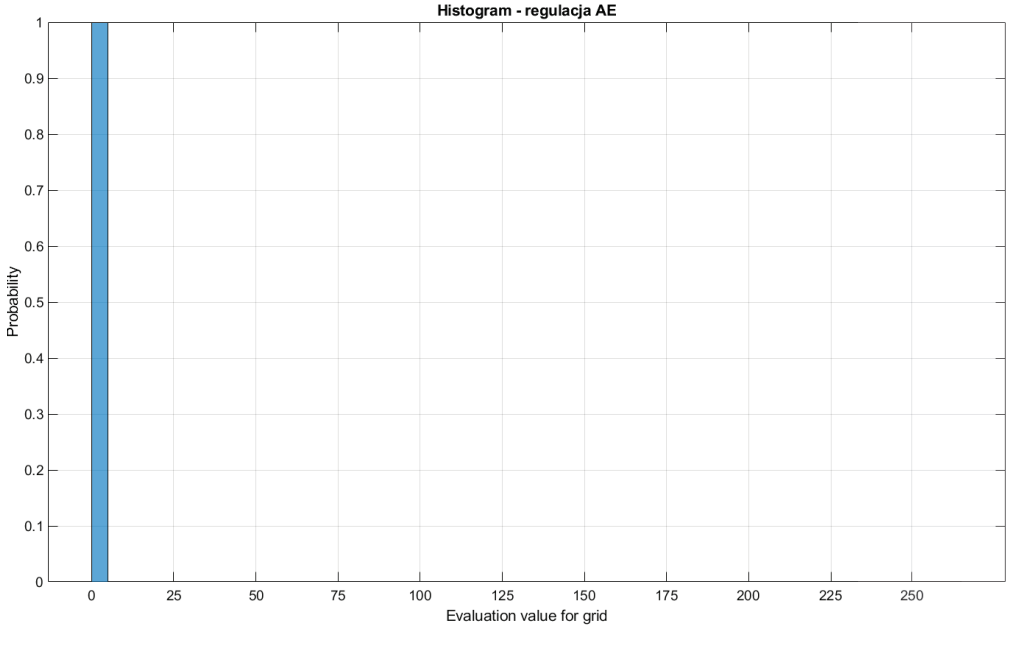


Figure 10. Histogram of evaluation value—voltage regulation optimization by means of an evolutionary algorithm and without reactive power compensation.

The following figures show the results for the independently operating voltage regulation system and independent reactive power compensation. For the case without voltage regulation, the reactive power compensation system improved the results. In other cases, the influence of reactive power compensation is not visible when analyzing all the results (Figures 11–15).

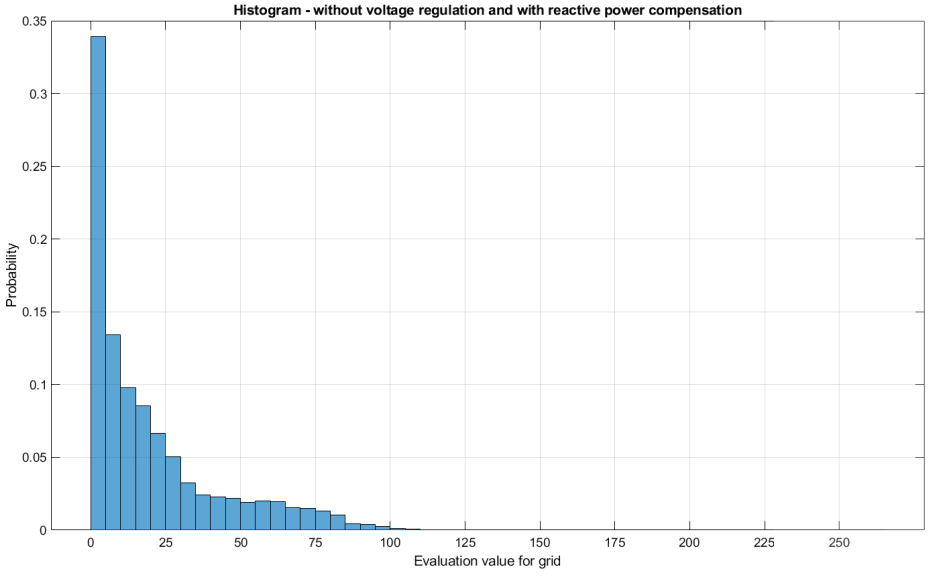


Figure 11. Histogram of evaluation value—without voltage regulation and with reactive power compensation.

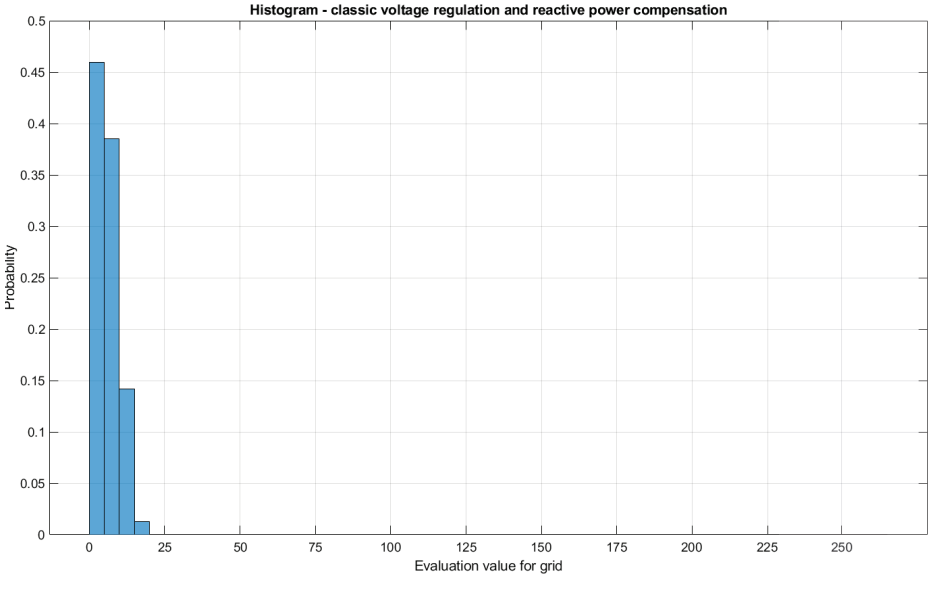


Figure 12. Histogram of evaluation value—classic voltage regulation and with reactive power compensation.

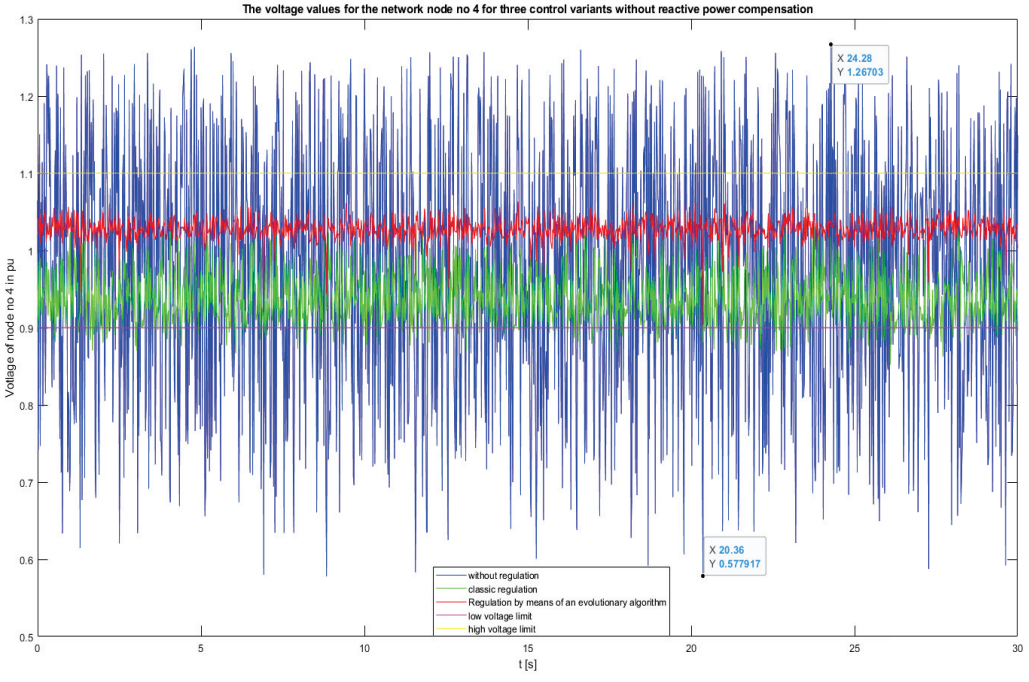


Figure 13. The voltage values for the network node no 4 for three control variants without reactive power compensation.

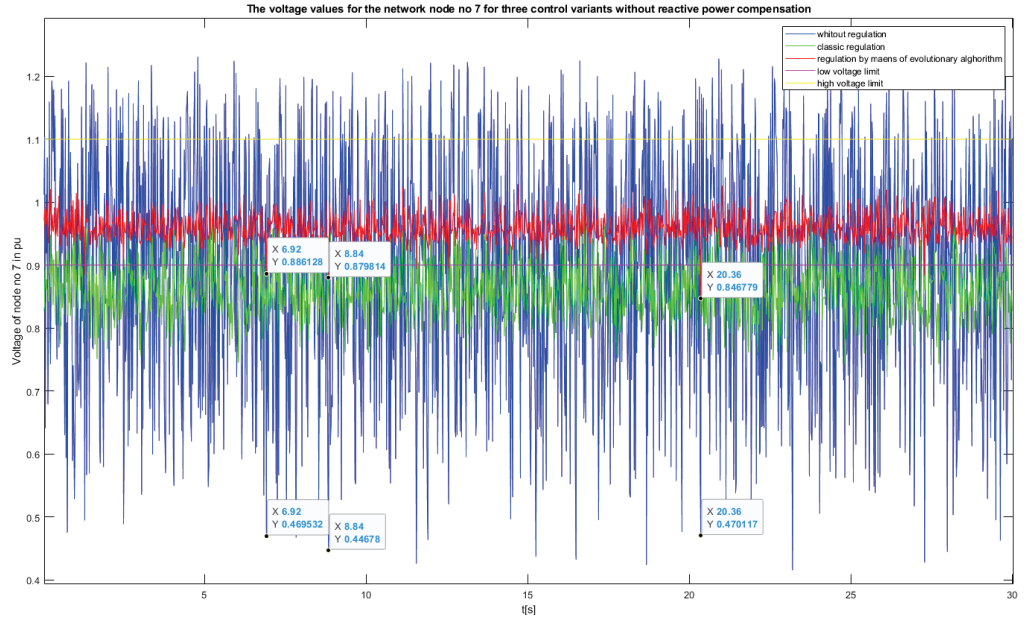


Figure 14. The voltage values for the network node no 7 for three control variants without reactive power compensation.

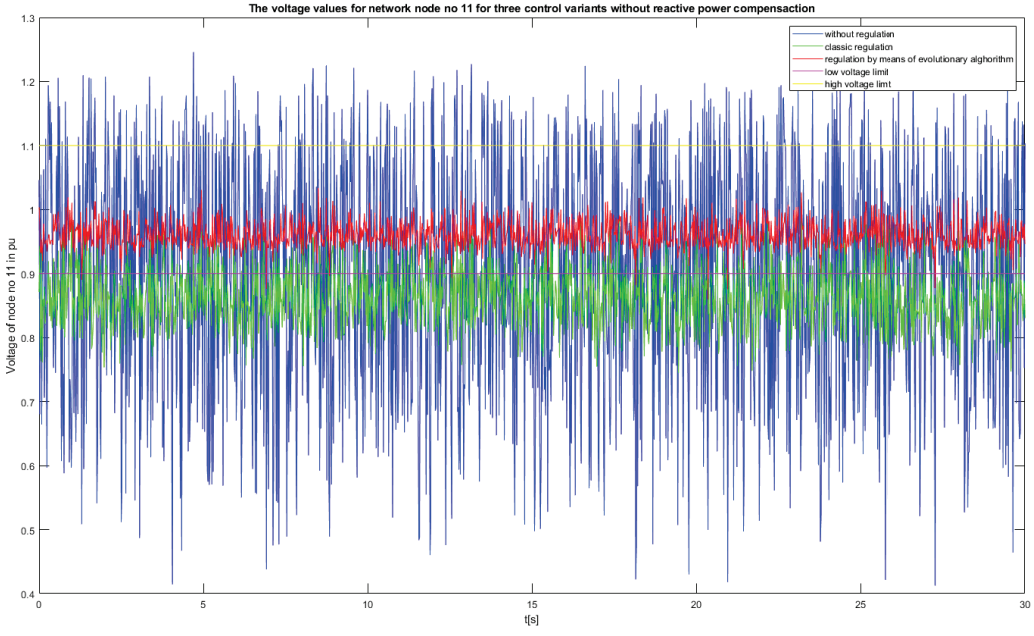


Figure 15. The voltage values for the network node no 11 for three control variants without reactive power compensation.

Histogram of evaluation value—voltage regulation optimization by means of an evolutionary algorithm and with reactive power compensation is identical to the histogram of evaluation value—voltage regulation optimization by means of an evolutionary algorithm and without reactive power compensation. This is due to the fact that no reactive power compensation was needed for the results obtained from the evolutionary algorithm.

The table below shows the maximum number of required capacitor banks for the three control variants without reactive power compensation (Table 2).

Table 2. The maximum number of required capacitor banks of 30 KVar.

Without Voltage Regulation	Classic Voltage Regulation	Voltage Regulation with an Evolutionary Algorithm
71	5	0

The voltage values for the selected network node for three control variants without reactive power compensation are presented below.

As you can see (Figure 13) in the variant without voltage regulation, it varies widely from 0.578 to 1.267, which is beyond the allowable range. With classic regulation, the voltage variability is smaller, but it exceeds the lower limit. In the case of regulation with the use of the evolutionary algorithm, the range of voltage changes is in the upper half of the allowable range and does not exceed it. It also results that in the most distant network nodes the voltage will decrease, which ensures voltage variability in them within the permissible range. Moreover, the voltage variation is the smallest.

Node 7 is at the end of one of the MV lines. As shown in Figure 14, the voltage is often below the lower voltage limit in classic regulation. In the case of regulation using the evolutionary algorithm, the lower voltage limit is rarely exceeded, after the regulation possibilities are exhausted. In order to verify this, a table with levels of regulation for selected time moments is presented.

As you can see (Table 3), when the lower voltage limit is exceeded, the tap changer was in the position to increase the voltage the most despite external conditions. With classic regulation, unfortunately, most of the time the voltage is below the lower limit.

Table 3. The state of OLTC (position tap changer) switch at selected time moments.

6.92 [s]	8.84 [s]	20.26 [s]
41	41	41

The table below shows the minimum, maximum, average and variance voltage values for the selected nodes (Table 4). The results of the statistical analysis for the three variants of voltage regulation confirm the conclusions of the presented voltage diagrams (Figures 11–17). When analyzing the minimum and maximum values for the three control variants, it is clear that in the case of no regulation, these values are outside the range of permissible values. In the case of classical regulation, there was an improvement. It is true that the minimum values exceed the lower limit of the permissible voltage range. Only the results obtained using the evolutionary algorithm with access to the current measurement values of the network nodes allowed for a significant improvement in the quality of voltage regulation. The minimum voltage is slightly below the permissible value, but it still doubles compared to the other variants. Variance is a measure of the volatility of a given. In the case of voltage regulation, despite the changes in the voltage supplying the substation and changes in the power consumed in stations 15/0.4, the system is designed to maintain the range of voltage changes within the permissible range. Moreover, it was shown that the voltage variability was about 100 times lower in all analyzed nodes in relation to the other control variants (evolution algorithm).

Table 4. Minimum, maximum, average and variance voltage values of selected nodes for three control variants without reactive power compensation.

Type of Voltage Regulation	Voltage in p.u.	Node No 4	Node No 7	Node No 11	Node No 13	Node 15
Without regulation	Minimum	0.564	0.393	0.390	0.469	0.370
	Maximum	1.270	1.245	1.246	1.261	1.240
	Average	0.976	0.899	0.900	0.942	0.900
	Variance	0.029	0.037	0.036	0.032	0.036
Classic regulation	Minimum	0.852	0.723	0.730	0.801	0.734
	Maximum	1.042	0.998	1.017	1.025	1.003
	Average	0.940	0.864	0.865	0.906	0.864
	Variance	0.001	0.002	0.002	0.002	0.002
Regulation by means of evolutionary algorithm	Minimum	0.903	0.847	0.847	0.870	0.824
	Maximum	1.070	1.029	1.035	1.056	1.033
	Average	1.028	0.960	0.961	0.997	0.961
	Variance	2.051×10^{-4}	4.84×10^{-4}	4.951×10^{-4}	2.956×10^{-4}	5.011×10^{-4}

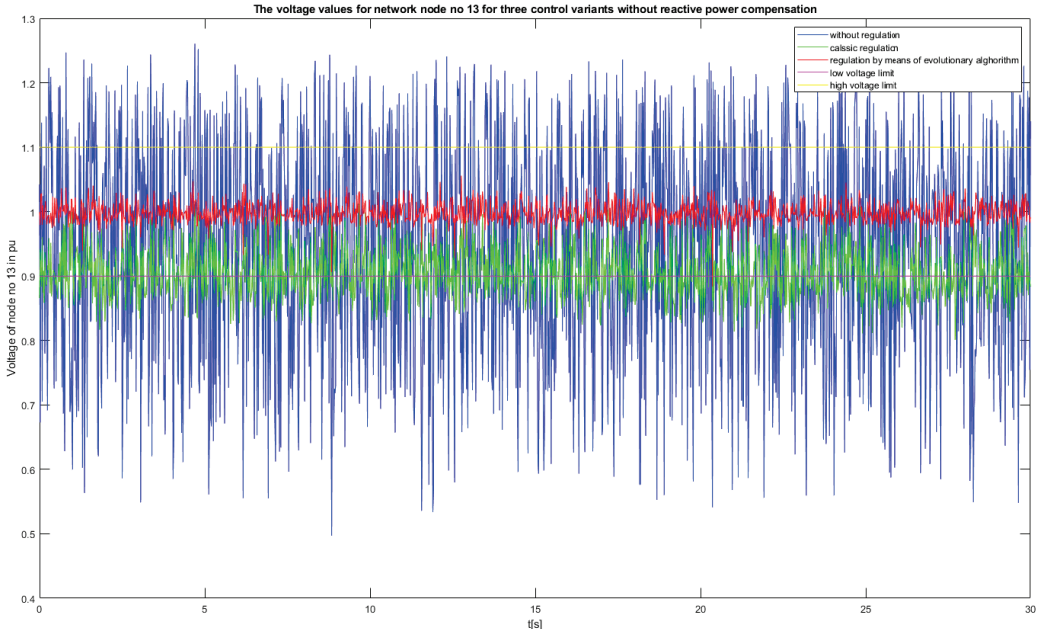


Figure 16. The voltage values for the network node no 13 for three control variants without reactive power compensation.

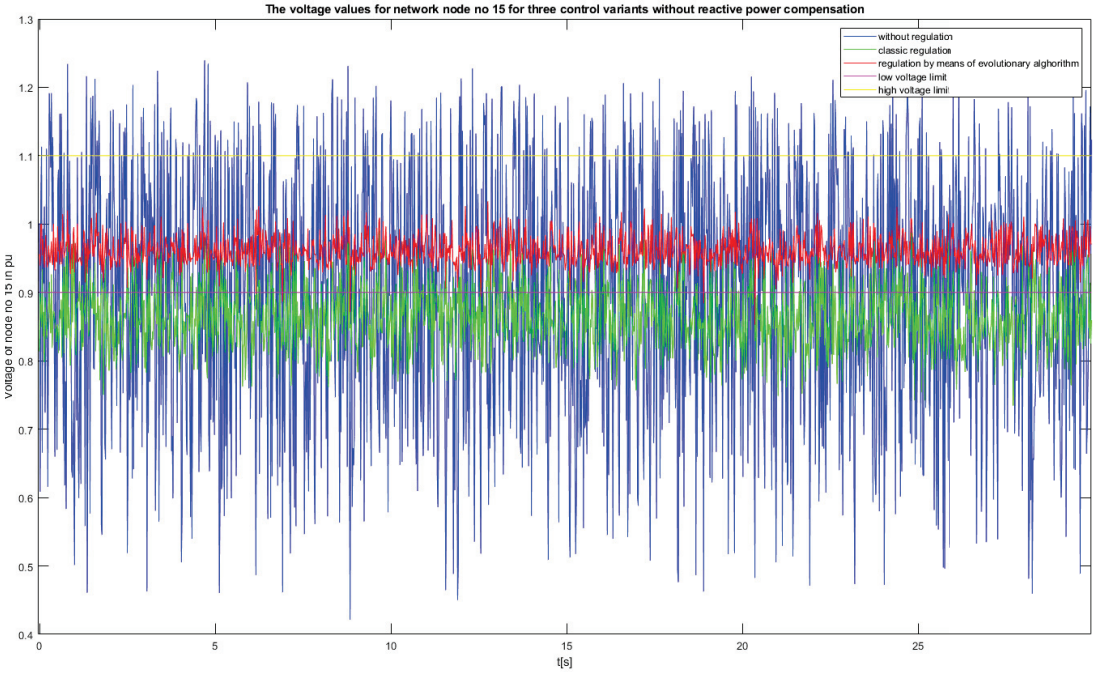


Figure 17. The voltage values for the network node no 15 for three control variants without reactive power compensation.

On this basis, it has been shown that the evolution algorithm using measurement data from all network nodes provides the best quality of voltage regulation. The presented results justify the need to use the measurements, e.g., voltages in stations 15/0.4 in order to significantly improve the quality of voltage regulation. Voltage regulation with the use of evolutionary algorithms maintains the voltage value in nodes most often in the range from 1 to 1.05 p.u. This prevents the voltage drops at the ends of the lines from dropping too much due to voltage drops.

There is one problem with building a voltage regulator. This regulator should work with a time resolution of at least one period of the mains voltage. Moreover, for the simulated data in the case of voltage regulation for the variant using the evolutionary algorithm, there was no need for reactive power compensation.

It follows that the evolutionary algorithm cannot be directly used to build the controller due to the fact that obtaining the results with its use required a long time.

In practice, reactive power compensation is often required in power stations. For this reason, additional simulation data was generated for which high reactive power compensation will be required. For this reason, another set of input data was prepared for the simulation. However, in this case, we have a problem of multi-criteria optimization. The reactive power at the node and the RMS voltage are strongly related.

3.2. Results of Simulation Research Using Power Flow Calculations in Power Network with High Reactive Power Consumption

The simulation tests were carried out in two variants. Application of an evolutionary algorithm to optimize voltage regulation. Then, the required number of connected capacitor banks was determined, and after such a change, the flow calculations were performed again. Table 5 shows the minimum, maximum and average number of capacitor banks required. Therefore, the reactive power compensation system should be designed for at least 170 capacitor banks. It was assumed that the reactive power compensation system would be able to switch on capacitor banks every 30 KVar with a maximum number of $Q_{max} = 200$.

Table 5. The minimum and maximum number of required capacitor banks of 30 KVar for voltage regulation with an evolutionary algorithm in high reactive consumption.

Minimum Capacitor Q_{min}	Maximum Capacitor Q_{max}	Average Capacitor Q_{avg}
0	170	54

Tables 6 and 7 present the results of the reactive power compensation influence on the voltage quality.

Table 6. Influence of independent reactive power compensation on the quality of voltage regulation.

The Number of the Second Dataset	Number of Times Reactive Power Compensation Was Required	Number of Cases Where the Voltage Quality Deteriorated Due to Reactive Power Compensation
6399	6368	4638

Implemented independently of the reactive power compensation voltage regulation, it decreased the evaluation function in 73% of cases. It follows that the reactive power compensation should be an element of the integrated voltage and reactive power regulation system (Table 6).

When the voltage is close to the upper allowable limit, connecting the capacitor banks additionally causes its increase, which results in deterioration of the quality of voltage regulation (Table 7).

Table 7. Influence of independent reactive power compensation on the quality of voltage regulation—one case.

Node No	Voltage Value before Compensation in p.u.	Voltage Value after Compensation in p.u.
3	1.129	1.191
4	1.037	1.107
5	0.990	1.064
6	0.962	1.038
7	0.943	1.021
8	1.074	1.140
9	1.047	1.116
10	1.031	1.101
11	1.015	1.087
12	1.043	1.112
13	1.000	1.073
14	0.974	1.050
15	0.955	1.032

3.3. Results of Simulation Research Using Power Flow Calculations in Pareto Multi-Criteria Optimizing

One of the solutions is presented below (Table 8). Out of 8200 possible solutions, the two-criteria optimization algorithm chose four (see Figure 18). Then the Pareto-front solution selection algorithm chose solution no 4.

Table 8. The result of two-criteria optimization with the indicator of the quality of regulation.

No	OLTC Position	Number Capacitor	J	J_Q	e_max—Maximum Absolute Value of Voltage Error for the Entire Network	$\lvert tg\,\varphi \rvert$
1	13	4	0.5	85.4	0.1113	0.74
2	14	26	0.5	61.7	0.113	0.64
3	15	48	0.5	38.3	0.1129	0.55
4	21	181	0.5	0	0.113	0.011

Then, the results of three simulations were compared for a dataset with high reactive power demand. The first one was carried out with the help of an evolutionary algorithm—single-criterion optimization. The second one, using the results from the first one, uses the classic algorithm for reactive power compensation. The last one was carried out with the use of two-criteria optimization (see Tables 9 and 10).

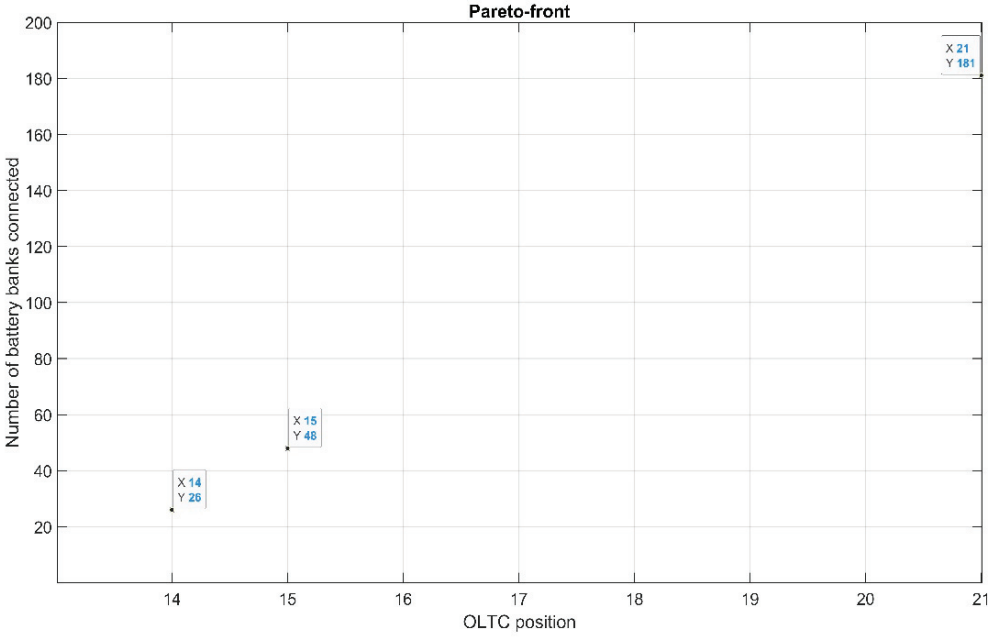


Figure 18. One of two-criteria optimization results.

Table 9. The minimum and maximum value of the voltage error for different voltage regulation versions from the entire simulation.

Type of Voltage Regulation	Minimal Voltage Deviation in p.u.	Maximal Voltage Deviation in p.u.
Voltage regulation with an evolutionary algorithm	−0.15	0.41
Voltage regulation with an evolutionary algorithm and independent compensation of reactive power	−0.21	0.22
With the use of two-criteria optimization and the Pareto-front solution selection algorithm	−0.14	0.12

Table 10. The minimum and maximum value of the $|tg\varphi|$ for different voltage regulation versions from the entire simulation.

Type of Voltage Regulation	Minimal $\|tg\varphi\|$	Maximal $\|tg\varphi\|$
Voltage regulation with an evolutionary algorithm	0.32	0.92
Voltage regulation with an evolutionary algorithm and independent compensation of reactive power	0	0.52
With the use of two-criteria optimization and the Pareto-front solution selection algorithm	0	0.4

In the case of voltage regulation with the use of evolutionary algorithms without reactive power compensation, there are large positive voltage errors. The maximum $tg\varphi$ factor significantly exceeds the permissible value. In the case of voltage regulation using evolutionary algorithms with independent reactive power compensation, the $tg\varphi$ range has improved, but it also exceeds the allowable value. The voltage deviations range from ±20% of Un. Only the reaction with multi-criteria optimization keeps the $tg\varphi$ in the correct range.

The range of voltage deviations slightly exceeds the permissible value by a maximum of 4% Un.

The first three figures show the frequency distribution of the voltage error. Figure 19 shows the voltage error for the evolution algorithm. The next Figure 20 shows the voltage error for the evolution algorithm with independent reactive power compensation. Figure 21 shows the voltage deviation for two-criteria optimization and the algorithm for selecting the Pareto front solution. For multi-criteria optimization, the obtained values were the smallest range of voltage errors and the highest frequency of errors close to zero. The charts above show that multi-criteria optimization works best. The next three figures refer to the absolute value of the $tg\varphi$ coefficient. Figure 22 shows the results of optimization of the evolution algorithm. The next Figure 23 shows $tg\varphi$ and the evolution algorithm with independent reactive power compensation. Figure 24 shows the $tg\varphi$ for two-criteria optimization and the algorithm for selecting a Pareto front solution. Only for the multi-criteria algorithm, the results of the $tg\varphi$ coefficient were obtained within the acceptable range.

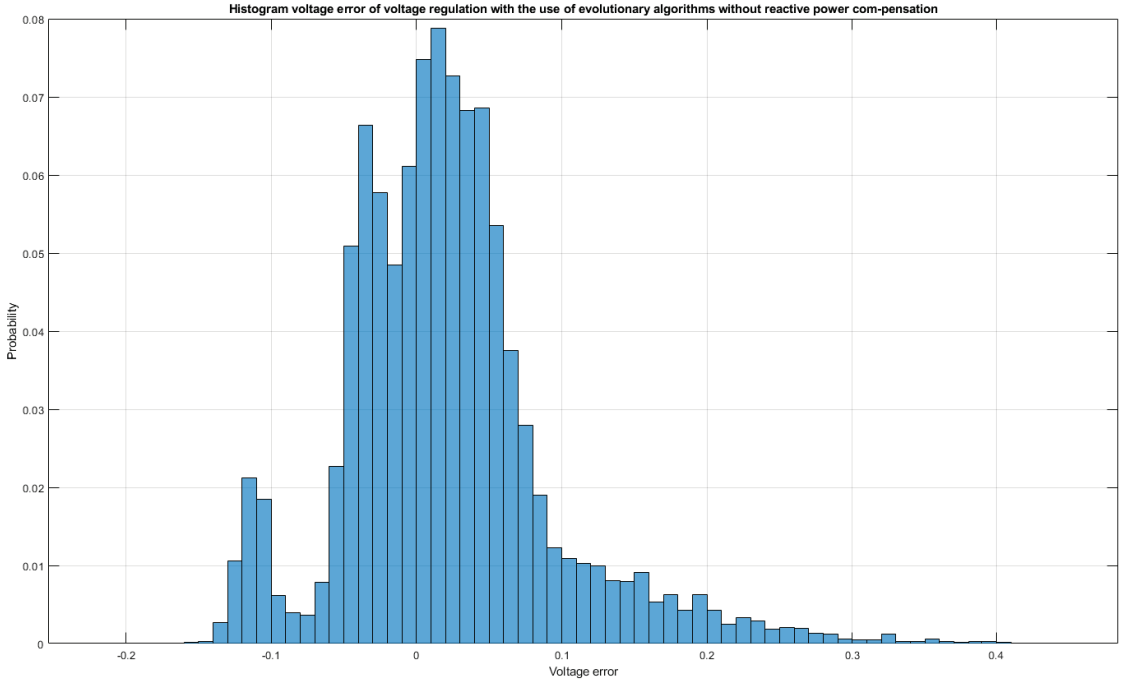


Figure 19. Histogram voltage error—evolutionary algorithm.

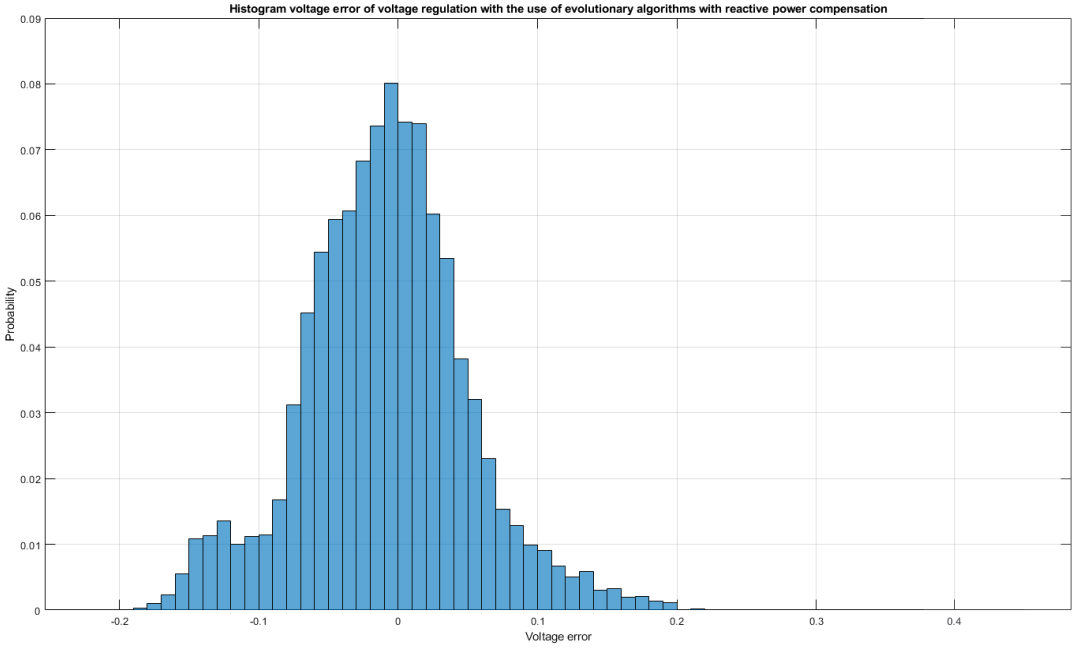


Figure 20. Histogram voltage error—evolutionary algorithm with reactive power compensation.

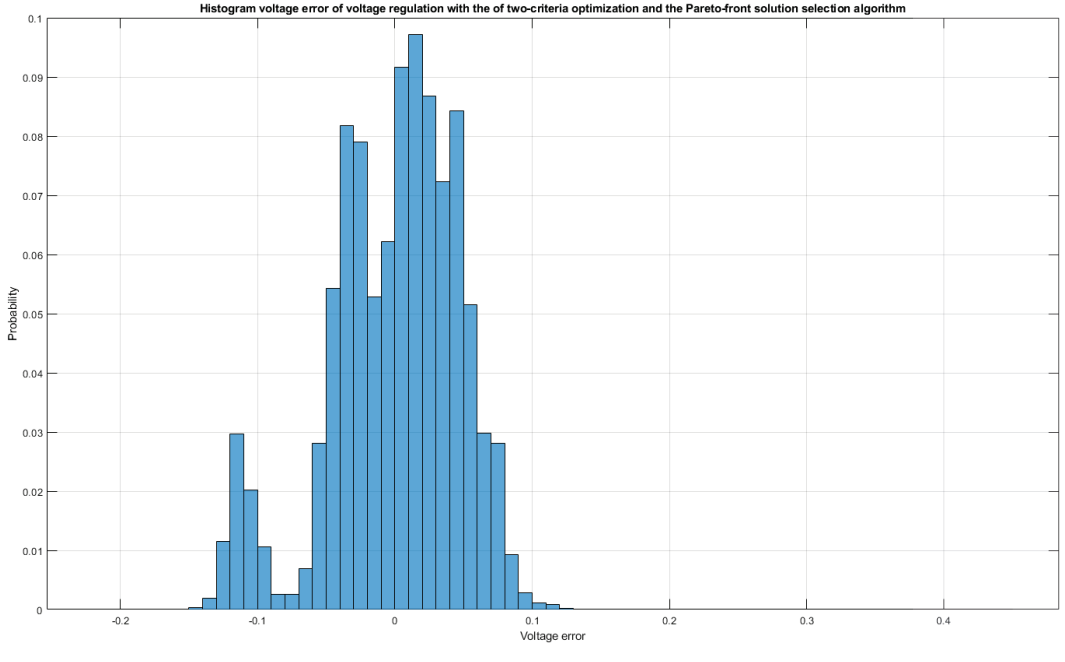


Figure 21. Histogram voltage error—use of two-criteria optimization and the Pareto-front solution selection algorithm.

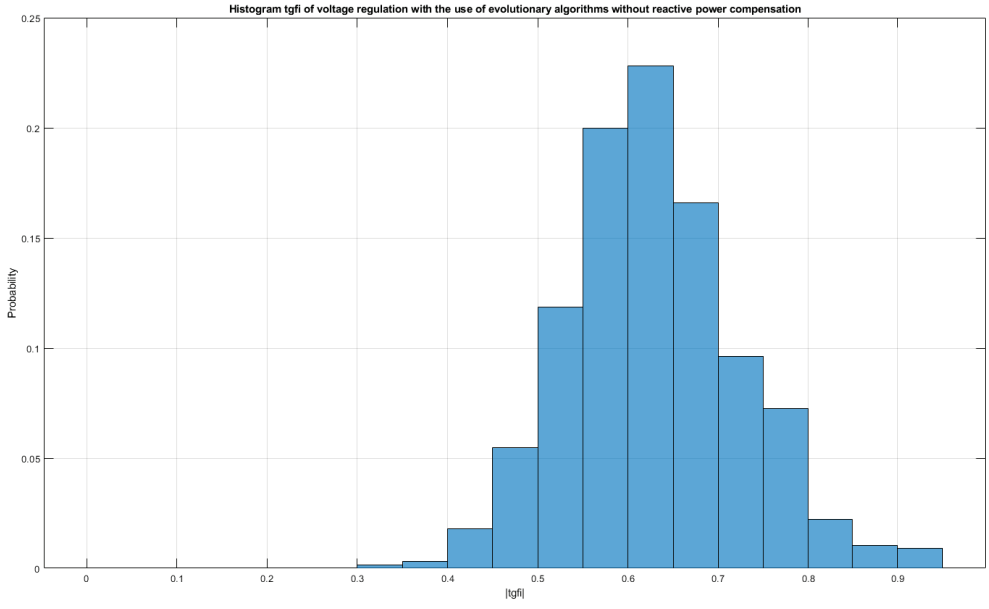


Figure 22. Histogram $tg\varphi$—evolutionary algorithm.

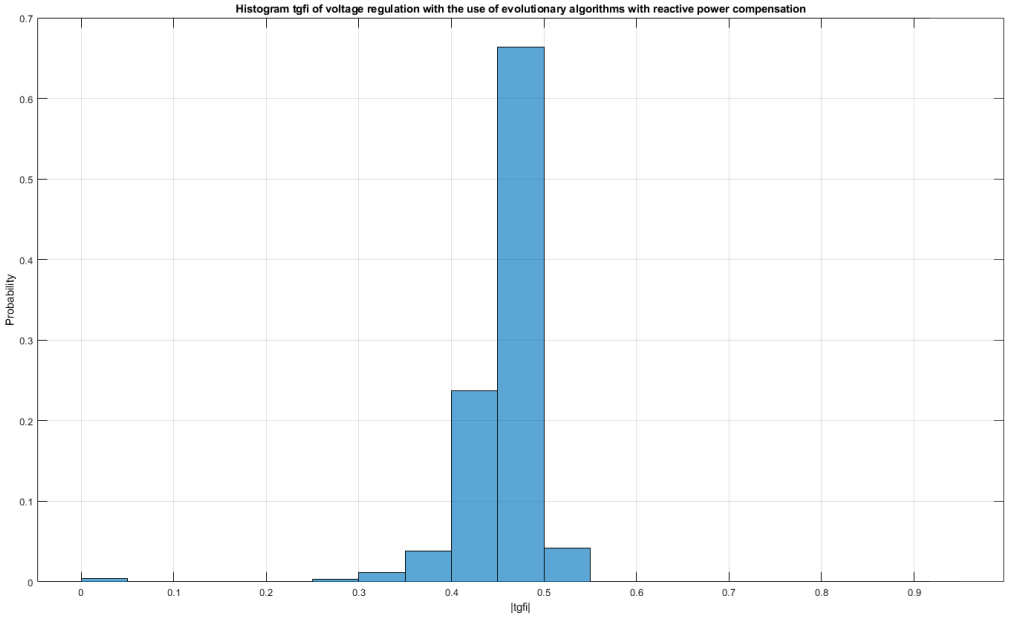


Figure 23. Histogram $tg\varphi$—evolutionary algorithm with reactive power compensation.

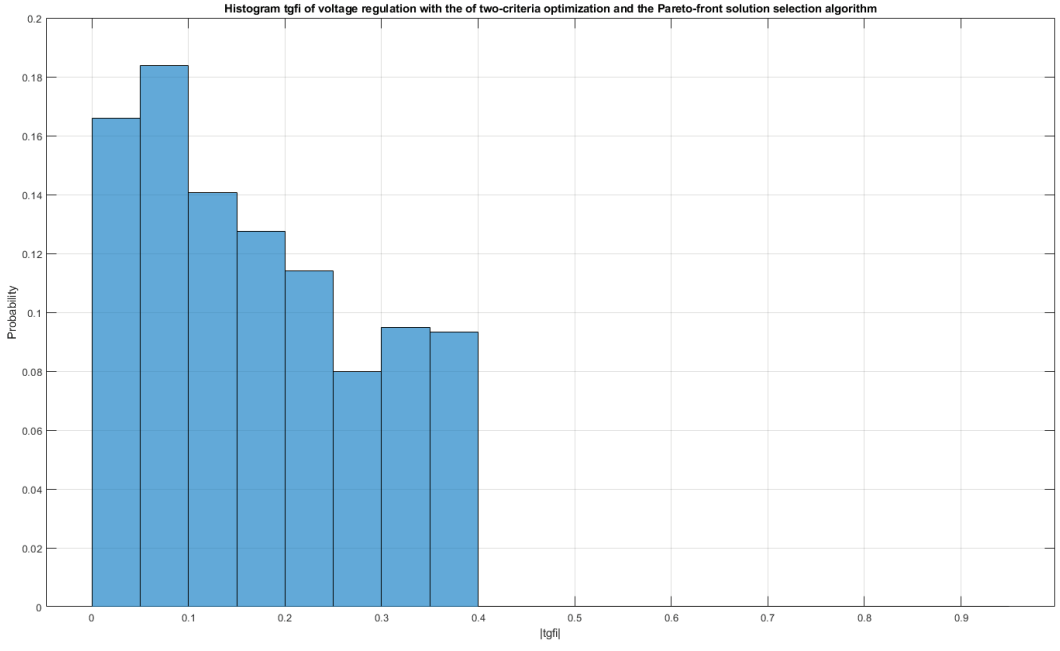


Figure 24. Histogram $tg\varphi$—use of two-criteria optimization and the Pareto-front solution selection algorithm.

4. Discussion and Conclusions

The use of voltage measurements from all MV/LV substations and the use of evolutionary algorithms significantly improve the quality of voltage regulation. Despite voltage changes in the reference node and power changes in load nodes, the voltage variance is several dozen smaller than in the case of classical regulation. The voltage range in nodes with the use of evolutionary algorithms has higher values than in the case of classical regulation. This is justified as there are greater voltage drops in distant nodes when there are no local energy sources. When analyzing the minimum and maximum values for the three control variants, it is clear that in the case of no regulation, these values are outside the range of permissible values. In the case of classical regulation, there was an improvement. It is true that the minimum values exceed the lower limit of the permissible voltage range. Only the results obtained using the evolutionary algorithm with access to the current measurement values of the network nodes allowed for a significant improvement in the quality of voltage regulation.

Independently conducted voltage regulation and reactive power compensation often cause deterioration of one of them. This is due to the fact that if the voltage on the MV side is close to the upper allowable limit and reactive power compensation is required, then the voltage value increases above the limit. For a data set with a higher reactive power, there were as many as 73% of such cases. Typically, reactive power compensation systems switch off all capacitor banks after exceeding the upper voltage limit. However, this causes a deterioration of the work quality of the reactive power compensation system. Exceeding $tg\varphi$ above 0.4 causes the necessity to pay additional charges, increase active power losses, and increase voltage drops. For this reason, it is required to build an integrated voltage regulation and reactive power compensation system. It follows that we have a multi-criteria optimization problem.

Classic voltage regulation systems in the power grid use only the transformer voltage on the lower voltage side. Due to the voltage drops at the ends of the lines, the voltage

value may exceed the lower allowable limit. For this reason, current compensation was implemented in voltage regulators. However, there are many lines fed from the same transformer. These lines are loaded differently. These lines can also have different sections. Therefore, it is difficult to choose an impedance value for current compensation. In practice, current compensation is turned off and the voltage setpoint is set to a value between 1 and 1.1 p.u. In the case of a voltage regulation system that uses voltage measurements from all powered stations, the problem of current compensation does not exist. For this reason, it is recommended to build an integrated voltage regulation and reactive power compensation system using voltage measurements from all substations supplied from this transformer.

The use of multi-criteria optimization together with the Pareto-front solution selection algorithm allows to obtain the correct settings of the semiconductor on-load tap-changer and the correct number connected of capacitor banks.

The obtained results enable the construction of a voltage regulator and reactive power compensation in the form of a neural network, a fuzzy regulator, or a neuro-fuzzy regulator. The obtained results will be used to train the neural network. The exported Matlab results will be used in the Anaconda/Phyton environment to create a neural network. The resulting network will be implemented on an STM32 microcontroller using Cube.AI.

Author Contributions: Conceptualization, J.K.; methodology, J.K. and M.M.-S., software, J.K.; validation, J.K. and M.M.-S., formal analysis, J.K. and M.M.-S., investigation, J.K. and M.M.-S.; resources, J.K. and M.M.-S.; data curation, J.K.; writing—J.K.; original draft preparation, J.K. and M.M.-S.; visualization, J.K.; supervision, M.M.-S.; project administration, M.M.-S.; funding acquisition, M.M.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a research project of Gdynia Maritime University in Poland, No. WE/2022/PZ/02.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Choi, J.-H.; Moon, S.-I. The Dead Band Control of LTC Transformer at Distribution Substation. *IEEE Trans. Power Syst.* **2009**, *24*, 319–326. [CrossRef]
2. Choi, J.-H.; Kim, J.-C. Advanced voltage regulation method of power distribution systems interconnected with dispersed storage and generation systems. *IEEE Trans. Power Deliv.* **2001**, *16*, 329–334. [CrossRef]
3. Faiz, J.; Siahkolah, B. New Solid-State on-Load Tap-Changer Topology for Distribution Transformers. *IEEE Power Eng. Rev.* **2002**, *22*, 71. [CrossRef]
4. Patel, N.R.; Lokhande, M.M.; Jamnani, J.G. Solid-State on Load Tap-Changer for Transformer Using Microcontroller. *Int. J. Eng. Dev. Res.* **2014**, *4*, 101–104.
5. Demirci, O.; Torrey, D.A.; Degeneff, R.C.; Schaeffer, F.K.; Frazer, R.H. A new approach to solid-state on load tap changing transformers. *IEEE Trans. Power Deliv.* **1998**, *13*, 952–961. [CrossRef]
6. De Oliveira Quevedo, J.; Cazakevicius, F.E.; Beltrame, R.C.; Marchesan, T.B.; Michels, L.; Rech, C.; Schuch, L. Analysis and Design of an Electronic On-Load Tap Changer Distribution Transformer for Automatic Voltage Regulation. *IEEE Trans. Ind. Electron.* **2017**, *64*, 883–894. [CrossRef]
7. Korpikiewicz, J.G. A Concept of New Current Compensation in a HV/MV Transformer's Semiconductor Tap-Changer Controller. *Acta Energ.* **2019**, *17*, 28–36.
8. Faiz, J.; Siahkolah, B. *Electronic Tap-Changer for Distribution Transformers*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 2.
9. Sanjay, M.A.; Raosaheb, T.S.; Ravindra, N.S. Solid State on Load Tap Changer for Transformer. *Resincap J. Sci. Eng.* **2021**, *5*, 4.
10. Abdou, M.S.; Mostafa, H.E.; Abdalla, Y.S. Solid State-Based On-Load Tap-Changer Control. *Port Said Eng. Res. J.* **2013**, *17*, 79–84.
11. Faiz, J.; Siahkolah, B. Solid-state tap-changer of transformers: Design, control and implementation. *Int. J. Electr. Power Energy Syst.* **2011**, *33*, 210–218. [CrossRef]
12. Shi, F.; Yin, Y.; Ding, B.P.; Gao, F.; Jia, P.F.; Hao, L.N.; Zhang, L. Development of 110 kV Thyristor Assisted Arc Extinguishing Hybrid OLTC. *E3S Web Conf.* **2021**, *243*, 01003. [CrossRef]
13. Faiz, J.; Siahkolah, B. Differences Between Conventional and Electronic Tap-Changers and Modifications of Controller. *IEEE Trans. Power Deliv.* **2006**, *21*, 1342–1349. [CrossRef]
14. Korpikiewicz, J.G.; Mysiak, P. Classical and Solid-state Tap-changers of HV/MV Regulating Transformers and their Regulators. *Acta Energ.* **2017**, *14*, 110–117.

15. Wei, T.; Yu, Z.; Chen, Z.; Zhang, X.; Wen, W.; Huang, Y.; Zeng, R. Design and test of the bidirectional solid-state switch for an 160 kV/9kA hybrid DC circuit breaker. In Proceedings of the 2018 IEEE Applied Power Electronics Conference and Exposition (APEC), San Antonio, TX, USA, 4–8 March 2018; pp. 141–148. [CrossRef]
16. Su, X.; Liu, J.; Tian, S.; Ling, P.; Fu, Y.; Wei, S.; SiMa, C. A Multi-Stage Coordinated Volt-Var Optimization for Integrated and Unbalanced Radial Distribution Networks. *Energies* **2020**, *13*, 4877. [CrossRef]
17. Beyer, K.; Beckmann, R.; Geißendörfer, S.; von Maydell, K.; Agert, C. Adaptive Online-Learning Volt-Var Control for Smart Inverters Using Deep Reinforcement Learning. *Energies* **2021**, *14*, 1991. [CrossRef]
18. Gubert, T.C.; Colet, A.; Casals, L.C.; Corchero, C.; Domínguez-García, J.L.; Sotomayor AA, D.; Alet, P.J. Adaptive Volt-Var Control Algorithm to Grid Strength and PV Inverter Characteristics. *Sustainability* **2021**, *13*, 4459. [CrossRef]
19. Jung, Y.; Han, C.; Lee, D.; Song, S.; Jang, G. Adaptive Volt–Var Control in Smart PV Inverter for Mitigating Voltage Unbalance at PCC Using Multiagent Deep Reinforcement Learning. *Appl. Sci.* **2021**, *11*, 8979. [CrossRef]
20. Go, S.-I.; Yun, S.-Y.; Ahn, S.-J.; Kim, H.-W.; Choi, J.-H. Heuristic Coordinated Voltage Control Schemes in Distribution Network with Distributed Generations. *Energies* **2020**, *13*, 2849. [CrossRef]
21. Hasan, E.O.; Hatata, A.Y.; Badran, E.A.; Yossef, F.M.H. A new strategy based on ANN for controlling the electronic on-load tap changer. *Int. Trans. Electr. Energ. Syst.* **2019**, *29*, e12069. [CrossRef]
22. Keshta, H.E.; Ali, A.A.; Malik, O.P.; Saied, E.M.; Bendary, F.M. Voltage Control of Islanded Hybrid Micro-grids Using AI Technique. In Proceedings of the 2020 IEEE Electric Power and Energy Conference (EPEC), Edmonton, AB, Canada, 9–10 November 2020; pp. 1–6. [CrossRef]
23. Bielecka, A.; Wojciechowski, D. Predykcyjne sterowanie równoleg\lym filtrem aktywnym ze sprzężeniem od prądu zasilającego. *Przegląd Elektrotechniczny* **2019**, 128–132. [CrossRef]
24. Bielecka, A.; Wojciechowski, D. Stability Analysis of Shunt Active Power Filter with Predictive Closed-Loop Control of Supply Current. *Energies* **2021**, *14*, 2208. [CrossRef]
25. Strzelecki, R.; Mysiak, P.; Sak, T. Solutions of inverter systems in Shore-to-Ship Power supply systems. In Proceedings of the 2015 9th International Conference on Compatibility and Power Electronics (CPE), Lisbon, Portugal, 24–26 June 2015; pp. 454–461.
26. Shuttleworth, R.; Tian, X.; Fan, C.; Power, A. New tap changing scheme. *IEE Proc. Electr. Power Appl.* **1996**, *143*, 108–112. [CrossRef]
27. Lebkowski, A. Evolutionary methods in the management of vessel traffic. In *Information, Communication and Environment: Marine Navigation and Safety of Sea Transportation*; CRC Press: Boca Raton, FL, USA, 2015; pp. 259–266.
28. Fadaee, M.; Radzi, M.A.M. Multi-objective optimization of a stand-alone hybrid renewable energy system by using evolutionary algorithms: A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3364–3369. [CrossRef]
29. Gu, F.; Liu, H.-L.; Tan, K.C. A multiobjective evolutionary algorithm using dynamic weight design method. *Int. J. Innov. Comput. Inf. Control* **2012**, *8*, 3677–3688.
30. Lazarowska, A. Ant colony optimization based navigational decision support system. *Procedia Comput. Sci.* **2014**, *35*, 1013–1022. [CrossRef]
31. Lazarowska, A. Swarm intelligence approach to safe ship control. *Pol. Marit. Res.* **2015**, *22*, 34–40. [CrossRef]

applied sciences

MDPI

Article

Techno-Economic Analysis of Commercial Size Grid-Connected Rooftop Solar PV Systems in Malaysia under the NEM 3.0 Scheme

Alaa A. F. Husain [1], Maryam Huda Ahmad Phesal [1,*], Mohd Zainal Abidin Ab Kadir [2] and Ungku Anisa Ungku Amirulddin [1]

[1] Institute of Power Engineering, Universiti Tenaga Nasional, Kajang 43000, Malaysia; alaa.a.f.husain@gmail.com (A.A.F.H.); anisa@uniten.edu.my (U.A.U.A.)
[2] Advanced Lightning, Power and Energy Research Centre (ALPER), Universiti Putra Malaysia (UPM), Serdang 43400, Malaysia; mzk@upm.edu.my
* Correspondence: hmaryam@uniten.edu.my

Abstract: Commercial grid-connected rooftop solar PV systems are widely applied worldwide as part of affordable and clean energy initiatives and viable long-term solutions for energy security. This is particularly true in a crowded city where space is a constraint and at the same time, there are unutilized rooftops. With the recently announced Net Energy Metering (NEM) 3.0, commercial buildings in Malaysia can apply up to 75% capacity of the maximum demand (MD), which can be connected to the grid. Apart from reducing electricity bills, the owner can offset energy for 10 years. This paper presents a design analysis with the details of the sizing of a rooftop PV system. The PVsyst software tool is used to estimate the energy produced by a 380 kWp system, and this study provides a financial analysis to evaluate the profitability of the system with a particular interest in commercial buildings under the NEM 3.0 policy, which has resulted in 8.4 years return of investment (ROI). PVsyst is a software used to size the PV system and provides technical, financial, and environmental analysis. This in-depth analysis could provide a useful case study for asset owners in deciding the way forward for sustainable energy production, cost saving, and combating the energy security issue, since Malaysia is blessed with an abundance of sunshine throughout the year.

Keywords: rooftop solar PV; net energy metering (NEM); maximum demand; PV software

Citation: Husain, A.A.F.; Phesal, M.H.A.; Ab Kadir, M.Z.A.; Ungku Amirulddin, U.A. Techno-Economic Analysis of Commercial Size Grid-Connected Rooftop Solar PV Systems in Malaysia under the NEM 3.0 Scheme. *Appl. Sci.* **2021**, *11*, 10118. https://doi.org/10.3390/app112110118

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 22 September 2021
Accepted: 19 October 2021
Published: 28 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Motivation

Energy plays a key role in the advancement and development of human activity and a brighter future. However, the current energy supply is not enough to cover the demand in the coming years, and it causes damage to the environment [1]. Solar energy is an abundant green source that can replace fossil fuel. Since the sun is available everywhere, solar energy is widely used by most countries, including developing countries. It is used in Malaysia under various government policies. These policies are implemented to encourage the use of green energy and support small and large projects [2,3]. Most implemented PV system are residential or large-scale PV system in Malaysia. Under NEM, residential fill in most of the quota.

Hybrid PV systems are not a popular practice in Malaysia. Due to the lack of analysis proving the variability of implementing a system that can satisfy the electrical needs of the consumer and at the same time reduce the cost of electricity, the consumer pays yearly. NEM 3.0 approved a rooftop PV system implementation that is connected to the utility grid for commercial buildings. The consumer that falls under this category only is able to implement 75% of the maximum demand according to the size of the PV power plant. This gives the motivation to test this policy profitability financially and assess its technical and environmental impact.

1.2. Brief on Solar PV Policies in Malaysia

Malaysia began to implement polices to encourage renewable energy storage in 2011 with Feed-in Tariff [4–11]. In 2016, the scheme changed to NEM, which was modified in 2019. These polices are implemented with the aim of meeting the target for installing renewable energy in Malaysia. In achieving the aspirational goal of having 20% renewable energy in the country's national installed capacity mix by 2025, the major renewable resource that will contribute to the RE mix is solar energy [12–14].

Three types of NEM have been implemented in recent years [14] i.e., NEM 2016 (NEM 1.0), NEM 2019 (NEM 2.0), and recently, NEM 2021 (NEM 3.0) [15–17]. In NEM 2016, Equation (1) is the total cost of the electricity bill generated monthly for the customer and is computed by deducting the price of the power generated by PV, which is MYR 0.31, from the price of the power consumed by the customer from the grid. The power generated by PV and the price of generated energy are computed using Equations (2) and (3), respectively.

$$\text{Bill price} = \text{price of consumed energy} - \text{price of generated energy} \quad (1)$$

$$\text{Price of generated energy} = \text{power generated by PV} \times \text{displaced cost by the grid} \quad (2)$$

$$\text{Price of consumed energy} = \text{power consumed by the customer} \times \text{price tariff} \quad (3)$$

However, in 2019, the NEM strategy changed, since not many customers installed PV systems based on the poor financial return from this policy. The policy now offsets the customer cost for every 1 kWh produced by the PV system with 1 kWh consumed from the grid. Equations (4) and (5) shows the monthly electricity bill computed after deducting the power generated by the customer from the power consumed:

$$\text{Total power} = \text{power consumed from grid} - \text{power generated by PV} \quad (4)$$

$$\text{Electricity bill} = \text{total power} \times \text{price tariff.} \quad (5)$$

Based on other case studies, this policy has a better financial return than NEM 2016 [10]. For commercial and industrial systems, the maximum capacity of the PV system installed is 1 MW or 75% of maximum demand (MD) of their existing installation, or 60% of the fuse rating of the transformer [12,13].

For domestic or residential consumers, the maximum capacity of the PV systems installed is less than or equal to 12 kW for a single phase or 72 kW for a three-phase system [11]. NEM 3.0 consists of three programs, each of which are assigned to a specific market sector. The first program is NEM Rakyat, which covers the residential segment [18,19]. A 100 MW capacity is allocated to this program, effective from 1 February 2021 [20]. The program applies the one-to-one NEM 3.0 policy for 10 years [21]. The allowed maximum PV installation capacity for the domestic consumer is 4 and 10 kW for single-phase and three-phase NEM consumers, respectively. The second program, called NEM GoMEn, covers government ministries and entities and has 100 MW allocated for solar power implementation. The program also applies a one-to-one offset for 10 years [22]. The maximum allowed PV installation capacity is 1000 kW per single account with a maximum of 75% from the MD, i.e., the average of the recorded MD of the past year or the declared MD for consumers with less than a one-year record. This is applicable for medium-voltage consumers and for low-voltage consumers not exceeding 60% of the fuse rating (for direct meters) or 60% of the current transformer (CT) rating of the metering current transformers, as shown in Table 1 [23].

Solar PV systems have many types based on different categories such as grid dependency. Some of the PV systems connect to the grid and use it as a battery storage alternative, and some systems are totally independent from the grid and use battery storage to store energy for nighttime use [16]. There are also hybrid systems using the grid as storage for excess PV energy or in case PV does not meet their demand, especially at night; at the same time, they use battery storage for a different reason [24,25]. This study focuses on

the grid-connected PV system [26]. Most of the available studies focus on the technical advancement on the designed under the old used policies for small or large-scale PV systems in Malaysia. This paper provides detailed financial analysis for a rooftop commercial-sized PV system under the recent policy announced in 2021. The paper analysis proves the profitability of such projects under the recent policy and clarifies the limitation of a system implemented under NEM3.0.

Table 1. PV policies in Malaysia applied to commercial buildings.

Policy	Year	Definition	Notes	Ref.
FiT	2011–2016	The concept of FiTs is that the yield of the user-generated photovoltaic system is sold to the utility grid at a price set by the utility network. Two meters are installed; one is used to count the electricity consumed by the user, and the second meter measures the kWh produced by the PV system and sent to the grid.	It was the first policy implemented in Malaysia.	[6–8]
NEM 1.0	2016–2019	NEM enables customers to produce and use solar energy to satisfy demand. The extra PV electricity will be exported to the grid. This surplus power is subsequently offset by a rate of MYR 0.31/kWh from the next electricity bill.	The energy produced by the PV system is consumed by the owner and the excess energy is exported to the grid. There is not money offset—only a reduction in the next electricity bill.	[15,20,27,28]
NEM 2.0	2019–2021	This policy allows a consumer who produces photovoltaic energy to export the excess energy to the grid, and each kWh is compensated by another kWh from the next electricity bill.	Each one kWh exported to the grid will be offset from the next electricity bill by deducting the value of one kWh starting with the highest tariff. In the previous NEM, the energy exported to the grid would only be paid at a displaced cost of MYR 0.31/kWh.	[16,21,29]
NEM 3.0	2021–2023	Similar concept to NEM 2.0, apart from that it permits indirect connection to commercial buildings. The allowed installed capacity is 75% of MD for commercial buildings.	A hybrid system/indirect connection is allowed	[30]

1.3. Grid-Connected PV System

The primary component of grid-connected PV systems is the power conditioning unit (PCU). The PCU converts the DC power produced by the PV array into AC power as per the voltage and power quality requirements of the utility grid. A bidirectional interface is made between the PV system, AC output circuits, and the electric utility network; typically, an onsite distribution panel or service entrance [30–32] allows the AC power produced by the PV system to either supply onsite electrical loads or to back feed the grid when the PV system output is greater than the onsite load demand. This safety feature is required in all grid-connected solar PV [33].

One of the important components of an on-grid system is net metering. Standard service meters are odometer-type counting wheels that record power consumption at a service point by means of a rotating disc, which is connected to the counting mechanism.

The rotating discs operate by an electro physical principle called the eddy current. Digital electric meters make use of digital electronic technology that registers power measurement by solid-state current and voltage sensing devices that convert analog measured values into binary values that are displayed on the meter using liquid crystal display (LCD) readouts [20]. Inverters are the main difference between a grid-connected system and a standalone system. The inverters must have the line frequency synchronization capability to deliver the excess power to the grid. Net meters have the capability to record consumed or generated power in an exclusive summation format. The recorded power registration is the net amount of power consumed—the total power used minus the amount of power that is produced by the solar power cogeneration system [34,35]. Net meters are supplied and installed by utility companies that provide grid-connection service systems [36]. Net metered solar PV power plants are subject to specific contractual agreements and are subsidized by state and municipal governmental agencies.

The usage profile (i.e., operation hours—24 h or not) of electricity is divided from Monday to Sunday into on-peak hours from 08:00 a.m. to 10:00 p.m. and off-peak hours from 10:00 p.m. to 08:00 a.m. Each period has a different tariff rate that is clarified in the next Table 2. In order to understand the PV system's financial benefit and policy rules, we must first identify the tariff that the user's electricity falls under. After that, the policy suited to the user is selected.

Table 2. Site Information.

Type	Information
Location	Bangi, Selangor (Klang Valley)
Slope azimuth	180
Roof type	Clay and concrete tiles

1.4. Contribution and Paper Organization

New polices have been implemented over the last few years in Malaysia. This is to encourage the use of renewable energy and increase its share in the energy mix. The latest implemented policy is NEM 3.0, which allows 75% of an MD rooftop PV system to be connected to the utility grid. Due to the fact that this policy is newly implemented for this category of system size, not much analysis is available that gives detailed technical and financial information. This paper provides a case study that guides the user to size the system under the recent NEM 3.0 policy and achieve final profit saving. Section 2 discusses the methodology involved to achieve the final design. Section 3 defines the load consumption and system configuration. Section 4 explains the load profile of the selected case study with the adopted method in sizing the PV system. Finally in Section 5, the PVsyst tool provides a technical and financial analysis with some discussions, and a conclusion provided at the end of the paper.

2. Methodology

Traditionally, the PV system size is obtained through the electricity profile of the consumer where the final design cost of the kWh produced using the PV system must be less than the price of the grid tariff. The PV system size must be adjusted by calculating the sun peak hours in the specific location and the average electricity used daily for a year of consumption, plus the available roof space to install the PV system; thus, the PV system size is computed. However, based on the recent NEM 3.0 policy in Malaysia, only 75% of MD is allowed to be installed and connected to the grid [37,38].

The components of the PV system are selected through the power rating decided by the manufacturers for the appliances, in conjunction with a careful estimation of how long each appliance will need power [39]. However, this can apply for small PV installation. In a large PV plant, the electricity demand based on the previous year of the electricity profile is thoroughly studied.

The PVsyst tool is commercially available software that is used to simulate solar PV projects. This study used a PV system for modeling purposes. The PVsyst software library contains detailed data about the most common photovoltaic modules, inverters, and all that is needed for a photovoltaic system project. Furthermore, it records losses due to the partial shadowing effects, mismatches between connected PV modules, wiring losses, inverter losses, and the effect of the ambient temperature variations on its electrical output power calculation. This functionality makes it a precise tool to estimate the amount of electrical energy produced by a designed system [40,41].

3. System Configuration

The grid-connected PV system configuration is simple compared with the off-grid PV system, which requires battery storage. It contains PV solar plants, an inverter, a meter, wiring, and a mounting system, as shown in Figure 1. The solar PV plant converts the photons in the sunlight into electricity that runs as DC [26,27]. The electricity enters the inverter, and then, it is transformed into AC to suit the appliances of the building [28]. The extra electricity is exported to the utility grid where it is measured by a meter. The meter also measures the electricity imported from the utility grid.

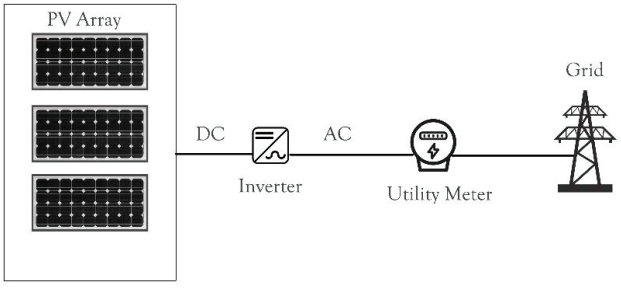


Figure 1. Grid-connected PV system configuration.

The site selected in this case study is denoted as building A, which is described in Table 2. The building is located in the Bangi area in Selangor, Malaysia and has a lot of empty spaces on the roof without any shading from nearby buildings or trees. The roof was sturdy and made of clay and concrete tiles, which make it possible to install a PV system that the roof will bear its weight.

4. Load Profile

In order to implement the NEM on the commercial building, load profile information was collected and analyzed. The collected data determined the size of the solar PV system. At present, the customer is using 100% electricity from the utility grid. From Table 3, it can be seen that the total yearly energy consumption was 3,194,184 kWh with an average monthly cost of MYR 125,787. It can be noticed that JAN and MAR are much lower than the other peak hours in other months due to system measurement dysfunction. These two months were excluded when computing the average peak hour.

Table 3. Customer electricity profile for the year 2019.

Month	Demand, kWh	MD, kW	MD, MYR	Cost, MYR
JAN	141,265	95	4285	47,957
FEB	127,680	552	24,895	43,727
MAR	346,075	95	4285	129,553
APRIL	100,093	502	22,640	100,094
MAY	357,702	494	22,279	108,422
JUN	283,652	499	22,505	113,256

Table 3. *Cont.*

Month	Demand, kWh	MD, kW	MD, MYR	Cost, MYR
JULY	344,294	514	23,181	133,032
AUG	342,547	503	22,685	342,158
SEP	299,892	502	22,640	118,990
OCT	331,863	491	22,144	128,770
NOV	335,701	498	22,460	130,132
DEC	283,513	496	22,370	113,350
TOTAL	3,194,184	-	-	1,509,441
Average	274,523	505 *	19,697	125,787

* exclude of Jan and March.

5. System Sizing

The system size was designed based on the customer's electrical profile during the year 2019. The peak sun hours of 2019 were 4.4 h on average based on Table 1. The total yearly electrical consumption was 3,194,184 kWh. The building falls under the 'C Tariff' with an electricity price during peak hours of MYR 0.365/kWh, an off-peak price of MYR 0.224/kWh, and an MD of MYR 45/kW.

In order to calculate the system size, we obtained the MD point of every month of in 2019 and calculated the average point, which was 505 kW.

$$\text{Size of PV system} = \text{MD} \times 75\% = 505\ \text{kW} \times 75\% = 378.75\ \text{kW} \cong 380\ \text{kW} \quad (6)$$

The allowed PV system size to connect to the utility grid is 380 kW, which is 16% of the total electricity demand. The annual global solar irradiation of the location of the Building A was measured using the PVsyst program, producing a value of 4.4144 kWh/m^2. The average ambient temperature was 27.4 °C. Table 4 shows the monthly global solar irradiation and the output of a one kW PV system at that location.

Table 4. Monthly global solar irradiation and the output of a 1 kW PV system.

Month	AC System Output (kWh)	Solar Radiation (kWh/m^2/day)	Irradiance (W/m^2)	DC Array Output (kWh)
1	103.19	4.44	137.60	107.85
2	95.95	4.55	127.47	100.29
3	104.86	4.52	140.15	109.62
4	94.78	4.19	125.55	99.10
5	89.40	3.80	117.80	93.58
6	85.31	3.77	113.08	89.36
7	89.52	3.80	117.91	93.78
8	94.33	4.01	124.28	98.72
9	95.52	4.21	126.16	99.85
10	98.78	4.21	130.44	103.25
11	95.13	4.20	125.87	99.45
12	95.95	4.05	125.55	100.34
Total	1142.72	49.74	1511.86	1195.19

After sizing the system according to 75% MD, the PV system production could only cover 380 kW of electricity demand. While this is seen lower compared to the 2671.2 kW PV plant where electricity demand was met 100%, this still could reduce the electricity bill and cost saving, particularly since the MD for electricity is recorded during the peak hours.

6. PV System Analysis Using PVsyst Tools

PVsyst is a design tool that provides optimization tools for sizing grid-connected, standalone, and pumping PV systems based on the location on the map. The tool also depends on the consumer electricity profile and the demands of electricity. It also provides a financial visibility for the designed project with environmental impact measured in

tonnes. The program contains the latest used technology in the market, which can be selected during the sizing process. It can also calculate the loss in the system [23]. On the other hand, there are many other online program tools that help calculate the system size based on the location and demand of the user such as HOMER and PVWATT. However, PVsyst is chosen in this work for its flexibility when choosing the modules used as well as more specification details as opposed to PVWATT. The annual energy yield for the proposed PV power plant was defined as the amount of energy fed into the grid after due consideration of all kinds of generation and distribution losses. The solar PV-based power plant comprises the optical energy input (which is essentially dependent on the geographical/seasonal/climatic and operating parameters with time) and the electrical output (which depends on the technical specifications of the electrical appliances in use). Industry standard software PVsyst V6.8.1 was used for the Energy Generation Assessment.

The system consisted of 810 units of 470 Wp PV modules that connected in 81 strings with 10 modules in each string connected in series. The total nominal power of the plant was 380 kWp, as shown in Table 5.

Table 5. PV module details.

Parameters	Values
PV module size	470 Wp
Number of modules	810 units
Nominal (STC)	380 kWp
Modules	81 strings × 10 in series
P_{mpp}	707 V
U_{mpp}	501 A
Module area	1751 m^2

The inverter used for this system design was a 100 kWac power inverter. Three inverters were used for the whole system. The inverters' operating system was 630–1000 V and the DC to AC power conversion ratio was 1.27, as shown in Table 6.

Table 6. Installed inverter details.

Parameters	Values
Unit Nom Power	100 kWac
Number of inverters	3 units
Total power	380 kWac
Operating voltage	630–1000 V
P_{nom} ratio (DC:AC)	1.27

The proposed 380 kW solar PV plant is expected to generate about 510 MWh of energy in the first year of operation at a net Capacity Utilization Factor (CUF) of 18% at the metering point, as per Table 7. Thereafter, an annual degradation factor of 2.5% for the first year and 0.7% thereafter in production has been considered for mono crystalline modules for financial calculations.

The CUF—Capacity Utilization Factor compares 380 kW solar power plants with other 380 kW power plants that run 24 × 7 for 365 days in terms of how much energy is generated. If the 380 kW solar power plant generates 100 s of Wh for 365 days running on 24 × 7 and we have some 50 s of Wh at the end of our solar plant, its capacity utilization would be 50%. The average ratio is 0.803. The PV system's nominal power output is 380 kW. Due to various factors such as site location and system losses, the power capacity of the system was reduced, as shown in Figure 2. The energy yield was calculated based on the south-facing array surface. The expected plant production for different probability scenarios (the probability of meeting a generation value) is also presented in Figure 2.

Table 7. Balance and main results.

	GlobHor kWh/m^2	DiffHor kWh/m^2	T_Amb °C	Globlnc kWh/m^2	GlobEff kWh/m^2	EArray MWh	E_Grid MWh	E_Grid MWh
JAN	133.6	79.58	27.21	140.4	137.0	44.38	43.10	0.806
FEB	131.6	81.85	27.84	135.4	132.0	42.72	41.55	0.806
MAR	153.8	88.56	28.27	153.6	149.9	48.24	46.89	0.802
APR	142.7	73.33	27.86	138.6	135.0	43.44	42.15	0.799
MAY	144.8	77.78	28.71	136.7	132.5	42.88	41.59	0.799
JUN	133.0	70.24	28.20	124.4	120.7	39.19	37.98	0.802
JUL	134.4	81.91	28.20	126.7	122.7	39.98	38.77	0.804
AUG	136.7	82.95	28.09	131.4	127.7	41.52	40.27	0.805
SEP	133.6	69,76	27.38	132.1	128.8	41.45	40.20	0.799
OCT	140.5	86.62	27.67	143.0	139.6	45.20	43.88	0.806
NOV	124.6	69.35	26.85	130.2	127.4	41.12	39.87	0.804
DEC	121.3	67.82	27.27	128.1	125.0	40.37	39.13	0.802
YEAR	1630.5	929.75	27.80	1620.5	1578.1	510.49	495.39	0.803

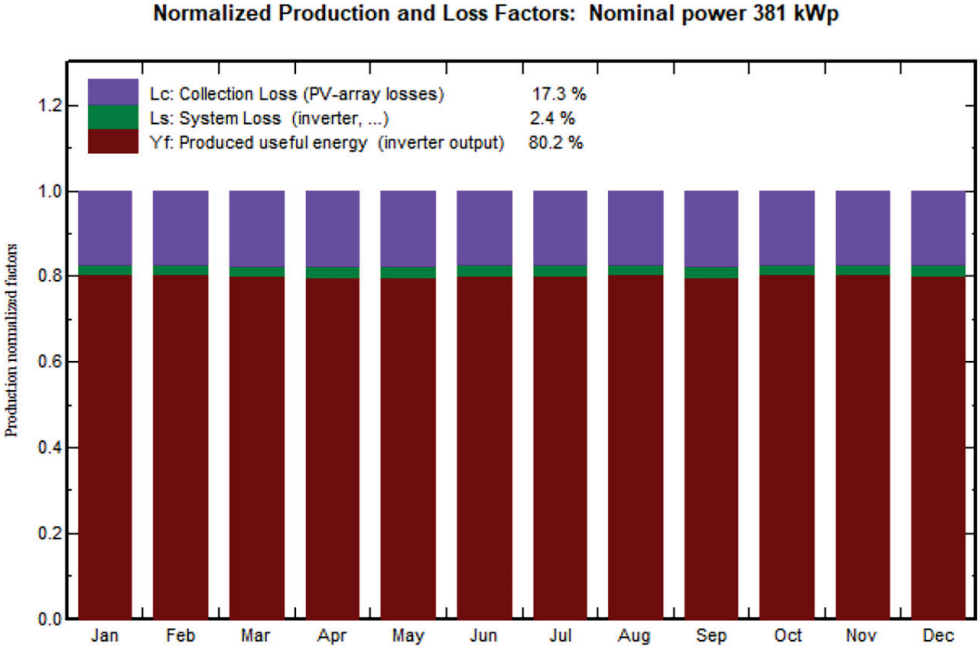


Figure 2. Normalized production (per installed kWp).

The yield factor was defined to be a factor consolidating all the system losses that occurred across this power plant. The major losses that occurred during the operation of the solar PV power plant were temperature loss, module mismatch loss, and DC to AC conversion losses. Figure 3 presents the loss diagram over the whole year. The diagram represents the energy flows in the system and the losses from every parts. Figure 3 shows that the largest losses came from temperature with 7.77%, while the inverter loss was 1.47%, which is expected from the manufacturing datils of the components.

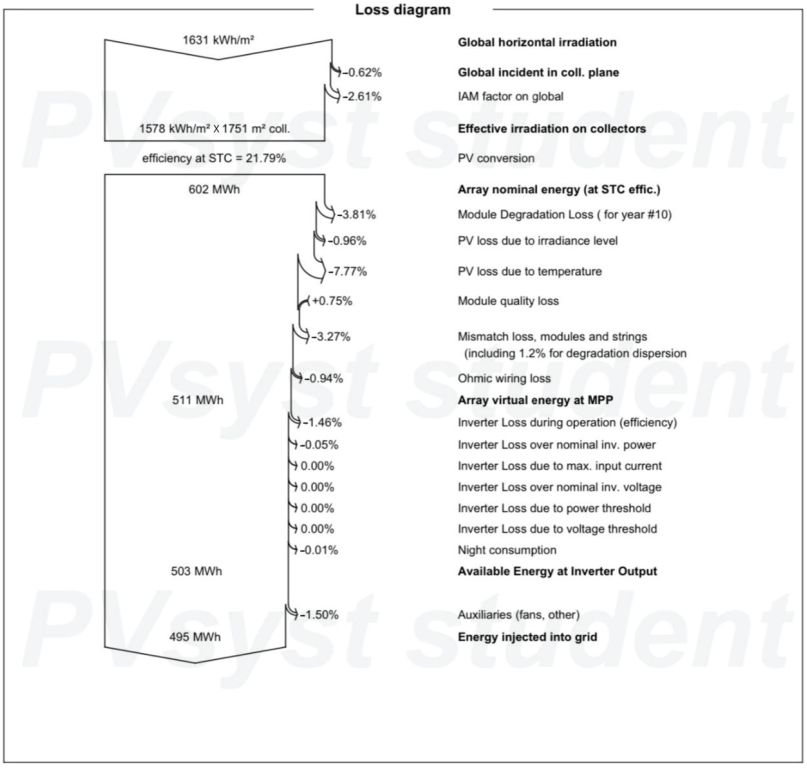


Figure 3. Loss diagram of energy for a year of expected production flowing through the PV system energy using PVsyt.

7. Financial Analysis

The objective of financial analysis is to assess the financial viability of the project from the perspective of a project developer so as to arrive at a suitable investment decision. As shown in Table 8, all of the costing is in Malaysian Ringgit (MYR). The system cost included the cost of the PV modules, inverters, mounting materials, other components (cables, mounting system, charge controller and utility meter, etc.), the balance of the system, and interconnection [29].

Table 8. Cost details of the PV system components.

Item	Quantity Units	Cost, MYR	Total, MYR
PV modules	810	932	755,121
Inverter	3	80,332	240,996
Other components	1	289,195	289,195
Installation	810	238	192,797
Operating cost		5141	5141
	Total		1,478,109

Based on Table 8, the total installation cost is MYR 1,478,108.8; the operating cost is MYR 5141.25/year produced energy, and the cost of the produced energy (LCOE) is MYR 0.130/kWh. After a detailed analysis, it was concluded that the project is financially viable and that it will have a project lifetime of 25 years (starting year 2022), financing cost at MYR 1,478,108.8, and a payback period of 8.4 years. The net present value (NPV) is MYR

2,459,445.1, and the return on investment (ROI) would be 166.4%, as shown in Figure 4. Due to the degradation of the PV system, the profit starts to decrease in the last years of the project lifetime.

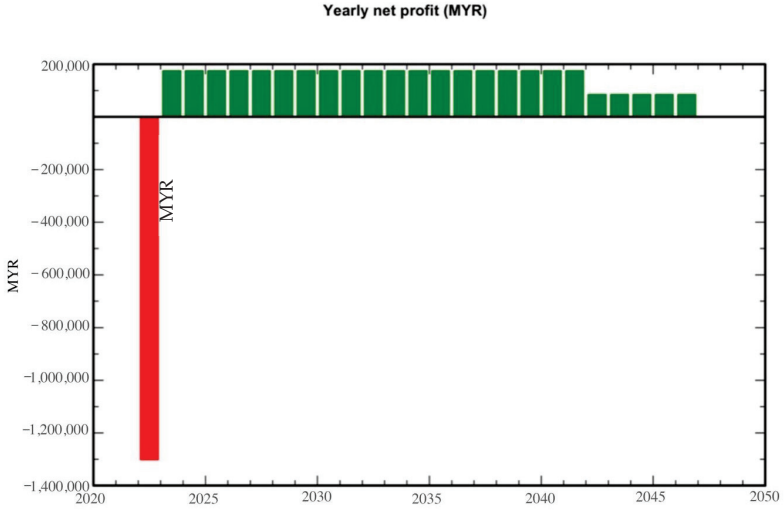


Figure 4. Yearly net profit in MYR during the lifetime of a 380 kW PV system.

From Figure 5, it can be observed that from the ninth year, the project will be in the positive *y*-axis, beginning to generate profit from the initial investment. Cash flow was calculated based on (1) the tariff rate for each unit (kWh) by the solar plant against (2) the fixed cost of the investment and (3) the operational cost annually. By calculating the difference as = (1) − (2 + 3), we derived the amount shown in the graph. All the costs were equalized for a lifespan of 25 years. For the first 8.4 years, the investment cost is greater than the profit amount; then, the breakeven point is reached, and positive values result in annual profits from the PV system.

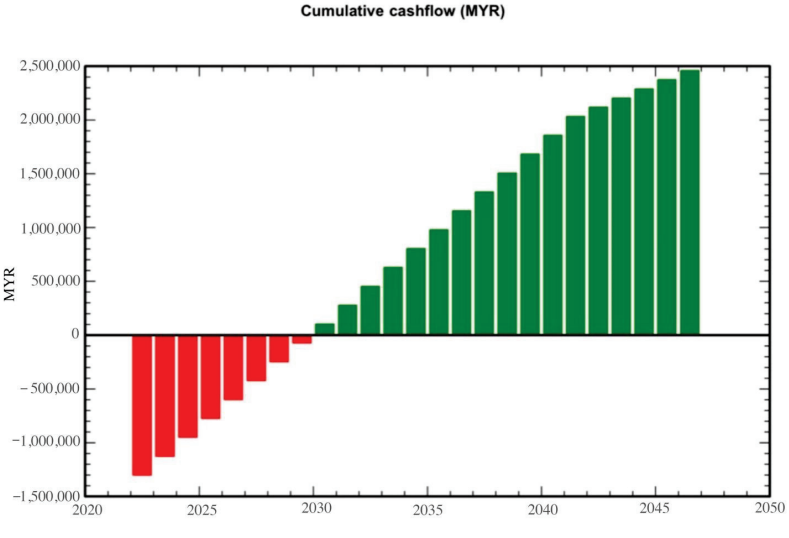


Figure 5. Cumulative cashflow in MYR during the lifetime of 380 kW PV system.

8. Environmental Impact

The implementation of solar power has been greatly encouraged due to its influence on reducing carbon emissions. Thus, it is important to measure the environmental benefit of using solar energy instead of fossil fuel. Figure 6 shows the amount of CO_2 emission saved during the lifetime of the project. The calculations depend majorly on the value of the life cycle emissions (LCA), which represent CO_2 emissions associated with a given component or quantity of energy. This includes the total life cycle of a component or the amount of energy, including production, operation, maintenance, disposal, etc. The rationale behind the carbon footprint tool is that the electricity generated by the photovoltaic system will replace the same amount of electricity in the existing grid. If the carbon footprint of the electricity generation on the grid is more than the PV system per kWh, carbon dioxide emissions will be reduced. From Figure 6, CO_2 emission is negative until the last 10 years of the project; then, it becomes positive due to the fact that the PV system installed satisfies only 16% of the consumer demand. The rest of the demand is met by the grid of utilities using traditional fossil fuel energy that causes CO_2 emissions in the air.

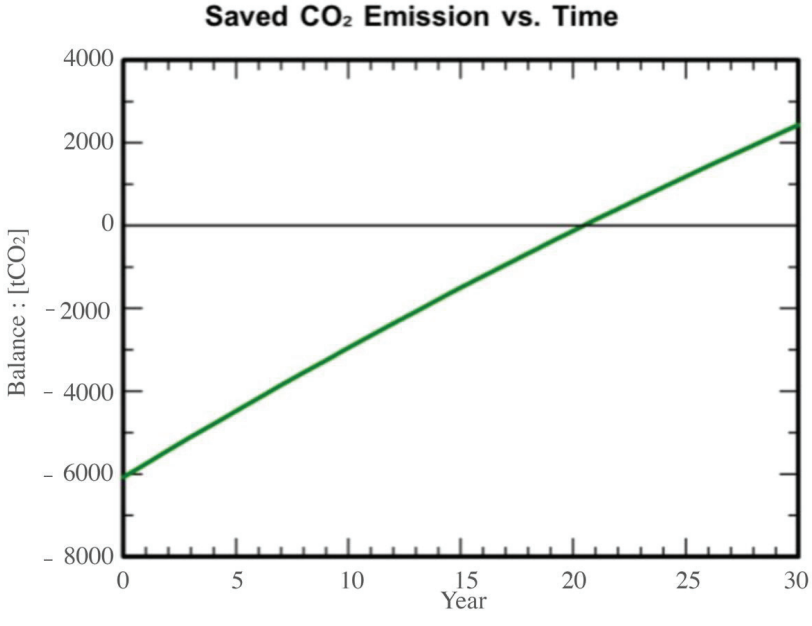


Figure 6. CO_2 emission reduction vs. time.

9. Discussion

The average annual solar radiation at the proposed site on the horizontal surface is about 1598 kWh/m^2 (as per NASA-SSE satellite data) [42], which is adequate for the installation of the solar PV ground-mounted utility scale system. Annual expected generation for the entire solar PV project is 510 MWh/year, where the evaluation is based on a probabilistic approach for the interpretation of the simulation results over several years.

The 380 kW grid-connected PV system consists of 810 PV solar modules grouped into 20 × 81 strings with each string containing 10 modules in series. The output of the strings is pooled in the array junction box through 4 mm^2 photovoltaic DC cables. The output from the junction box is fed to the three grid-tied inverters. The grid-connected inverter is used as a power conditioning unit. DC and AC distribution cabinets contain protective components for the safety of the system.

Based on the results obtained in PVsyst tools, one day of energy production by the PV system plant was measured and compared with the total electricity demand in one day, as

shown in Figure 2. From the figure, it can be seen that only 16% of the electricity demand is covered. This is due to the restricted rule of NEM that only allows 75% of the MD of electricity to be connected to the utility grid.

Although the PV system plant would only cover 16% of the total electricity demand, the reduction in the electricity bill would be significant since the solar PV plant produces energy during the on-peak hours. The electricity rates are MYR 0.365/kWh and MYR 0.224/kWh during on-peak and off-peak hours, respectively. Moreover, the MD point during the day can also be covered by the PV plant or can be reduced significantly. Since this plant is small compared with the electricity demand of the building, it will not export electricity except for on the weekends. It is expected that the electricity will be used on a daily basis, as shown in Figure 7.

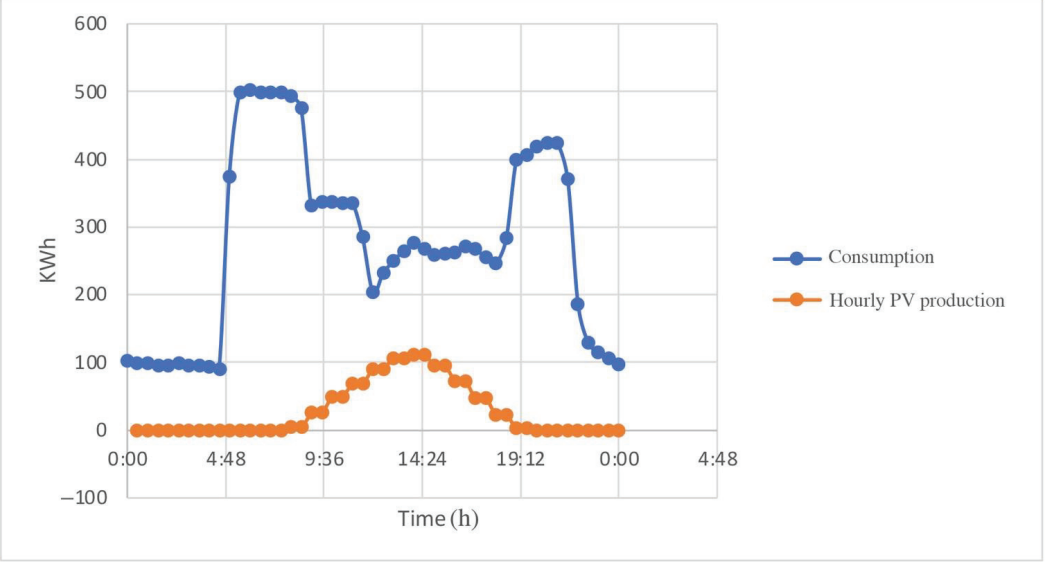


Figure 7. One-day energy consumption and PV production.

10. Conclusions

A techno-economic analysis of commercial-size grid-connected rooftop solar PV systems has been presented in this paper, taking into account the recently announced NEM 3.0 scheme in 2021. The availability of the solar irradiance at that location is relatively high i.e., 4.4144 kWh/m^2 day throughout the year. The size of the PV plant was calculated based on the consumer electricity profile of one year. The average MD of 505 kW was considered for sizing 75% of the system. Due to the fact that the power output of the PV system is 510 MWh/year, the PV system size would only cover 16% of the yearly electricity demand. After sizing the PV system, it was modeled and simulated using the PVsyst program. The system performance showed promising results of 510.49 MWh produced yearly from the plant. A financial analysis was carried out with an estimation of the system price based on percentage and rates from the original country of the building location. Based on the analysis, the project's ROI is in 8.4 years, and the LCOE is at MYR 0.130/kWh, which is less than both tariffs from the utility during on-peak and off-peak hours. The net present value (NPV) is MYR 2,459,445.1, and ROI is positive and equal to 166.4%. The system would have a positive environmental impact with a reduction in the emission of CO_2 at least until the end of the tenure agreement. The results from this project analysis show good promise

and benefits not just for energy and cost savings but also in terms of utilizing the dead spaces on the rooftop of buildings.

Author Contributions: Conceptualization; methodology; resources; data curation; writing—original draft preparation, A.A.F.H.; visualization; supervision, M.H.A.P.; supervision, writing—review and editing, M.Z.A.A.K.; project administration, U.A.U.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universiti Tenaga Nasional through UNITEN BOLD Scholarship.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Tsoutsos, T.; Frantzeskaki, N.; Gekas, V. Environmental impacts from the solar energy technologies. *Energy Policy* **2005**, *33*, 289–296. [CrossRef]
2. Vaka, M.; Walvekar, R.; Rasheed, A.K.; Khalid, M. A review on Malaysia's solar energy pathway towards carbon-neutral Malaysia beyond COVID'19 pandemic. *J. Clean. Prod.* **2020**, *273*, 122834. [CrossRef]
3. SEDA. Sustainable Energy Development Authority Malaysia. "Grid Parity; Displaced Cost". Available online: http://www3.seda.gov.my (accessed on 10 October 2021).
4. Gomesh, N.; Daut, I.; Irwanto, M.; Irwan, Y.; Fitra, M. Study on Malaysian's Perspective towards Renewable Energy Mainly on Solar Energy. *Energy Procedia* **2013**, *36*, 303–312. [CrossRef]
5. Fayaza, H.; Rahimb, N.A.; Saidura, R.; Solangi, K.H.; Niaz, H.; Hossaina, M.S. Solar Energy Policy: Malaysia vs. De-veloped Countries. In Proceedings of the IEEE Conference on Clean Energy and Technology (CET), Kuala Lumpur, Malaysia, 27–29 June 2011; pp. 374–378.
6. Zhang, H.L.; Van Gerven, T.; Baeyens, J.; Degrève, J. Photovoltaics: Reviewing the European Feed-in-Tariffs and Changing PV Efficiencies and Costs. *Sci. World J.* **2014**, *2014*, 1–10. [CrossRef] [PubMed]
7. Chua, S.C.; Oh, T.H.; Goh, W.W. Feed-in tariff outlook in Malaysia. *Renew. Sustain. Energy Rev.* **2011**, *15*, 705–712. [CrossRef]
8. Wong, S.L.; Ngadi, N.; Abdullah, T.A.T.; Inuwa, I. Recent advances of feed-in tariff in Malaysia. *Renew. Sustain. Energy Rev.* **2015**, *41*, 42–52. [CrossRef]
9. Solangi, K.H.; Saidur, R.; Rahim, N.A.; Islam, M.R.; Fayaz, H. Current solar energy policy and potential in Malay-sia, 3rd. In Proceedings of the International Conference on Science and Technology, Pulau Pinang, Malaysia, 12–13 December 2008.
10. Saadatian, O.; Haw, L.C.; Mat, S.B.; Sopian, K. Perspective of sustainable development in Malaysia. *Int. J. Energy Environ.* **2012**, *6*, 260–267.
11. New Economic Model for Malaysia Part 1. Available online: https://www.pmo.gov.my/dokumenattached/NEM_Report_I.pdf (accessed on 10 October 2021).
12. Raza, M.; Alshebami, A.S.; Sibghatullah, A. Factors Influencing Renewable Energy Technological Innovation in Malaysia. *Int. J. Energy Econ. Policy* **2020**, *10*, 573–579. [CrossRef]
13. Islam, S.Z.; Othman, M.L.; Saufi, M.; Omar, R.; Toudeshki, A. Photovoltaic modules evaluation and dry-season energy yield prediction model for NEM in Malaysia. *PLoS ONE.* **2020**, *15*, e0241927. [CrossRef] [PubMed]
14. Husain, A.A.; Phesal, M.H.A.; Kadir, M.Z.A.; Amirulddin, U.A.U. Short Review on recent solar PV policies in Malaysia. *E3s Web Conf.* **2020**, *191*, 1002. [CrossRef]
15. Poullikkas, A.; Kourtis, G.; Hadjipaschalis, I. A review of net metering mechanism for electricity renewable energy sources. *Int. J. Energy Environ.* **2013**, *4*, 975–1002.
16. Abdullah, W.S.W.; Osman, M.; Ab Kadir, M.Z.A.; Verayiah, R. The potential and status of renewable energy de-velopment in Malaysia. *Energies* **2019**, *12*, 2437. [CrossRef]
17. Christoforidis, G.C.; Chrysochos, A.; Papagiannis, G.; Hatzipanayi, M.; Georghiou, G.E. Promoting PV energy through net metering optimization: The PV-NET Project. In Proceedings of the 2013 International Conference on Renewable Energy Research and Applications (ICRERA), San Diego, CA, USA, 17–22 October 2013.
18. Razali, A.H.; Abdullah, P.; Hassan, M.Y.; Hussin, F. Comparison of New and Previous Net Energy Metering (NEM) Scheme in Malaysia. *Elektr. J. Electr. Eng.* **2019**, *18*, 36–42. [CrossRef]
19. Dutta, S.; Ghosh, D.; Mohanta, D.K. Optimum solar panel rating for net energy metering environment. In Proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 3–5 March 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2900–2904. [CrossRef]

20. Abdul, R.D. Sustainable Energy, Suruhanjaya Tenaga Energy Commission. Available online: http://gtalcc.gov.my/wp-content/uploads/2021/08/Sustainable-Energy-Energy-Efficiency-Renewable-Energy.pdf (accessed on 21 October 2021).
21. Zainudin, N.; Sharifuddin, N.S.I.; Osman, S.; Jusoh, Z.M.; Paim, L.; Zainalaludin, Z.; Nordin, N. Validating of So-lar Energy Acceptance Measurements Among Malaysian Households. *Int. J. Soc. Sci. Res.* **2020**, *2*, 20–31.
22. Tan, R.H.; Chow, T. A Comparative Study of Feed in Tariff and Net Metering for UCSI University North Wing Campus with 100 kW Solar Photovoltaic System. *Energy Procedia* **2016**, *100*, 86–91. [CrossRef]
23. Husain, A.A.F.; Phesal, M.H.A.; Ab Kadir, M.Z.A.; Amirulddin, U.A.U.; Junaidi, A.H.J. A Decade of Transitioning Malaysia toward a High-Solar PV Energy Penetration Nation. *Sustainability* **2021**, *13*, 9959. [CrossRef]
24. Chaurey, A.; Deambi, S. Battery storage for PV power systems: An overview. *Renew. Energy* **1992**, *2*, 227–235. [CrossRef]
25. Ip, A.H.; Thon, S.; Hoogland, S.; Voznyy, O.; Zhitomirsky, D.; Debnath, R.K.; Levina, L.; Rollny, L.R.; Carey, G.H.; Fischer, A.H.; et al. Hybrid passivated colloidal quantum dot solids. *Nat. Nanotechnol.* **2012**, *7*, 577–582. [CrossRef] [PubMed]
26. Alwaeli, A.H.A.; Sopian, K.; Kazem, H.A.; Chaichan, M.T. Photovoltaic/Thermal (PV/T) systems: Status and future prospects. *Renew. Sustain. Energy Rev.* **2017**, *77*, 109–130. [CrossRef]
27. Marańda, W.; Piotrowicz, M. Sizing of photovoltaic array for low feed-in tariffs. In Proceedings of the 21st International Conference Mixed Design of Integrated Circuits and Systems (MIXDES), Lublin, Poland, 19–21 June 2014; pp. 405–408.
28. Murdan, A.P.; Jeetun, A.K. Simulation of a Single Phase Grid-tied PV System under Net-Metering Scheme. In Proceedings of the 2021 IEEE Power and Energy Conference at Illinois (PECI), Urbana, IL, USA, 1–2 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6. [CrossRef]
29. Ratnam, E.L.; Weller, S.R.; Kellett, C. Scheduling residential battery storage with solar PV: Assessing the benefits of net metering. *Appl. Energy* **2015**, *155*, 881–891. [CrossRef]
30. For Solar Photovoltaic Installation under the Programme of NEM Rakyat and NEM GoMEn in Peninsular Malaysia. Available online: http://www.seda.gov.my/reportal/nem/ (accessed on 10 October 2021).
31. Zainuddin, H.; Salikin, H.R.; Shaari, S.; Hussin, M.Z.; Manja, A. Revisiting solar photovoltaic roadmap of tropical malaysia: Past, present and future. *Pertanika J. Sci. Technol.* **2021**, *29*, 1567–1578. [CrossRef]
32. Gitizadeh, M.; Fakharzadegan, H. Battery capacity determination with respect to optimized energy dispatch schedule in grid-connected photovoltaic (PV) systems. *Energy* **2014**, *65*, 665–674. [CrossRef]
33. Seepromting, K.; Chatthaworn, R.; Khunkitti, P.; Kruesubthaworn, A.; Siritaratiwat, A.; Surawanitkun, C. Distribution company investment cost reduction analysis with grid-connected solar PV allocation in power distribution system. *Int. J. Smart Grid Clean Energy* **2020**. [CrossRef]
34. Subramaniam, U.; Vavilapalli, S.; Padmanaban, S.; Blaabjerg, F.; Holm-Nielsen, J.B.; Almakhles, D. A Hybrid PV-Battery System for ON-Grid and OFF-Grid Applications—Controller-In-Loop Simulation Validation. *Energies* **2020**, *13*, 755. [CrossRef]
35. Yusoff, N.F.; Zakaria, N.Z.; Zainuddin, H.; Shaari, S. Mounting Configuration factor for building integrated photovol-taic and retrofitted grid-connected photovoltaic system. *Sci. Lett.* **2017**, *11*, 1–6.
36. Watts, D.; Valdés, M.F.; Jara, D.; Watson, A. Potential residential PV development in Chile: The effect of Net Metering and Net Billing schemes for grid-connected PV systems. *Renew. Sustain. Energy Rev.* **2015**, *41*, 1037–1051. [CrossRef]
37. Humada, A.M.; Aaref, A.M.; Hamada, H.M.; Sulaiman, M.H.; Amin, N.; Mekhilef, S. Modeling and characterization of a grid-connected photovoltaic system under tropical climate conditions. *Renew. Sustain. Energy Rev.* **2017**, *82*, 2094–2105. [CrossRef]
38. Mekhilef, S.; Barimani, M.; Safari, A.; Salam, Z. Malaysia's renewable energy policies and programs with green aspects. *Renew. Sustain. Energy Rev.* **2014**, *40*, 497–504. [CrossRef]
39. Mansur, T.M.N.T.; Baharudin, N.H.; Ali, R. A Comparative Study for Different Sizing of Solar PV System under Net Energy Metering Scheme at University Buildings. *Bull. Electr. Eng. Inform.* **2018**, *7*, 450–457. [CrossRef]
40. Irwan, Y.; Amelia, A.; Irwanto, M.; Fareq, M.; Leow, W.; Gomesh, N.S.I. Stand-Alone Photovoltaic (SAPV) System Assessment using PVSYST Software. *Energy Procedia* **2015**, *79*, 596–603. [CrossRef]
41. Gharakhani Siraki, A.; Pillay, P.D. *Comparison of PV System Design Software Packages for Urban Applications*; IEEE: Piscataway, NJ, USA, 2010.
42. NASA-SSE Satellite Data. Available online: http://eosweb.larc.nasa.gov/sse/ (accessed on 10 October 2021).

entropy

MDPI

Article

Event-Triggered Fixed-Time Integral Sliding Mode Control for Nonlinear Multi-Agent Systems with Disturbances

Xue Li, Zhiyong Yu * and Haijun Jiang

College of Mathematics and System Sciences, Xinjiang University, Urumqi 830046, China; lixuejiayouya@163.com (X.L.); jianghai@xju.edu.cn (H.J.)
* Correspondence: yzygsts@163.com or yzygsts@xju.edu.cn

Abstract: In this paper, the leader-following consensus problem of first-order nonlinear multi-agent systems (FONMASs) with external disturbances is studied. Firstly, a novel distributed fixed-time sliding mode manifold is designed and a new static event-triggered protocol over general directed graph is proposed which can well suppress the external disturbances and make the FONMASs achieve leader-following consensus in fixed-time. Based on fixed-time stability theory and inequality technique, the conditions to be satisfied by the control parameters are obtained and the Zeno behavior can be avoided. In addition, we improve the proposed protocol and propose a new event-triggering strategy for the FONMASs with multiple leaders. The systems can reach the sliding mode surface and achieve containment control in fixed-time if the control parameters are designed carefully. Finally, several numerical simulations are given to show the effectiveness of the proposed protocols.

Keywords: multi-agent systems; sliding mode control; leader-following consensus; fixed-time

Citation: Li, X.; Yu, Z.; Jiang, H. Event-Triggered Fixed-Time Integral Sliding Mode Control for Nonlinear Multi-Agent Systems with Disturbances. *Entropy* **2021**, *23*, 1412. https://doi.org/10.3390/e23111412

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 29 September 2021
Accepted: 25 October 2021
Published: 27 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the past several years, more and more researchers are interested in cooperative control of multi-agent systems (MASs) because of its robustness, flexible deployment and high efficiency. Cooperative control is widely used in various research fields to solve engineering and non-engineering problems, such as formation of robots [1], sensor networks [2], attitude alignment [3] and so on. Among multitudinous cooperative control objectives, consensus is a basic problem in MASs. Its purpose is to design a controller which can ensure that all members agree on an interest signal according to local information. Therefore, the information exchange between agents on the shared network is regulated by the consensus algorithm or protocol.

Based on observation of nature, the emergence of leaders in animal groups led to the development of the leader-following problem in collective behavior of MASs. In the distributed consensus problem, the existing results of MASs can be roughly divided into three categories according to the number of leaders: leaderless consensus [4–6], leader-following consensus [7–9] and containment control of multiple leaders [10,11]. In [4], the leaderless consensus of discrete-time MASs was studied by considering the connectivity of the network. In [5], the leaderless consensus of model-independent MASs was considered. In [6], the leaderless consensus of fractional-order MASs was investigated. In the case of single leader, the leader-following bipartite consensus problem was investigated for linear MASs in [7]. The leader-following consensus for MASs with Lipshitz-type node dynamics was considered in [8]. Furthermore, by using distributed impulsive control method, the authors studied the leader-following consensus of nonlinear MASs in [9]. In the case of multiple leaders, the reduplicative learning control problem for nonlinear heterogeneous MASs was investigated in [10]. In [11], a completely distributed control protocol was proposed to study the time-varying group formation tracking problem for linear MASs with multiple leaders.

In the consensus analysis of MASs, the convergence rate is an important index to evaluate the effectiveness of the proposed protocol. Most of existing results mainly concerned with the asymptotic convergence of the system. Due to the rapid development of finite-time theory, some researchers developed the finite-time consensus protocols [12–14]. In [12], the authors investigated the practical finite-time consensus of second-order heterogeneous switched nonlinear MASs. In [13], the authors investigated the distributed finite-time tracking control problem for second-order MASs, and proposed a novel observer-based control algorithm. In [14], the finite-time control law for continuous FONMAS was proposed, which ensures that the obstacles in the way can be passed by all agents, and the relative position between two agents reaches a constant value in finite-time. In proposed finite-time protocols, the estimation of convergence time depends on the initial values of MASs. To overcome this shortcoming, the researchers developed the fixed-time consensus protocols. In [15], the fixed-time leader-following flocking for second-order MASs was studied. For fixed-time consensus of heterogeneous MASs, the protocol based on neighbors' states was proposed in [16], and the state observation control protocol was designed in [17].

Compared with the traditional continuous control, the sampling control can effectively reduce the communication update frequency, so as to reduce the control cost. Therefore, the sampled-data control method was applied to study the issue of resilient reliable dissipativity performance index for systems including actuator faults and probabilistic time-delay signals in [18]. However, the traditional sampled-data control method is time-dependent, which requires the controller to be updated regularly even if the control goal is achieved. This method also leads to some unnecessary waste of computational and communication resources. The event-triggered protocol is used as an efficient method to further reduce communication and computing load. Therefore, the event-triggered controller was designed to study the consensus of first-order MASs in [19]. The consensus of linear MASs was studied via observer-based event-triggered control and two novel schemes were proposed in [20]. The containment control of second-order nonlinear MASs was considered based on event-triggered method in [21]. The consensus problem for a kind of stochastic MASs was studied and an adaptive output feedback approach based on event-triggered was proposed in [22]. To our knowledge, there is little research on the fixed-time consensus of MASs under event-triggered control protocol.

Most of the work mentioned above considers the ideal environment, but agents may face various disturbances signals or noise in communication. As pointed out in [23], disturbances signals or noise can destroy some good properties of a system. Therefore, the consensus of MASs under imperfect environment is worth considering. In [24], the consensus problem for linear MASs with the heterogeneous disturbances generated by the Brown motion was investigated. In [25], the authors investigated the distributed finite-time optimization issue for second-order MASs with matched interferences. By using disturbance rejection strategy, the event-triggered output consensus for MASs with time-varying disturbances was considered in [26]. Furthermore, the fixed-time event-triggered consensus for high-order and second-order MASs with uncertain disturbances was studied in [27,28], respectively.

Since the sliding mode technology can achieve a fast convergence rate to suppress disturbances, sliding mode control (SMC) method is widely used in the control of MASs with disturbances. In [29], a sliding mode estimator was given to accomplish distributed consensus for MASs. The sliding mode controllers were proposed for second-order MASs with mismatched uncertainties in [30]. The adaptive SMC protocols were designed to study the consensus of MASs with unknown disturbances in [31]. In order to improve the convergence rate, the finite-time SMC protocol was proposed in [32]. Furthermore, by using integral terminal SMC, the fixed-time consensus tracking issue for second-order MASs was investigated under the influence of interference signals in [33]. In these works [29–33], all control protocols were continuously updated. In order to reduce the control costs, considering the external interference, the event-triggered integral SMC was proposed to study the time-varying formation control of high-order MASs in [34]. Moreover, the event-

triggered finite-time consensus for multirobot systems with disturbances was considered via integral SMC strategy in [35]. The finite-time consensus for nonholonomic MASs with disturbances was studied by using event-triggered integral SMC method in [36]. However, in the existing research, the fixed-time event-triggered integral SMC method for single leader and multiple leaders of MASs with disturbances are rarely considered.

Inspired by the above considerations, this paper studies the fixed-time consensus problem for FONMASs with single leader and multiple leaders by using integral SMC and the theory of fixed-time stability. Firstly, to study the fixed-time leader-following consensus of FONMASs with external disturbances, a new event-triggered integral SMC protocol is devised for each agent. We show that both the systems can get to the sliding mode surface and all agents will achieve consensus in fixed-time under the proposed protocol. Moreover, the FONMAS with multiple leaders and external disturbances is considered. By generalizing our proposed event-triggered integral SMC protocol, it is proved that the containment control can also be achieved and the disturbance can be effectively suppressed in fixed-time. Compared with the existing works, the main contributions of the paper are at least the following three points:

1. In existing works [27,28], the disturbance rejection method was applied to study the fixed-time consensus of MASs with external disturbances. However, the integral SMC method combination with event-triggered control mechanism are introduced to design the distributed protocol in this paper, which can effectively suppress the disturbances and achieve consensus in fixed-time.
2. In [35,36], the finite-time event-triggered integral SMC protocols were proposed, in which the estimation of settling time was associated with the initial conditions. To overcome this disadvantage, the fixed-time event-triggered integral SMC protocols are proposed in this paper. According to the stability theory of fixed-time, we can prove that the consensus can be reached in fixed-time and the upper bound estimation of settling time is regardless of the initial conditions of MASs.
3. The containment control for FONMASs with multiple leaders is also considered, in which a generalized event-triggered integral SMC protocol is designed and the controller is updated only at some discrete instants. The sliding surface and the containment control can be reached in fixed-time. The Zeno phenomenon can be avoided.

The remainder of this paper is organized as follows. Section 2 introduces some preliminaries including graph theory, definitions, lemmas and problem formulation. In Section 3, the consensus protocols based on SMC technique are proposed, and some theorems are proved. In Section 4, the effectiveness of the proposed control protocols is verified by numerical simulations. Some conclusions are given in Section 5.

Notations. In this paper, $\mathbf{R}^n$ denotes the n-dimensional Euclidean space. I_n denotes n dimensional unit matrix. For $q = [q_1, q_2, \cdots, q_N]^T$, $\|q\|_1$, $\|q\|$ and $\|q\|_\infty$ represent the 1-norm, 2-norm and ∞-norm of q, respectively. $\mathrm{sgn}(q) = [\mathrm{sgn}(q_1), \cdots, \mathrm{sgn}(q_N)]^T$, $\mathrm{sig}^\alpha(q) = [|q_1|^\alpha \mathrm{sgn}(q_1), \cdots, |q_N|^\alpha \mathrm{sgn}(q_N)]^T$ where $\alpha > 0$ is a constant, $\mathrm{sgn}(\cdot)$ represents the sign function. For a matrix $A \in R^{N\times N}$, let A^T represent its transpose, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ represent the maximum eigenvalue and minimum eigenvalue of A, respectively. The symbol $\otimes$ denotes the Kronecker product of matrices. $\mathrm{diag}(\cdot)$ represents the diagonal matrix.

2. Preliminaries

2.1. Algebraic Graph Theory

A graph that consists of N nodes is represented by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ where $\mathcal{V} = \{v_1, \cdots, v_N\}$ is the set of nodes, $\mathcal{E}$ denotes the edges set, in which $(i, j) \in \mathcal{E}$ if there exists an edge between v_i and v_j. The weighted adjacency matrix is denoted as $\mathcal{A} = [a_{ij}] \in \mathbf{R}^{N\times N}$, where $a_{ij} > 0$ if $(j, i) \in \mathcal{E}$, and $a_{ij} = 0$, otherwise. The set of neighbors of agent i is denoted by $N_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$. The graph $\mathcal{G}$ is called directed and strongly connected if there exists a directed path between each pair of nodes. The graph $\mathcal{G}$ contains a directed

spanning tree if there exists at least one root. The Laplacian matrix $L = [l_{ij}]_{N\times N}$ is defined by $l_{ij} = -a_{ij}$ for $i \neq j$, and $l_{ii} = \sum_{j\neq i}^{N} a_{ij}$.

2.2. Definitions and Lemmas

Consider the following differential equation

$$\dot{x}(t) = f(x(t)), \qquad x(0) = x_0, \tag{1}$$

where $x(t) \in \mathbf{R}^n$, and $f : \mathbf{R}^n \mapsto \mathbf{R}^n$ is a nonlinear function with $f(0) = 0$. The following definitions and lemmas are given.

Definition 1 ([37]). *For any solution $x(t, t_0, x_0)$ of system (1), if there exists a positive number $T(x_0)$ such that $x(t, t_0, x_0) = 0$ for all $t \geq t_0 + T(x_0)$, then the solution $x = 0$ is said to be globally uniformly finite-time stable. $T(x_0)$ is called the settling time. Moreover, $x = 0$ is said to be globally fixed-time stable if $T(x_0)$ is independent of the initial value x_0.*

Lemma 1 ([38]). *For system (1), if there is a regular, positive definite and radially unbounded function $W(x) : \mathbf{R}^n \to \mathbf{R}$ such that any solution of (1) satisfies the inequality*

$$\dot{W}(x(t)) \leq -(\tau W^p(x(t)) + \phi W^q(x(t)))^{\varrho}, \;\; x(t) \in \mathbf{R}^n \setminus 0,$$

where $\tau, \phi, p, \varrho > 0$, $q \geq 0$, $p\varrho > 1$, $q\varrho < 1$, then solution $x = 0$ of system (1) is fixed-time stable, and the settling time $T(x_0)$ is estimated by

$$T(x_0) \leq \frac{1}{\phi^{\varrho}}(\frac{\phi}{\tau})^{\frac{1-q\varrho}{p-q}}(\frac{1}{1-q\varrho} + \frac{1}{p\varrho - 1}).$$

2.3. Problem Formulation

Consider a FONMAS consisting of N followers and a virtual leader indexed by $i = 1, 2, \cdots, N$ and 0, respectively. The dynamics is described by

$$\begin{aligned} \dot{x}_i(t) &= f(x_i(t)) + u_i(t) + w_i(t), \qquad i = 1, 2, \cdots, N, \\ \dot{x}_0(t) &= f(x_0(t)) + u_0(t), \end{aligned} \tag{2}$$

where $x_i(t) \in \mathbf{R}^n$, $u_i(t) \in \mathbf{R}^n$ and $w_i(t) \in \mathbf{R}^n$ are the state, the bounded control input and the external disturbance of the ith agent, respectively. $f(x_i(t))$ is a nonlinear function, which represents the inherent dynamics. In addition, we assume that the external disturbance is bounded, which satisfies $\|w_i(t)\|_\infty \leq D < \infty$, for $D > 0$. $x_0(t) \in \mathbf{R}^n$ and $u_0(t) \in \mathbf{R}^n$ are the state and the bounded control input of the leader, respectively. $f(x_0(t))$ is a nonlinear function, which also represents the inherent dynamics.

The communication topology among followers is expressed as directed graph $\hat{G}$, and the corresponding Laplacian matrix is described by the weighted matrix $\hat{L}$. We use b_i to represent the communication weight between the leader and the ith agent, in which $b_i > 0$ if the ith agent can receive information from the leader, $b_i = 0$ otherwise. In addition, we denote $B = \mathrm{diag}(b_0, \cdots, b_N)$.

Definition 2. *For the FONMAS (2), the fixed-time leader-following consensus is achieved for any initial conditions, if the following equations hold*

$$\lim_{t\to\mathcal{T}} \|x_i(t) - x_0(t)\| = 0, \qquad \|x_i(t) - x_0(t)\| \equiv 0, \;\; t \geq \mathcal{T}, \;\; i = 1, 2, \cdots, N,$$

where $\mathcal{T} > 0$ is called the settling time.

Assumption 1. *For the nonlinear function $f(\cdot)$, there exists a positive constant $l_1 > 0$ such that*

$$\|f(z_1(t)) - f(z_2(t))\| \leq l_1\|z_1(t) - z_2(t)\|, \tag{3}$$

where $z_1(t), z_2(t) \in \mathbf{R}^n$.

Assumption 2. *The communication between the leader and all followers is represented by graph G which contains a directed spanning tree with the leader as the root. In addition, the communication topology $\hat{G}$ is directed.*

3. Main Results

3.1. Fixed-Time Consensus with Single Leader

In this section, in order to achieve consensus between leader and followers, the integral SMC protocol will be designed for FONMAS described by (2). Before moving on, we define the following error variables

$$\begin{aligned} &\widetilde{x}_i(t) = x_i(t) - x_0(t), \\ &\widetilde{u}_i(t) = u_i(t) - u_0(t), \qquad i = 1, 2, \cdots, N. \end{aligned} \tag{4}$$

Since the disturbances exist in the follower agent dynamics, the integral SMC technique is applied. Then, we define the following integral type sliding mode variable

$$\sigma_i(t) = \widetilde{x}_i(t) - \int_0^t (\chi_i^{\eta}(s) + \mathrm{sgn}(\chi_i(s)))ds, \qquad i = 1, 2, \cdots, N, \tag{5}$$

where $\sigma_i(t) = [\sigma_{i1}(t), \sigma_{i2}(t), \cdots, \sigma_{in}(t)]^T$, $\chi_i(t) = -[\sum_{j\in N_i} a_{ij}(\widetilde{x}_i(t) - \widetilde{x}_j(t)) + b_i(\widetilde{x}_i(t))]$, and $\mathrm{sgn}(\chi_i(t)) = [\mathrm{sgn}(\chi_{i1}(t)), \mathrm{sgn}(\chi_{i2}(t)), \cdots, \mathrm{sgn}(\chi_{in}(t))]^T$. η is the ratio of two positive odd numbers and $\eta > 1$. When the sliding mode surface is reached, $\sigma_i(t) = 0$ and $\dot{\sigma}_i(t) = 0$. Hence, it has

$$\dot{\widetilde{x}}_i(t) = \chi_i^{\eta}(t) + \mathrm{sgn}(\chi_i(t)), \qquad i = 1, 2, \cdots, N. \tag{6}$$

In order to reduce the control cost and increase the rate of convergence, the event-triggered consensus protocol is designed as follows

$$\begin{aligned} \widetilde{u}_i(t) = {} & \chi_i^{\eta}(t_k^i) + \mathrm{sgn}(\chi_i(t_k^i)) - K\mathrm{sgn}(\sigma_i(t_k^i)) - K_3\mathrm{sig}^{\beta+1}(\sigma_i(t_k^i)) \\ & - K_4\|\widetilde{x}_i(t_k^i)\|\mathrm{sgn}(\sigma_i(t_k^i)), \quad t \in [t_k^i, t_{k+1}^i), \end{aligned} \tag{7}$$

where $\beta > 0$, $K = K_1 + K_2$, K_1, K_2, K_3, K_4 are constants to be determined. t_k^i is the triggering instant. Then, the novel measurement error is designed as

$$\begin{aligned} e_i(t) = {} & \chi_i^{\eta}(t_k^i) + \mathrm{sgn}(\chi_i(t_k^i)) - K\mathrm{sgn}(\sigma_i(t_k^i)) - K_3\mathrm{sig}^{\beta+1}(\sigma_i(t_k^i)) \\ & - K_4\|\widetilde{x}_i(t_k^i)\|\mathrm{sgn}(\sigma_i(t_k^i)) - \Big(\chi_i^{\eta}(t) + \mathrm{sgn}(\chi_i(t)) - K\mathrm{sgn}(\sigma_i(t)) \\ & - K_3\mathrm{sig}^{\beta+1}(\sigma_i(t)) - K_4\|\widetilde{x}_i(t)\|\mathrm{sgn}(\sigma_i(t))\Big). \end{aligned} \tag{8}$$

In this paper, a distributed event-triggered sampling control is proposed. The trigger instant of each agent only depends on its trigger function. Based on the zero order hold, the control input is a constant in each trigger interval. In order to make FONMAS (2) achieve leader-following consensus under the proposed protocol (7), the following theorem is given.

Theorem 1. *Suppose that Assumptions 1 and 2 hold for the FONMAS (2). Under the protocol (7), the leader-following consensus can be achieved in fixed-time, if the following conditions are satisfied*

$$K_1 \geq D, K_2 > \max_{1\leq i\leq N}\{\xi_i\}, K_3 > 0, K_4 \geq l_1, \tag{9}$$

where $\xi_i > 0$ *for* $i = 1, 2, \cdots, N$. *The triggering condition is defined as*

$$t^i_{k+1} = \inf\left\{t > t^i_k \mid \|e_i(t)\| - \xi_i \geq 0\right\}, \ \ i = 1, 2, \cdots, N. \tag{10}$$

Proof. Firstly, we prove that the sliding mode surface $\sigma_i(t) = \dot{\sigma}_i(t) = 0$ for $i = 1, 2, \cdots, N$ can be achieved in fixed-time. Consider the Lyapunov function as

$$V_i(t) = \frac{1}{2}\sigma_i^T(t)\sigma_i(t), \qquad i = 1, 2, \cdots, N. \tag{11}$$

Take the time derivative of $V_i(t)$ for $t \in [t^i_k, t^i_{k+1})$, we have

$$\begin{aligned}
\dot{V}_i(t) &= \sigma_i^T(t)\dot{\sigma}_i(t) \\
&= \sigma_i^T(t)(\dot{\tilde{x}}_i(t) - \chi_i^{\eta}(t) - \mathrm{sgn}(\chi_i(t))) \\
&= \sigma_i^T(t)(\dot{x}_i(t) - \dot{x}_0(t) - \chi_i^{\eta}(t) - \mathrm{sgn}(\chi_i(t))) \\
&= \sigma_i^T(t)(f(x_i(t)) + u_i(t) + w_i(t) - f(x_0(t)) - u_0(t) - \chi_i^{\eta}(t) - \mathrm{sgn}(\chi_i(t))) \\
&= \sigma_i^T(t)(f(x_i(t)) - f(x_0(t)) + \widetilde{u}_i(t) + w_i(t) - \chi_i^{\eta}(t) - \mathrm{sgn}(\chi_i(t))) \\
&= \sigma_i^T(t)(f(x_i(t)) - f(x_0(t)) + e_i(t) + w_i(t) - K\mathrm{sgn}(\sigma_i(t)) \\
&\quad - K_3\mathrm{sig}^{\beta+1}(\sigma_i(t)) - K_4\|\widetilde{x}_i(t)\|\mathrm{sgn}(\sigma_i(t))).
\end{aligned} \tag{12}$$

Based on Assumption 1, it has

$$\begin{aligned}
&\sigma_i^T(t)(f(x_i(t)) - f(x_0(t))) \leq \|\sigma_i(t)\|l_1\|x_i(t) - x_0(t)\| \leq l_1\|\sigma_i(t)\|\|\widetilde{x}_i(t)\|, \\
&\sigma_i^T(t)(w_i(t) - K_1\mathrm{sgn}(\sigma_i(t))) \leq D\|\sigma_i(t)\|_1 - K_1\|\sigma_i(t)\|_1.
\end{aligned}$$

Based on conditions (9), we can get

$$\dot{V}_i(t) \leq \|e_i(t)\|\|\sigma_i(t)\| - K_3\|\sigma_i(t)\|^{\beta+2} - K_2\|\sigma_i(t)\|. \tag{13}$$

According to triggering condition (10), we have

$$\begin{aligned}
\dot{V}_i(t) &\leq -(K_2 - \xi_i)\|\sigma_i(t)\| - K_3\|\sigma_i(t)\|^{\beta+2} \\
&= -(K_2 - \xi_i)(2V_i(t))^{\frac{1}{2}} - K_3(2V_i(t))^{\frac{\beta+2}{2}}.
\end{aligned} \tag{14}$$

The closed-loop system will get to the sliding mode surface in fixed-time, which can be obtained according to Lemma 1. The settling time can be computed as

$$T_i \leq \frac{1}{\sqrt{2}(K_2 - \xi_i)}\left(\frac{K_2 - \xi_i}{K_3 2^{\frac{\beta+1}{2}}}\right)^{\frac{1}{\beta+1}}(2 + \frac{2}{\beta}). \tag{15}$$

Define $T = \max_{1\leq i\leq N}\{T_i\}$. Then, it is proved that the sliding mode surface $\sigma_i(t) = 0$ can be reached for any $t > T$.

Secondly, we will prove that the leader-following consensus can be achieved in fixed-time. For convenience, $\chi_i(t)$ for $i = 1, 2, \cdots, N$ can be rewritten in the following compact form $\chi(t) = -((\hat{L} + B) \otimes I_n)\widetilde{x}(t)$ and $\|\mathrm{sgn}(\chi(t))\| \leq \sqrt{Nn}$.

Let $\hat{L} + B = H$. Based on Assumption 2, the matrix H is invertible and all eigenvalues have positive real parts. Therefore, there exists a positive symmetric matrix P such that

$Q = PH + H^T P > 0$. Define the Lyapunov function as $\tilde{V}(t) = \chi^T(t)(P \otimes I_n)\chi(t)$, then taking the time derivative of $\tilde{V}(t)$ for $t > T$ yields

$$\begin{aligned}\dot{\tilde{V}}(t) &= -2\chi^T(t)(P \otimes I_n)(H \otimes I_n)(\chi^\eta(t) + \mathrm{sgn}(\chi(t))) \\ &= -\chi^T(t)(Q \otimes I_n)\chi^\eta(t) - \chi^T(t)(Q \otimes I_n)\mathrm{sgn}(\chi(t)) \\ &\leq -\lambda_{\min}(Q)\|\chi\|^{\eta+1} - \lambda_{\min}(Q)\|\chi\| \\ &\leq -\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}\tilde{V}^{\frac{\eta+1}{2}}(t) - \frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}\tilde{V}^{\frac{1}{2}}(t). \end{aligned} \tag{16}$$

By Lemma 1, we can conclude that the closed-loop system will achieve consensus in fixed-time. The settling time can be computed as

$$\tilde{T} \leq T + \frac{\lambda_{\max}(P)}{\lambda_{\min}(Q)}(2 + \frac{2}{\eta - 1}). \tag{17}$$

□

Remark 1. *In this paper, the general directed network topology is considered, so the matrix H is asymmetric. We need to select the positive definite matrix P to make it symmetric. In particular, if the network topology $\hat{G}$ is undirected, the matrix P corresponds to the identity matrix, and the construction of Lyapunov function $\tilde{V}(t)$ can be simplified. This reduces the computational burden.*

Remark 2. *In [12–14], the finite-time consensus problem of MASs was studied. Compared with these literatures, we propose a fixed-time consensus protocol. Based on (17), we can find that the estimation of settling time is independent of initial values. In [15,16], the fixed-time consensus of MASs under ideal environment was studied. However, this paper considers a more complex environment in which agents of MASs are affected by external disturbances. We propose a new fixed-time consensus protocol based on integral sliding mode technique, which can suppress the disturbances better and improve the closed-loop performance of the system.*

Remark 3. *There are generally three methods to deal with disturbances, namely internal made method, disturbances observation and sliding mode control. In [27,28], the disturbance rejection method was applied to eliminate the influence of disturbances in the protocols. However, in this paper, we adopt the integral sliding mode technique combined with event-triggered to suppress disturbances. Our research enriches the design method of control protocol and theoretical results. In [35,36], although the consensus of FONMASs with external disturbances was discussed by using integral sliding mode technique, only finite-time convergence was analyzed, and the estimation of settling time related to the initial conditions of system. To overcome this disadvantage, this paper proposes a new fixed-time event-triggered integral SMC protocol, in which the sliding mode surface can be reached and the consensus can be achieved in fixed-time.*

Theorem 2. *Consider the FONMAS (2) with the event-triggered control protocol (7). If the triggering condition is defined by (10) and the conditions of Theorem 1 hold, then the Zeno behavior can be eliminated.*

Proof. The proof is divided into two parts, before and after reaching the sliding mode surface.

On the one hand, we show the Zeno behavior does not exist before the systems achieve the sliding mode surface. Through the analysis of Theorem 1, the sliding mode surface will be reached when $t > T$. Therefore, we need to eliminate the Zeno behavior in the closed interval $t \in [0, T]$. Since $\chi_i(t)$ is a continuous function, it must exist a maximum value. Define $\tau_i = \max_{0 \leq t \leq T}\{\|\chi_i^\eta(t)\|\}$ and $\phi_i = \max_{0 \leq t \leq T}\{\|\mathrm{diag}(\chi_i^{\eta-1}(t))\|\}$.

Take the time derivative of $\|e_i(t)\|$, it has

$$\begin{aligned}\frac{d}{dt}\|e_i(t)\| \leq & \left\| \frac{d}{dt}\left[\chi_i^{\eta}(t) + \mathrm{sgn}(\chi_i(t)) - K\mathrm{sgn}(\sigma_i(t)) - K_3\mathrm{sig}^{\beta+1}(\sigma_i(t))\right.\right. \\ & \left.\left. - K_4\|\widetilde{x}_i(t)\|\mathrm{sgn}(\sigma_i(t))\right]\right\| \\ \leq & \|\frac{d}{dt}\chi_i^{\eta}(t)\| + \|\frac{d}{dt}K_3\mathrm{sig}^{\beta+1}(\sigma_i(t))\| + \left\|\frac{d}{dt}(K_4\|\widetilde{x}_i(t)\|\mathrm{sgn}(\sigma_i(t)))\right\| \\ \leq & \eta\|\mathrm{diag}(\chi_i^{\eta-1}(t))\|\|\dot{\chi}_i(t)\| + K_3(\beta+1)\|\mathrm{diag}(\sigma_i^{\beta}(t))\|\|\dot{\sigma}_i(t)\| + K_4\sqrt{n}\|\dot{\widetilde{x}}_i(t)\| \\ \leq & \left[\eta\phi_i\sqrt{Nn}H_{ii} + K_3(\beta+1)\|\mathrm{diag}(\sigma_i^{\beta}(0))\| + K_4\sqrt{n}\right]\|\dot{\widetilde{x}}_i(t)\| + K_3(\beta+1)\tau_i \\ & \times \|\mathrm{diag}(\sigma_i^{\beta}(0))\| + K_3(\beta+1)\sqrt{n}\|\mathrm{diag}(\sigma_i^{\beta}(0))\| \\ \leq & R_i l_1\|\widetilde{x}_i(t)\| + R_i D + R_i\|\widetilde{u}_i(t)\| + K_3(\beta+1)\tau_i\|\mathrm{diag}(\sigma_i^{\beta}(0))\| \\ & + K_3(\beta+1)\sqrt{n}\|\mathrm{diag}(\sigma_i^{\beta}(0))\| \\ \leq & R_i l_1\bar{x}_i + R_i D + R_i\bar{u}_i + K_3(\beta+1)\tau_i\|\mathrm{diag}(\sigma_i^{\beta}(0))\| \\ & + K_3(\beta+1)\sqrt{n}\|\mathrm{diag}(\sigma_i^{\beta}(0))\|, \end{aligned} \tag{18}$$

where $R_i = \eta\phi_i\sqrt{Nn}H_{ii} + K_3(\beta+1)\|\mathrm{diag}(\sigma_i^{\beta}(0))\| + K_4\sqrt{n}$, H_{ii} is the element of i-th row and column of matrix H, $\bar{x}_i = \max_{0\leq t\leq T}\{\|\widetilde{x}_i(t)\|\}$ and $\bar{u}_i = \max_{0\leq t\leq T}\{\|\widetilde{u}_i(t)\|\}$. Combination with $e_i(t_k^i) = 0$, it yields

$$\begin{aligned}\|e_i(t)\| \leq & \left[R_i l_1\bar{x}_i + R_i D + R_i\bar{u}_i + K_3(\beta+1)\tau_i\|\mathrm{diag}(\sigma_i^{\beta}(0))\|\right. \\ & \left. + K_3(\beta+1)\sqrt{n}\|\mathrm{diag}(\sigma_i^{\beta}(0))\|\right](t - t_k^i). \end{aligned} \tag{19}$$

Using the triggering condition (10), the next trigger instant satisfies $\|e_i(t_{k+1}^i)\| = \xi_i$. Therefore,

$$\begin{aligned}\xi_i \leq & \left[R_i l_1\bar{x}_i + R_i D + R_i\bar{u}_i + K_3(\beta+1)\tau_i\|\mathrm{diag}(\sigma_i^{\beta}(0))\|\right. \\ & \left. + K_3(\beta+1)\sqrt{n}\|\mathrm{diag}(\sigma_i^{\beta}(0))\|\right](t_{k+1}^i - t_k^i). \end{aligned} \tag{20}$$

Denote $Q_{1i} = R_i l_1\bar{x}_i + R_i D + R_i\bar{u}_i + K_3(\beta+1)\tau_i\|\mathrm{diag}(\sigma_i^{\beta}(0))\| + K_3(\beta+1)\sqrt{n}\|\mathrm{diag}(\sigma_i^{\beta}(0))\|$, and $\Delta T_k^i = t_{k+1}^i - t_k^i$, we can get $\Delta T_k^i \geq \frac{\xi_i}{Q_{1i}} > 0$.

On the other hand, when the sliding mode surface is reached, $\sigma_i(t) = 0$. Similar to the above proof, we can obtain

$$\begin{aligned}\frac{d}{dt}\|e_i(t)\| \leq & \left\| \frac{d}{dt}\left[\chi_i^{\eta}(t) + \mathrm{sgn}(\chi_i(t)) - K\mathrm{sgn}(\sigma_i(t)) - K_3\mathrm{sig}^{\beta+1}(\sigma_i(t))\right.\right. \\ & \left.\left. - K_4\|\widetilde{x}_i(t)\|\mathrm{sgn}(\sigma_i(t))\right]\right\| \\ \leq & \eta\|\mathrm{diag}(\chi_i^{\eta-1}(t))\|\|\dot{\chi}_i(t)\| \\ \leq & \eta\phi_i\sqrt{Nn}H_{ii}(\frac{1}{\lambda_{\min}(P)})^{\frac{\eta}{2}}\tilde{V}(0)^{\frac{\eta}{2}} + \eta\phi_i NnH_{ii}. \end{aligned} \tag{21}$$

Combination with $e_i(t_k^i) = 0$, one has

$$\|e_i(t)\| \leq \left[\eta\phi_i\sqrt{Nn}H_{ii}(\frac{1}{\lambda_{\min}(P)})^{\frac{\eta}{2}}\tilde{V}(0)^{\frac{\eta}{2}} + \eta\phi_i NnH_{ii}\right](t - t_k^i). \tag{22}$$

Using the triggering condition (10), on can obtain

$$\Delta T_k^i \geq \frac{\xi_i}{Q_{2i}} > 0, \tag{23}$$

where $Q_{2i} = \eta\phi_i\sqrt{Nn}H_{ii}(\frac{1}{\lambda_{\min}(P)})^{\frac{\eta}{2}}\tilde{V}(0)^{\frac{\eta}{2}} + \eta\phi_i NnH_{ii}$, and $\Delta T_k^i = t_{k+1}^i - t_k^i$. Based on the above analysis, the Zeno behavior can be avoided in control process. □

Remark 4. *Since the trigger mechanism exists in the whole control process, the proof of Theorem 2 divided into two parts, i.e., before and after the system reaches the sliding mode surface. In this paper, a static distributed event-triggered strategy is developed. In order to reduce the number of triggers more effectively, we will consider the dynamic event-triggered control strategy in our future work.*

3.2. Fixed-Time Containment Control with Multiple Leaders

In this section, we consider the MASs with multiple leaders. The main aim is to make MASs realize containment control in fixed-time by designing appropriate control protocol. That means all follower agents' states converge to the convex combination of leaders' states in fixed-time. In particular, if the MASs has only one leader, the containment control will degenerate into leader-following consensus.

For the sake of generality, we hypothesize that the FONMAS consisting of N followers and M leaders indexed by indexed by $i = 1, \cdots, N$ and $j = N+1, \cdots, N+M$, respectively. The dynamics of the FONMAS is described by

$$\begin{aligned} \dot{x}_i(t) &= f(x_i(t)) + u_i(t) + w_i(t), \qquad i = 1, \cdots, N, \\ \dot{x}_j(t) &= f(x_j(t)) + u_j(t) + w_j(t), \qquad j = N+1, \cdots, N+M, \end{aligned} \tag{24}$$

where $x_i(t) \in \mathbf{R}^n$, $u_i(t) \in \mathbf{R}^n$ and $w_i(t) \in \mathbf{R}^n$ are the state, the bounded control input and the external disturbance of the ith agent, respectively. $f(x_i(t))$ is a nonlinear function which represents the inherent dynamics. $x_j(t) \in \mathbf{R}^n$, $u_j(t) \in \mathbf{R}^n$ and $w_j(t) \in \mathbf{R}^n$ are the state, the bounded control input and the internal disturbance of the jth leader, respectively. $f(x_j(t))$ is a nonlinear function, which also represents the inherent dynamics. In addition, we assume that the disturbances are bounded, which satisfy $\|w_i(t)\|_\infty \leq B < \infty$, $\|w_j(t)\|_\infty \leq F < \infty$ for $B > 0$ and $F > 0$.

Assumption 3. *Suppose that the communication among the leaders and followers is represented by graph G. For each follower, there exists at least one leader that has a directed path to it.*

Assumption 4. *Given scalars $\rho_1, \rho_2, \cdots, \rho_M$, satisfying $\sum_{j=1}^M \rho_j = 1$ and $\rho_j \geq 0$. There exists a constant $l_2 > 0$ such that for $x_i(t), x_j(t) \in \mathbf{R}^n$,*

$$\|f(x_i(t)) - \sum_{j=1}^M \rho_j f(x_j(t))\| \leq l_2\|x_i(t) - \sum_{j=1}^M \rho_j x_j(t)\|.$$

Under Assumption 3, the Laplcain matrix of graph G is denoted by L, which can be decomposed into $L = \begin{bmatrix} L_1 & L_2 \\ 0 & 0 \end{bmatrix}$, where L_1 is a nonsingular matrix, $L_2 \in R^{N\times M}$ has at least one positive entry and $-L_1^{-1}L_2 1_{M\times 1} = 1_{N\times 1}$.

Before moving on, we define the following error variables

$$\begin{aligned}\widetilde{X}(t) &= (L_1 \otimes I_n)X_1(t) + (L_2 \otimes I_n)X_2(t),\\ \widetilde{U}(t) &= (L_1 \otimes I_n)U_1(t) + (L_2 \otimes I_n)U_2(t),\\ \widetilde{W}(t) &= (L_1 \otimes I_n)W_1(t) + (L_2 \otimes I_n)W_2(t),\end{aligned} \tag{25}$$

where

$\widetilde{X}(t) = [\widetilde{X}_1^T(t), \widetilde{X}_2^T(t), \cdots, \widetilde{X}_N^T(t)]^T$, $\widetilde{U}(t) = [\widetilde{U}_1^T(t), \widetilde{U}_2^T(t), \cdots, \widetilde{U}_N^T(t)]^T$, $\widetilde{W}(t) = [\widetilde{W}_1^T(t), \cdots, \widetilde{W}_N^T(t)]^T$,
$X_1(t) = [x_1^T(t), \cdots, x_N^T(t)]^T, X_2(t) = [x_{N+1}^T(t), \cdots, x_{N+M}^T(t)]^T$,
$U_1(t) = [u_1^T(t), u_2^T(t), \cdots, u_N^T(t)]^T, U_2(t) = [u_{N+1}^T(t), u_{N+2}^T(t), \cdots, u_{N+M}^T(t)]^T$,
$W_1(t) = [w_1^T(t), w_2^T(t), \cdots, w_N^T(t)]^T, W_2(t) = [w_{N+1}^T(t), w_{N+2}^T(t), \cdots, w_{N+M}^T(t)]^T$.

Combination with Assumption 3 and the property of Laplacian matrix L, we can easily obtain that the containment control is achieve in fixed-time if and only if there exists a $\mathcal{T} > 0$ such that $\lim_{t\to\mathcal{T}} \|\widetilde{X}(t)\| = 0$ and $\|\widetilde{X}(t)\| \equiv 0$ for $t > \mathcal{T}$.

Considering the disturbances in the system, the consensus protocol can employ sliding mode approach. The integral type sliding variable is defined as follows

$$\sigma_i(t) = \widetilde{X}_i(t) - \int_0^t (\tilde{\chi}_i^{\eta}(s) + \mathrm{sgn}(\tilde{\chi}_i(s)))ds, \tag{26}$$

where $\tilde{\chi}_i(t) = -\widetilde{X}_i(t)$, η is the ratio of two positive odd numbers and $\eta > 1$. The sliding mode manifold (26) is given by following comport form

$$\sigma(t) = \widetilde{X}(t) - \int_0^t (\tilde{\chi}^{\eta}(s) + \mathrm{sgn}(\tilde{\chi}(s)))ds. \tag{27}$$

When the sliding mode surface is reached, $\sigma(t) = 0$ and $\dot{\sigma}(t) = 0$. Hence, it has

$$\dot{\widetilde{X}}(t) = \tilde{\chi}^{\eta}(t) + \mathrm{sgn}(\tilde{\chi}(t)). \tag{28}$$

In order to reduce the control cost and increase the rate of convergence, the event-triggered sample-data control protocol is presented as

$$\begin{aligned}\widetilde{U}_i(t) =& \tilde{\chi}_i^{\eta}(t_k) + \mathrm{sgn}(\tilde{\chi}_i(t_k)) - K\mathrm{sgn}(\sigma_i(t_k)) - K_3\mathrm{sig}^{\beta+1}(\sigma_i(t_k))\\ &- K_4\|\widetilde{X}(t_k)\|\mathrm{sgn}(\sigma_i(t_k)), \qquad t \in [t_k, t_{k+1}),\end{aligned} \tag{29}$$

where $\beta > 0$, $K = K_1 + K_2$, K_1, K_2, K_3, K_4 are constants to be determined. t_k is the triggering instant.

Similarly, the controller (29) can be rewritten in the following comport form

$$\begin{aligned}\widetilde{U}(t) =& \tilde{\chi}^{\eta}(t_k) + \mathrm{sgn}(\tilde{\chi}(t_k)) - K\mathrm{sgn}(\sigma(t_k)) - K_3\mathrm{sig}^{\beta+1}(\sigma(t_k))\\ &- K_4\|\widetilde{X}(t_k)\|\mathrm{sgn}(\sigma(t_k)), \qquad t \in [t_k, t_{k+1}).\end{aligned} \tag{30}$$

Then, the novel measurement error for the system (24) is designed as

$$\begin{aligned}e(t) =& \tilde{\chi}^{\eta}(t_k) + \mathrm{sgn}(\tilde{\chi}(t_k)) - K\mathrm{sgn}(\sigma(t_k)) - K_3\mathrm{sig}^{\beta+1}(\sigma(t_k))\\ &- K_4\|\widetilde{X}(t_k)\|\mathrm{sgn}(\sigma(t_k)) - \Big(\tilde{\chi}^{\eta}(t) + \mathrm{sgn}(\tilde{\chi}(t)) - K\mathrm{sgn}(\sigma(t))\\ &- K_3\mathrm{sig}^{\beta+1}(\sigma(t)) - K_4\|\widetilde{X}(t)\|\mathrm{sgn}(\sigma(t))\Big).\end{aligned} \tag{31}$$

Theorem 3. *Suppose that Assumptions 3 and 4 hold for the FONMAS (24). Under the protocol (30), the containment control can be achieved in fixed-time, if the following inequalities are satisfied:*

$$K_1 \geq \|L_1\|B + \|L_2\|F, K_2 > \xi, K_3 > 0, K_4 \geq l_2\|L_1\|\|L_1^{-1}\|. \tag{32}$$

The triggering condition is defined as

$$t_{k+1} = \inf\{t > t_k \mid \|e(t)\| - \xi \geq 0\}, \tag{33}$$

where $\xi > 0$.

Proof. Consider the Lyapunov function as

$$V(t) = \frac{1}{2}\sigma^T(t)\sigma(t). \tag{34}$$

For $t \in [t_k, t_{k+1})$, the derivative of $V(t)$ is

$$\begin{aligned}
\dot{V}(t) =& \sigma^T(t)\dot{\sigma}(t) \\
=& \sigma^T(t)(\dot{\tilde{X}}(t) - \tilde{\chi}^{\eta}(t) - \mathrm{sgn}(\tilde{\chi}(t))) \\
=& \sigma^T(t)((L_1 \otimes I_n)F_1 + (L_2 \otimes I_n)F_2 + \widetilde{U}(t) + \widetilde{W}(t) - \tilde{\chi}^{\eta}(t) - \mathrm{sgn}(\tilde{\chi}(t))) \\
=& \sigma^T(t)((L_1 \otimes I_n)F_1 + (L_2 \otimes I_n)F_2 + e(t) + \widetilde{W}(t) - K\mathrm{sgn}(\sigma(t)) \\
& - K_3\mathrm{sig}^{\beta+1}(\sigma(t)) - K_4\|\widetilde{X}(t)\|\mathrm{sgn}(\sigma(t))).
\end{aligned} \tag{35}$$

Define $F_1 = [f^T(x_1(t)), \cdots, f^T(x_N(t))]^T$, $F_2 = [f^T(x_{N+1}(t)), \cdots, f^T(x_{N+M}(t))]^T$. Let $-L_1^{-1}L_2 \triangleq (\rho_1^T, \rho_2^T, \cdots, \rho_N^T)^T$, where $\rho_i = (\rho_{i1}, \cdots, \rho_{iM})$. From Assumption 4,

$$\begin{aligned}
&\|F_1 + (L_1^{-1}L_2 \otimes I_n)F_2\| \\
&= \left\| [(f(x_1(t)) - \sum_{j=1}^{M}\rho_{1j}f(x_j(t)))^T, \cdots, (f(x_N(t)) - \sum_{j=1}^{M}\rho_{Nj}f(x_j(t)))^T]^T \right\| \\
&= \left\| \left(\|(f(x_1(t)) - \sum_{j=1}^{M}\rho_{1j}f(x_j(t)))\|, \cdots, \|(f(x_N(t)) - \sum_{j=1}^{M}\rho_{Nj}f(x_j(t)))\| \right) \right\| \\
&\leq \left\| (l_2\|x_1(t) - \sum_{j=1}^{M}\rho_{1j}x_j(t)\|, \cdots, l_2\|x_N(t) - \sum_{j=1}^{M}\rho_{Nj}x_j(t))\| \right\| \\
&= l_2\|(L_1^{-1} \otimes I_n)\widetilde{X}(t)\| \leq l_2\|L_1^{-1}\|\|\widetilde{X}(t)\|,
\end{aligned}$$

$$\sigma^T(t)(\widetilde{W}(t) - K_1\mathrm{sgn}(\sigma(t))) \leq (\|L_1\|B + \|L_2\|F)\|\sigma(t)\|_1 - K_1\|\sigma(t)\|_1.$$

Based on (32), we can get

$$\dot{V}(t) \leq \|e(t)\|\|\sigma(t)\| - K_3\|\sigma(t)\|^{\beta+2} - K_2\|\sigma(t)\|. \tag{36}$$

According to (33), we have

$$\begin{aligned}
\dot{V}(t) &\leq -(K_2 - \xi)\|\sigma(t)\| - K_3\|\sigma(t)\|^{\beta+2} \\
&= -(K_2 - \xi)(2V(t))^{\frac{1}{2}} - K_3(2V(t))^{\frac{\beta+2}{2}}.
\end{aligned} \tag{37}$$

According to Lemma 1, the closed-loop system (24) will get to the sliding mode surface in fixed-time. The settling time can be estimated by

$$\bar{T} \leq \frac{1}{\sqrt{2}(K_2-\xi)}\left(\frac{K_2-\xi}{K_3 2^{\frac{\beta+1}{2}}}\right)^{\frac{1}{\beta+1}}(2+\frac{2}{\beta}). \tag{38}$$

Then, it is proved that $\sigma(t)=0$ is reached for $t>\bar{T}$.

Then, we will prove that the containment control can be achieved in fixed-time. Define the Lyapunov function as $\hat{V}(t)=\tilde{\chi}^T(t)\tilde{\chi}(t)$. Taking the time derivative of $\hat{V}(t)$ for $t>\bar{T}$ yields

$$\begin{aligned}\dot{\hat{V}}(t) &= -\tilde{\chi}^T(t)(\tilde{\chi}^\eta(t)+\mathrm{sgn}(\tilde{\chi}(t)))\\ &= -\|\tilde{\chi}(t)\|^{\eta+1}-\|\tilde{\chi}(t)\|_1\\ &\leq -\hat{V}^{\frac{\eta+1}{2}}(t)-\hat{V}^{\frac{1}{2}}(t).\end{aligned} \tag{39}$$

By Lemma 1, we can conclude that the closed-loop system will achieve containment control in fixed-time. The settling time can be computed as

$$\hat{T} \leq \bar{T}+(2+\frac{2}{\eta-1}). \tag{40}$$

The proof is finished. □

Remark 5. *In [27], the fixed-time consensus problem of MASs with nonlinear dynamics and indeterminate disturbances was considered based on event-triggered method. Compared with [27], we introduce the integral sliding mode technique to deal with disturbances, and consider the containment control problem in the case of multiple leaders. In addition, the event-triggered strategy applied in this paper can greatly save computation and communication resources.*

Theorem 4. *Consider the FONMAS (24) with the event-triggered control protocol (30). If the triggering condition is defined by (33) and all conditions of Theorem 3 are satisfied, then the Zeno behavior can be avoided.*

Proof. Similar to the proof of Theorem 2, the proof is divided into two parts.

First, we show that the Zeno behavior does not exist before the systems reach the sliding mode surface. Through the analysis of Theorem 3, we know that sliding mode surface will be reached when $t>\bar{T}$. Therefore, we need to eliminate the Zeno behavior in the closed interval $[0,\bar{T}]$. Since $\chi(t)$ is a continuous function, it must exist a maximum value. Define $\varepsilon=\max_{0\leq t\leq\bar{T}}\{\|\tilde{\chi}^\eta(t)\|\}$ and $\gamma=\max_{0\leq t\leq\bar{T}}\{\|\mathrm{diag}(\tilde{\chi}^{\eta-1}(t))\|\}$.

Take the time derivative of $\|e(t)\|$, we have

$$\begin{aligned}\frac{d}{dt}\|e(t)\| \leq& \left\|\frac{d}{dt}\Big[\tilde{\chi}^\eta(t)+\mathrm{sgn}(\tilde{\chi}(t))-K\mathrm{sgn}(\sigma(t))-K_3\mathrm{sig}^{\beta+1}(\sigma(t))\right.\\ &\left.-K_4\|\widetilde{X}(t)\|\mathrm{sgn}(\sigma(t))\Big]\right\|\\ \leq& \psi l_2\|L_1\|\|L_1^{-1}\|\bar{X}+\psi(\|L_1\|B+\|L_2\|F)+\psi\bar{U}+K_3(\beta+1)\varepsilon\\ &\times\|\mathrm{diag}(\sigma^\beta(0))\|+K_3(\beta+1)\sqrt{Nn}\|\mathrm{diag}(\sigma^\beta(0))\|,\end{aligned} \tag{41}$$

where $\psi = \eta\gamma + K_3(\beta+1)\|\mathrm{diag}(\sigma^\beta(0))\| + K_4\sqrt{Nn}$, $\bar{X} = \max_{0\le t\le\bar{T}}\{\|\widetilde{X}(t)\|\}$ and $\bar{U} = \max_{0\le t\le\bar{T}}\{\|\widetilde{U}(t)\|\}$. Based on $e(t_k) = 0$, it has

$$\|e(t)\| \le \Big[\psi l_2\|L_1\|\|L_1^{-1}\|\bar{X} + \psi(\|L_1\|B + \|L_2\|F) + \psi\bar{U} + K_3(\beta+1)\varepsilon\|\mathrm{diag}(\sigma^\beta(0))\| + K_3(\beta+1)\sqrt{Nn}\|\mathrm{diag}(\sigma^\beta(0))\|\Big](t-t_k). \tag{42}$$

Applying the triggering mechanism (33), it has $\|e(t_{k+1})\| = \xi$. Therefore,

$$\xi \le \Big[\psi l_2\|L_1\|\|L_1^{-1}\|\bar{X} + \psi(\|L_1\|B + \|L_2\|F) + \psi\bar{U} + K_3(\beta+1)\varepsilon \times \|\mathrm{diag}(\sigma^\beta(0))\| + K_3(\beta+1)\sqrt{Nn}\|\mathrm{diag}(\sigma^\beta(0))\|\Big](t_{k+1}-t_k). \tag{43}$$

Denote $\pi_1 = \psi l_2\|L_1\|\|L_1^{-1}\|\bar{X} + \psi(\|L_1\|B + \|L_2\|F) + \psi\bar{U} + K_3(\beta+1)\varepsilon\|\mathrm{diag}(\sigma^\beta(0))\| + K_3(\beta+1)\sqrt{Nn}\|\mathrm{diag}(\sigma^\beta(0))\|$, and $\Delta T_k = (t_{k+1} - t_k)$, we can get $\Delta T_k \ge \frac{\xi}{\pi_1} > 0$.

Next, we prove that the Zeno behavior can be avoided when the sliding mode surface is reached. Similar to the above proof, we can obtain

$$\begin{aligned}\frac{d}{dt}\|e(t)\| \le& \Big\|\frac{d}{dt}\Big[\tilde{\chi}^\eta(t) + \mathrm{sgn}(\tilde{\chi}(t)) - K\mathrm{sgn}(\sigma(t)) - K_3\mathrm{sig}^{\beta+1}(\sigma(t)) \\ &- K_4\|\widetilde{X}(t)\|\mathrm{sgn}(\sigma(t))\Big]\Big\| \\ \le& \eta\gamma\hat{V}(0)^{\frac{\eta}{2}} + \eta\gamma\sqrt{Nn}.\end{aligned} \tag{44}$$

Combination with $e(t_k) = 0$, it yields

$$\|e(t)\| \le \Big[\eta\gamma\hat{V}(0)^{\frac{\eta}{2}} + \eta\gamma\sqrt{Nn}\Big](t-t_k). \tag{45}$$

When the event next event is triggered, it has $\|e(t)\| = \xi$. Therefore,

$$\xi \le \Big[\eta\gamma\hat{V}(0)^{\frac{\eta}{2}} + \eta\gamma\sqrt{Nn}\Big](t_{k+1}-t_k). \tag{46}$$

Let $\pi_2 = \eta\gamma\hat{V}(0)^{\frac{\eta}{2}} + \eta\gamma\sqrt{Nn}$ and $\Delta T_k = t_{k+1} - t_k$, we can get $\Delta T_k \ge \frac{\xi}{\pi_2} > 0$. Based on above analysis, the lower bound of event-triggered interval is positive, then Zeno phenomenon is eliminated in the whole control process. □

4. Numerical Example

In this section, two numerical examples are presented to demonstrate the effectiveness of the control protocols.

Example 1. *Consider the FONMAS (2) with one leader and four followers. Figure 1a shows the directed communication topology between the leader and all followers. Obviously, Assumption 2 is satisfied. The nonlinear function is defined as follows*

$$f(x_i(t)) = \begin{pmatrix} -x_{i1}(t) + 2g(x_{i1}(t)) - 1.2g(x_{i2}(t)) \\ -x_{i2}(t) + 1.2g(x_{i1}(t)) + 2g(x_{i2}(t)) \end{pmatrix},\ i = 0, 1, \cdots, 4.$$

where $g(x_{ij}(t)) = 0.5(|x_{ij}(t)+1| - |x_{ij}(t)-1|) + 0.01sgn(x_{ij}(t)), i = 0, 1, \cdots, 4, j = 1, 2$. *Then, Assumption 1 holds. The external disturbances are defined as* $w_{11}(t) = w_{12}(t) = 0.05\sin(t) + 0.1\cos(t)$, $w_{21}(t) = w_{22}(t) = 0.05\sin(t) + 0.1\cos(t)$, $w_{31}(t) = w_{32}(t) = 0.05\sin(t)$, *and*

$w_{41}(t) = w_{42}(t) = 0.1\cos(t)$. *It follows that* $\|w_i(t)\|_\infty \leq 0.2, i = 1,2,3,4$. *The control input of leader is* $u_{01}(t) = u_{02}(t) = 0.1\sin(t) + 0.1\cos(t)$. *We choose the controller parameters* $K_1 = 0.2$, $K_2 = 1.7$, $K_3 = 1.5$, $K_4 = 2$, $\eta = \frac{7}{5}$, $\beta = 1.5$, $\xi_i = 0.2$ *for* $i = 1,2,3,4$ *and implement the proposed control protocol (7). Through the analysis, all conditions (9) of Theorem 1 are satisfied.*

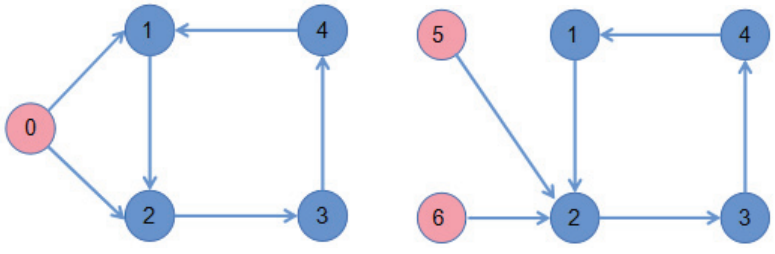


Figure 1. The network topology. (**a**) Topology with 1 leader. (**b**) Topology with 2 leaders.

The simulation results are presented in Figures 2–4. Specifically, Figure 2 depicts the states of all followers and the leader. It can be seen that all followers can track the leader's state in fixed-time under the proposed sliding mode control protocol (7) and the setting time is $\tilde{T} \leq 13.86$. Based on analysis of Theorem 1, the sliding mode variable $\sigma(t)$ converges to zero in fixed-time, and the setting time is $T \leq 2$, which is verified in Figure 3. The event-triggered instants under the triggering mechanism (10) are shown in Figure 4. It is shown that the event-triggered instants of each agent are different. Therefore, the results of Theorem 1 are feasible and the proposed sliding mode control protocol (7) can effectively suppress the external disturbances and realize leader-following consensus in fixed-time.

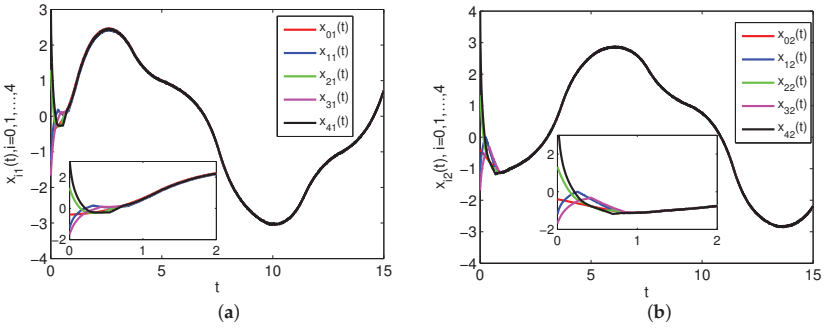


Figure 2. The states of $x_i(t)$, $i = 0,1,\cdots,4$. (**a**) The states of $-x_{i1}(t) + 2z(x_{i1}(t)) - 1.2z(x_{i2}(t))$; (**b**) The states of $-x_{i2}(t) + 1.2z(x_{i1}(t)) + 2z(x_{i2}(t))$.

Example 2. *Consider the FONMAS (24) with two leaders and four followers. The directed communication topology between two leaders and all followers are given in Figure 1b. The nonlinear function is defined as follows*

$$f(x_i(t)) = \begin{pmatrix} -x_{i1}(t) + 2z(x_{i1}(t)) - 1.2z(x_{i2}(t)) \\ -x_{i2}(t) + 1.2z(x_{i1}(t)) + 2z(x_{i2}(t)) \end{pmatrix}, \ i = 1,2,\cdots,6.$$

where $z(x_{ij}(t)) = 0.5(|x_{ij}(t)+1| - |x_{ij}(t)-1|) + 0.01sgn(x_{ij}(t)), i = 1,2,\cdots,6, j = 1,2$. *The external disturbances are defined as* $w_{11}(t) = w_{12}(t) = 0.05\sin(t) + 0.1\cos(t)$, $w_{21}(t) = w_{22}(t) = 0.05\sin(t) + 0.1\cos(t)$, $w_{31}(t) = w_{32}(t) = 0.05\sin(t)$, $w_{41}(t) = w_{42}(t) = 0.1\cos(t)$, $w_{51}(t) = w_{52}(t) = 0.03\sin(t) + 0.2\cos(t)$, *and* $w_{61}(t) = w_{62}(t) = 0.06\sin(t)$.

It has $\|w_i(t)\|_\infty \leq 0.2, i = 1,2,3,4$, $\|w_j(t)\|_\infty \leq 0.2, j = 5,6$. *The control input are* $u_{j1}(t) = u_{j2}(t) = 0.1\sin(t) + 0.1\cos(t)$, $j = 5,6$. *We choose the controller parameters* $K_1 = 1.7$, $K_2 = 2$, $K_3 = 1.5$, $K_4 = 24.5$, $\eta = \frac{7}{5}$, $\beta = 1.2$, $\xi = 1$ *and implement the proposed control protocol (30). Through simple calculation, we can verify that all conditions of Theorem 3 are satisfied.*

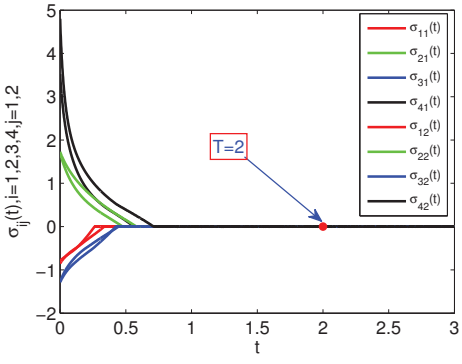


Figure 3. The state of $\sigma(t)$.

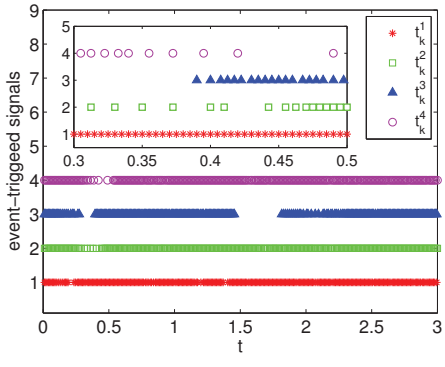


Figure 4. The event-triggered instants.

The simulation results are presented in Figures 5–7. Specifically, Figure 5 shows the states of four followers and two leaders. It can be seen that all followers' states gradually achieve consensus and fall into the convex hull of the leaders' states in fixed-time and the setting time is $\hat{T} \leq 11.3$. Figure 6 shows the evolution of sliding mode variable $\sigma(t)$. The sliding mode surface can be reached in fixed-time, and the settling time is $\bar{T} \leq 4.3$. The triggering interval under the event-triggered mechanism (33) is presented in Figure 7. In order to show the event-triggered intervals more clearly, we only give the simulation result for a short period of time, from which we can see that the Zeno phenomenon can be excluded. Different from the distributed event triggering condition (10), we employ a centralized trigger function, which also can ensure the reachability of the consensus. In particular, if the FONMAS (24) with one leader, the containment control can be reduced into leader-following tracking problem.

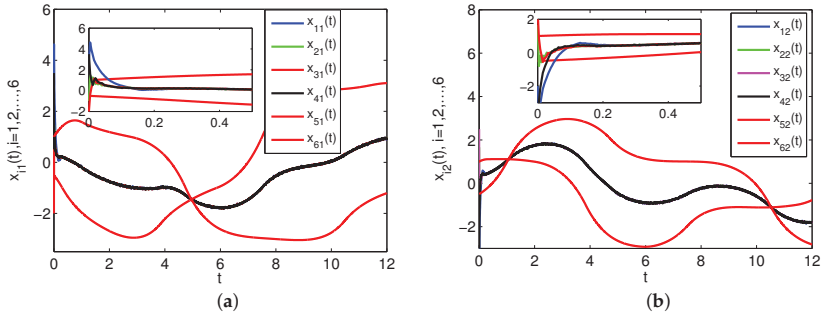


Figure 5. The states of $x_i(t)$, $i = 1, 2, \cdots, 6$. (**a**) The states of $-x_{i1}(t) + 2z(x_{i1}(t)) - 1.2z(x_{i2}(t))$; (**b**) The states of $-x_{i2}(t) + 1.2z(x_{i1}(t)) + 2z(x_{i2}(t))$.

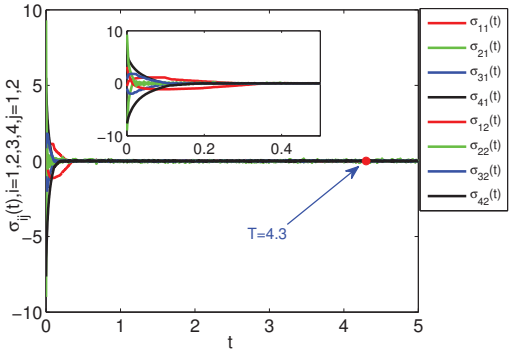


Figure 6. The state of $\sigma(t)$.

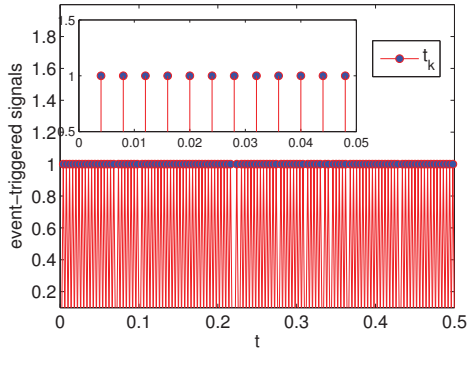


Figure 7. The event-triggered instants.

5. Conclusions

In this paper, considering external disturbances, the leader-following consensus and containment control of FONMASs are studied. Two kinds of event-triggered integral SMC protocols are designed, which can well suppress the external disturbances and make the FONMASs achieve consensus in fixed-time. Based on fixed-time stability theory and inequality technique, some criteria are obtained and the Zeno behavior can be avoided. The effectiveness of the proposed protocols are verified by several numerical simulations.

In the future work, the consensus of higher-order MASs with dynamic event-triggered communication mechanism will be considered.

Author Contributions: Conceptualization, X.L. and Z.Y.; methodology, X.L. and Z.Y.; software, X.L.; validation, X.L., Z.Y. and H.J.; formal analysis, X.L.; investigation, X.L.; resources, Z.Y.; data curation, Z.Y.; writing—original draft preparation, X.L.; writing—review and editing, Z.Y.; visualization, Z.Y.; supervision, H.J.; project administration, Z.Y.; funding acquisition, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (62003289), in part by the China Postdoctoral Science Foundation (2021M690400), in part by the Doctoral Foundation of Xinjiang University (BS180207), in part by the Tianshan Youth Program (2018Q068), and in part by the Tianshan Innovation Team Program (2020D14017).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MASs	Multi-agent systems
FONMASs	First-order nonlinear multi-agent systems
SMC	Sliding mode control

References

1. Wang, Z.; Wang, L.; Zhang, H.; Chen, Q.; Liu, J. Distributed regular polygon formation control and obstacle avoidance for non-holonomic wheeled mobile robots with directed communication topology. *IET Control. Theory Appl.* **2020**, *14*, 1113–1122. [CrossRef]
2. Halakarnimath, B.; Sutagundar, A. Multi-agent-based acoustic sensor node deployment in underwater acoustic wireless sensor networks. *J. Inf. Technol. Res.* **2020**, *13*, 136–155. [CrossRef]
3. Rezaee, H.; Abdollahi, F. Robust attitude alignment in multispacecraft systems with stochastic links failure. *Automatica* **2020**, *118*, 109033. [CrossRef]
4. Angeli, D.; Bliman, P. Stability of leaderless discrete-time multi-agent systems. *Math. Control Signals Syst.* **2006**, *18*, 293–322. [CrossRef]
5. Ren, W. Distributed leaderless consensus algorithms for networked Euler-Lagrange systems. *Int. J. Control* **2009**, *82*, 2137–2149. [CrossRef]
6. Bai, J.; Wen, G.; Rahmani, A. Leaderless consensus for the fractional-order nonlinear multi-agent systems under directed interaction topology. *Int. J. Syst. Sci.* **2018**, *49*, 954–963. [CrossRef]
7. Wen, G.; Wang, H.; Yu, X.; Yu, W. Bipartite tracking consensus of linear multi-agent systems with a dynamic leader. *IEEE Trans. Circuits Syst. II Express Briefs* **2017**, *65*, 1204–1208. [CrossRef]
8. Wen, G.; Duan, Z.; Li, Z.; Chen, G. Consensus tracking of nonlinear multi-agent systems with switching directed topologies. In Proceedings of the International Conference on Control Automation, Guangzhou, China, 5–7 December 2012; pp. 889–894.
9. He, W.; Chen, G.; Han, Q.; Qian, F. Network-based leader-following consensus of nonlinear multi-agent systems via distributed impulsive control. *Inf. Sci.* **2017**, *380*, 145–158.
10. Fu, Q. Iterative learning control for nonlinear heterogeneous multi-agent systems with multiple leaders. *Trans. Inst. Meas. Control.* **2020**, *43*, 854–861.
11. Hu, J.; Bhowmick, P.; Lanzon, A. Distributed adaptive time-varying group formation tracking for multiagent systems with multiple leaders on directed graphs. *IEEE Trans. Control Netw. Syst.* **2020**, *7*, 140–150.
12. Zou, W.; Shi, P.; Xiang, Z.; Shi, Y. Finite-time consensus of second-order switched nonlinear multi-agent systems. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 1757–1762. [CrossRef]
13. Zhao, Y.; Duan, Z.; Wen, G.; Zhang, Y. Distributed finite-time tracking control for multi-agent systems: An observer-based approach. *Syst. Control Lett.* **2013**, *62*, 22–28. [CrossRef]
14. Li, W.; Sun, S.; Xia, C. Finite-time stability of multi-agent system in disturbed environment. *Nonlinear Dyn.* **2012**, *67*, 2009–2016.
15. Xu, Z.; Liu, H.; Liu, Y. Fixed-time leader-following flocking for nonlinear second-order multi-agent systems. *IEEE Access* **2020**, *8*, 86262–86271. [CrossRef]

16. Zou, W.; Qian, K.; Xiang, Z. Fixed-time consensus for a class of heterogeneous nonlinear multiagent systems. *IEEE Trans. Circuits Syst. II Express Briefs* **2020**, *67*, 1279–1283. [CrossRef]
17. Sun, F.; Liu, P.; Li, H.; Zhu, W. Fixed-time consensus of heterogeneous multi-agent systems based on distributed observer. *Int. J. Syst. Sci.* **2021**, *52*, 1780–1789. [CrossRef]
18. Manivannan, R.; Samidurai, R.; Cao, J.; Perc, M. Design of resilient reliable dissipativity control for systems with actuator faults and probabilistic time-delay signals via sampled-data approach. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 4243–4255. [CrossRef]
19. Dimarogonas, D.; Frazzoli, E.; Johansson, K. Distributed event-triggered control for multi-agent systems. *IEEE Trans. Autom. Control* **2012**, *57*, 1291–1297. [CrossRef]
20. Zhang, H.; Feng, G.; Yan, H.; Chen, Q. Observer-based output feedback event-triggered control for consensus of multi-agent systems. *IEEE Trans. Ind. Electron.* **2014**, *61*, 4885–4894. [CrossRef]
21. Zou, W.; Xiang, Z. Event-triggered containment control of second-order nonlinear multi-agent systems. *J. Frankl. Inst.* **2019**, *356*, 10421–10438. [CrossRef]
22. Yang, Z.; Zheng, S.; Liang, B.; Xie, Y. Event-triggered finite-time consensus for stochastic multi-agent systems. *Trans. Inst. Meas. Control* **2020**, *43*, 1–10. [CrossRef]
23. Liu, L.; Perc, M.; Cao, J. Aperiodically intermittent stochastic stabilization via discrete time or delay feedback control. *Sci. China Inf. Sci.* **2019**, *62*, 072201. [CrossRef]
24. Wei, Q.; Wang, X.; Zhong, X.; Wu, N. Consensus control of leader-following multi-agent systems in directed topology with heterogeneous disturbances. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 423–431. [CrossRef]
25. Wang, X.; Wang, G. Distributed finite-time optimisation algorithm for second-order multi-agent systems subject to mismatched disturbances. *IET Control. Theory Appl.* **2020**, *14*, 2977–2988. [CrossRef]
26. Sun, J.; Yang, J.; Li, S.; Wang, X.; Li, G. Event-triggered output consensus disturbance rejection algorithm for multi-agent systems with time-varying disturbances. *J. Frankl. Inst.* **2020**, *357*, 12870–12885. [CrossRef]
27. Liu, J.; Yu, Y.; Sun, J.; Sun, C. Distributed event-triggered fixed-time consensus for leader-follower multiagent systems with nonlinear dynamics and uncertain disturbances. *Int. J. Robust Nonlinear Control* **2018**, *28*, 3543–3559. [CrossRef]
28. Zhou, D.; Zhang, A.; Yang, P. Fixed-time event-triggered consensus of second-order multi-agent systems with fully continuous communication free. *IET Control Theory Appl.* **2020**, *14*, 2385–2394. [CrossRef]
29. Bai, J.; Wen, G.; Rahmani, A.; Yu, Y. Consensus for the fractional-order double-integrator multi-agent systems based on the sliding mode estimator. *IET Control Theory Appl.* **2018**, *12*, 621–628. [CrossRef]
30. Wang, Q.; Sun, C.; Chat, X.; Yao, Y. Disturbance observer-based sliding mode control for multi-agent systems with mismatched uncertainties. *Assem. Autom.* **2018**, *38*, 606–614. [CrossRef]
31. Yu, Z.; Yu, S.; Jiang, H.; Hu, C. Distributed consensus for multi-agent systems via adaptive sliding mode control. *Int. J. Robust Nonlinear Control* **2021**, *31*, 7125–7151. [CrossRef]
32. Park, D.; Moon, J.; Han, S. Finite-time sliding mode controller design for formation control of multi-agent mobile robots. *J. Korea Robot. Soc.* **2017**, *12*, 339–349. [CrossRef]
33. Wang, C.; Wen, G.; Peng, Z.; Zhang, X. Integral sliding-mode fixed-time consensus tracking for second-order non-linear and time delay multi-agent systems. *J. Frankl. Inst.* **2019**, *35*, 3692–3710. [CrossRef]
34. Wang, J.; Xu, Y.; Xu, Y.; Yang, D. Time-varying formation for high-order multi-agent systems with external disturbances by event-triggered integral sliding mode control. *Appl. Math. Comput.* **2019**, *359*, 333–343. [CrossRef]
35. Nair, R.; Behera, L.; Kumar, S. Event-triggered finite-time integral sliding mode controller for consensus-based formation of multirobot systems with disturbances. *IEEE Trans. Control Syst. Technol.* **2017**, *27*, 39–47. [CrossRef]
36. Wang, J.; Zhang, Y.; Li, X.; Zhao, Y. Finite-time consensus for nonholonomic multi-agent systems with disturbances via event-triggered integral sliding mode controller. *J. Frankl. Inst.* **2020**, *357*, 7779–7795. [CrossRef]
37. Polyakov, A. Nonlinear feedback design for fixed-time stabilization of linear control systems. *IEEE Trans. Autom. Control* **2012**, *57*, 2106–2110. [CrossRef]
38. Yu, Z.; Yu, S.; Jiang, H.; Mei, X. Distributed fixed-time optimization for multiagent systems over a directed network. *Nonlinear Dyn.* **2021**, *103*, 775–789. [CrossRef]

sustainability

MDPI

Article

Research on Precipitation Forecast Based on LSTM–CP Combined Model

Yan Guo [1,2], Wei Tang [1,2], Guanghua Hou [1], Fei Pan [1], Yubo Wang [1] and Wei Wang [3,*]

[1] College of Information Engineering, Sichuan Agricultural University, Ya'an 625000, China; 14403@sicau.edu.cn (Y.G.); sau_tangwei@126.com (W.T.); 201803671@stu.sicau.edu.cn (G.H.); fei.pan@sicau.edu.cn (F.P.); 201902255@stu.sicau.edu.cn (Y.W.)
[2] Key Laboratory of Agricultural Information Engineering of Sichuan Province, Sichuan Agricultural University, Ya'an 625000, China
[3] College of Management, Sichuan Agricultural University, Ya'an 625000, China
* Correspondence: wangwei@sicau.edu.cn

Abstract: The tremendous progress made in the field of deep learning allows us to accurately predict precipitation and avoid major and long-term disruptions to the entire socio-economic system caused by floods. This paper presents an LSTM–CP combined model formed by the Long Short-Term Memory (LSTM) network and Chebyshev polynomial (CP) as applied to the precipitation forecast of Yibin City. Firstly, the data are fed into the LSTM network to extract the time-series features. Then, the sequence features obtained are input into the BP (Back Propagation) neural network with CP as the excitation function. Finally, the prediction results are obtained. By theoretical analysis and experimental comparison, the LSTM–CP combined model proposed in this paper has fewer parameters, shorter running time, and relatively smaller prediction error than the LSTM network. Meanwhile, compared with the SVR model, ARIMA model, and MLP model, the prediction accuracy of the LSTM–CP combination model is significantly improved, which can aid relevant departments in making disaster response measures in advance to reduce disaster losses and promote sustainable development by providing them data support.

Keywords: precipitation forecast; long short-term memory network; Chebyshev polynomial; BP neural network

Citation: Guo, Y.; Tang, W.; Hou, G.; Pan, F.; Wang, Y.; Wang, W. Research on Precipitation Forecast Based on LSTM–CP Combined Model. *Sustainability* **2021**, *13*, 11596. https://doi.org/10.3390/su132111596

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 26 September 2021
Accepted: 12 October 2021
Published: 20 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Disasters caused by natural hazards can often lead to significant and long-lasting disruptions of the whole socioeconomic system. One catastrophic event, such as a flood, can destroy multi-infrastructure systems, lead to cascading failures and substantial socioeconomic damages, and hinder development. A large amount of precipitation will directly lead to floods and waterlogging disasters and make crops impossible to harvest, as well as easily cause secondary disasters [1], such as collapses, landslides, mudslides, and waterlogging. The causes of precipitation are highly complex [2–4] due to the comprehensive influence of monsoons, topography, urban distribution, temperature, and evaporation, leading to more difficulties in predicting precipitation. In addition, rainfall also has some fixed characteristics, and its influencing factors, such as terrain, urban distribution, and temperature, will not change greatly in a short time. Precipitation also shows a high degree of regularity.

With the continuous progress of technology, artificial intelligence (AI) has become an important driving force in various fields, including sustainable development. Deep learning can improve the ability to deal with complex problems and help us increase our understanding of variables and sources that affect rainfall. At present, there is a myriad of existing studies on precipitation prediction, among which forecasts based on regression analysis and forecasts based on time series are two classic forecasting approaches.

Forecasts based on regression analysis mainly include autoregressive models, moving average models, autoregressive moving average models, and differential autoregressive moving average models [5]. Prediction methods based on time series can be mainly divided into grey systems [6], Markov models [7], and set pair analysis [8]. These methods are simple and widely used, but the accuracy of precipitation prediction is low, which cannot accurately describe the trend of precipitation development and change. With the rapid improvement of computer computing power and the development of big data [9], deep learning technology has become more and more widely used in recent years [10]. For one thing, deep learning is highly suitable for processing multi-dimensional and complex data, with no requirement for the physical modeling [11] of data; for another, deep learning has multiple levels, where low-level features are combined to form more abstract high-level features, and nonlinear network structure can achieve complex function approximation, showing powerful dataset representation capabilities. Therefore, using deep learning technology to predict precipitation has become a very practical value and challenging problem.

Among numerous deep learning technologies, BP and LSTM are two widely used deep learning neural networks [12,13]. The neural network has been put to work in many ways, including fitting, classification, and pattern recognition, since the BP algorithm was proposed [14]. For example, Ferreira et al. [15] evaluated the potential of deep learning and traditional machine learning models to predict daily reference evapotranspiration (seven days). The results show that the performance of the deep learning model is slightly better than that of the machine learning model. Granata et al. [16] established three models based on a recurrent neural network to predict short-term future actual evapotranspiration. The results show that the model based on deep learning can predict the actual evapotranspiration very accurately, but the performance of the model will be significantly affected by the local climate conditions. There is a myriad of improvements in BP neural networks made by researchers, one of which is to change the excitation function of the BP neural network. For example, Zhang et al. [17] took the sine function as the excitation function of the BP neural network. CP is a set of orthogonal polynomials that is often used for function approximation. Previous studies have shown that orthogonal polynomials perform better in fitting functions, and in comparison to ordinary polynomials [18–20], orthogonal polynomials have better fitting stability and fitting ability. CP already has a wide range of applications in neural networks. Zhang et al. [21,22] proposed a variety of neural network structures for classification, achieved by applying CP in a feedforward neural network and combing with the direct weight determination method, as well as the cross-validation method. Based on Zhang's research, Jin et al. [23,24] further improved the research as applied to wine region classification and breast cancer classification, respectively, and achieved good classification results. Unlike the BP neural network, the recurrent neural network (RNN) is a network dedicated to processing sequence data. The original RNN has poor processing capacity for sequence data due to its limited memory capacity, such that many improvements have been made on RNN by researchers. LSTM [25] is the most widely used network among many variants of RNN, with its ability to effectively alleviate the disadvantages of RNN, such as gradient disappearance and weak memory ability, making RNN widely applied in various fields. For example, Kratzert et al. [26] explored the potential of using a long-term and short-term memory network (LSTM) to simulate meteorological observation runoff, and verified by practice that its prediction accuracy is comparable to that of the perfect baseline hydrological model. Xiang et al. [27] used the prediction model based on LSTM and seq2seq structure to predict hourly rainfall runoff. The results show that the prediction accuracy of the LSTM-seq2seq model is higher than that of other models such as ordinary LSTM. This method is used to improve the accuracy of short-term flood prediction.

At present, researchers have applied the above two kinds of neural networks to the prediction of precipitation. The prediction approach of the neural network can effectively extract the random characteristics of a nonlinear sequence, which achieves a high prediction precision and has good research and application value. For example, according to the

meteorological data of Jingdezhen from 2008 to 2018, J. Kang et al. used the long-term and short-term memory neural network (LSTM) model to predict precipitation. The experimental results show that the LSTM model can be well applied to precipitation prediction [28]. Y. Zhou [29] used an improved BP neural network model to predict typhoon precipitation and typhoon precipitation events. By analyzing the difference in candidate predictors between normal years and years with a large prediction error, this method proposed a new predictor for the BP model in each iteration, and the precipitation prediction accuracy was better than that of the original BP neural network. In addition to predicting precipitation through deep learning methods, precipitation can also be predicted through satellite cloud images and radar detection. For example, Zahraei et al. [30] introduced a pixel algorithm for short-term quantitative precipitation forecasting (SQPF) using radar rainfall data, and proposed a pixel-based nowcasting (PBN) algorithm, which uses a hierarchical grid tracking algorithm. The image captures the high-resolution advection of storms in space and time. The results show that the proposed algorithm can effectively track and predict severe storm events in the next few hours. Bowler et al. [31] proposed a new Gandalf system precipitation prediction scheme based on advection. The method does not need to divide the radar analysis into continuous rain areas (CRA) and uses smoothing constraints to diagnose the block advection velocity in rainfall analysis by using the idea of optical flow. This scheme is compared with the old Gandolf advection scheme based on CRA, and the new scheme performs better in cases related to severe floods and in a continuous validation period of 3 months. Pham et al. [32] compared several advanced artificial intelligence (AI) models for predicting daily precipitation, and the results showed that support vector machine is the best method for predicting precipitation, and it was also found to be the most robust and effective prediction model. Banadkooki et al. [33] applied the flow pattern optimization algorithm (FRA) to the optimization of the multilayer perceptron neural network (MLP) and support vector regression (SVR), and established the precipitation prediction model. The results show that the performance of the proposed MLP-FRA model is better than all other models and has a stronger rainfall prediction ability. Wang et al. [34] combined satellite and radar observation data, and through proper orthogonal decomposition and assimilation of the data, the effect of precipitation forecasting was improved.

There have been many studies on precipitation prediction from the perspective of relevant studies at home and abroad, and an excellent application of LSTM in the prediction of sequence data has been achieved. However, the LSTM network structure is more complicated, and the number of network unit parameters is relatively large. A slight increase in the network depth will lead to a rapid increase in the number of parameters. The huge amount of parameters increases the difficulty of calculation. For medium and large datasets, higher performance equipment is required to perform calculations [35]. In addition, although the LSTM network overcomes the problem of gradient disappearance to a certain extent, the memory function of the LSTM network still depends on the long sequence. When the sequence is too long, the problem of gradient disappearance may still occur, which greatly affects the performance of LSTM [36], and the gradient vanishing problem has not been completely solved. At the same time, the LSTM network training model is more complicated and the training time is longer [37].

Given the above situation, this paper proposes to combine the Long Short-Term Memory (LSTM) [38] network and the Chebyshev polynomial (CP) [39], aiming to form an LSTM–CP combined model for rainfall prediction. From a theoretical point of view, this model combines CP and LSTM networks for the first time. Firstly, the LSTM network is used to extract the time-series features in the original data. Then, the BP (Back Propagation) neural network [40] with CP as the activation function is used to process the time-series features. This approach can effectively reduce the number of parameters, with the premise of ensuring accuracy, and has stronger characterization capabilities for sequence data, which provides a new idea for researchers in the field of neural networks. In the prediction of rainfall using a machine learning algorithm, the ARIMA model has low accuracy in predicting non-stationary or fluctuating time series [41]. The number of parameters in

the SVR model is usually very large [42]. The MLP network needs a large number of patterns and iterations to realize effective learning so it needs more execution time [43]. Compared with these classical machine learning algorithms, the LSTM–CP combined model proposed in this paper has higher accuracy, fewer parameters, and faster operation speed in the prediction of precipitation. The prediction results of the model in monthly units are relatively accurate, basically reflecting the changing trend of precipitation. It is helpful to provide a data reference for areas prone to floods and drought disasters, as well as help relevant departments to prepare in advance, reducing local economic losses. The model is capable of shortening the running time more effectively when dealing with large and medium-sized datasets as it can effectively reduce the use of parameters, making the process of sequence data more efficient.

This article is structured as follows: Introduction, where the importance and necessity of accurate precipitation forecasts are addressed and the existing precipitation forecasting methods and the existing problems are listed. The method section gives a detailed introduction to the related models and theoretical methods used, and compares and analyzes the parameters of different models. In the experimental evaluation section, the prediction models of LSTM, LSTM–BP, and LSTM–CP are constructed, respectively, and the parameter setting process of the LSTM–CP combined model is elaborated. The experimental results show that, compared with the ordinary LSTM neural network model, the LSTM–CP combined model proposed in this paper has fewer parameters, shorter running time, and relatively smaller prediction error than the LSTM network. At the same time, this paper also compares the LSTM–CP combined model with the traditional rainfall prediction SVR model, ARIMA model, and MLP model, finding that the prediction accuracy of the LSTM–CP combined model is significantly improved. Finally, the discussion of results and conclusions is presented, showing the ability to predict precipitation through the LSTM–CP combination model.

2. Materials and Methods

2.1. LSTM–CP Combined Model Framework

The LSTM network has inherent advantages in processing sequential data on account of its powerful memory [44–46]. In this paper, the LSTM network is used as the basic network of sequence data prediction, with the BP neural network combined to use its excellent function fitting ability. We can obtain an LSTM–CP combination model by using CP to improve BP neural networks. CP and LSTM networks are combined for the first time in this model, where the LSTM network is first used to extract the time-series features in the original data. Then, the BP (Back Propagation) neural network of CP as the activation function is used to process the time-series features, with the specific process shown in Figure 1.

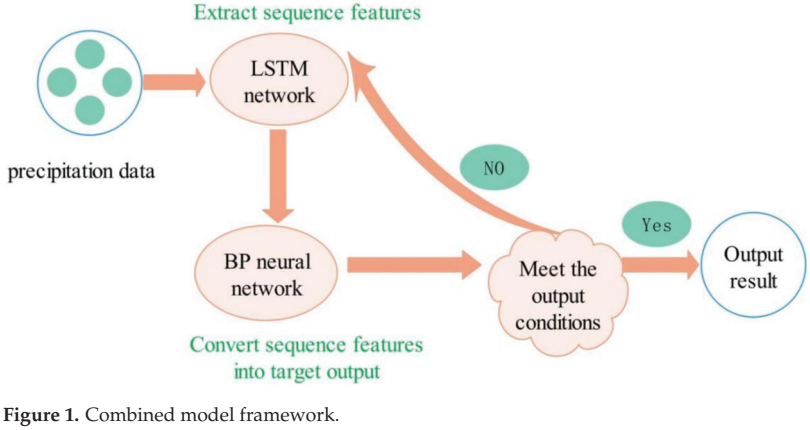


Figure 1. Combined model framework.

2.2. Feature Extraction Based on LSTM Network

Because the LSTM network has a strong memory capacity, it has a natural advantage in processing sequence data. This article employs the LSTM network as the basic network for sequence data prediction. In practical applications, RNN has been able to process some simple correlation information while its memory capacity is not strong. When the sequence is too long, error back propagation will cause larger gradient dispersion and gradient explosion problems, which can be effectively alleviated by introducing a "gate" mechanism [47,48] and memory unit [25,49] in the LSTM network.

2.2.1. Basic Idea

Only two factors, the current round of input x_t and the last round of output h_{t-1}, affect the traditional RNN network unit. Since there is only one tanh excitation unit in the network, the network output is:

$$h_t = \tanh(W_t[h_{t-1}, x_t] + b_t) \quad (1)$$

Therefore, RNN is sensitive to short-term input, making it difficult to solve the long sequence problem, as shown in Figure 2.

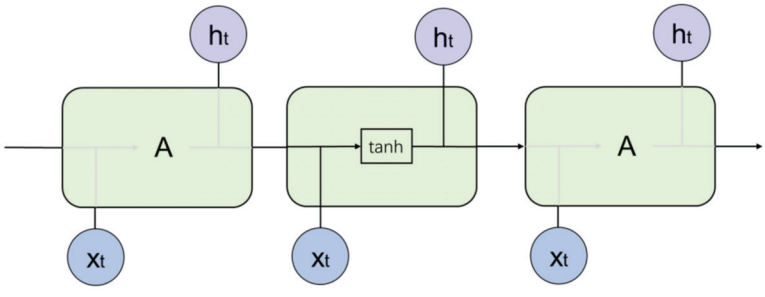


Figure 2. RNN network structure.

The LSTM network introduces a unit state and "gate" mechanism, which enhances the network's ability to remember long-term information, as shown in Figure 3. The current cell state C_t consists of the previous cell state C_{t-1}, the previous cell output h_{t-1}, as well as the current input x_t. The forget gate and input gate process the output h_{t-1} of the previous round and the input x_t of the current unit, and then combine with the current unit state C_t to form the output h_t of the current round through the output gate.

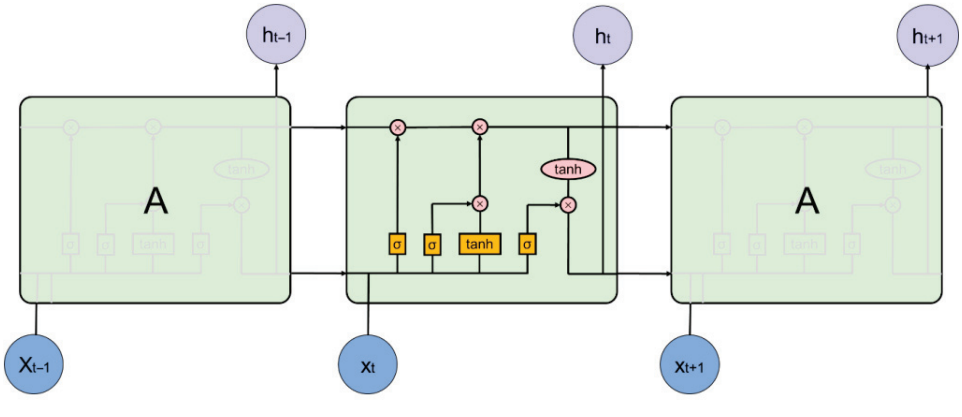


Figure 3. LSTM network structure.

It can be seen from Figure 3 that the unit state C, which runs through the whole LSTM network, constantly transfers information from the previous layer to the next layer, realizing the long-term memory retention function. In the LSTM network, there are three gate switches: input gate, forgetting gate, and output gate, through which the LSTM network can determine whether the current output depends on the early output, recent output, or current input.

2.2.2. Forgetting Gate

The first problem that the LSTM network solves is to determine the information that can pass through the current neuron, which is determined by the forgetting gate in LSTM. In the forgetting gate, the output h_{t-1} at the previous moment is dot multiplied with the input x_t at the current moment, and then the output f_t at this moment inside the neuron is obtained through the Sigmoid function [50], which is:

$$f_t = \sigma\Big(W_f[h_{t-1}, x_t] + b_f\Big) \tag{2}$$

where W_f represents the weight matrix and b_f represents the bias term.

2.2.3. Input Gate and Unit Status

After confirming the reserved information, LSTM needs to determine how much of the current input needs to be stored in the cell state, with this function implemented by the input gate in LSTM. In the input gate, the current input x_t together with the previous round of output h_{t-1} are point multiplied and then passed through the function of Sigmoid, with the purpose to determine which inputs are updated; the current input x_t and the previous round of output h_{t-1} are subjected to a dot product operation and then passed through the tanh function, aiming to form alternative update information.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i). \tag{3}$$

$$C_t = \tanh(W_C[h_{t-1}, x_t] + b_C). \tag{4}$$

where W_i and W_C are the weight matrix, respectively, and b_i and b_C are the bias items, respectively. The current cell state C_t is starting to be updated after obtaining the results of the forgetting gate and the input gate. The output f_t of the current time in the neuron is point multiplied with the result C_{t-1} of the previous round of the memory unit, while at the same time the two internal update information points i_t and $\widetilde{C}_t$ perform the dot product operation, and finally the new unit state is obtained by adding them together.

$$C_t = f_t * C_{t-1} + i_t * C_t. \tag{5}$$

2.2.4. Output Gate

For LSTM, it is necessary to determine how to output the current information when the unit state is determined, and with this function determined by the output gate, the unit output is jointly determined by x_t, h_{t-1}, and C_t. o_t is obtained through the function of the Sigmoid function after x_t and s_{t-1} are dot multiplied with C_t, which passes through the tanh function dot multiplied by o_t, and finally output s_t is obtained:

$$o_t = \sigma(W_o[s_{t-1}, x_t] + b_o). \tag{6}$$

$$s_t = o_t * \tanh(C_t). \tag{7}$$

where W_o is the weight matrix and b_o is the offset term.

2.3. Convert Sequence Features into Target Output

2.3.1. CP Combined with BP Neural Network

The Chebyshev polynomial is an important special function named after the famous Russian mathematician Tschebyscheff. It originates from the cosine function of multiple angles and the expansion of the cosine function. It is divided into the first kind of Chebyshev polynomial and the second kind of Chebyshev polynomial. Chebyshev polynomials used in this paper belong to the first category. Chebyshev polynomials play a very important role in approximate calculation in mathematics, physics, and technical science, such as the injection continuous function approximation problem, impedance transformation problem, and so on. The roots of the first kind of Chebyshev polynomials (called Chebyshev nodes) can be used for polynomial interpolation. The corresponding interpolation polynomials can minimize the Runge phenomenon and provide the best uniform approximation of polynomials in continuous functions. In practical application, it is often necessary to solve a known complex function $f(x)$, and in order to simplify the calculation, it is usually necessary to find a function $Q_n(x)$ to minimize the error between the two in a certain metric sense. In the Chebyshev best uniform approximation theory, the function $Q_n(x)$ is a Chebyshev polynomial and it satisfies that the difference between and in an interval $[a, b]$ is the smallest of all polynomials $Q_n(x)$ and $f(x)$ in the interval, as shown in the following formula:

$$\max_{a \le x \le b} |Q_n(x) - f(x)| = \min \left| \max \left| Q(x) - f(x) \right| \right| \tag{8}$$

The function approximation theory of Chebyshev polynomials shows that such polynomials $Q_n(x)$ exist and are unique: let $D_x = \max_{a \le x \le b} |Q_n(x) - f(x)|$, D_x has at least $n + 2$ interleaving points $[x_1 \ \cdots \ x_{n+2}] (a \le x_1 < \cdots < x_{n+2} \le b)$ on $[a, b]$, so that $D(x_i) = \pm D_n$, among them, $i \in [1 \ , n + 2]$, $Q_n(x)$ is the best uniform approximation of $f(x)$.

Chebyshev polynomials are a series of orthogonal polynomials [51], which can approximate any continuous function. Neural networks based on CP have excellent capabilities in fitting as well as generalization. The Chebyshev polynomial is defined in a recursive manner, where CP can be expressed by the following recursive expression when the variable has a value range between -1 and 1, for an n-th order CP:

$$T_0(x) = 1, \tag{9}$$

$$T_1(x) = x, \tag{10}$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \tag{11}$$

According to the theory of orthogonal polynomial approximation, a set of Chebyshev polynomials can approach any objective function, when the variables belong to -1 to 1 and the number of polynomials R is large enough. As follows:

$$f(x) \approx \sum_{r=0}^{R} w_r T_r(x). \tag{12}$$

where R is the number of Chebyshev polynomials used to fit $f(x)$, $T_r(x)$ represents the r-th polynomial, and w_r represents the weight of the r-th polynomial.

It can be seen from Equation (12) that the objective function $f(x)$ is obtained by the weighted sum of R CPs. To express this more intuitively, this paper adopts the method of lexicographical sorting to express Chebyshev polynomials, and sorts them according to the order of each polynomial. Given two different basis functions $\varphi_q(x) = \mu_{i_1}(x_1) \ldots \mu_{i_N}(x_N)$ and $\varphi_{\hat{q}}(x) = \mu_{j_1}(x_1) \ldots \mu_{j_N}(x_N)$ in the condition of $q \ne \hat{q}$. Let $Q = [i_1, i_2, \ldots, i_N]$, $|Q| = [i_1 + i_2 + \ldots i_N]$, $\hat{Q} = [\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_N]$, and $|\hat{Q}| = [\hat{i}_1 + \hat{i}_2 + \ldots \hat{i}_N]$, $q > \hat{q}$ is established when any of the following conditions are met:

Condition 1: $|Q| > |\hat{Q}|$;
Condition 2: $|Q| = |\hat{Q}|$, and the first non-0 element of $Q - \hat{Q} = [i_1 - \hat{i}_1, i_2 - \hat{i}_2, \ldots i_N - \hat{i}_N]$ is positive.

The BP neural network usually consists of a myriad of layers, including one input layer, several hidden layers, and one output layer. It has already been proved that the BP neural network, with a single hidden layer, can approach any continuous function in the closed interval with arbitrary precision [52]. The BP neural network of a single hidden layer is combined with the LSTM network in this paper, and the topology diagram of the common single hidden layer BP neural network is shown in Figure 4.

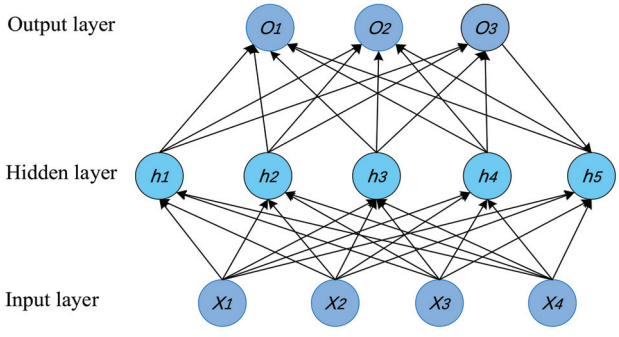


Figure 4. The topology of BP neural network.

In this network, where the input is $x_1 \ldots x_N$, the actual output is $o_1 \ldots o_k$ and the target output is $y_1 \ldots y_k$, each neuron in the input layer is fully connected with each neuron in the hidden layer, and each neuron in the hidden layer is fully connected with each neuron in the output layer. The weight between the i-th neuron in the input layer and the j-th neuron in the hidden layer is represented by w_{ij}, and the bias term is a_j. The weight between the j-th neuron in the hidden layer and the k-th neuron in the output layer is represented by v_{jk}, and the bias term is represented by a_j. The learning process of the BP neural network includes two steps, where the first step is the forward spread of information and the second step is the error back propagation. In the stage of the forward spread of information, information is transmitted forward, and the data are transferred from the input layer to the output layer through a weighted sum, with each neuron in the hidden layer and the output layer that can be, respectively, expressed as:

$$f(z_j) = f\left(\sum_{i=1}^{I} w_{ij} x_i - a_j\right), \tag{13}$$

$$o_k = f\left(\sum_{j=1}^{J} v_{jk} f(z_j) - \beta_k\right). \tag{14}$$

For the error back propagation stage, by computing the error and gradually correcting the weight and bias value through the gradient descent approach, the error and weight adjustment can be expressed as:

$$E = \frac{1}{2}\sum_{k=1}^{K} (o_k - y_k), \tag{15}$$

$$v_{jk} = v_{jk} - \eta \frac{\partial E}{\partial f(z_j)}, \tag{16}$$

$$v_{ij} = v_{ij} - \eta \frac{\partial E}{\partial f(z_j)} \times \frac{\partial f(z_j)}{\partial x_i}. \tag{17}$$

The "learning" process of the BP neural network is to gradually correct the weight and bias value according to the input data until the accuracy is satisfied or the maximum number of iterations is reached.

Hecht-Nielsen [52] has proved that a feedforward neural network with three layers can approximate any nonlinear continuous function in a closed interval with arbitrary precision. However, BP neural networks have some inherent shortcomings, such as slow convergence, ease of falling into a local minimum, and ease of falling into a saddle point, etc. The excitation function adopted by the traditional BP neural network is usually sigmoid, tanh, and ReLU, while this paper employs a set of linear independent orthogonal polynomials, which are Chebyshev polynomials instead of Sigmoid function, as the excitation function, as shown in Figure 5.

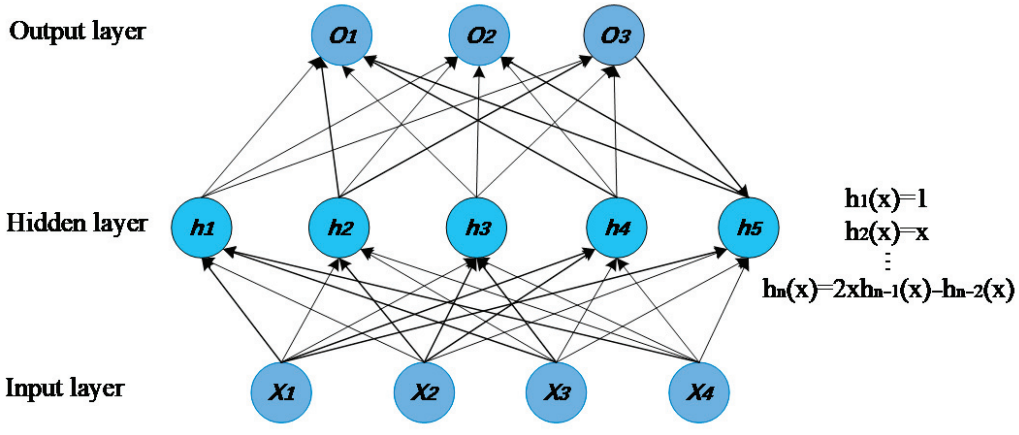
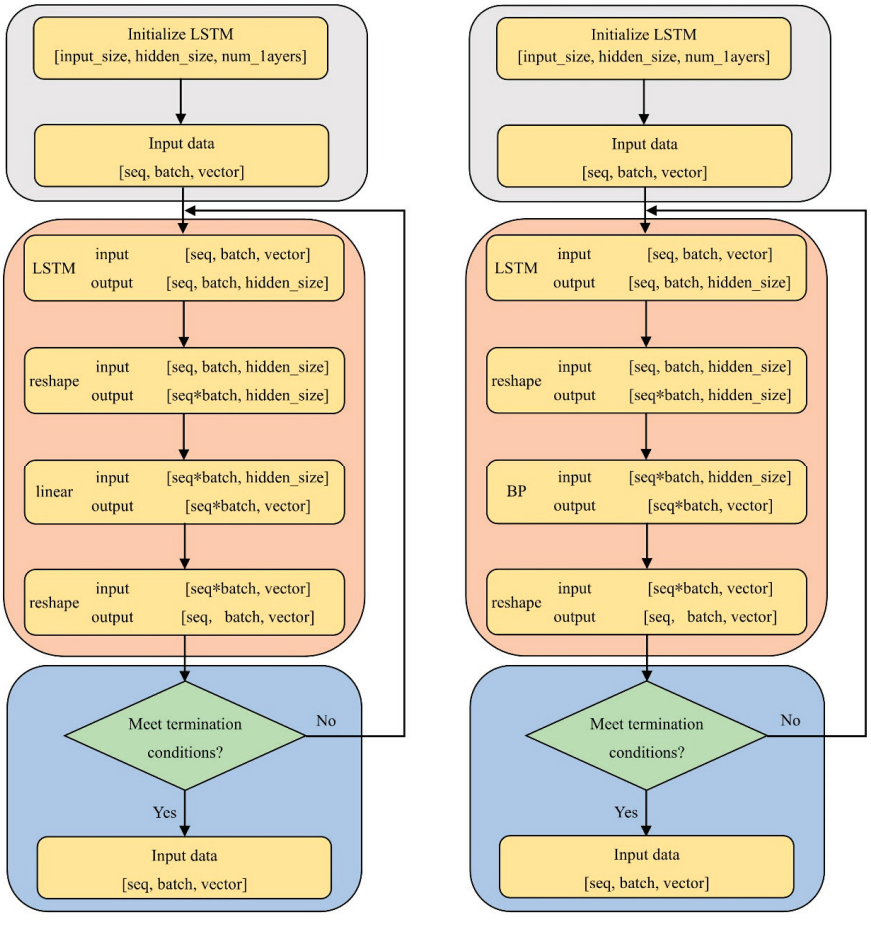


Figure 5. BP neural network based on CP.

A large amount of literature [17–19] has verified that using Chebyshev polynomials as the excitation function can effectively optimize the BP neural network. In the experiment of this paper, in contrast to networks and LSTM networks that use Sigmoid as the excitation function, the error of the network using CP as the excitation function declines faster and more steadily, and the prediction accuracy is also higher at the same time.

2.3.2. LSTM Combined with BP Neural Network

RNN is a typical feedback neural network whose network structure takes the time dimension into account, which can achieve excellent performance in processing data with timing laws. The structure of a single-layer RNN is shown in Figure 2, where each unit will receive the output of the previous unit and the input of this unit, and then the output can be given. Because the longer RNN is accompanied by the problems of gradient explosion and gradient disappearance, it has a limited memory capacity, which makes it unable to deal with long sequence data. In the actual operation of the LSTM network, the data need feeding into a linear layer to change the data dimension after passing through the LSTM network. This linear layer will transform the output of the LSTM network into a target output. Adopting the BP neural network to replace the linear layer of the LSTM network is considered in this paper, as shown in Figure 6.



(**a**) (**b**)

Figure 6. Network operation process. (**a**) LSTM operation process. (**b**) LSTM–BP operation process.

Firstly, the LSTM network is employed to process the original data, with the purpose to extract the time-series features of the data, and then the characteristic data are fed into the BP neural network with CP as the excitation function. In this way, the prediction ability of LSTM for sequence data and the function fitting ability of the BP neural network are used at the same time. This can effectively overcome the shortcomings of the BP neural network, such as slow convergence and ease of falling into a local minimum and local saddle point, through changing the excitation function of the BP neural network to CP.

2.4. Parameter Analysis/Complexity Analysis

The parameters of the LSTM–CP combined model consist of two parts: one is the parameters of the LSTM network and the other is the parameters of the BP neural network. Jin et al. [24] have proved that, for a fully connected feedforward neural network, the computational complexity of CP as an excitation function is lower than that of a Sigmoid. Therefore, this paper only focuses on the parameters of LSTM that combined with the BP neural network with CP as the excitation function.

The LSTM network has a total of three "gates" and a unit state, where each of the three gates generates some parameters. In contrast, the unit state does not generate any new parameters, and some parameters are also generated in the BP neural network. The number of parameters of different networks will be analyzed next according to the network operation flow chart in Figure 6.

For the forget gate, h_{t-1} is the output at the previous moment and the length is m; x_t is the input at the current moment and the length is n; W_f represents the weight matrix and the matrix size is $[m+n, m]$; and b_f represents the offset term and the length is m. For the input gate, h_{t-1} is the output at the previous moment and the length is m; x_t is the input at the current moment and the length is n; W_i and W_C are the weight matrices, whose matrix sizes are both $[m+n, m]$; and b_i and b_C are, respectively, offset terms and the length is m. For the output gate, h_{t-1} is the output at the previous moment and the length is m; x_t is the input at the current moment and the length is n; W_o is the weight matrix and the matrix size is $[m+n, m]$; and b_o is the bias term and the length is m. For the memory unit, it only performs a dot multiplication operation between the current output f_t in the neuron and the last round of memory unit result C_{t-1}, and the two internal update information points i_t and $\widetilde{C}_t$ perform the dot multiplication operation, with no new parameters generated. Therefore, the parameters of the forget gate, input gate, and output gate are, respectively:

$$s1 = (m+n) * m + m, \tag{18}$$

$$s2 = 2 * ((m+n) * m + m), \tag{19}$$

$$s3 = (m+n) * m + m. \tag{20}$$

When the number of LSTM network layers is Q, the total parameter quantity of the LSTM network is determined by the number of network layers, as well as the number of parameters of the three "gates", and the total parameter quantity of the LSTM network is:

$$s = Q_1 * 4 * ((m+n) * m + m). \tag{21}$$

In each round of parameter training, the parameters of the LSTM network and BP neural network will be updated at the same time, and the input of the BP neural network is determined by the LSTM network output. According to the above analysis, for the BP neural network, the input is h_t and the length is m. Let the number of neurons in the BP network be R, the output be h_t, and the length be m. The LSTM–BP combined model in this paper adopts the BP neural network to replace the linear layer of the LSTM network, with CP as the excitation function of the BP neural network, and each neuron is fully connected to the output, as shown in Figure 5. Then, the number of parameters of the BP neural network is:

$$s4 = m * R + m * R = 2mR. \tag{22}$$

The number of parameters in the LSTM–CP combination model is:

$$S = s + s4 = Q_2 * 4 * ((m+n) * m + m) + 2mR. \tag{23}$$

In summary, the parameters of each network are listed in Table 1.

This paper studies the precipitation data of 784 months in Yibin City from 1951 to 2017. The precipitation data of 1971 are ignored due to the missing data from January to June in 1971. In each round, 90% of the data are selected for training, and then the length is $n = 703$. The experiment of Section 3.1 shows that in LSTM the length of h_t is $m = 16$ and the number of LSTM network layers is $Q1 = 2$, and in LSTM–CP the number of LSTM network layers is $Q2 = 1$, the length of h_t is $m = 32$, and the number of BP neural network neurons is $R = 6$. The various parts and overall parameters of the network used in this article are shown in Table 2.

Table 1. Network parameter quantity function.

Structure	Parameter Quantity
forgetting gate	$s1 = (m + n) * m + m$
input gate	$s2 = 2 * ((m + n) * m + m)$
output gate	$s3 = (m + n) * m + m$
memory unit	0
LSTM	$s = Q_1 * 4 * ((m + n) * m + m)$
CP	$s4 = 2mR$
LSTM–CP	$S = Q_2 * 4 * ((m + n) * m + m) + 2mR$

Table 2. Network parameters.

Structure	Parameter
forgetting gate	$s1 = 11,520$
input gate	$s2 = 23,040$
output gate	$s3 = 11,520$
LSTM	$s = 92,160$
CP	$s4 = 384$
LSTM–CP	$S = 46,464$

The derivative of the function $f(x)$ at x_0 represents the slope of $y = f(x)$ at x_0, that is, the rate of change of $f(x)$ at x_0. The larger the derivative, the faster the change, that is, the faster the function grows. Because a variety of function variables that represent the parameter quantity occur in this article, drawing seems to be more difficult, such that the derivative is used for comparing multiple functions, as shown in Table 3.

Table 3. Parameter quantity derivative.

Structure	Parametric Function Derivative
LSTM	$S1'(Q) = 4 * ((m + n) * m + m)$
LSTM–CP	$S2'(R) = 2m$

As can be seen from Table 3, when m and n are constant, the parameters of the LSTM network increase in a square form as the number of LSTM network layers Q increases, while the number of parameters in the BP neural network increases linearly when the number of neurons in the BP neural network (R) increases. Therefore, the LSTM–CP combination model can effectively reduce the number of parameters with the use of the approach presented in this paper.

3. Results

Yibin City is located in the southeastern part of Sichuan Province, China, with an area of 13,300 square kilometers. The city is located between 103°36′–105°20′ east longitude and 27°50′–29°16′ north latitude. Yibin is 298.7 km away from Chengdu, the capital of Sichuan Province in the north, and 583.5 km away from Kunming, the capital of Yunnan Province in the south, and the brief geographical location is shown in Figure 7. It is an important city from Sichuan to the middle and lower reaches of the Yangtze River and coastal areas. The terrain of Yibin City is dominated by hills and middle–low mountains, accounting for 91.9% of the city's total area. It belongs to a subtropical humid monsoon climate, and the annual average temperature is about 17.9 °C, the average temperature in January is 7.8 °C, and the average temperature in July is 26.8 °C. The water system of Yibin City is very complex and intertwined. The rivers in Yibin City are mainly the Yangtze River, the river network is dense, and the total water resources and hydropower resources are relatively abundant. The annual average precipitation is 1050–1618 mm, which is a typical humid area. The rainy season is concentrated in the summer and autumn. The precipitation in these two seasons accounts for 81.7% of the annual precipitation. The main flood season is

mainly July, August, and September. The precipitation in these three months accounts for about 51% of the annual precipitation.

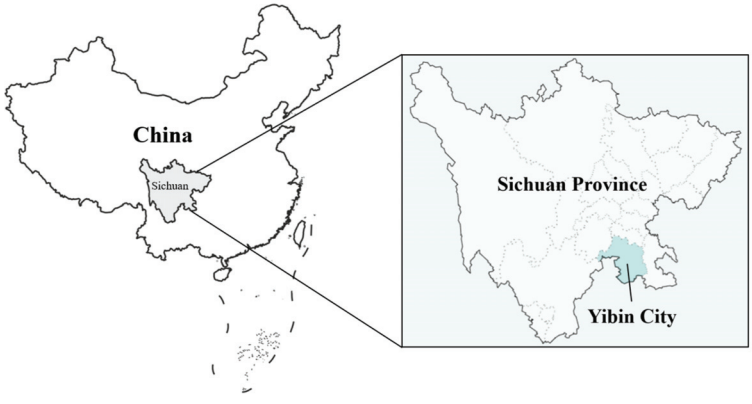


Figure 7. General location of Yibin City.

To eliminate the adverse effects of single sample data, improve the operation speed and accuracy as much as possible, and facilitate the operation of the model, it is necessary to first normalize the precipitation data [50] of Yibin City and map the original data to the interval [0, 1]:

$$x_i^{norm} = (x_i - x_{\min})/(x_{\max} - x_{\min}). \quad (24)$$

Mapping the input LSTM data to [0, 1] can aid in speeding up the convergence of the model. The use of CP for function fitting requires that the value of the data is supposed to be located in the interval $[-1, 1]$, and the output data of LSTM will pass through the output gate, that is, through Equations (6) and (7), such that the output value of LSTM must be in the range of interval $[-1, 1]$, meeting the requirements of the value range of the data fitted by the CP function.

In this paper, the common mean square error [53,54] is selected as the loss function of the training model, which is also adopted to calculate the validation set error of the model, and its formula is:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2. \quad (25)$$

The Adam optimizer [55] has a fast convergence speed and can adjust the learning rate adaptively according to the data distribution, which is why the Adam optimizer is selected to optimize the error function in this paper. Additionally, Dropout [56] is added to the network to reduce the influence of over fitting [57,58].

The experimental environment adopted in this paper is as follows: the training platform is Windows 10 Home 64-bit operating system, the computer memory is 4G, the processor model is Intel(R) Core(TM) I5-6300HQ CPU @ 2.30ghz, and the graphics card model is NVIDIA GeFosrce GTX 960M 2G. With Anaconda as the development environment and Python3.7 as the programming language, the PyTorch 1.2.0 [59] deep learning framework is used as the development framework, and the Nvidia CUDA 10.0 computing platform is used for accelerated computing.

3.1. LSTM Parameter Setting

When LSTM is used to predict precipitation, network parameters of LSTM should be determined first. First, the learning rate [60,61] is fixed to 0.01 to determine the remaining parameters, and then the parameters including the number of LSTM network layers and the size of the hidden layer in the LSTM network will be changed. For LSTM networks

with different parameters, take the first NUM minimum errors, and the average prediction error is shown in Figure 8.

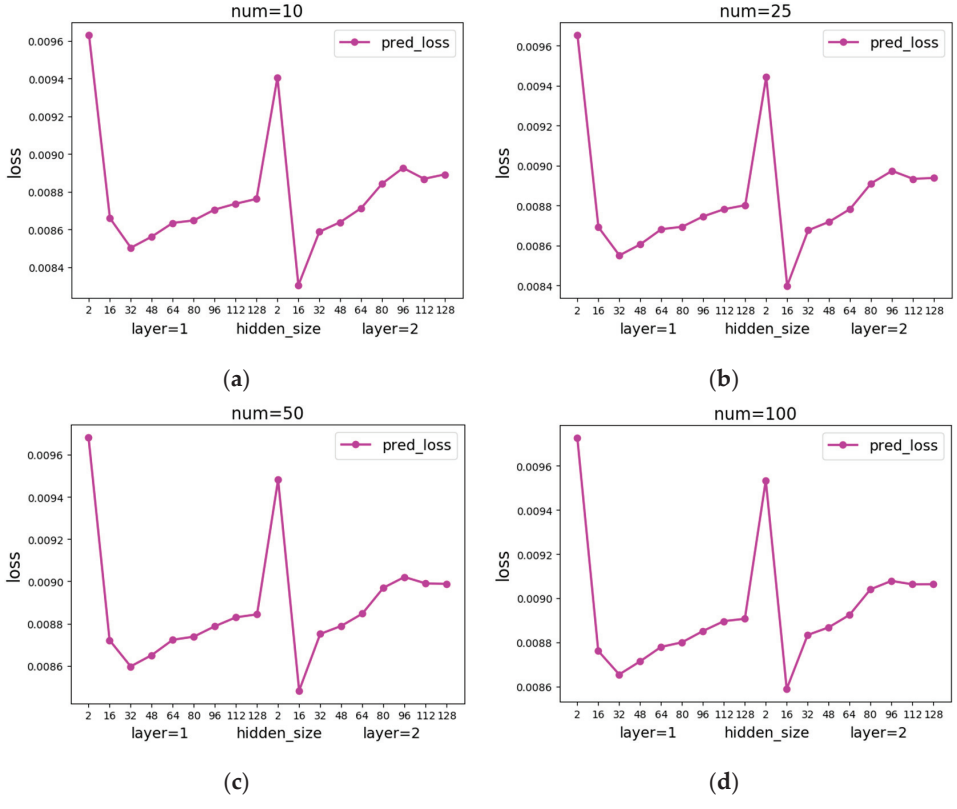
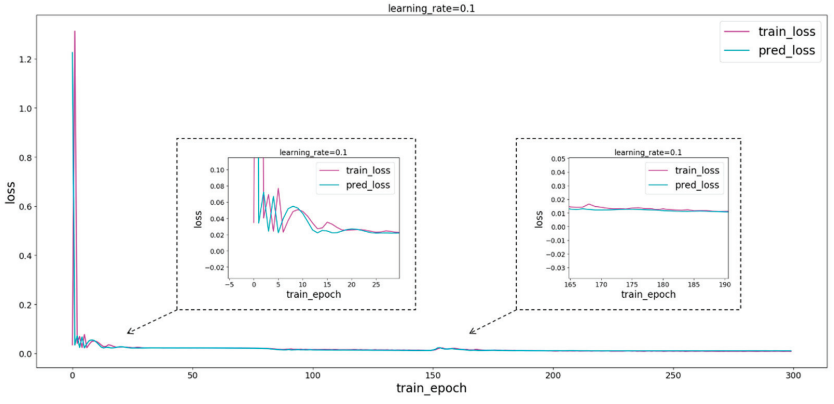
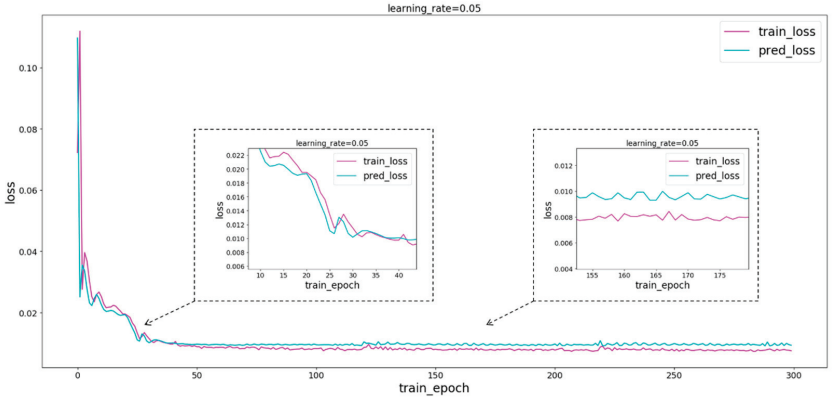


Figure 8. LSTM error curves of different layers and hidden_size. (**a**) Average error when Num is 10. (**b**) Average error when Num is 25. (**c**) Average error when Num is 50. (**d**) Average error when Num is 100.

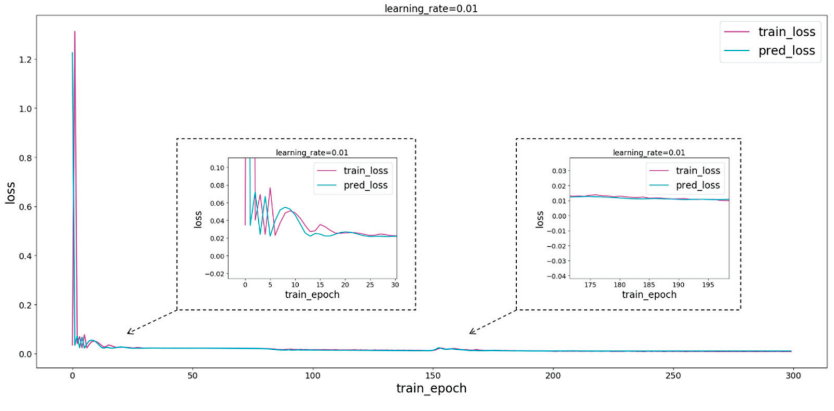
It is not difficult to see from the error curve in Figure 8 that when layer = 2 and hidden_size = 16 of LSTM, loss reaches the minimum. Therefore, layer = 2 and hidden_size = 16 are taken in this paper. Then, test the learning rate, take layer = 2, hidden_size = 16, and take the learning rate 0.1, 0.05, and 0.01, respectively, for the experiment. As can be seen from Figure 9, the loss will eventually stabilize at 0.01 when the learning rate is 0.1, but the error will decrease slowly; when the learning rate is 0.05, the loss will eventually stabilize around 0.01, but it is not stable; when the learning rate is 0.01, loss quickly drops to 0.01 and remains stable all the time. In summary, this article sets the number of layers of the LSTM network to 2, the hidden features to 16, and the learning rate to 0.01.



(**a**)



(**b**)



(**c**)

Figure 9. LSTM error curve for different learning rates. (**a**) LSTM error curve when the learning rate is 0.1. (**b**) LSTM error curve when the learning rate is 0.05. (**c**) LSTM error curve when the learning rate is 0.01.

3.2. LSTM–CP Parameter Setting

At the end of the LSTM network, there will be a linear layer, which can convert the time-series characteristic data extracted by the LSTM network into the target output [62,63]. The combination of CP and LSTM networks is supposed to use BP neural networks instead of this linear layer. With the Sigmoid function used as the common excitation function of the BP neural network, the Sigmoid function and CP function are, respectively, used as excitation functions to carry out comparative experiments in this paper. Figures 8 and 9 show that when layer = 2 and hidden_size = 16, the LSTM network performs the best. When layer = 1 and hidden_size = 32, the network is simpler but the model performance is relatively better. Therefore, this paper sets the basic LSTM network layer = 1, hidden_size = 32, learning_rate = 0.01 in the LSTM–CP combination model.

The next step is to determine the number of neurons in the BP neural network. Zhang Y proposed a two-stage approach to determine the number of neurons. In the first stage, the number of neurons is increased to a large extent to determine the approximate value range of neurons. In the second stage, the number of neurons is increased one by one to determine the exact value of the number of neurons. This method can effectively determine the number of neurons in the BP neural network. In order to obtain the approximate value range of neurons quickly and determine the number of neurons accurately, the initial number of neurons is set to 5 in the first stage of this paper, with a step size of 5 to increase neurons. First, Sigmoid is tested as the excitation function, and the prediction error of the minimum NUM among the error values of different neurons is taken as the average error curve.

It can be seen from Figure 10 that the error fluctuates as the number of neurons changes. When the number of neurons is about 70, the error is small, which means that the loss of the network is relatively small and stable when the number of neurons is about 70, and the optimal number of neurons is about 70. Therefore, the number of neurons is set from 66 to 74 for the experiment. As can be seen from Figure 11, when the number of neurons is 67, the prediction error is the smallest, that is, when the Sigmoid function is used as the excitation function of the BP neural network, the optimal number of neurons is 67.

Then, it is necessary to determine how many neurons should be used as the excitation function of the BP neural network, and the minimum NUM prediction errors among the error values of different numbers of neurons are still taken as the average error curve. The first-order CP transforms all the input into 1, while the second-order CP is actually a linear function. Since the first-order and second-order CP do not have nonlinear characteristics, they are not suitable for excitation functions of neural networks, so the CP order is at least 3 in this paper. A good fitting effect can be obtained at a lower order due to CP's strong fitting ability, which makes it unnecessary to use the two-stage method to determine the order of CP, and the order can be increased from 3 to 3. Figure 12 shows the experimental results of CP with different orders as the excitation function. It is not difficult to see that the error is low when the CP order is 6. Therefore, this paper sets the CP order of LSTM–CP to 6, that is, the number of neurons in the hidden layer of the BP neural network is 6.

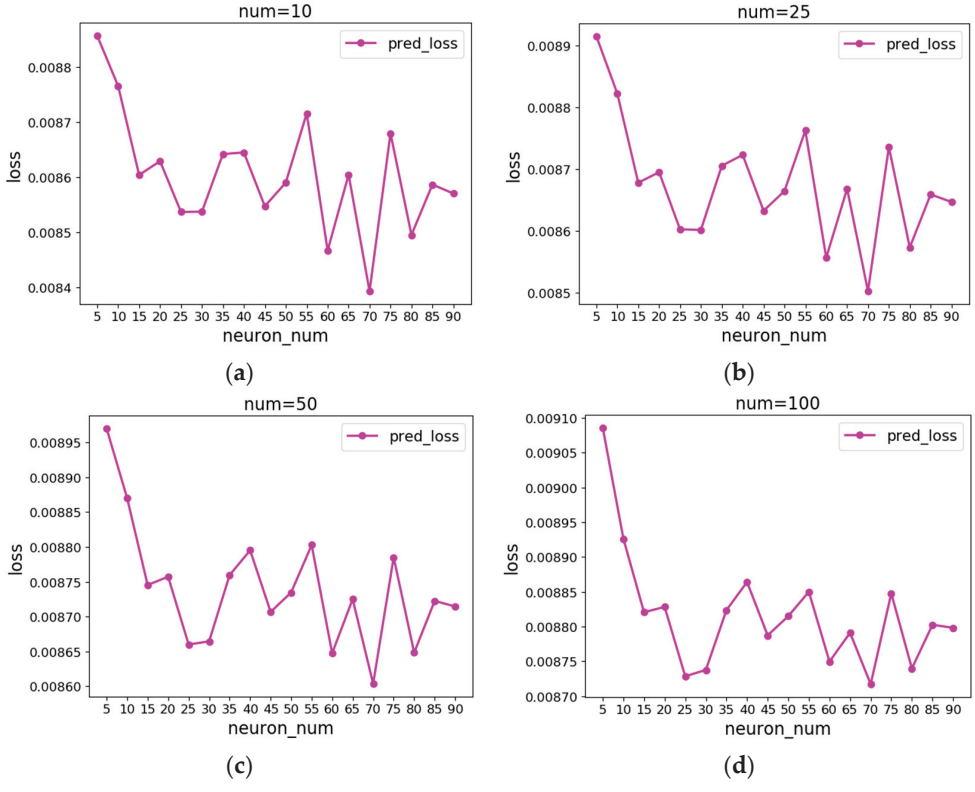


Figure 10. LSTM–BP (Sigmoid): the first-stage error curve. (**a**) Average error curve when Num is 10. (**b**) Average error curve when Num is 25. (**c**) Average error curve when Num is 50. (**d**) Average error curve when Num is 100.

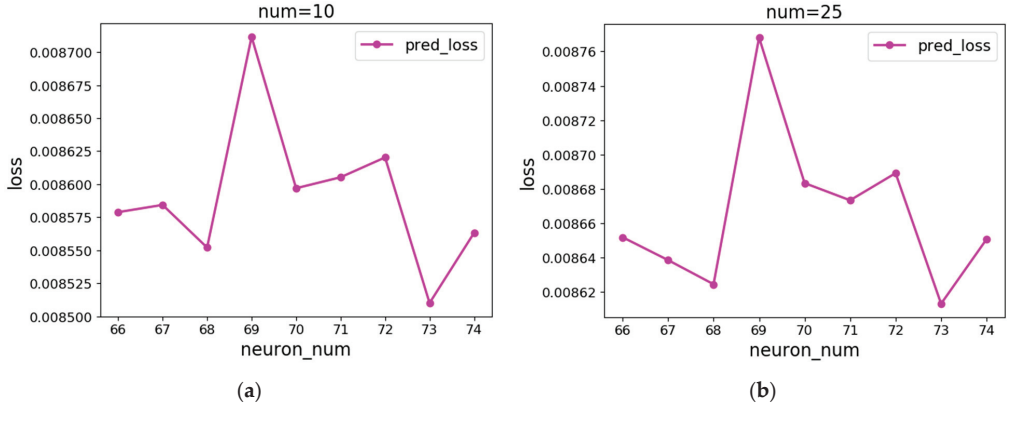
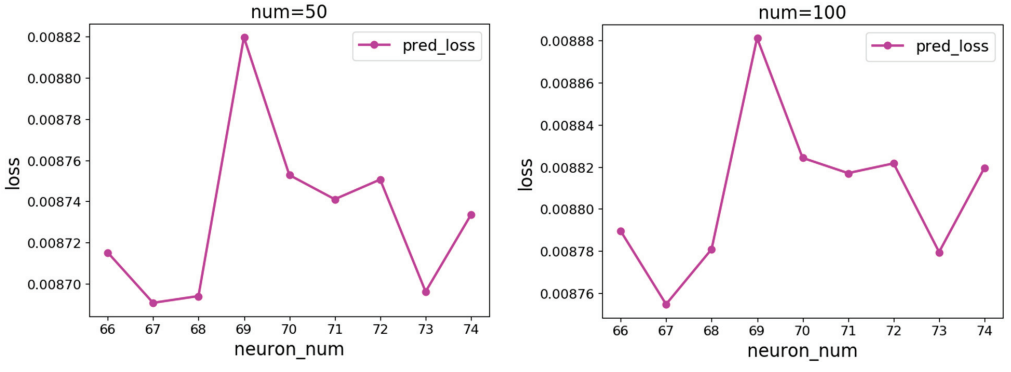
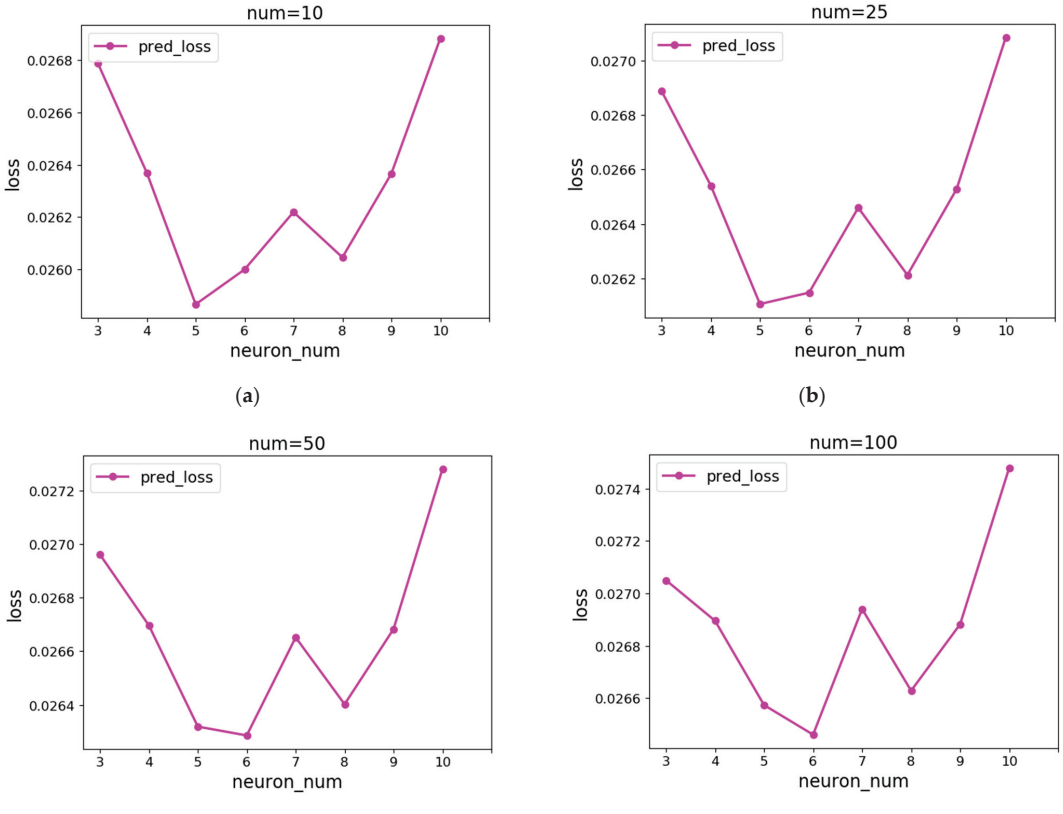


Figure 11. *Cont.*



(**c**) (**d**)

Figure 11. LSTM–BP (Sigmoid): the second-stage error curve. (**a**) Average error curve when Num is 10. (**b**) Average error curve when Num is 25. (**c**) Average error curve when Num is 50. (**d**) Average error curve when Num is 100.



(**a**) (**b**)

(**c**) (**d**)

Figure 12. LSTM–CP: Error curve. (**a**) Average error curve when Num is 10. (**b**) Average error curve when Num is 25. (**c**) Average error curve when Num is 50. (**d**) Average error curve when Num is 100.

3.3. Comparative Analysis

The optimal parameters of different networks are given in Section 3.1, while this section will make a detailed comparative analysis of the performance of each model. Figure 13 shows the error curves of different models, and Figure 14 shows the prediction results of different models. It can be easily seen from the error in Figure 13 that the LSTM network has a stable error of around 0.01 after 100 rounds of training, LSTM–CP has a stable error of around 0.01 after 100 rounds of training, and LSTM–BP (Sigmoid) also has a stable error of around 0.01 after 100 rounds of training. The reason for this is that the LSTM network has a strong ability to process sequence data and can quickly extract sequence features. We ran each model separately and obtained the prediction results of different models, as shown in Figure 14. It is not difficult to see from Figure 14 that the prediction results obtained by all networks are very close to the original data.

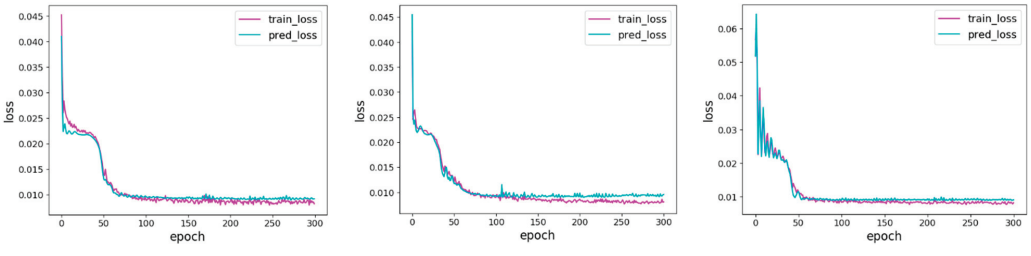


Figure 13. Training and prediction error curves of different networks. (**a**) LSTM. (**b**) LSTM–CP. (**c**) LSTM–BP (Sigmoid).

According to Figures 13 and 14, each model has better performance. Listing the detailed data of each model in Table 4, it is not difficult to see that the training error, prediction error, and training time of LSTM–CP are less than those of an ordinary LSTM network. In particular, if the excitation function of the BP neural network is operated on the CPU, the running time of CP will be shorter than that of Sigmoid. Therefore, using CP as the excitation function can obtain the lowest training error, the prediction error is smaller, and the running speed is better than the LSTM network.

Table 4. Comparison of model effects.

Model	Training Error	Prediction Error	Running Speed
LSTM	0.0078	0.0091	4.95
LSTM–BP (Sigmoid)	0.0079	0.0090	3.19
LSTM–CP	0.0076	0.0090	4.62

Note: The running speed is s/100 times.

Next, by using the ARMA linear model, SVR model, and MLP model to predict the precipitation at the same time, we calculated the evaluation indexes of each model, and compared the results with the LSTM–CP model proposed in this paper, as shown in Table 5. It is not difficult to see that, compared with other models, the values of MAE (mean absolute error), MSE (mean square error), and MAPE (mean percentage error) of the LSTM–CP network model are smaller than other models, which indicates that the LSTM–CP network model proposed in this paper has higher consistency and accuracy in rainfall prediction.

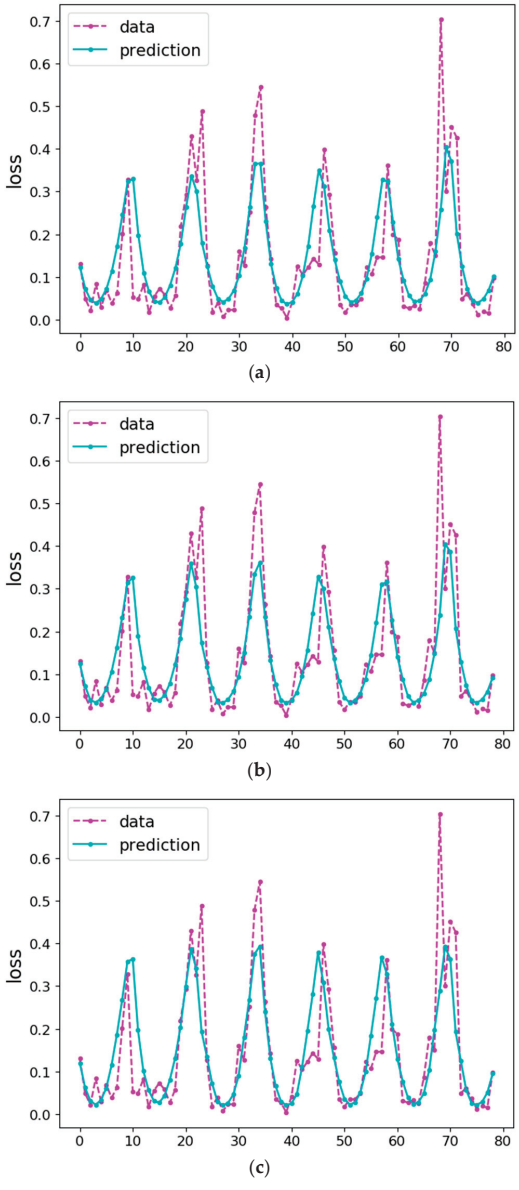


Figure 14. Forecast results. (**a**) LSTM. (**b**) LSTM–CP. (**c**) LSTM–BP (Sigmoid).

Table 5. Comparison results of prediction models.

Model	MAE	MSE	MAPE
ARMIA	0.0836	0.0120	55.051
SVR	0.0925	0.0172	65.731
MLP	0.1101	0.0191	75.210
LSTM–CP	0.0601	0.0090	53.121

4. Discussion

Due to the complex and diverse causes of precipitation and the interaction of various factors [64], it is very difficult to establish a mathematical model [65] of precipitation. Deep learning can automatically extract the low-level features of data and form abstract high-level features, without the need for the physical modeling of data, and it can easily deal with complex data structures because of its strong nonlinear ability [66]. The LSTM network is often employed to process time-series data. Its strong memory ability makes it have natural advantages in processing time-series data.

As can be seen from (a) in Figure 14, the precipitation value predicted by the LSTM neural network model is basically consistent with the real value of precipitation data, and LSTM can accurately extract the time-series features hidden in precipitation data. However, the LSTM network has some problems such as complex structure and gradient disappearance. Therefore, this paper proposes the LSTM–CP combination model by combining LSTM and CP, which makes full use of LSTM's ability to predict series data and CP's powerful function fitting ability in order to ensure the accuracy of the model, reduce the parameters of the network, and reduce the complexity of the precipitation prediction model.

Table 2 shows that using the LSTM–CP combination model can effectively reduce the number of network parameters. The amount of LSTM network model parameters is 92160, while the amount of LSTM–CP combination model parameters is only 46464, which greatly reduces the complexity of the model and is suitable for processing large and medium-sized datasets. At the same time, Table 4 shows that compared with the single LSTM model and the traditional precipitation prediction model, the LSTM–CP combined model has a smaller training error and test error, higher prediction accuracy, and is more suitable for precipitation prediction research. Because the model can reduce the use of parameters to a higher degree, it can more effectively reduce the running time when dealing with large and medium datasets, and make the processing of sequence data more efficient.

Rainfall is affected by the fluctuations of sea and land locations, topography, latitude, and human factors, but in this study, we ignored these changes. In future research, LSTM–CP can be applied to the scene with complex and huge data, such as text, music, and other sequence data processing. In this case, the number of network layers and hidden layers is larger when LSTM is used alone, and the combination of LSTM and CP can make the parameters have a larger space to decline, and it is not easy to overfit. In addition, the derivative function can be determined in advance according to the order of CP without using the deep learning framework for automatic derivation, which improves the computational efficiency.

5. Conclusions

Natural disasters often lead to major and long-term damage to the entire socio-economic system, such as floods, which may damage multiple infrastructure systems, lead to cascading failures and major socio-economic losses, and hinder development. Therefore, reducing the risk of precipitation disaster is closely related to sustainable development. With the progress of technology in recent years, artificial intelligence has become the main driver in various fields including sustainable development. Deep learning improves the ability to deal with complexity and increases our understanding of the variables and sources that affect rainfall. This paper proposed the LSTM–CP model to predict the precipitation of Yibin City. Firstly, the BP neural network is combined with LSTM to form a combined model where the LSTM network is used to extract the sequence features of the precipitation data. Then, the BP neural network is used to process the sequence features to obtain the target output. Because the traditional BP neural network has the disadvantages of easily falling into local minimums and saddle points, this article considers using CP as the excitation function to replace the Sigmoid function in the BP neural network, with the powerful function fitting ability of CP to process sequence features.

Through experimental tests and comparative analysis, the LSTM–CP combination model proposed in this paper has fewer parameters, a shorter running time, and smaller prediction error than the LSTM network. At the same time, compared with the SVR model, ARIMA model, and MLP model, the prediction accuracy of the LSTM–CP combined model is significantly improved, which improves the accuracy of rainfall prediction and makes the model more applicable. It can reflect the change trend of precipitation and help provide a data reference in areas prone to floods and drought disasters to help relevant departments prepare in advance, reduce local economic losses, and better achieve sustainable development. Furthermore, the rainfall prediction model can be incorporated into the regional early warning system to help better plan and manage water resources and reduce the risk of flooding. Finally, the application of artificial intelligence to precipitation prediction provides new ideas and methods for the current precipitation prediction research, and opens up a broader space for realizing the goal of sustainable development.

Author Contributions: Y.G. conducted the experiments and the whole article. W.T. collected and sorted out the data, and revised and improved the article. G.H. constructed the framework of the whole paper and wrote the review. F.P. designed the experiment and methodology. Y.W. wrote the original draft preparation. W.W. provided formal analysis and experimental tools. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Key Laboratory of Agricultural Information Engineering of Sichuan Province and Social Science Foundation of Sichuan Province in 2019, grant number SC19C032.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the research of this paper comes from the website: http://dataju.cn/Dataju/web/home (accessed on 15 October 2021).

Acknowledgments: Thanks for the help of the Key Laboratory of Agricultural Information Engineering of Sichuan Province.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Li, Y.; Yu, G.; Zhang, J. A three-stage stochastic model for emergency relief planning considering secondary disasters. *Eng. Optim.* **2021**, *53*, 551–575. [CrossRef]
2. Seager, R.; Naik, N.; Baethgen, W.; Robertson, A.; Kushnir, Y.; Nakamura, J.; Jurburg, S. Tropical Oceanic Causes of Interannual to Multidecadal Precipitation Variability in Southeast South America over the Past Century. *J. Clim.* **2010**, *23*, 5517–5539. [CrossRef]
3. Bishop, D.A.; Williams, A.P.; Seager, R.; Fiore, A.M.; Cook, B.I.; Mankin, J.S.; Singh, D.; Smerdon, J.E.; Rao, M.P. Investigating the Causes of Increased Twentieth-Century Fall Precipitation over the Southeastern United States. *J. Clim.* **2019**, *32*, 575–590. [CrossRef]
4. Hodnebrog, Ø.; Myhre, G.; Forster, P.M.; Sillmann, J.; Samset, B.H. Local biomass burning is a dominant cause of the observed precipitation reduction in southern Africa. *Nat. Commun.* **2016**, *7*, 11236. [CrossRef] [PubMed]
5. Zhao, J.; Liu, X. A hybrid method of dynamic cooling and heating load forecasting for office buildings based on artificial intelligence and regression analysis. *Energy Build.* **2018**, *174*, 293–308. [CrossRef]
6. Tien, T.-L. A research on the grey prediction model GM(1,n). *Appl. Math. Comput.* **2012**, *218*, 4903–4916. [CrossRef]
7. Fu, G.; Charles, S.P.; Kirshner, S. Daily rainfall projections from general circulation models with a downscaling nonhomogeneous hidden Markov model (NHMM) for south-eastern Australia. *Hydrol. Process.* **2013**, *27*, 3663–3673. [CrossRef]
8. Wang, D.; Borthwick, A.G.; He, H.; Wang, Y.; Zhu, J.; Lu, Y.; Xu, P.; Zeng, X.; Wu, J.; Wang, L.; et al. A hybrid wavelet de-noising and Rank-Set Pair Analysis approach for forecasting hydro-meteorological time series. *Environ. Res.* **2018**, *160*, 269–281. [CrossRef]
9. Chen, M.; Mao, S.; Liu, Y. Big data: A survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [CrossRef]
10. Wong, T.Y.; Bressler, N.M. Artificial Intelligence with Deep Learning Technology Looks Into Diabetic Retinopathy Screening. *JAMA J. Am. Med. Assoc.* **2016**, *316*, 2366–2367. [CrossRef]
11. Lee, J. Physical modeling of charge transport in conjugated polymer field-effect transistors. *J. Phys. D Appl. Phys.* **2021**, *54*, 143002. [CrossRef]

12. Nanda, T.; Sahoo, B.; Beria, H.; Chatterjee, C. A wavelet-based non-linear autoregressive with exogenous inputs (WNARX) dynamic neural network model for real-time flood forecasting using satellite-based rainfall products. *J. Hydrol.* **2016**, *539*, 57–73. [CrossRef]
13. Kashiwao, T.; Nakayama, K.; Ando, S.; Ikeda, K.; Lee, M.; Bahadori, A. A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency. *Appl. Soft Comput.* **2017**, *56*, 317–330. [CrossRef]
14. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back Propagating Errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
15. Ferreira, L.B.; da Cunha, F.F. Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Comput. Electron. Agric.* **2020**, *178*, 105728. [CrossRef]
16. Granata, F.; Nunno, F.D. Forecasting evapotranspiration in different climates using ensembles of recurrent neural networks. *Agric. Water Manag.* **2021**, *255*, 107040. [CrossRef]
17. Zhang, Y.; Qu, L.; Liu, J.; Guo, D.; Li, M. Sine neural network (SNN) with double-stage weights and structure determination (DS-WASD). *Soft Comput.* **2016**, *20*, 211–221. [CrossRef]
18. Tian, C.; Liu, S. Demodulation of two-shot fringe patterns with random phase shifts by use of orthogonal polynomials and global optimization. *Opt. Express* **2016**, *24*, 3202–3215. [CrossRef]
19. Mahmmod, B.M.; Ramli, A.B.D.R.; Baker, T.; Al-Obeidat, F.; Abdulhussain, S.H.; Jassim, W.A. Speech Enhancement Algorithm Based on Super-Gaussian Modeling and Orthogonal Polynomials. *IEEE Access* **2019**, *7*, 103485–103504. [CrossRef]
20. Lin, H.; Cao, D.; Shao, C. An admissible function for vibration and flutter studies of FG cylindrical shells with arbitrary edge conditions using characteristic orthogonal polynomials. *Compos. Struct.* **2018**, *185*, 748–763. [CrossRef]
21. Zhang, Y.; Yin, Y.; Guo, D.; Yu, X.; Xiao, L. Cross-validation based weights and structure determination of Chebyshev-polynomial neural networks for pattern classification. *Pattern Recognit.* **2014**, *47*, 3414–3428. [CrossRef]
22. Zhang, Y.; Yu, X.; Guo, D.; Yin, Y.; Zhang, Z. Weights and structure determination of multiple-input feed-forward neural network activated by Chebyshev polynomials of Class 2 via cross-validation. *Neural Comput. Appl.* **2014**, *25*, 1761–1770. [CrossRef]
23. Jin, L.; Huang, Z.; Li, Y.; Sun, Z.; Li, H.; Zhang, J. On Modified Multi-Output Chebyshev-Polynomial Feed-Forward Neural Network for Pattern Classification of Wine Regions. *IEEE Access* **2019**, *7*, 1973–1980. [CrossRef]
24. Jin, L.; Huang, Z.; Chen, L.; Liu, M.; Li, Y.; Chou, Y.; Yi, C. Modified single-output Chebyshev-polynomial feedforward neural network aided with subset method for classification of breast cancer. *Neurocomputing* **2019**, *350*, 128–135. [CrossRef]
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
26. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [CrossRef]
27. Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [CrossRef]
28. Kang, J.; Wang, H.; Yuan, F.; Wang, Z.; Huang, J.; Qiu, T. Prediction of Precipitation Based on Recurrent Neural Networks in Jingdezhen, Jiangxi Province, China. *Atmosphere* **2020**, *11*, 246. [CrossRef]
29. Zhou, Y.; Li, Y.; Jin, J.; Zhou, P.; Zhang, D.; Ning, S.; Cui, Y. Stepwise Identification of Influencing Factors and Prediction of Typhoon Precipitation in Anhui Province Based on the Back Propagation Neural Network Model. *Water* **2021**, *13*, 550. [CrossRef]
30. Zahraei, A.; Hsu, K.; Sorooshian, S.; Gourley, J.J.; Lakshmanan, V.; Hong, Y.; Bellerby, T. Quantitative precipitation nowcasting: A Lagrangian pixel-based approach. *Atmos. Res.* **2012**, *118*, 418–434. [CrossRef]
31. Bowler, N.E.H.; Pierce, C.E.; Seed, A. Development of a precipitation nowcasting algorithm based upon optical flow techniques. *J. Hydrol.* **2004**, *288*, 74–91. [CrossRef]
32. Pham, B.T.; Le, L.M.; Le, T.T.; Bui, K.T.T.; Le, V.M.; Ly, H.B.; Prakash, I. Development of advanced artificial intelligence models for daily rainfall prediction. *Atmos. Res.* **2020**, *237*, 104845. [CrossRef]
33. Banadkooki, F.B.; Ehteram, M.; Ahmed, A.N.; Fai, C.M.; Afan, H.A.; Ridwan, W.M.; Sefelnasr, A.; Elshafie, A. Precipitation forecasting using multilayer neural network and support vector machine optimization based on flow regime algorithm taking into account uncertainties of soft computing models. *Sustainability* **2019**, *11*, 6681. [CrossRef]
34. Wang, J.; Zhang, L.; Guan, J.; Zhang, M. Evaluation of combined satellite and radar data assimilation with POD-4DEnVar method on rainfall forecast. *Appl. Sci.* **2020**, *10*, 5493. [CrossRef]
35. Li, Y.; Zhu, Z.; Kong, D.; Han, H.; Zhao, Y. EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowl.-Based Syst.* **2019**, *181*, 104785. [CrossRef]
36. Wang, Y.; Zhang, X.; Lu, M.; Wang, H.; Choe, Y. Attention augmentation with multi-residual in bidirectional LSTM. *Neurocomputing* **2020**, *385*, 340–347. [CrossRef]
37. Liu, J.; Gong, X. Attention mechanism enhanced LSTM with residual architecture and its application for protein-protein interaction residue pairs prediction. *BMC Bioinform.* **2019**, *20*, 609. [CrossRef]
38. Zhao, R.; Yan, R.; Wang, J.; Mao, K. Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks. *Sensors* **2017**, *17*, 273. [CrossRef]
39. Ahmadian, A.; Salahshour, S.; Chan, C.S. Fractional Differential Systems: A Fuzzy Solution Based on Operational Matrix of Shifted Chebyshev Polynomials and Its Applications. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 218–236. [CrossRef]

40. Cui, K.; Qin, X. Virtual reality research of the dynamic characteristics of soft soil under metro vibration loads based on BP neural networks. *Neural Comput. Appl.* **2018**, *29*, 1233–1242. [CrossRef]
41. Su, Z.; Wang, J.; Lu, H.; Zhao, G. A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting. *Energy Convers. Manag.* **2014**, *85*, 443–452. [CrossRef]
42. Juang, C.F.; Hsieh, C.D. TS-fuzzy system-based support vector regression. *Fuzzy Set Syst.* **2009**, *160*, 2486–2504. [CrossRef]
43. Ölmez, T.; Dokur, Z. Classification of heart sounds using an artificial neural network. *Pattern Recognit. Lett.* **2003**, *24*, 617–629. [CrossRef]
44. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.Y.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [CrossRef]
45. Chang, Z.; Zhang, Y.; Chen, W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy* **2019**, *187*, 115804. [CrossRef]
46. Zhou, S.; Zhou, L.; Mao, M.; Tai, H.M.; Wan, Y. An Optimized Heterogeneous Structure LSTM Network for Electricity Price Forecasting. *IEEE Access* **2019**, *7*, 108161–108173. [CrossRef]
47. Wu, P.; Lei, Z.; Zhou, Q.; Zhu, R.; Chang, X.; Sun, J.; Zhang, W.; Guo, Y. Multiple premises entailment recognition based on attention and gate mechanism. *Expert Syst. Appl.* **2020**, *147*, 113214. [CrossRef]
48. Murayama, Y.; Uhlmann, F. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell* **2015**, *163*, 1628–1640. [CrossRef]
49. Wang, J.; Zhang, L.; Guo, Q.; Yi, Z. Recurrent Neural Networks with Auxiliary Memory Units. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 1652–1661. [CrossRef]
50. Schwarzenbach, H.; da Silva, A.M.; Calin, G.; Pantel, K. Data Normalization Strategies for MicroRNA Quantification. *Clin. Chem.* **2015**, *61*, 1333–1342. [CrossRef]
51. Berrone, S.; Borio, A. Orthogonal polynomials in badly shaped polygonal elements for the Virtual Element Method. *Finite Elem. Anal. Des.* **2017**, *129*, 14–31. [CrossRef]
52. Hecht-Nielsen, R. Theory of the backpropagation neural network. *IEEE IJCNN* **1989**, *1*, 593–605.
53. Wang, Z.; Bovik, A.C. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **2009**, *26*, 98–117. [CrossRef]
54. Rougier, J. Ensemble Averaging and Mean Squared Error. *J. Clim.* **2016**, *29*, 8865–8870. [CrossRef]
55. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
56. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *Comput. Sci.* **2012**, *3*, 212–223.
57. Hawkins, D.M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [CrossRef]
58. Liu, R.; Gillies, D.F. Overfitting in linear feature extraction for classification of high-dimensional image data. *Pattern Recognit.* **2016**, *53*, 73–86. [CrossRef]
59. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026–8037.
60. Takase, T.; Oyama, S.; Kurihara, M. Effective neural network training with adaptive learning rate based on training loss. *Neural Netw.* **2018**, *101*, 68–78. [CrossRef]
61. Chandra, B.; Sharma, R.K. Deep learning with adaptive learning rate using laplacian score. *Expert Syst. Appl.* **2016**, *63*, 1–7. [CrossRef]
62. Liu, J.; Wang, Z.; Xu, M. DeepMTT: A deep learning maneuvering target-tracking algorithm based on bidirectional LSTM network. *Inf. Fusion* **2020**, *53*, 289–304. [CrossRef]
63. Oehmcke, S.; Zielinski, O.; Kramer, O. Input quality aware convolutional LSTM networks for virtual marine sensors. *Neurocomputing* **2018**, *275*, 2603–2615. [CrossRef]
64. Liu, R.; Liu, S.C.; Cicerone, R.J.; Shiu, C.J.; Li, J.; Wang, J.; Zhang, Y. Trends of Extreme Precipitation in Eastern China and Their Possible Causes. *Adv. Atmos. Sci.* **2015**, *32*, 1027–1037. [CrossRef]
65. Granich, R.M.; Gilks, C.F.; Dye, C.; De Cock, K.M.; Williams, B.G. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: A mathematical model. *Lancet* **2009**, *373*, 48–57. [CrossRef]
66. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef]

entropy

MDPI

Article

Beware the Black-Box: On the Robustness of Recent Defenses to Adversarial Examples

Kaleel Mahmood [1,*], Deniz Gurevin [2], Marten van Dijk [3] and Phuoung Ha Nguyen [4]

[1] Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06269, USA
[2] Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA; deniz.gurevin@uconn.edu
[3] CWI, 1098 XG Amsterdam, The Netherlands; Marten.van.Dijk@cwi.nl
[4] eBay, San Jose, CA 95125, USA; phuongha.ntu@gmail.com
* Correspondence: kaleel.mahmood@uconn.edu

Abstract: Many defenses have recently been proposed at venues like NIPS, ICML, ICLR and CVPR. These defenses are mainly focused on mitigating white-box attacks. They do not properly examine black-box attacks. In this paper, we expand upon the analyses of these defenses to include adaptive black-box adversaries. Our evaluation is done on nine defenses including Barrage of Random Transforms, ComDefend, Ensemble Diversity, Feature Distillation, The Odds are Odd, Error Correcting Codes, Distribution Classifier Defense, K-Winner Take All and Buffer Zones. Our investigation is done using two black-box adversarial models and six widely studied adversarial attacks for CIFAR-10 and Fashion-MNIST datasets. Our analyses show most recent defenses (7 out of 9) provide only marginal improvements in security (<25%), as compared to undefended networks. For every defense, we also show the relationship between the amount of data the adversary has at their disposal, and the effectiveness of adaptive black-box attacks. Overall, our results paint a clear picture: defenses need both thorough white-box and black-box analyses to be considered secure. We provide this large scale study and analyses to motivate the field to move towards the development of more robust black-box defenses.

Keywords: adversarial machine learning; black-box attacks; security

Citation: Mahmood, K.; Gurevin, D.; van Dijk, M.; Nguyen, P.H. Beware the Black-Box: On the Robustness of Recent Defenses to Adversarial Examples. *Entropy* **2021**, *23*, 1359. https://doi.org/10.3390/e23101359

Academic Editor: Luis Hernández-Callejo

Received: 16 September 2021
Accepted: 14 October 2021
Published: 18 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Convolutional Neural Networks (CNNs) are widely used for image classification [1,2] and object detection. Despite their widespread use, CNNs have been shown to be vulnerable to adversarial examples [3]. Adversarial examples are clean images which have malicious noise added to them. This noise is small enough so that humans can visually recognize the images, but CNNs misclassify them.

Adversarial examples can be created through white-box or black-box attacks, depending on the assumed adversarial model. White-box attacks create adversarial examples by directly using information about the trained parameters in a classifier (e.g., the weights of a CNN). Black-box attacks on the other hand, assume an adversarial model where the trained parameters of the classifier are secret or unknown. In black-box attacks, the adversary generates adversarial examples by exploiting other information such as querying the classifier [4–6], or using the original dataset the classifier was trained on [7–10]. We can also further categorize black-box attacks based on whether the attack tries to tailor the adversarial example to specifically overcome the defense (adaptive black-box attacks), or if the attack is fixed regardless of the defense (non-adaptive black-box attacks). In terms of attacks, we focus on adaptive black-box adversaries. A natural question is why do we choose this scope?

(1) White-box robustness does not automatically mean black-box robustness. In security communities such as cryptology, black-box attacks are considered strictly weaker than

white-box attacks. This means that if a defense is shown to be secure against a white-box adversary, it would also be secure against a black-box adversary. In the field of adversarial machine learning, this principle does NOT always hold true. Why does this happen? In adversarial machine learning, white-box attacks use gradient information directly to create adversarial examples. It is possible to obfuscate this gradient, an effect known as gradient masking [9] and thus make white-box attacks fail. Black-box attacks do not directly use gradient information. As a result, black-box attacks may still be able to work on defenses that have gradient masking. This means adversarial machine learning defenses need to be analyzed under both white-box AND black-box attacks.

(2) White-box adversaries are well studied in most defense papers [11–18] as opposed to black-box adversaries. Less attention has been given to black-box attacks, despite the need to test defenses on both types of attacks (as mentioned in our first point). This paper offers a unique perspective by testing defenses under adaptive black-box attacks. By combining the white-box analyses already developed in the literature with the black-box analyses we present here, we give a full security picture.

Having explained our focus for the type of attacks, we next explain why we chose the following 9 defenses to investigate:

(1) Each defense is unique in the following aspect: No two defenses use the exact same set of underlying methods to try and achieve security. We illustrate this point in Table 1. Further in Section 3 we go into specifics about why each individual defense is chosen. As a whole, this diverse group of defenses allows us to evaluate many different competing approaches to security.

(2) Most of the defenses we analyze have been published at NIPS, ICML, ICLR or CVPR. This indicates the machine learning community and reviewers found these approaches worthy of examination and further study.

Major Contributions, Related Literature and Paper Organization

Having briefly introduced the notion of adversarial machine learning attacks and explained the scope of our work, we discuss several other important introductory points. First, we list our major contributions. Second, we discuss literature that is related but distinct from our work. Finally, we give an overview of the organization of the rest of our paper. Major contributions:

1. *Comprehensive black-box defense analysis*—Our experiments are comprehensive and rigorous in the following ways: we work with 9 recent defenses and a total of 12 different attacks. Every defense is trained on the same dataset and with the same base CNN architecture whenever possible. Every defense is attacked under the same adversarial model. This allows us to directly compare defense results. It is important to note some papers use different adversarial models which makes comparisons across papers invalid [19].
2. *Adaptive adversarial strength study*—In this paper we are the first (to the best of our knowledge) to show the relationship between each of the 9 defenses and the strength of an adaptive black-box adversary. Specifically, for every defense we are able to show how its security is effected by varying the amount of training data available to an adaptive black-box adversary (i.e., 100%, 75%, 50%, 25% and 1%).
3. *Open source code and detailed implementations*—One of our main goals of this paper is to help the community develop stronger black-box adversarial defenses. To this end, we publicly provide code for our experiments: https://github.com/MetaMain/BewareAdvML (accessed on 20 May 2021). In addition, in Appendix A we give detailed instructions for how we implemented each defense and what experiments we ran to fine tune the hyperparameters of the defense.

Related Literature: There are a few works that are related but distinctly different from our paper. We briefly discuss them here. As we previously mentioned, the field of adversarial machine learning has mainly been focused on white-box attacks on defenses. Works that consider white-box attacks and/or multiple defenses include [20–24].

In [20] the authors test white-box and black-box attacks on defenses proposed in 2017, or earlier. It is important to note, all the defenses in our paper are from 2018 or later. There is no overlap between our work and the work in [20] in terms of defenses studied. In addition, in [20], while they do consider a black-box attack, it is not adaptive because they do not give the attacker access to the defense training data.

In [21], an ensemble is studied by trying to combine multiple weak defenses to form a strong defense. Their work shows that such a combination does not produce a strong defense under a white-box adversary. None of the defenses covered in our paper are used in [21]. Also [21] does not consider a black-box adversary like our work.

In [23], the authors also do a large study on adversarial machine learning attacks and defenses. It is important to note that they do not consider adaptive black-box attacks, as we define them (see Section 2). They do test defenses on CIFAR-10 like us, but in this case only one defense (ADP [11]) overlaps with our study. To reiterate, the main threat we are concerned with is adaptive black-box attacks which is not covered in [23].

One of the closest studies to us is [22]. In [22] the authors also study adaptive attacks. However, unlike our analyses which use black-box attacks, they assume a white-box adversary. Our paper is a natural progression from [22] in the following sense: If the defenses studied in [22] are broken under an adaptive white-box adversary, could these defenses still be effective under under a weaker adversarial model? In this case, the model in question would be one that disallows white-box access to the defense, i.e., a black-box adversary. Whether these defenses are secure against adaptive black-box adversaries is an open question, and one of the main questions our paper seeks to answer.

Lastly, adaptive black-box adversaries have also been studied before in [24]. However, they do not consider variable strength adaptive black-box adversaries as we do. We also cover many defenses that are not included in their paper (Error Correcting Codes, Feature Distillation, Distribution Classifier, K-Winner Take All and ComDefend). Finally, the metric we use to compare defenses is fundamentally different from the metric proposed in [24]. They compare results using a metric that balances clean accuracy and security. In this paper, we study the performance of a defense relative to no defense (i.e., a vanilla classifier).

Paper Organization: Our paper is organized as follows: in Section 2, we describe the goal of the adversary mathematically, the capabilities given in different adversarial models and the categories of black-box attacks. In Section 3, we break down the defenses used in this paper in terms of their underlying defense mechanisms. We also explain why each individual defense was selected for analysis in this paper. In Section 4, we discuss the principal experimental results and compare the performances of the defenses. In Section 5, we analyze and discuss each defense individually. We also show the relationship between the security of each defense and the strength (amount of training data) available to an adaptive black-box adversary. We offer concluding remarks in Section 6. Lastly, full experimental details and defense implementation instructions are given in the Appendix A.

Table 1. Defenses analyzed in this paper and the corresponding defense mechanisms they employ. For definitions of the each defense mechanism see Section 3.

Defense Mechanism	Ensemble Diversity (ADP) [11]	Error Correcting Codes (ECOC) [12]	Buffer Zones (BUZz) [24]	Com Defend [13]	Barrage (BaRT) [14]	Distribution Classifier (DistC) [16]	Feature Distillation (FD) [18]	Odds Are Odd [17]	K-Winner (k-WTA) [15]
Mutiple Models	✓	✓	✓						
Fixed Input Transformation			✓	✓			✓		
Random Input Transformation				✓	✓	✓		✓	
Adversarial Detection			✓					✓	
Network Retraining	✓	✓			✓	✓			✓
Architecture Change		✓				✓			✓

2. Attacks

2.1. Attack Setup

The general setup for an attack in adversarial machine learning can be defined in the following way [25]: The adversary is given a trained classifier F which outputs a class label l for a given input x such that $F(x) = l$. In this paper, the classifiers we consider are deep Convolutional Neural Networks (CNN), and the inputs (x) are images. The goal of the adversary is to create an adversarial example from the original input x by adding a small noise η. The adversarial example that is created is a perturbed version of the original input: $x' = x + \eta$. There are two criteria for the attack to be considered successful:

1. The adversarial example x' must make the classifier produce a certain class label: $F(x') = c$. Here the certain class label c depends on whether the adversary is attempting a targeted, or untargeted type of attack. In a targeted attack c is a specific wrong class label (e.g., a picture of cat MUST be recognized as a dog by the classifier). On the other hand, if the attack is untargeted, the only criteria for c is that it must not be the same as the original class label: $c \neq l$ (e.g., as long as a picture of a cat is labeled by the classifier as anything except a cat, the attack is successful).
2. The noise η used to create the adversarial image x' must be barely recognizable by humans. This constraint is enforced by limiting the size of perturbation η such that the difference between the original input x and the perturbed input x' is less than a certain distance d. This distance d is typically measured [19] using the l_p norm: $\|x' - x\|_p \leq d$

In summary, an attack is considered successful if the classifier produces an output label desired by the adversary $F(x') = c$ and the difference between the original input x and the adversarial sample x' is small enough, $\|x' - x\|_p \leq d$.

2.2. Adversarial Capabilities

In this subsection, we go over what information the adversary can use to create adversarial examples. Specifically, the adversarial model defines what information is available to the attacker to assist them in crafting the perturbation η. In Table 2 we give an overview of the attacks and the adversarial capabilities need to run the attack. Such abilities can be broadly grouped into the following categories:

1. Having knowledge of the trained parameters and architecture of the classifier. For example, when dealing with CNNs (as is the focus of this paper) knowing the architecture means knowing precisely which type of CNN is used. Example CNN

architectures include VGG-16, ResNet56 etc. Knowing the trained parameters for a CNN means the values of the weights and biases of the network (as well as any other trainable parameters) are visible to the attacker [19].

2. Query access to the classifier. If the architecture and trained parameters are kept private, then the next best adversarial capability is having query access to the target model as a black-box. The main concept here is that the adversary can adaptively query the classifier [26] with different inputs to help create the adversarial perturbation η. Query access can come in two forms. In the stronger version, when the classifier is queried, the entire probability score vector is returned (i.e., the softmax output from a CNN). Naturally this gives the adversary more information to work with because the confidence in each label is given. In the weaker version, when the classifier is queried, only the final class label is returned (the index of the score vector with the highest value).
3. Having access to (part of the) training or testing data. In general, for any adversarial machine learning attack, at least one example must be used to start the attack. Hence, every attack requires some input data. However, how much input data the adversary has access to depends on the type of attack (or parameters in the attack). Knowing part or all of the training data used to build the classifier can be especially useful when the architecture and trained parameters of the classifier are not available. This is because the adversary can try to replicate the classifier in the defense, by training their own classifier with the given training data [8].

2.3. Types of Attacks

The types of attacks in machine learning can be grouped based on the capabilities the adversary needs to conduct the attack. We described these different capabilities in Section 2.2. In this section, we describe the attacks and what capabilities the adversary must have to run them.

White-box attacks: Examples of white-box attacks include the Fast Gradient Sign Method (FGSM) [3], Projected Gradient Descent (PGD) [27] and Carlini & Wagner (C&W) [28] to name a few. They require having knowledge of the trained parameters and architecture of the classifier, as well as query access. In white-box attacks like FGSM and PGD, having access to the classifier's trained parameters allows the adversary to use a form of backpropagation. By calculating the gradient with respect to the input, the adversarial perturbation η can be estimated directly. In some defenses, where directly backpropagating on the classifier may not be applicable or yield poor results, it is possible to create attacks tailored to the defense that are more effective. These are referred to as adaptive attacks [22]. In general, white-box attacks and defenses against them have been heavily focused on in the literature. In this paper, our focus is on black-box attacks. Hence, we only give a brief summary of the white-box attacks as mentioned above.

Black-Box Attacks: The biggest difference between white-box and black-box attacks is that black-box attacks lack access to the trained parameters and architecture of the defense. As a result, they need to either have training data to build a synthetic model, or use a large number of queries to create an adversarial example. Based on these distinctions, we can categorize black-box attacks as follows:

1. Query only black-box attacks [26]. The attacker has query access to the classifier. In these attacks, the adversary does not build any synthetic model to generate adversarial examples or make use of training data. Query only black-box attacks can further be divided into two categories: score based black-box attacks and decision based black-box attacks.
    - Score based black-box attacks. These are also referred to as zeroth order optimization based black-box attacks [5]. In this attack, the adversary adaptively queries the classifier with variations of an input x and receives the output from the softmax layer of the classifier $f(x)$. Using $x, f(x)$ the adversary attempts to approximate the gradient of the classifier ∇f and create an adversarial example.

SimBA is an example of one of the more recently proposed score based black-box attacks [29].

- Decision based black-box attacks. The main concept in decision based attacks is to find the boundary between classes using only the hard label from the classifier. In these types of attacks, the adversary does not have access to the output from the softmax layer (they do not know the probability vector). Adversarial examples in these attacks are created by estimating the gradient of the classifier by querying using a binary search methodology. Some recent decision based black-box attacks include HopSkipJump [6] and RayS [30].

2. Model black-box attacks. In model black-box attacks, the adversary has access to part or all of the training data used to train the classifier in the defense. The main idea here is that the adversary can build their own classifier using the training data, which is called the synthetic model. Once the synthetic model is trained, the adversary can run any number of white-box attacks (e.g., FGSM [3], BIM [31], MIM [32], PGD [27], C&W [28] and EAD [33]) on the synthetic model to create adversarial examples. The attacker then submits these adversarial examples to the defense. Ideally, adversarial examples that succeed in fooling the synthetic model will also fool the classifier in the defense. Model black-box attacks can further be categorized based on how the training data in the attack is used:
   - Adaptive model black-box attacks [4]. In this type of attack, the adversary attempts to adapt to the defense by training the synthetic model in a specialized way. Normally, a model is trained with dataset X and corresponding class labels Y. In an adaptive black-box attack, the original labels Y are discarded. The training data X is re-labeled by querying the classifier in the defense to obtain class labels $\hat{Y}$. The synthetic model is then trained on $(X, \hat{Y})$ before being used to generate adversarial examples. The main concept here is that by training the synthetic model with $(X, \hat{Y})$, it will more closely match or adapt to the classifier in the defense. If the two classifiers closely match, then there will (hopefully) be a higher percentage of adversarial examples generated from the synthetic model that fool the classifier in the defense. To run adaptive black-box attacks, access to at least part of the training data and query access to the defense is required. If only a small percentage of the training data is known (e.g., not enough training data to train a CNN), the adversary can also generate synthetic data and label it using query access to the defense [4].
   - Pure black-box attacks [7–10]. In this type of attack, the adversary also trains a synthetic model. However, the adversary does not have query access to make the attack adaptive. As a result, the synthetic model is trained on the original dataset and original labels (X, Y). In essence this attack is defense agnostic (the training of the synthetic model does not change for different defenses).

Table 2. Adversarial machine learning attacks and the adversarial capabilities required to execute the attack. For a full description of these capabilities, see Section 2.2.

	Adversarial Capabilities			
	Training/Testing Data	**Hard Label Query Access**	**Score Based Query Access**	**Trained Parameters**
White-Box		✓	✓	✓
Score Based Black-Box		✓	✓	
Decision Based Black-Box		✓		
Adaptive Black-Box	✓	✓		
Pure Black-Box	✓			

2.4. Our Black-Box Attack Scope

We focus on black-box attacks, specifically the adaptive black-box and pure black-box attacks. Why do we refine our scope in this way? First of all we don't focus on white-box attacks as mentioned in Section 1 as this is well documented in the current literature. In addition, simply showing white-box security is not enough in adversarial machine learning. Due to gradient masking [9], there is a need to demonstrate both white-box and black-box robustness. When considering black-box attacks, as we explained in the previous subsection, there are query only black-box attacks and model black-box attacks. Score based query black-box attacks can be neutralized by a form of gradient masking [19]. Furthermore, it has been noted that a decision based query black-box attack represents a more practical adversarial model [34]. However, even with these more practical attacks there are disadvantages. It has been claimed that decision based black-box attacks may perform poorly on randomized models [19,23]. It has also been shown that even adding a small Gaussian noise to the input may be enough to deter query black-box attacks [35]. Due to their poor performance in the presence of even small randomization, we do not consider query black-box attacks.

Focusing on black-box adversaries and discounting query black-box attacks, leaves model black-box attacks. In our analyses, we first use the pure black-box attack because this attack has no adaptation and no knowledge of the defense. In essence it is the least capable adversary. It may seem counter-intuitive to start with a weak adversarial model. However, by using a relatively weak attack we can see the security of the defense under idealized circumstances. This represents a kind of best-case defense scenario.

The second type of attack we focus on is the adaptive black-box attack. This is the strongest model black-box type of attack in terms of the powers given to the adversary. In our study on this attack, we also vary its strength by giving the adversary different amounts of the original training data (1%, 25%, 50%, 75% and 100%). For the defense, this represents a stronger adversary, one that has query access, training data and an adaptive way to try and tailor the attack to break the defense. In short, we chose to focus on the pure and adaptive black-box attacks. We do this because they do not suffer from the limitations of the query black-box attacks, and they can be used as an efficient and nearly universally applicable security test.

3. Defense Summaries, Metrics and Datasets

In this paper we investigate 9 recent defenses, Barrage of Random Transforms (BaRT) [14], End-to-End Image Compression Models (ComDefend) [13], The Odds are Odd (Odds) [17], Feature Distillation (FD) [18], Buffer Zones (BUZz) [24], Ensemble Diversity (ADP) [11], Distribution Classifier (DistC) [16], Error Correcting Output Codes (ECOC) [12] and K-Winner-Take-All (k-WTA) [15].

In Table 1. , we decompose these defenses into the underlying methods they use to try to achieve security. This is by no means the only way these defenses can be categorized and the definitions here are not absolute. We merely provide this hierarchy to provide a basic overview and show common defense themes. The defense themes are categorized as follows:

1. Multiple models—The defense uses multiple classifiers' for prediction. The classifiers outputs may be combined through averaging (i.e., ADP), majority voting (BUZz) or other methods (ECOC).
2. Fixed input transformation—A non-randomized transformation is applied to the input before classification. Examples of this include, image denoising using an autoencoder (Comdefend), JPEG compression (FD) or resizing and adding (BUZz).
3. Random input transformation—A random transformation is applied to the input before classification. For example both BaRT and DistC randomly select from multiple different image transformations to apply at run time.

4. Adversarial detection—The defense outputs a null label if the sample is considered to be adversarially manipulated. Both BUZz and Odds employ adversarial detection mechanisms.
5. Network retraining—The network is retrained to accommodate the implemented defense. For example BaRT and BUZz require network retraining to achieve acceptable clean accuracy. This is due to the significant transformations both defenses apply to the input. On the other hand, different architectures mandate the need for network retraining like in the case of ECOC, DistC and k-WTA. Note network retraining is different from adversarial training. In the case of adversarial training, it is a fundamentally different technique in the sense that it can be combined with almost every defense we study. Our interest however is not to make each defense as strong as possible. Our aim is to understand how much each defense improves security on its own. Adding in techniques beyond what the original defense focuses on is essentially adding in confounding variables. It then becomes even more difficult to determine from where security may arise. As a result, we limit the scope of our defenses to only consider retraining when required and do not consider adversarial training.
6. Architecture change—A change in the architecture which is made solely for the purposes of security. For example k-WTA uses different activation functions in the convolutional layers of a CNN. ECOC uses a different activation function on the output of the network.

3.1. Barrage of Random Transforms

Barrage of Random Transforms (BaRT) [14] is a defense based on applying image transformations before classification. The defense works by randomly selecting a set of transformations and a random order in which the image transformations are applied. In addition, the parameters for each transformation are also randomly selected at run time to further enhance the entropy of the defense. Broadly speaking, there are 10 different image transformation groups: JPEG compression, image swirling, noise injection, Fourier transform perturbations, zooming, color space changes, histogram equalization, grayscale transformations and denoising operations.

Prior security studies: In terms of white-box analyses, the original BaRT paper tests PGD and FGSM. They also test a combined white-box attack designed to deal with randomization. This combinational white-box attack is composed of expectation over transformation [36] and backward pass differentiable approximation [9]. No analysis of the BaRT defense with black-box adversaries is done.

Why we selected it: In [19], they claim gradient free attacks (i.e., black-box attacks) most commonly fail due to randomization. Therefore BaRT is a natural candidate to test for black-box security. Also in the original paper, BaRT is only tested with ImageNet. We wanted to see if this defense could be expanded to work on other datasets.

3.2. End-to-End Image Compression Models

ComDefend [13] is a defense where image compression/reconstruction is done using convolutional autoencoders before classification. ComDefend consists of two modules: a compression convolutional neural network (ComCNN) and a reconstruction convolutional neural network (RecCNN). The compression network transforms the input image into a compact representation by compressing the original 24 bit pixels into compact 12 bit representations. Gaussian noise is then added to the compact representation. Decompression is then done using the reconstruction network and the final output is fed to the classifier. In this defense, retraining of the classifier on reconstructed input data is not required.

Prior security studies: White-box attacks such as FGSM, BIM and C&W are run on ComDefend. They also vary their threat model between using the l_∞ norm and l_2 norm to create white-box adversarial examples that have different constraints. No black-box attacks are ever presented for the defense.

Why we selected it: Other autoencoder defenses have fared poorly [37]. It is worth studying new autoencoder defenses to see if they work, or if they face the same vulnerabilities as older defense designs. Since ComDefend [13] does not study black-box adversaries, our analysis also provides new insight on this defense.

3.3. The Odds Are Odd

The Odds are Odd [17] is a defense based on a statistical test. This test is motivated by the following observation: the behaviors of benign and adversarial examples are different at the logits layer (i.e., the input to the softmax layer). The test works as follows: for a given input image, multiple copies are created and a random noise is added to each copy. This creates multiple random noisy images. The defense calculates the logits values of each noisy image and uses them as the input for the statistical test.

Prior security studies: In the original Odds paper, the statistical test is done in conjunction with adversarial examples generated using PGD (a white-box attack). Further white-box attacks on the Odds were done in [22]. The authors in [22] use PGD and a custom objective function to show the flaws in the statistical test under white-box adversarial model. To the best of our knowledge, no work has been done on the black-box security of the Odds defense.

Why we selected it: In [22], they mention that Odds is based on the common misconception that building a test for certain adversarial examples will then work for all adversarial examples. However, in the black-box setting this still brings up an interesting question: if the attacker is unaware of the type of test, can they still adaptively query the defense and come up with adversarial examples that circumvent the test?

3.4. Feature Distillation

Feature Distillation (FD) implements a unique JPEG compression and decompression technique to defend against adversarial examples. Standard JPEG compression/decompression preserves low frequency components. However, it is claimed in [18] that CNNs learn features which are based on high frequency components. Therefore, the authors propose a compression technique where a smaller quantization step is used for CNN accuracy-sensitive frequencies and a larger quantization step is used for the remaining frequencies. The goal of this technique is two-fold. First, by maintaining high frequency components, the defense aims to preserve clean accuracy. Second, by reducing the other frequencies, the defense tries to eliminate the noise that make samples adversarial. Note this defense does have some parameters which need to be selected through experimentation. For the sake of brevity, we provide the experiments for selecting these parameters in the Appendix A.

Prior security studies: In the original FD paper, the authors test their defense against standard white-box attacks like FGSM, BIM and C&W. They also analyze their defense against the backward pass differentiable approximation [9] white-box attack. In terms of black-box adversaries, they do test a very simple black-box attack. In this attack, samples are generated by first training a substitute model. However, this black-box adversary cannot query the defense to label its training data, making it extremely limited. Under our attack definitions, this is not an adaptive black-box attack.

Why we selected it: A common defense theme is the utilization of multiple image transformations like in the case of BaRT, BUZz and DistC. However, this requires a cost in the form of network retraining and/or clean accuracy. If a defense could use only one type of transformation (as done in FD), it may be possible to significantly reduce those costs. To the best of our knowledge, so far no single image transformation has accomplished this, which makes the investigation of FD interesting.

3.5. Buffer Zones

Buffer Zones (BUZz) employs a combination of techniques to try and achieve security. The defense is based on unanimous majority voting using multiple classifiers. Each

classifier applies a different fixed secret transformation to its input. If the classifiers are unable to agree on a class label, the defense marks the input as adversarial. The authors also note that a large drop in clean accuracy is incurred due to the number of defense techniques employed.

Prior security studies: BUZz is the only defense on our list that experiments with a similar black-box adversary (one that has access to the training data and can query the defense). However, as we explain below, their study has room to further be expanded upon.

Why we selected it: We selected this defense to study because it specifically claims to deal with the exact adversarial model (adaptive black-box) that we work with. However, in their paper they only use a single strength adversary (i.e., one that uses the entire training dataset). We test across multiple strength adversaries (see Section 5) to see how well their defense holds up.

3.6. Improving Adversarial Robustness via Promoting Ensemble Diversity

Constructing ensembles of enhanced networks is one defense strategy to improve the adversarial robustness of classifiers. However, in an ensemble model, the lack of interaction among individual members may cause them to return similar predictions. This defense proposes a new notion of ensemble diversity by promoting the diversity among the predictions returned by members of an ensemble model using an adaptive diversity promoting (ADP) regularizer, which works with a logarithm of ensemble diversity term and an ensemble entropy term [11]. The ADP regularizer helps non-maximal predictions of each ensemble member to be mutually orthogonal, while the maximal prediction is still consistent with the correct label. This defense employs a different training procedure where the ADP regularizer is used as the penalty term and the ensemble network is trained interactively.

Prior security studies: ADP has widely been studied in the context of white-box security in [11,22,23]. In the original paper in which ADP was proposed, they tested the defense against white-box attacks like FGSM, BIM, PGD, C&W and EAD. In [22], they use different attack parameters (more iterations) in order to show the defense was not as robust as previously thought. These results are further supported by white-box attacks done on ADP and reported in [23].They use FGSM, BIM and MIM (as well as others white-box attacks) in [23] to further analyze the robustness of ADP. They also test some black-box attacks on ADP in [23], but these attacks are transfer based and boundary based. They do not test our adaptive type of black-box attack in [23].

Why we selected it: It has been shown that adversarial samples can have high transferability [4]. Model black-box attacks have a basic underlying assumption: adversarial samples that fool the synthetic model will also fool the defense. ADP trains networks to specifically enhance diversity which could mitigate the transferability phenomena. If the adversarial transferability between networks is indeed really mitigated, then black-box attacks should not be effective.

3.7. Enhancing Transformation-Based Defenses against Adversarial Attacks with a Distribution Classifier

The basic idea of this defense is that if the input is adversarial, basing the predicted class on the softmax output may yield a wrong result. Instead in this defense the input is randomly transformed multiple times, to create many different inputs. Each transformed input yields a softmax output from the classifier. Prediction is then done on the distribution of the softmax outputs [16]. To classify the softmax distributions, a separate distributional classifier is trained.

Prior security studies: In [16], white-box attacks on the defense were done using methods like FGSM, IFGSM and C&W. Query only black-box attacks were also studied, but by our definition, no adaptive black-box attacks were ever considered for this defense.

Why we selected it: In [16], the defense is tested with query only black-box attacks as we previously mentioned. However, it does not test any model black-box attacks. This defense is built on [38] which was initially a promising randomization defense that was

broken in [9]. Whether the combination of a new classification scheme and randomization can achieve model black-box security is an open question.

3.8. Error Correcting Output Codes

The Error Correcting Output Codes (ECOC) [12] defense uses the idea of coding theory and changes the output representation in a network to codewords. There are three main ideas of the defense. First, is the use of a special sigmoid decoding activation function instead of the softmax function. This function allocates the non-trivial volume in logit space to uncertainty. This makes the attack surface smaller to the attacker who tries to craft adversarial examples. Second, a larger Hamming distance between the codewords is used to increase the distance between two high-probability regions for a class in logit space. This forces the adversary to use larger perturbations in order to succeed. Lastly, the correlation between outputs is reduced by training an ensemble model.

Prior security studies: In [12], the authors test ECOC against white-box attacks like PGD and C&W. A further white-box analysis of ECOC is done in [22], where PGD with a custom loss function is used. Through this modified PGD, the authors in [22] are able to significantly reduce the robustness of the ECOC defense in the white-box setting. No black-box analyses of ECOC are ever considered in [22] or [12].

Why we selected it: Much like ADP, this method relies on an ensemble of models. However unlike ADP, this defense is based on coding theory and the original paper does not consider a black-box adversary. The authors in [22] were only able to come up with an effective attack on ECOC in the white-box setting. Therefore, exploring the black-box security of this defense is of interest.

3.9. k-Winner-Take-All

In k-Winner-Take-All (k-WTA) [15] a special activation function is used that is C^0 discontinuous. This activation function mitigates white-box attacks through gradient masking. The authors claim this architecture change is nearly free in terms of the drop in clean accuracy.

Prior security studies: In the original k-WTA paper [15] the authors test their defense against white-box attacks like PGD, MIM and C&W. They also test against a weak transfer based black-box attack that is not adaptive. They do not consider a black-box adversary that has access to the entire training dataset and query access like we assume in our adversarial model. Further white-box attacks against k-WTA were done in [22]. The authors in [22] used PGD with more iterations (400) and also considered a special averaging technique to better estimate the gradient of the network.

Why we selected it: The authors of the defense claim that k-WTA performs better under model black-box attacks than networks that use ReLU activation functions. If this claim is true, this would be the first defense in which gradient masking could mitigate both white-box and black-box attacks. In [22], they already showed the vulnerability of this defense to white-box attacks. Additionally, in [22] they hypothesize a black-box adversary that queries the network may work well against this defense, but do not follow up with any experiments. Therefore, this indicates k-WTA still lacks proper black-box security experiments and analyses.

3.10. Defense Metric

In this paper, our goal is to demonstrate what kind of gain in security can be achieved by using each defense against a black-box adversary. Our aim is not to claim any defense is broken. To measure the improvement in security, we use a simple metric: Defense accuracy improvement.

Defense accuracy improvement is the percent increase in correctly recognized adversarial examples gained when implementing the defense as compared to having no defense. The formula for defense accuracy improvement for the ith defense is defined as:

$$A_i = D_i - V \tag{1}$$

We compute the defense accuracy improvement A_i by first conducting a specific black-box attack on a vanilla network (no defense). This gives us a vanilla defense accuracy score V. The vanilla defense accuracy is the percent of adversarial examples the vanilla network correctly identifies. We run the same attack on a given defense. For the ith defense, we will obtain a defense accuracy score of D_i. By subtracting V from D_i we essentially measure how much security the defense provides as compared to not having any defense on the classifier.

For example if $V \approx 99\%$, then the defense accuracy improvement A_i can be 0, but at the very least should not be negative. If $V \approx 85\%$, then a defense accuracy improvement of 10% may be considered good. If $V \approx 40\%$, then we want at least a 25% defense accuracy improvement, for the defense to be considered effective (i.e. the attack fails more than half of the time when the defense is implemented). While sometimes an improvement is not possible (e.g. when $V \approx 99\%$) there are many cases where attacks works well on the undefended network and hence there are places where large improvements can be made. Note to make these comparisons as precise as possible, almost every defense is built with the same CNN architecture. Exceptions to this occur in some cases, which we fully explain in the Appendix A.

3.11. Datasets

In this paper, we test the defenses using two distinct datasets, CIFAR-10 [39] and Fashion-MNIST [40]. CIFAR-10 is a dataset comprised of 50,000 training images and 10,000 testing images. Each image is $32 \times 32 \times 3$ (a 32×32 color image) and belongs to 1 of 10 classes. The 10 classes in CIFAR-10 are airplane, car, bird, cat, deer, dog, frog, horse, ship and truck. Fashion-MNIST is a 10 class dataset with 60,000 training images and 10,000 test images. Each image in Fashion-MNIST is 28×28 (grayscale image). The classes in Fashion-MNIST correspond to t-shirt, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag and ankle boot.

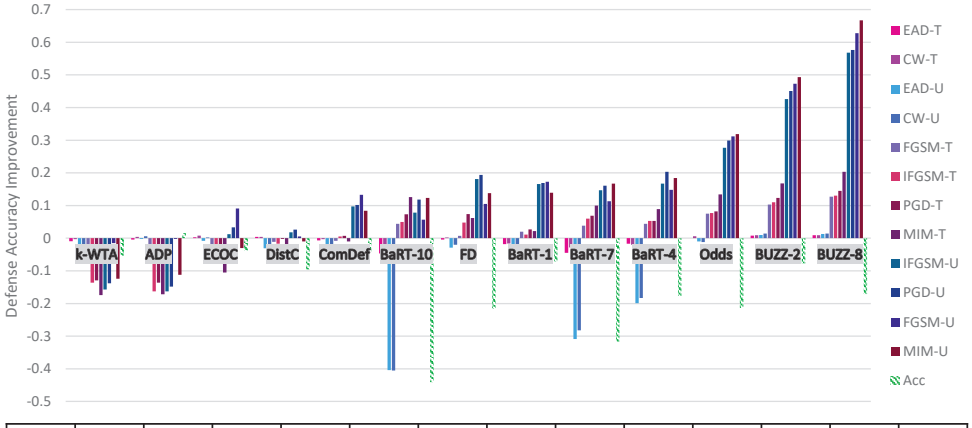
Why we selected them: We chose the CIFAR-10 defense because many of the existing defenses had already been configured with this dataset. Those defenses already configured for CIFAR-10 include ComDefend, Odds, BUZz, ADP, ECOC, the distribution classifier defense and k-WTA. We also chose CIFAR-10 because it is a fundamentally challenging dataset. CNN configurations like ResNet do not often achieve above 94% accuracy on this dataset [41]. In a similar vein, defenses often incur a large drop in clean accuracy on CIFAR-10 (which we will see later in our experiments with BUZz and BaRT for example). This is because the amount of pixels that can be manipulated without hurting classification accuracy is limited. For CIFAR-10, each image only has in total 1024 pixels. This is relatively small when compared to a dataset like ImageNet [42], where images are usually $224 \times 224 \times 3$ for a total of 50,176 pixels (49 times more pixels than CIFAR-10 images). In short, we chose CIFAR-10 as it is a challenging dataset for adversarial machine learning and many of the defenses we test were already configured with this dataset in mind.

For Fashion-MNIST, we primarily chose it for two main reasons. First, we wanted to avoid a trivial dataset on which all defenses might perform well. For example, CNNs can already achieve a clean accuracy of 99.7% on a dataset like MNIST [40]. Testing on such types of datasets would not work towards the main aim of our paper, which is to distinguish defenses that perform significantly better in terms of security and clean accuracy. The second reason we chose Fashion-MNIST is for its differences from CIFAR-10. Specifically, Fashion-MNIST is a non-color dataset and contains very different types of images than CIFAR-10. In addition, many of the defenses we tested were not originally designed for Fashion-MNIST. This brings up an interesting question, can previously proposed defenses be readily adapted to work with different datasets. To summarize, we chose Fashion-MNIST for its difficult to learn and its differences from CIFAR-10.

4. Principal Experimental Results

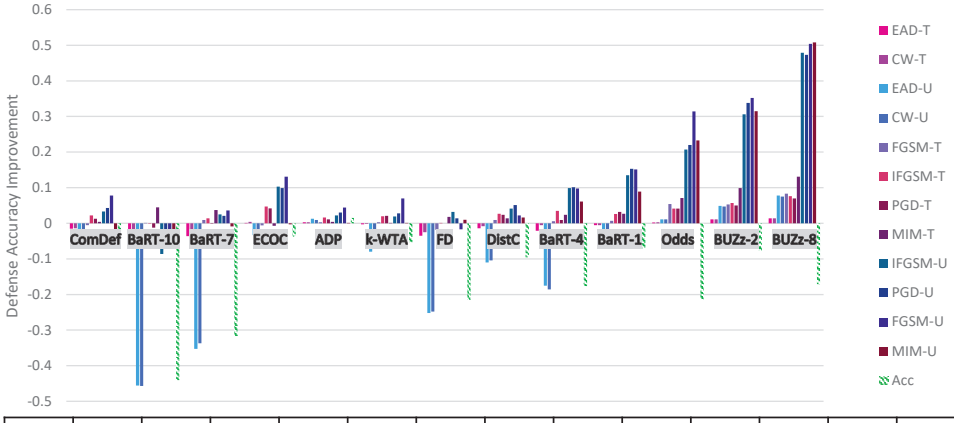
In this section, we conduct experiments to test the black-box security of the 9 defenses. We measure the results using the metric defense accuracy improvement (see Section 3.10). For each defense, we test it under a pure black-box adversary, and five different strength adaptive black-box adversaries. The strength of the adaptive black-box adversary is determined by how much of the original training dataset they are given access to (either 100%, 75%, 50%, 25% or 1%). For every adversary, once the synthetic model is trained, we use 6 different methods (FGSM [3], BIM [31], MIM [32], PGD [27], C&W [28] and EAD [33]) to generate adversarial examples. We test both targeted and untargeted styles of attack. In these experiments we use the l_∞ norm with maximum perturbation $\epsilon = 0.05$ for CIFAR-10 and $\epsilon = 0.1$ for Fashion-MNIST. Further attack details can be found in our Appendix A.

Before going into a thorough analysis of our results, we briefly introduce the figures and tables that show our experimental results. Figures 1 and 2 illustrate the defense accuracy improvement of all the defenses under a 100% strength adaptive black-box adversary (Figure 1) and a pure black-box adversary (Figure 2) for the CIFAR-10 dataset. Likewise, for Fashion-MNIST, Figure 3 shows the defense accuracy improvement under a 100% strength adaptive black-box adversary and Figure 4 shows the defense accuracy improvement under a pure black-box adversary. For each of these figures, we report the vanilla accuracy numbers in a chart below the graph. Figure 5 through Figure 6 show the relationship between the defense accuracy and the strength of the adversary (how much training data the adversary has access to). Figure 5 through Figure 6 show this relationship for each defense, on both CIFAR-10 and Fashion-MNIST. The corresponding values for the figures are given in Table A4 through Table A15.



	EAD-T	CW-T	EAD-U	CW-U	FGSM-T	IFGSM-T	PGD-T	MIM-T	IFGSM-U	PGD-U	FGSM-U	MIM-U	Acc
Vanilla	0.991	0.991	0.987	0.986	0.866	0.861	0.848	0.777	0.387	0.374	0.334	0.259	0.9278

Figure 1. CIFAR-10 adaptive black-box attack on each defense. Here the U/T refers to whether the attack is untargeted/targeted. Negative values means the defense performs worse than the no defense (vanilla) case. The Acc value refers to the drop in clean accuracy incurred by implementing the defense. The chart below the graph gives the vanilla defense accuracy numbers.



	EAD-T	CW-T	EAD-U	CW-U	FGSM-T	IFGSM-T	PGD-T	MIM-T	IFGSM-U	PGD-U	FGSM-U	MIM-U	Acc
Vanilla	0.984	0.984	0.919	0.923	0.902	0.917	0.924	0.853	0.453	0.455	0.443	0.384	0.9278

Figure 2. CIFAR-10 pure black-box attack on each defense. Here the U/T refers to whether the attack is untargeted/targeted. Negative values means the defense performs worse than the no defense (vanilla) case. The Acc value refers to the drop in clean accuracy incurred by implementing the defense. The chart below the graph gives the vanilla defense accuracy numbers. For all the experimental numbers see Table A4.

Considering the range of our experiments (9 defenses, 6 adversarial models, 6 methods to generate adversarial samples and 2 datasets), it is infeasible to report all the results and experimental details in just one section. Instead, we organize our experimental analysis as follows. In this section, we present the most pertinent results in Figures 1 and 3 and give the principal takeaways. For readers interested in a specific defense or attack results, in Section 5 we give a comprehensive break down of the results for each defense, dataset and attack. For anyone wishing to recreate our experimental results, we give complete implementation details for every attack and defense in the Appendix A.

Principal Results

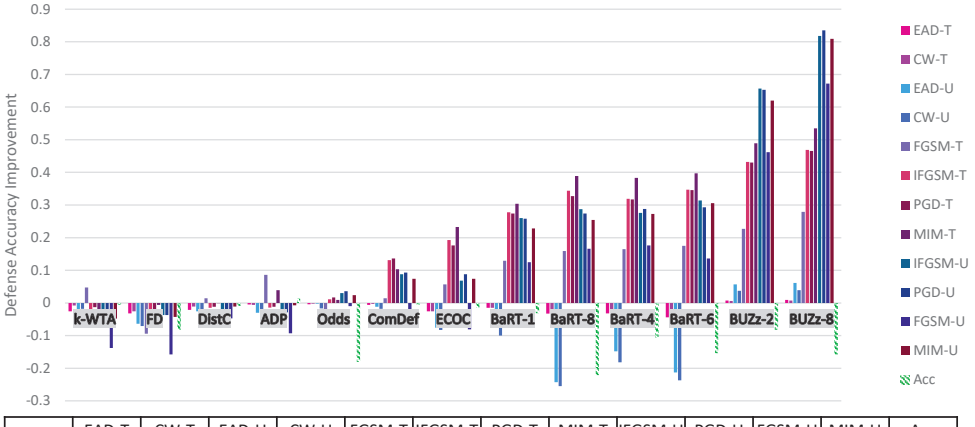
1. **Marginal or negligible improvements over no defense**: Figure 1 shows the defense results for CIFAR-10 with a 100% strength adaptive black-box adversary. In this figure, we can clearly see 7 out of 9 defenses give marginal (less than 25%) increases in defense accuracy for any attack. BUZz and the Odds defense are the only ones to break this trend for CIFAR-10. For example, BUZz-8 gives a 66.7% defense accuracy improvement for the untargeted MIM attack. Odds gives a 31.9% defense accuracy improvement for the untargeted MIM attack. Likewise, for Fashion-MNIST again, 7 out of 9 defenses give only marginal improvements (see Figure 3). BUZz and BaRT are the exceptions for this dataset.
2. **Security is not free (yet)**: Thus far, no defense we experimented with that offers significant (greater than 25% increase) improvements comes for free. For example, consider the defenses that give significant defense accuracy improvements. BUZz-8 drops the clean accuracy by 17% for CIFAR-10. BaRT-6 drops the clean accuracy by 15% for Fashion-MNIST. As defenses improve, we expect to see this trade-off between clean accuracy and security become more favorable. However, our experiments show we have not reached this point with the current defenses.
3. **Common defense mechanisms**: It is difficult to decisively prove any one defense mechanism guarantees security. However, among the defenses that provide more than marginal improvements (Odds, BUZz and BaRT), we do see common defense trends. Both Odds and BUZz use adversarial detection. This indirectly deprives the adaptive black-box adversary of training data. When an input sample is marked as adversarial, the black-box attacker cannot use it to train the synthetic model. This is because the synthetic model has

no adversarial class label. It is worth noting that in the Appendix A, we also argue why a synthetic model should not be trained to output an adversarial class label.

Along similar lines, both BaRT and BUZz offer significant defense accuracy improvements for Fashion-MNIST. Both employ image transformations so jarring that the classifier must be retrained on transformed data. The experiments show that increasing the number of the transformations only increases security up to a certain point though. For example, BaRT-8 does not perform better than BaRT defenses that use less image transformations (see BaRT-6 and BaRT-4 in Figure 3).

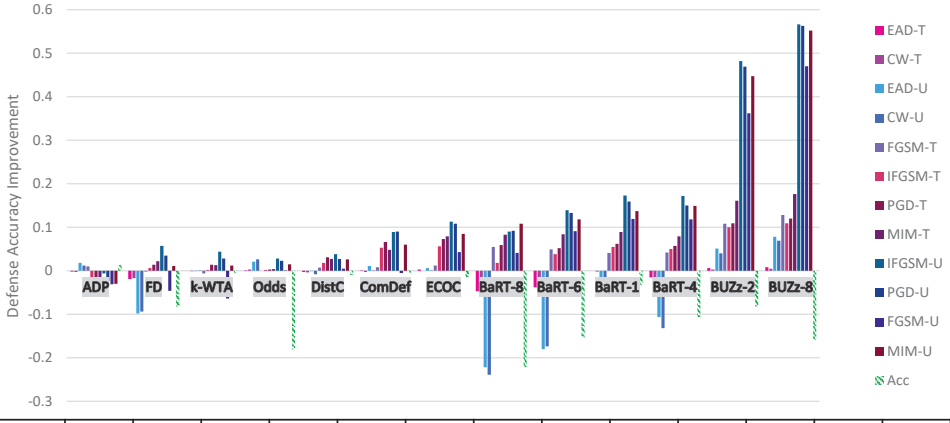
4. **Adaptive and pure black-box follow similar trends.** In Figures 2 and 4 we show results for the pure black-box attack for CIFAR-10 and Fashion-MNIST. Just like for the adaptive black-box attack, we see similar trends in terms of which defenses provide the highest security gains. For CIFAR-10, the defenses that give at least 25% greater defense accuracy than the vanilla defense include BUZz and Odds. For Fashion-MNIST, the only defense that gives this significant improvement is BUZz.

5. **Future defense analyses should be broad**: From our first point in this subsection, it is clear that a majority of these defenses give marginal improvements or less. This brings up an important question, what impact does our security study have for future defenses? The main lesson is future defense designers need to test against a broad spectrum of attacks. From the literature, we see the majority of the 9 defenses already considered white-box attacks like PGD or FGSM and some weak black-box attacks. However, in the face of adaptive attacks, these defenses perform poorly. Future defense analyses at the very least need white-box attacks *and* adaptive black-box attacks. By providing our paper's results and code we hope to help future defense designers perform these analyses and advance the field of adversarial machine learning.



	EAD-T	CW-T	EAD-U	CW-U	FGSM-T	IFGSM-T	PGD-T	MIM-T	IFGSM-U	PGD-U	FGSM-U	MIM-U	Acc
Vanilla	0.991	0.993	0.939	0.961	0.707	0.529	0.531	0.46	0.123	0.118	0.234	0.111	0.9356

Figure 3. Fashion-MNIST adaptive black-box attack on each defense. Here the U/T refers to whether the attack is untargeted/targeted. Negative values means the defense performs worse than the no defense (vanilla) case. The Acc value refers to the drop in clean accuracy incurred by implementing the defense. The chart below the graph gives the vanilla defense accuracy numbers.



	EAD-T	CW-T	EAD-U	CW-U	FGSM-T	IFGSM-T	PGD-T	MIM-T	IFGSM-U	PGD-U	FGSM-U	MIM-U	Acc
Vanilla	0.992	0.995	0.905	0.914	0.865	0.889	0.879	0.817	0.363	0.374	0.429	0.351	0.9356

Figure 4. Fashion-MNIST pure black-box attack on each defense. Here the U/T refers to whether the attack is untargeted/targeted. Negative values means the defense performs worse than the no defense (vanilla) case. The Acc value refers to the drop in clean accuracy incurred by implementing the defense. The chart below the graph gives the vanilla defense accuracy numbers. For all the experimental numbers see Table A10.

5. Individualized Experimental Defense Results

In the previous section, we discussed the overarching themes represented in the adaptive black-box attack experimental results. In this section, we take a more fine grained approach and consider each defense individually.

Both the 100% adaptive black-box and pure black-box attack have access to the entire original training dataset. The difference between them lies in the fact that the adaptive black-box attack can generate synthetic data, and label the training data by querying the defense. Since both attacks are similar in terms of how much data they start with, a question arises. How effective is the attack if the attacker doesn't have access to the full training dataset? In the following subsections, we seek to answer that question by considering each defense under a variable strength adversary in the adaptive black-box setting. Specifically we test out adversaries that can query the defense but only have 75%, 50%, 25% or 1% of the original training dataset.

To simplify things with the variable strength adaptive black-box adversary, we only consider the untargeted MIM attack for generating adversarial examples. We use the MIM attack because it is the best performing attack on the vanilla (no defense) network for both datasets. Therefore, this attack represents the place where the most improvement in security can be made. For the sake of completeness, we do report all the defense accuracies for all six types of attacks for the variable strength adaptive black-box adversaries in the tables at the end of this section.

After discussing defense results, we also present brief experiment and discussion on why the adaptive black-box attack is actually considered *adaptive*. We do this by comparing the attack success rate of the adaptive attack to the non-adaptive pure black-box attack while simultaneously fixing the underlying method to generate adversarial examples, fixing the dataset and fixing the amount of training data available to the attacker.

5.1. Barrage of Random Transforms Analysis

The adaptive black-box attack with variable strength for BaRT defenses is shown in Figure 5. There are several interesting observations that can be made about this defense. First, for CIFAR-10, the maximum transformation defense (BaRT-10) actually performs worse than the vanilla defense in most cases. BaRT-1, BaRT-4 and BaRT-7 perform approxi-

mately the same as the vanilla defense. These statements hold except for the 100% strength adaptive black-box adversary. Here, all BaRT defenses show a 12% or greater improvement over the vanilla defense.

Where as the performance of BaRT is rather varied for CIFAR-10, for Fashion-MNIST this is not the case. All BaRT defenses show improvement for the MIM attack for adversaries with 25% strength or greater.

When examining the results of BaRT on CIFAR-10 and Fashion-MNIST, we see a clear discrepancy in performance. One possible explanation is as follows: the image transformations in a defense must be selected in a way that does not greatly impact the original clean accuracy of the classifier. In the case of BaRT-10 (the maximum number of transformations) for CIFAR-10, it performs much worse than the vanilla case. However, BaRT-8 for Fashion-MNIST (again the maximum number of transformations) performs much better than the vanilla case. If we look at the clean accuracy of BaRT-10, it is approximately 48% on CIFAR-10. This is a drop of more than 40% as compared to the vanilla clean accuracy. For BaRT-8, the clean accuracy is approximately 72% on Fashion-MNIST which is a drop of about 21%. Here we do not use precise numbers when describing the clean accuracy because as a randomized defense, the clean accuracy may drop or rise a few percentage points every time the test set is evaluated.

From the above stated results, we can make the following conclusion: A defense that employs random image transformations cannot be applied naively to every dataset. The set of image transformations must be selected per dataset in such a manner that the clean accuracy is not drastically impacted. In this sense, while random image transformations may be a promising defense direction, it seems they may need to be designed on a per dataset basis.

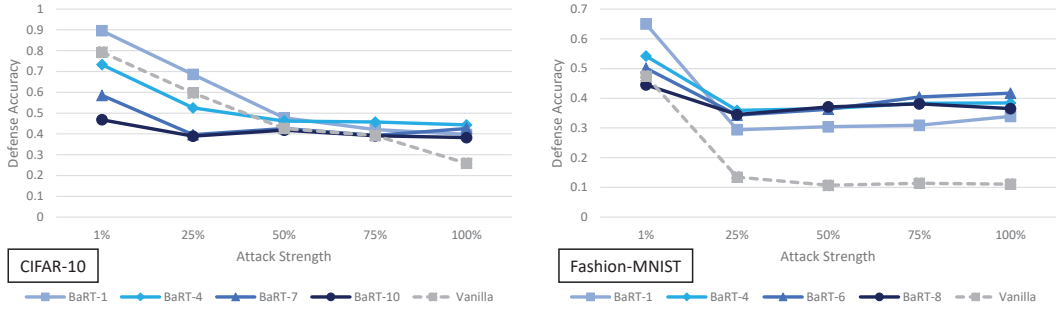


Figure 5. Defense accuracy of barrage of random transforms defense on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

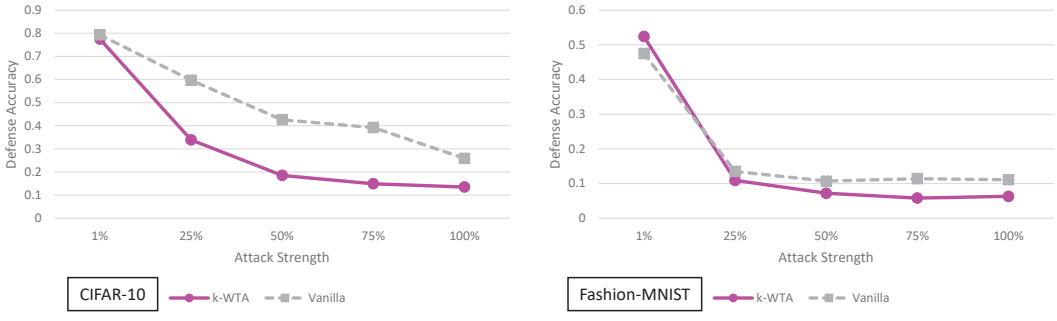


Figure 6. Defense accuracy of the k-Winners-Take-All defense on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

5.2. End-to-End Image Compression Models Analysis

The adaptive black-box attack with variable strength for ComDefend is shown in Figure 7. For CIFAR-10, we see the defense performs very close to the vanilla network (and sometimes slightly worse). On the other hand, for Fashion-MNIST, the defense does offer a modest average defense accuracy improvement of 8.84% across all adaptive black-box adversarial models.

In terms of understanding the performance of ComDefend, it is important to note the following: In general it has been shown that more complex architectures (e.g., deeper networks) can better resist transfer based adversarial attacks [10]. In essence an autoencoder/decoder setup can be viewed as additional layers in the CNN and hence a more complex model. Although this concept was shown for ImageNet [10], it may be a phenomena that occurs in other datasets as well.

This more complex model can partially explain why ComDefend slightly outperforms the vanilla defense in most cases. In short, a slightly more complex model is slightly more difficult to learn and attack. Of course this begs the question, if a more complex model yields more security, why does the model complexity even have come from an autoencoder/decoder? Why not use ResNet164 or ResNet1001?

These are all valid questions which are possible directions of future studies. While ComDefend itself does not yield significant (greater than 25%) improvements in security, it does bring up an interesting question: Under a black-box adversarial model, to what extent can increasing model complexity also increase defense accuracy? We leave this as an open question for possible future work.

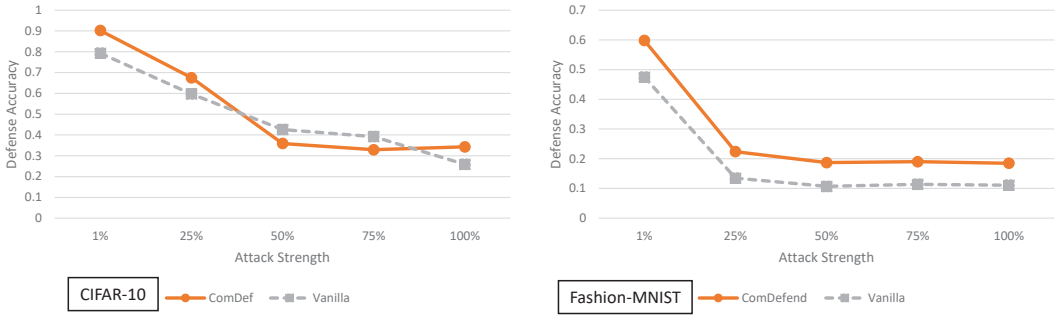


Figure 7. Defense accuracy of ComDefend on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

5.3. The Odds Are Odd Analysis

In Figure 8, the adaptive black-box attack with different strengths is shown for the Odds defense. For CIFAR-10 the Odds has an average improvement of 19.3% across all adversarial models. However, for Fashion-MNIST the average improvement over the vanilla model is only 2.32%. As previously stated, this defense relies on the underlying assumption that creating a test for one set of adversarial examples will then generalize to all adversarial examples.

When the test used in the Odds does provide security improvements (as in the case for CIFAR-10), it does highlight one important point. If the defense can mark some samples as adversarial, it is possible to deprive the adaptive black-box adversary of data to train the synthetic model. This in turn weakens the overall effectiveness of the adaptive black-box attack. We stress however that this occurs only when the test is accurate and does not greatly hurt the clean prediction accuracy of the classifier.

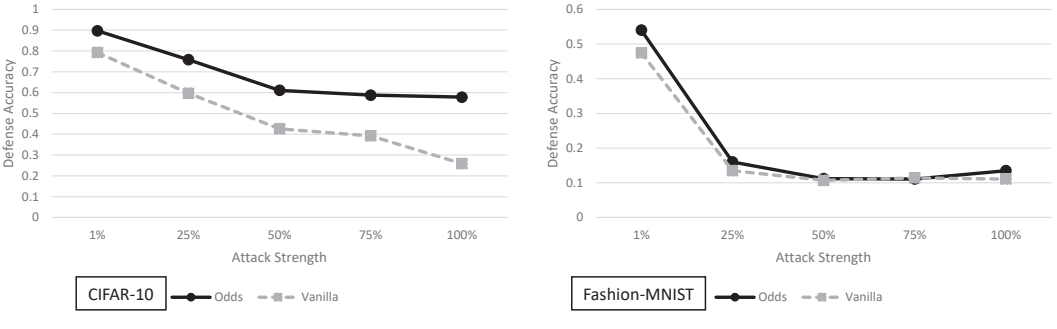


Figure 8. Defense accuracy of the odds defense on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

5.4. Feature Distillation Analysis

Figure 9 shows the adaptive black-box with a variable strength adversary for the feature distillation defense. In general feature distillation performs worse than the vanilla network for all Fashion-MNIST adversaries. It performs worse or roughly the same for all CIFAR-10 adversaries, except for the 100% case where it shows a marginal improvement of 13.8%.

In the original feature distillation paper the authors claim that they test a black-box attack. However, our understanding of their black-box attack experiment is that the synthetic model used in their experiment was not trained in an adaptive way. To be specific, the adversary they use does not have query access to the defense. Hence, this may explain why when an adaptive adversary is considered, the feature distillation defense performs roughly the same as the vanilla network.

As we stated in the main paper, it seems unlikely a single image transformation would be capable of providing significant defense accuracy improvements. Thus far, the experiments on feature distillation support that claim for the JPEG compression/decompression transformation. The study of this image transformation and the defense are still very useful. The idea of JPEG compression/decompression when combined with other image transformations may still provide a viable defense, similar to what is done in BaRT.

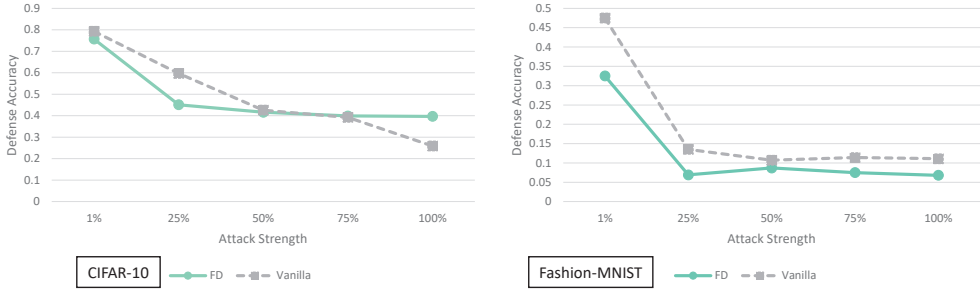


Figure 9. Defense accuracy of feature distillation on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

5.5. Buffer Zones Analysis

The results for the buffer zone defense in regards to the adaptive black-box variable strength adversary are given in Figure 10. For all adversaries, and all datasets we see an improvement over the vanilla model. This improvement is quite small for the 1% adversary for the CIFAR-10 dataset at only a 10.3% increase in defense accuracy for BUZz-2. However, the increases are quite large for stronger adversaries. For example, the difference between the BUZz-8 and vanilla model for the Fashion-MNIST full strength adversary is 80.9%.

As we stated earlier, BUZz is one of the defenses that does provide more than marginal improvements in defense accuracy. This improvement comes at a cost in clean accuracy however. To illustrate: BUZz-8 has a drop of 17.13% and 15.77% in clean testing accuracy for CIFAR-10 and Fashion-MNIST respectively. An ideal defense is one in which the clean accuracy is not greatly impacted. In this regard, BUZz still leaves much room for improvement. The overall idea presented in BUZz of combining adversarial detection and image transformations does give some indications of where future black-box security may lie, if these methods can be modified to better preserve clean accuracy.

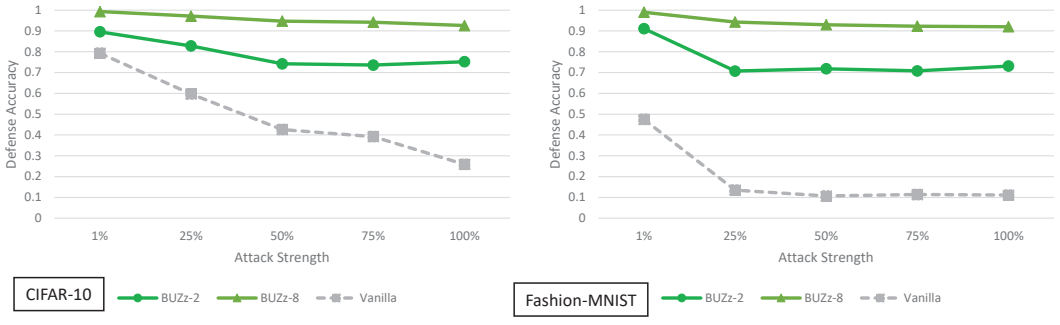


Figure 10. Defense accuracy of the buffer zones defense on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

5.6. Improving Adversarial Robustness via Promoting Ensemble Diversity Analysis

The ADP defense and its performance under various strength adaptive black-box adversaries is shown in Figure 11. For CIFAR-10, the defense does slightly worse than the vanilla model. For Fashion-MNIST, the defense does almost the same as the vanilla model.

It has also been shown before in [24] that using multiple vanilla networks does not yield significant security improvements against a black-box adversary. The adaptive black-box attacks presented here support these claims when it comes to the ADP defense. At this time we do not have an adequate explanation as to why the ADP defense performs worse on CIFAR-10 given its clean accuracy is actually slightly higher than the vanilla model. We would expect slightly higher clean accuracy would result in slightly higher defense accuracy but this is not the case. Overall though, we do not see significant improvements in defense accuracy when implementing ADP against adaptive black-box adversaries of varying strengths for CIFAR-10 and Fashion-MNIST.

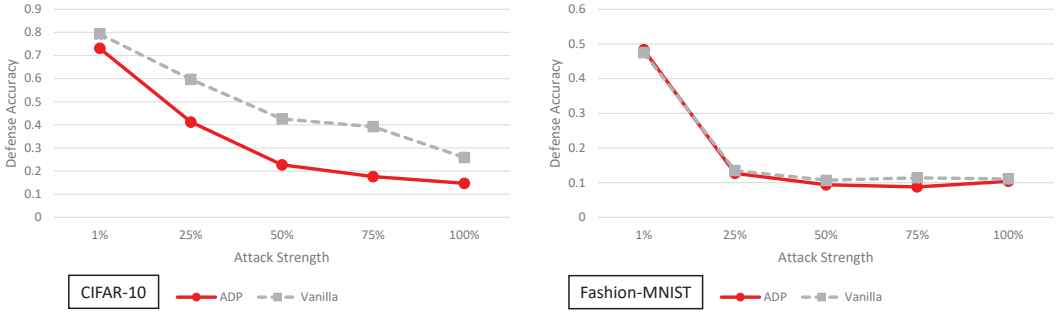


Figure 11. Defense accuracy of the ensemble diversity defense on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

5.7. Enhancing Transformation-Based Defenses against Adversarial Attacks with a Distribution Classifier Analysis

The distribution classifier defense [16] results for adaptive black-box adversaries of varying strength are shown in Figure 12. This defense does not perform significantly better than the vanilla model for either CIFAR-10 or Fashion-MNIST. This defense employs randomized image transformations, just like BaRT. However, unlike BaRT, there is no clear improvement in defense accuracy. We can attribute this to two main reasons. First, the number of transformations in BaRT are significantly larger (i.e., 10 different image transformation groups in CIFAR-10, 8 different image transformation groups in Fashion-MNIST). In the distribution classifier defense, only resizing and zero padding transformations are used. Second, BaRT requires retraining the entire classifier to accommodate the transformations. This means all parts of the network from the convolutional layers, to the feed forward classifier are modified (retrained). The distribution classifier defense only retrains the final classifier after the soft-max output. This means the feature extraction layers (convolutional layers) between the vanilla model and the distributional classifier are virtually unchanged. If two networks have the same convolutional layers with the same weights, it is not surprising that the experiments show that they have similar defense accuracies.

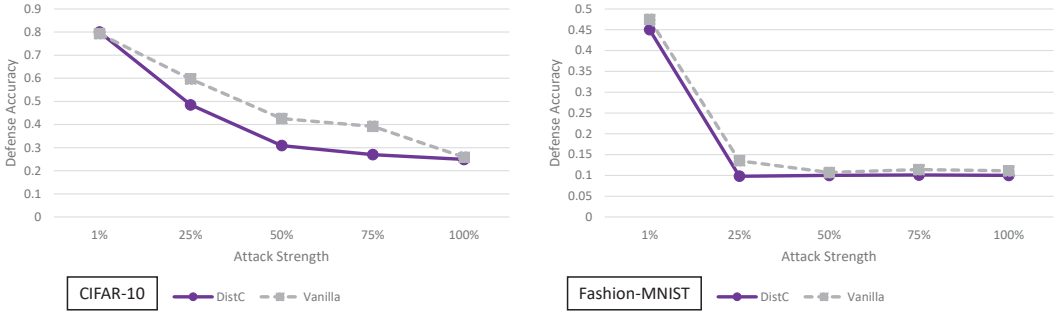


Figure 12. Defense accuracy of the distribution classifier defense on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

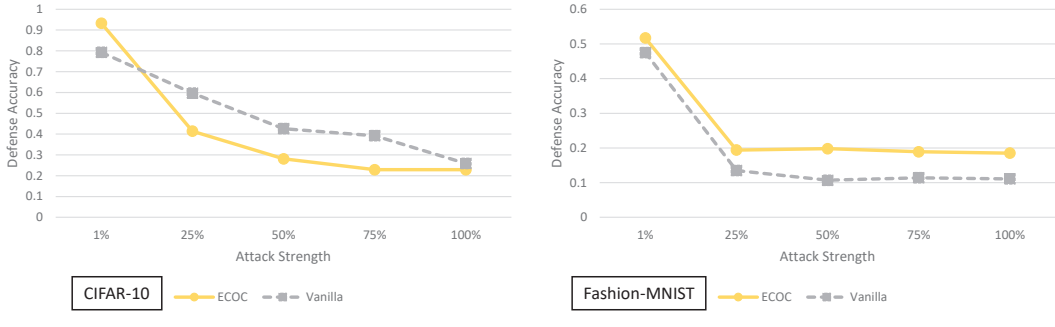


Figure 13. Defense accuracy of the error correcting output code defense on various strength adaptive black-box adversaries for CIFAR-10 and Fashion-MNIST. The defense accuracy in these graphs is measured on the adversarial samples generated from the untargeted MIM adaptive black-box attack. The % strength of the adversary corresponds to what percent of the original training dataset the adversary has access to. For full experimental numbers for CIFAR-10, see Table A5 through Table A9. For full experimental numbers for Fashion-MNIST, see Table A11 through Table A15.

5.8. Error Correcting Output Codes Analysis

In Figure 13, we show the ECOC defense for the adaptive black-box adversaries with varied strength. For CIFAR-10, ECOC performs worse than the vanilla defense in all cases except for the 1% strength adversary. For Fashion-MNIST, the ECOC defense performs only slightly better than the vanilla model. ECOC performs 6.82% greater in terms of defense accuracy on average when considering all the different strength adaptive black-box adversaries for Fashion-MNIST. In general, we don't see significant improvements (greater than 25% increases) in defense accuracy when implementing ECOC.

5.9. k-Winner-Take-All Analysis

The results for the adaptive black-box variable strength adversary for the k-WTA defense are given in Figure 6. We can see that the k-WTA defense performs approximately the same or slightly worse than the vanilla model in almost all cases.

The slightly worse performance on CIFAR-10 can be attributed to the fact that the clean accuracy of the k-WTA ResNet56 is slightly lower than the clean accuracy of the vanilla model. We go into detailed explanations about the lower accuracy in the Appendix A. In short, the k-WTA defense is implemented in PyTorch while the vanilla ResNet56 is implemented in Keras. The slightly lower accuracy is due to implementation differences between Keras and PyTorch. It is not necessarily a direct product of the defense.

Regardless of the slight clean accuracy discrepancies, we see that this defense does not offer any significant improvements over the vanilla defense. From a black-box attacker perspective, this makes sense. Replacing an activation function in the network while still making it have almost identical performance on clean images should not yield security. The only exception to this would be if the architecture change fundamentally alters the way the image is processed in the CNN. In the case of k-WTA, the experiments support the hypothesis that this is not the case.

5.10. On the Adaptability of the Adaptive Black-Box Attack

The adaptive black-box is aptly named because it *adapts* to the defense it is attacking. It does this by training the synthetic model on the output labels from the defense, as opposed to using the original training data labels. While this claim is intuitive in this subsection we give experimental proof to support our claim.

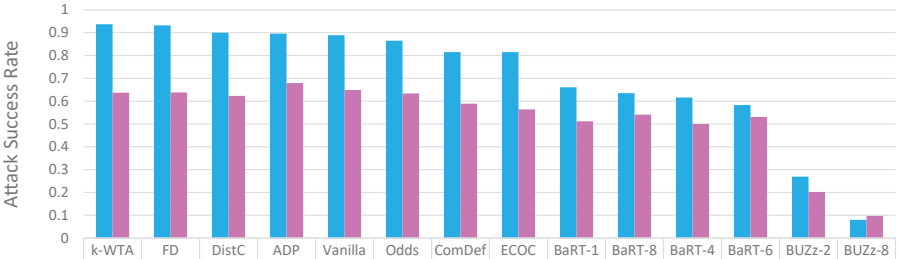
To show the advantage of the adaptive black-box attack, we compare it to the pure black-box attack (which is non-adaptive). The pure black-box attack is not considered adaptive because the adversarial examples generated in the pure black-box attack are defense agnostic. Specifically, this means the *same* set of adversarial examples are used, regardless of which defense is being attacked.

To compare attack results, we setup the following simple experiment: we use the Fashion-MNIST dataset, assuming an untargeted attack with respect to the l_∞ norm and maximum perturbation $\epsilon = 0.1$. We give both attacks access to 100% of the training data and we use the MIM method for generating adversarial examples once the synthetic model in each attack has been trained. For both the pure and adaptive black-box attack, we use the same synthetic model (see the Appendix A for further model details).

Having fixed the dataset, attack generation method, synthetic model and the amount of data available to the attacker, we report the attack success rate on vanilla classifiers and all 9 defenses in Figure 14. For each defense, we use 1000 clean examples and measure the percent of adversarial examples created from the clean examples that are misclassified. For almost every case, we can see that the adaptive black-box attack does better than the pure black-box attack, demonstrating the notion of adaptability. For example, the adaptive black-box attack has a 20% or greater improvement in attack success rate over the pure black-box attack on k-WTA, FD, DistC, ADP, Odds, ComDefend and ECOC. It should be worth noting the improvement is smaller but still there for all the BaRT defenses and every BUZZ defense except for BUZz-8. We conjecture this may be due to the fact the adaptive black-box attack does not train on null label data, something that the BUZz-8 defense

outputs. Hence the lack of training data when attacking the BUZz-8 defense may cause the attack to be weaker. We discuss this notion of adaptive attacks on null label defenses in greater detail in the Appendix A.

Overall, our results in this subsection give strong experimental evidence to support the adaptability claim for the adaptive black-box attack. It can clearly be seen that in almost every case, the adaptive attack is able to make use of querying the defense to produce a higher attack success rate. When compared to a static black-box attack like the pure black-box attack, the adaptive black-box attack does better against the majority of the defenses analyzed in this work.



	k-WTA	FD	DistC	ADP	Vanilla	Odds	ComDef	ECOC	BaRT-1	BaRT-8	BaRT-4	BaRT-6	BUZz-2	BUZz-8
Adaptive	0.937	0.932	0.9	0.896	0.889	0.865	0.815	0.815	0.661	0.635	0.616	0.583	0.269	0.08
Pure	0.637	0.638	0.623	0.679	0.649	0.634	0.589	0.564	0.512	0.541	0.5	0.531	0.202	0.097

Figure 14. Adaptive black-box attack (100% strength) vs pure black-box attack on the vanilla classifier and all 9 defenses for Fashion-MNIST. It can clearly be seen that in almost every case, the adaptive black-box attack outperforms the pure black-box attack.

6. Conclusions

In this paper, we investigated and rigorously experimented with adaptive black-box attacks on recent defenses. Our paper's results span nine defenses, two adversarial models, six different attacks, and two datasets. From our vast set of experiments, we derive several principal results to advance the field of adversarial machine learning. We show that most defenses (7 out of 9 for each dataset) offer less than a 25% improvement in defense accuracy for an adaptive black-box adversary. We demonstrate that currently no defense gives significant black-box robustness without sustaining a drop in clean accuracy. While the defenses we cover generally provide marginal or less than marginal robustness, there are several common defense trends among the stronger defenses we analyzed. The common effective defense trends include using detection methods to mark suspicious samples as adversarial and using image transformations so large in magnitude that retraining of the classifier is required. Lastly, our experiments highlight the need for proper black-box attack testing. Simply building white-box defenses and only testing against white-box attacks can result in highly misleading claims about robustness. Overall, we complete the security picture for currently proposed defense with our experiments and give future defense designers insight and direction with our analyses.

Author Contributions: Conceptualization, investigation, experimentation, writing, K.M.; Experimentation, D.G.; Investigation, writing; M.v.D.; Experimentation, investigation: P.H.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CIFAR-10 dataset can be found at: https://www.cs.toronto.edu/~kriz/cifar.html (accessed on 1 May 2020). The Fashion-MNIST dataset can be found at: https://github.com/zalandoresearch/fashion-mnist (accessed 1 May 2020).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Black-Box Settings

We describe the detailed setup of our black box attacks in this paper. We strictly follow the setup of the black box attacks as described in [24]. This setup is carefully chosen by the authors to allow them to properly analyze the security of many defenses under the notion of pure black box attacks and adaptive black box attacks. For the sake of completeness, we re-introduce the setup used in [24].

Algorithm 1 describes the oracle based black-box attack from [4]. The oracle $\mathcal{O}$ represents black-box query access to the target model f and only returns the final class label $F(f(x))$ for a query x (and not the score vector $f(x)$). Initially, the adversary is given (a part of) the training data set $\mathcal{X}$, i.e., he knows $\mathcal{D} = \{(x, F(f(x))) : x \in \mathcal{X}_0\}$ for some $\mathcal{X}_0 \subseteq \mathcal{X}$.

Algorithm 1 Construction of synthetic network g in Papernot's oracle based black-box attack [24]

1: **Input**:
2: $\mathcal{O}$ represents black-box access to $F(f(\cdot))$ for target model f with output function F;
3: $\mathcal{X}_0 \subseteq \mathcal{X}$, where $\mathcal{X}$ is the training data set of target model f;
4: substitute architecture S
5: training method M;
6: constant λ;
7: number N of synthetic training epochs
8: **Output:**
9: synthetic model s defined by parameters θ_s
10: (s also has output function F which selects the max confidence score;
11: s fits architecture S)
12:
13: **Algorithm:**
14: **for** N iterations **do**
15: $\mathcal{D} \leftarrow \{(x, \mathcal{O}(x)) : x \in \mathcal{X}_t\}$
16: $\theta_s = \mathrm{M}(S, \mathcal{D})$
17: $\mathcal{X}_{t+1} \leftarrow \{x + \lambda \cdot \mathrm{sgn}(J_{\theta_s}(x)[\mathcal{O}(x)]) : x \in \mathcal{X}_t\} \cup \mathcal{X}_t$
18: **end for**
19: Output θ_s

Let S and θ_s be an a-priori synthetic architecture and the parameter of the synthetic network, respectively. θ_s is trained using Algorithm 1, i.e., the image-label pairs in $\mathcal{D}$ are used to train θ_s using a training method M (e.g., Adam [43]). The data augmentation method (i.e., Jacobian) is used to increase the samples in training dataset $\mathcal{X}_t$ as described in line 17. Algorithm 1 runs N iterations before outputting the final trained parameters θ_s.

Table A1. Training parameters used in the experiments [24].

Training Parameter	Value
Optimization Method	ADAM
Learning Rate	0.0001
Batch Size	64
Epochs	100
Data Augmentation	None

Table A2. Adaptive black-box attack parameters [24].

	$\lvert\mathcal{X}_0\rvert$	N	λ
CIFAR-10	50,000	4	0.1
Fashion-MNIST	60,000	4	0.1

Table A3. Architectures of synthetic neural network s from [24,28].

Layer Type	Fashion-MNIST and CIFAR-10
Convolution + ReLU	$3 \times 3 \times 64$
Convolution + ReLU	$3 \times 3 \times 64$
Max Pooling	2×2
Convolution + ReLU	$3 \times 3 \times 128$
Convolution + ReLU	$3 \times 3 \times 128$
Max Pooling	2×2
Fully Connected + ReLU	256
Fully Connected + ReLU	256
Softmax	10

Tables A1–A3 from [24] describe the setup of our experiments in this paper. Table A1 presents the setup of the optimization algorithm used for training in Algorithm 1. The architecture of the synthetic model is described in Table A3 and the main parameters for Algorithm 1 for CIFAR-10 and Fashion-MNIST are presented in Table A2.

Appendix A.2. The Adapative Black-Box Attack on Null Class Label Defenses

For the adaptive black-box attack, there is a special case to consider when applying this attack to defenses that have the option of outputting a *null class label*. We study two of these defenses, Odds and BUZz. Here we define the null class label l as a label the defense gives to an input x' when it considers the input to be manipulated by the adversary. This means for a 10 class problem like CIFAR-10, the defense actually has the option of outputting 11 class labels (with class label 11 being the adversarial label). In the context of the adaptive black-box attack, two changes occur. The first change is outside the control of the attacker, and is in regards to the definition of a successful attack. On a defense, that does not output a null class label, the attacker has to satisfy the following output condition: $\mathcal{O}(x') = y'$. We further specify $y' = y_t$ for a targeted attack or $y' \neq y$ for an untargeted attack. Also, we define $\mathcal{O}$ as the oracle in the defense, x' as the adversarial example, y as the original class label and y_t as the target class label. The above formulation only holds when the defense does not employ any detection method (such as adversarial labeling). When adversarial labeling is employed the conditions change slightly. Now a successful attack must be misclassified by the defense and not be the null class label. Formally, we can write this as: $\mathcal{O}(x') = y' \wedge \mathcal{O}(x') \neq l$. While this first change is straightforward, there is another major change in the attack which we describe next.

The second main change in the adaptive black-box attack on null class label defenses comes from training the synthetic model. In the main paper, we mention that training the synthetic model is done with data labeled from the defense $\mathcal{O}(x) = y$. However, we do not use data which has a null class label l, i.e. $\mathcal{O}(x) = l$. We ignore this type of data because this would require modifying the untargeted attack in an unnecessary way. The untargeted attack tries to find the malicious (wrong) label. If the synthetic network is outputting null labels, it is possible for the untargeted attack to produce an adversarial sample that will have a null label. In essence, the attack would fail under those circumstances. To prevent this, the objective function of every untargeted attack would need to be modified, such that the untargeted attack produces the malicious label and it is not the null label. To avoid needlessly complicating the attack, we simply do not use null labeled data. It

is an open question of whether using null labeled data to train the synthetic network and the specialized untargeted attack we describe, would actually yield any meaningful performance gains.

Appendix A.3. Vanilla Model Implementation

CIFAR-10: We train a ResNet56 [44] for 200 epochs with ADAM. We accomplish this using Keras :https://github.com/keras-team/keras (accessed on 1 May 2020) and the ResNet56 version 2 implementation: https://keras.io/examples/cifar10_resnet/ (accessed on 1 May 2020). In terms of the dataset, we use 50,000 samples for training and 10,000 samples for testing. All images are normalized in the range [0,1] with a shift of -0.5 so that they are in the range $[-0.5, 0.5]$. We also use the built in data augmentation technique provided by Keras during training. With this setup our vanilla network achieves a testing accuracy of 92.78%.

Fashion-MNIST: We train a VGG16 network [2] for 100 epochs using ADAM. We use 60,000 samples for training and 10,000 samples for testing. All images are normalized in the range [0,1] with a shift of -0.5 so that they are in the range $[-0.5, 0.5]$. For this dataset we do not use any augmentation techniques. However, our VGG16 network has a built in resizing layer that transforms the images from 28×28 to 32×32. We found this process slightly boosts the clean accuracy of the network. On testing data we achieve an accuracy of 93.56%.

Appendix A.4. Barrage of Random Transforms Implementation

The authors of BaRT [14] do not provide source code for their defense. We contacted the authors and followed their recommendations as closely as possible to re-implement their defense. However, some implementation changes had to be made. For the sake of the reproducibility of our results, we enumerate the changes made here.

Image transformations: In the appendix for BaRT, they provide code snippets which are configured to work with scikit image package version 14.0.0. However, due to compatibility issues, the closest version we could implement with our other existing packages was scikit image 14.4.0. Due to the different scikit version, two parts of the defense had to be modified. The original denoising wavelet transformation code in the BaRT appendix had invalid syntax for version 14.4.0, so we had to modify it and run it with different less random parameters.

The second defense change we made was due to error handling. In extremely rarely cases, certain sequences of image transformations return images with NAN values. When contacting the authors they acknowledged that their code failed when using newer versions of sci-kit. As a result, in sci-kit 14.4.0 when we encounter this error, we randomly pick a new sequence of random transformations for the image. We experimentally verified that this has a negligible impact on the entropy of the defense. For example, in CIFAR-10 for the 5 transformation defense, we encounter this error 47 times when running all 50,000 training samples. That means roughly only 0.094% of the possible transformations sequences cannot be used in sci-kit 14.4.0.

It is worth noting one other change we made to the Fashion-MNIST version of this defense. The original BaRT defense was only implemented for ImageNet, a three color channel (RGB) dataset. Fashion-MNIST is a single color channel (grayscale) dataset. As a results two transformation groups are not usable for the Fashion-MNIST BaRT defense (the color space change group and grayscale transformation group).

Training BaRT: In [14] the authors start with a ResNet model pre-trained on ImageNet and further train it on transformed data for 50 epochs using ADAM. The transformed data is created by transforming samples in the training set. Each sample is transformed T times, where T is randomly chosen from distribution $U(0, 5)$. Since the authors did not experiment with CIFAR-10 and Fashion-MNIST, we tried two approaches to maximize the accuracy of the BaRT defense. First, we followed the author's approach and started with a ResNet56 pre-trained for 200 epochs on CIFAR-10 with data-augmentation. We then further trained this model on transformed data for 50 epochs using ADAM. For CIFAR-10, we

were able to achieve an accuracy of 98.87% on the training dataset and a testing accuracy of 62.65%. Likewise, we tried the same approach for training the defense on the Fashion-MNIST dataset. We started with a VGG16 model that had already been trained with the standard Fashion-MNIST dataset for 100 epochs using ADAM. We then generated the transformed data and trained it for an additional 50 epochs using ADAM. We were able to achieve a 98.84% training accuracy and a 77.80% testing accuracy. Due to the relatively low testing accuracy on the two datasets, we tried a second way to train the defense.

In our second approach we tried training the defense on the randomized data using untrained models. For CIFAR-10 we trained ResNet56 from scratch with the transformed data and data augmentation provided by Keras for 200 epochs. We found the second approach yielded a higher testing accuracy of 70.53%. Likewise for Fashion-MNIST, we trained a VGG16 network from scratch on the transformed data and obtained a testing accuracy of 80.41%. Due to the better performance on both datasets, we built the defense using models trained using the second approach.

Appendix A.5. Improving Adversarial Robustness via Promoting Ensemble Diversity Implementation

The original source code for the ADP defense [11] on MNIST and CIFAR-10 datasets was provided on the author's Github page: https://github.com/P2333/Adaptive-Diversity-Promoting (accessed on 1 May 2020). We used the same ADP training code the authors provided, but trained on our own architecture. For CIFAR-10, we used the ResNet56 model mentioned in subsection Appendix A.3 and for Fashion-MNIST, we used the VGG16 model mentioned in Appendix A.3. We used K = 3 networks for ensemble model. We followed the original paper for the selection of the hyperparameters, which are $\alpha = 2$ and $\beta = 0.5$ for the adaptive diversity promoting (ADP) regularizer. In order to train the model for CIFAR-10, we trained using the 50,000 training images for 200 epochs with a batch size of 64. We trained the network using ADAM optimizer with Keras data augmentation. For Fashion-MNIST, we trained the model for 100 epochs with a batch size of 64 on the 60,000 training images. For this dataset, we again used ADAM as the optimizer but did not use any data augmentation.

We constructed a wrapper for the ADP defense where the inputs are predicted by the ensemble model and the accuracy is evaluated. For CIFAR-10, we used 10,000 clean test images and obtained an accuracy of 94.3%. We observed no drop in clean accuracy with the ensemble model, but rather observed a slight increase from 92.78% which is the original accuracy of the vanilla model. For Fashion-MNIST, we tested the model with 10,000 clean test images and obtained an accuracy of 94.86%. Again for this dataset we observed no drop in accuracy after training with the ADP method.

Appendix A.6. Error Correcting Output Codes Implementation

The training and testing code for ECOC defense [12] on CIFAR-10 and MNIST datasets was provided on the Github page of the authors: https://github.com/Gunjan108/robust-ecoc/ (accessed on 1 May 2020). We employed their "TanhEns32" method which uses 32 output codes and the hyperbolic tangent function as sigmoid function with an ensemble model. We choose this model because it yields better accuracy with clean and adversarial images for both CIFAR-10 and MNIST than the other ECOC models they tested, as reported in the original paper.

For CIFAR-10, we used the original training code provided by the authors. Unlike the other defenses, we did not use a ResNet network for this defense because the models used in their ensemble predict individual bits of the error code. As a result these models are much less complex than ResNet56 (fewer trainable parameters). Due to the lower model complexity of each individual model in the ensemble, we used the default CNN structure the authors provided instead of our own. We did this to avoid over parameterization of the ensemble. We used 4 individual networks for the ensemble model and trained the

network with 50,000 clean images for 400 epochs with a batch size of 200. We used data augmentation (with Keras) and batch normalization during training.

We used the original MNIST training code to train Fashion-MNIST by simply changing the dataset. Similarly, to avoid over parameterization, we again used the CNNs the authors used with lower complexity instead of using our VGG16 architecture. We trained the ensemble model with 4 networks for 150 epochs and with a batch size of 200. We did not use data augmentation for this dataset.

For our implementation, we constructed our own wrapper class where the input images are predicted and evaluated using the TanhEns32 model. We tested the defense with 10,000 clean testing images for both CIFAR-10 and Fashion-MNIST, and obtained 89.08% and 92.13% accuracy, respectively.

Appendix A.7. Distribution Classifier Implementation

For the distribution classifier defense [16], we used random resize and pad (RRP) [38] and a DRN [45] as distribution classifier. The authors did not provide a public code for their complete working defense. However, the DRN implementation by the same author was previously released on Github: https://github.com/koukl/drn (accessed on 1 May 2020). We also contacted the authors, followed their recommendations for the training parameters and used the DRN implementation they sent to us as a blueprint.

In order to implement RRP, we followed the resize ranges the paper suggested, specifically for IFGSM attack. Therefore, we chose the resize range as 19 pixels to 25 pixels for CIFAR-10 and 22 pixels to 28 pixels for Fashion-MNIST and used these parameters for all of our experiments.

As for the distribution classifier, the DRN consists of fully connected layers and each node encodes a distribution. We use one hidden layer of 10 nodes. For the final layer, there are 10 nodes (representing each class) and there are two bins representing the logit output for each class. In this type of network the output from the layers are 2D. For the final classification, we convert from 2D to 1D by taking the output from the hidden layer and simply discarding the second bin each time. The distribution classifier then performs the final classification and outputs the class label.

Training: We followed the parameters the paper suggested to prepare training data. First, we collected 1000 correctly classified training clean images for Fashion-MNIST and 10,000 correctly classified clean images for CIFAR-10. Therefore, with no transformation, the accuracy of the networks is 100%. For Fashion-MNIST, we used N = 100 transformation samples and for CIFAR-10, we used N = 50 samples, as suggested in the original paper. After collecting N samples from the RRP, we fed them into our main classifier network and collected the softmax probabilities for each class. Finally, for each class, we made an approximation by computing the marginal distributions using kernel density estimation with a Gaussian kernel (kernel width = 0.05). We used 100 discretization bins to discretize the distribution. For each image, we obtain 100 distribution samples per class. For further details of this distribution, we refer the readers to [16].

We trained the model with the previously collected distribution of 1000 correctly classified Fashion-MNIST images for 10 epochs as the authors suggested. For CIFAR-10, we trained the model with the distributions collected from 10,000 correctly classified images for 50 epochs. For both of the datasets, we used a learning rate of 0.1 and a batch size of 16. The cost function is the cross entropy loss on the logits and the distribution classifier is optimized using backpropagation with ADAM.

Testing: We first tested the RRP defense alone with 10,000 clean test images for both CIFAR-10 and Fashion-MNIST to see the drop in clean accuracy. We observed that this defense resulted in approximately 71% for CIFAR-10 and 82% for Fashion-MNIST. Compared to the clean accuracies we obtain without the defense (93.56% for Fashion-MNIST and 92.78% for CIFAR-10), we observe drops in accuracy after random resizing and padding.

We tested the full implementation with RRP and DRN. In order to compare our results with the paper, we collected 5000 correctly classified clean images for both datasets and

collected distributions after transforming images using RRP (N = 50 for Fashion-MNIST and N = 100 for CIFAR-10) like we did for training. We observed a clean test accuracy of 87.48% for CIFAR-10 and 97.76% Fashion-MNIST, which is consistent with the results reported by the original paper. Clearly, if we test all of the clean testing data (10,000 images), we obtain lower accuracy (approximately 83% for CIFAR-10 and 92% for Fashion-MNIST) since there is also some drop in accuracy caused by the CNN. On the other hand, it can be seen that there is a smaller drop in clean accuracy as compared to the basic RRP implementation.

Appendix A.8. Feature Distillation Implementation

Background: The human visual system (HVS) is more sensitive to high frequency parts of the image and less sensitive to the low frequency parts. The standard JPEG compression is based on this understanding, so the standard JPEG quantization table compresses less sensitive frequency parts of the image (i.e. low frequency components) more than other parts. In order to defend against images, a higher compression rate is needed. However, since the CNNs work differently than the HVS, the testing accuracy and defense accuracy both suffer if a higher compression rate is used across all frequencies. In the Feature Distillation defense, as mentioned in Section 3, a crafted quantization technique is used as a solution to this problem. A large quantization step (QS) can reduce adversarial perturbations but also cause more classification errors. Therefore, the proper selection of QS is needed. In the crafted quantization technique, the frequency components are separated as Accuracy Sensitive (AS) band and Malicious Defense (MD) band. A higher quantization step (QS_1) is applied to the MD band to mitigate adversarial perturbations while a lower quantization step (QS_2) is used for AS band to enhance clean accuracy. For more details of this technique, we refer the readers to [18].

Implementation: The implementation of the defense can be found on the author's Github page: https://github.com/zihaoliu123 (accessed on 1 May 2020). However, this defense has only been implemented and tested for the ImageNet dataset by the authors. In order to fairly compare our results with the other defenses, we implemented and tested this defense for CIFAR-10 and Fashion-MNIST datasets.

This defenses uses two different methods: A one-pass process and a two-pass process. The one-pass process uses the proposed quantization/dequantization only in the decompression of the image. The two-pass process, on the other hand, uses the proposed quantization/dequantization in compression followed by one-pass process. In our experiments, we use the two-pass method as it has better defense accuracy than the one-pass process [18].

In the original paper, experiments were performed in order to find a proper selection of (QS_1) and (QS_2) for the AS and MD bands. At the end of these experiments, they set ($QS_1 = 30$) and ($QS_2 = 20$). However, these experiments were performed on ImageNet images where the images are much larger than CIFAR-10 and Fashion-MNIST images. Therefore, we performed experiments in order to properly select QS_1 and QS_2 for the Fashion-MNIST and CIFAR-10 datasets. For each dataset we start with the vanilla classifier (see Appendix A.3). For each vanilla CNN we first do a one-pass and then generate 500 adversarial samples using untargeted FGSM. For CIFAR-10 we use $\epsilon = 0.05$ and for Fashion-MNIST we use $\epsilon = 0.15$. Here we use FGSM to do the hyperparameter selection for the defense because this is how the authors designed the original defense for ImageNet.

After generating the adversarial examples for each QS combination, we do a grid search over the possible hyperparameters QS_1 and QS_2. Specifically, we test 100 defense combinations by varying QS_1 from 10 to 100 and varying QS_2 from 10 to 100. For every possible combination of QS_1 and QS_2 we measure the accuracy on the clean test set and on the adversarial examples. The results of these experiments are shown in Figure A1.

In Figure A1 for the CIFAR-10 dataset, there is an intersection where both the green dots and red dots overlap. This region represents a defense with both higher clean accuracy and higher defense accuracy (the idealized case). There are multiple different combinations of QS_1 and QS_2 that we could choose that give a decent trade-off. Here we arbitrarily select

from among these better combinations $QS_1 = 70$ and $QS_2 = 40$ which gives a clean score of 71.4% and a defense accuracy of 21.2%.

In Figure A1 for the Fashion-MNIST dataset, there is no region in which both the clean accuracy and defense accuracy are high. This may show a limitation in the use of feature distillation as a defense for some datasets, as here no ideal trade-off exists. We pick $QS_1 = 70$ and $QS_2 = 40$ which gives a clean score of 89.34% and a defense accuracy of 9%. We picked these values because this combination gave the highest defense accuracy out of all possible hyperparameter choices.

Appendix A.9. End-to-End Image Compression Models Implementation

The original source code for defenses on Fashion-MNIST and ImageNet were provided by the authors of ComDefend [13] on their Github page: https://github.com/jiaxiaojunQAQ/Comdefend (accessed on 1 May 2020). In addition, they included their trained compression and reconstruction models for Fashion-MNIST and CIFAR-10 separately.

Since this defense is a pre-processing module, it does not require modifications to the classifier network [13]. Therefore, in order to perform the classification, we used our own models as described in Section A.3 and we combined them with this pre-processing module.

According to the authors of ComDefend, ComCNN and RecCNN were trained on 50,000 clean (not perturbed) images from the CIFAR-10 dataset for 30 epochs using a batch size of 50. In order to use their pre-trained models, we had to install the canton package v0.1.22 for Python. However, we had incompatibility issues with canton and the other Python packages installed in our system. Therefore, instead of installing this package directly, we downloaded the source code of the canton library from its Github page and added it to our defense code separately. We constructed a wrapper for ComDefend, where the type of dataset (Fashion-MNIST or CIFAR-10) is indicated as input so that the corresponding classifier can be used (either ResNet56 or VGG16). We tested the defense with the testin data of CIFAR-10 and Fashion-MNIST and we were able to achieve an accuracy of 88% and 93% respectively.

Appendix A.10. The Odds Are Odd Implementation

Mathematical background: Here we give a detailed description of the defense based on the statistical test derived from the logits layer. For given image x, we denote $\phi(x)$ as the logits layer (i.e., the input to the softmax layer) of a classifier, $f_y = \langle w_y, \phi(x) \rangle$ where w_y is the weight vector for the class $y, y \in \{1, \cdots, K\}$. The class label is determined by $F(x) = \text{argmax}_y f_y(x)$. We define pair-wise log-odds between class y and z as

$$f_{y,z}(x) = f_z(x) - f_y(x) = \langle w_z - w_y, \phi(x) \rangle. \tag{A1}$$

We denote $f_{y,z}(x + \eta)$ the noise-perturbed log-odds where the noise η is sampled from a distribution $\mathcal{D}$. Moreover, we define the following formulas for a pair (y, z):

$$\begin{aligned} g_{y,z} &:= f_{y,z}(x + \eta) - f_{y,z}(x) \\ \mu_{y,z} &:= \mathrm{E}_{x|y}\mathrm{E}_\eta[g_{y,z}(x, \eta)] \\ \sigma_{y,z} &:= \mathrm{E}_{x|y}\mathrm{E}_\eta[(g_{y,z}(x, \eta) - \mu_{y,z})^2] \\ \bar{g}_{y,z}(x, \eta) &:= [g_{y,z}(x, \eta) - \mu_{y,z}]/\sigma_{y,z} \end{aligned} \tag{A2}$$

For the original training data set, we compute $\mu_{y,z}$ and $\sigma_{y,z}$ for all (y, z). We apply the untargeted white-box attack (PGD [27]) to generate the adversarial dataset. After that, we compute $\mu^{adv}_{y,z}$ and $\sigma^{adv}_{y,z}$ using the adversarial dataset. We denote $\tau_{y,z}$ as the threshold to control the false positive rate (FPR) and it is computed based on $\mu^{adv}_{y,z}$ and $\sigma^{adv}_{y,z}$. The distribution of clean data and the distribution of adversarial data are represented by (μ, σ) and $(\mu^{adv}, \sigma^{adv})$, respectively. These distributions are supposed to be separated and τ is used to control the FPR.

For a given image x, the statistical test is done as follows. First, we calculate the expected perturbed log-odds $\bar{g}_{y,z}(x) = \mathrm{E}_\eta[\bar{g}_{y,z}(x,\eta)]$ where y is the predicted class label of image x given by the vanilla classifier. The test will determine the image x with the label $y = F(x)$ as adversarial (malicious) if

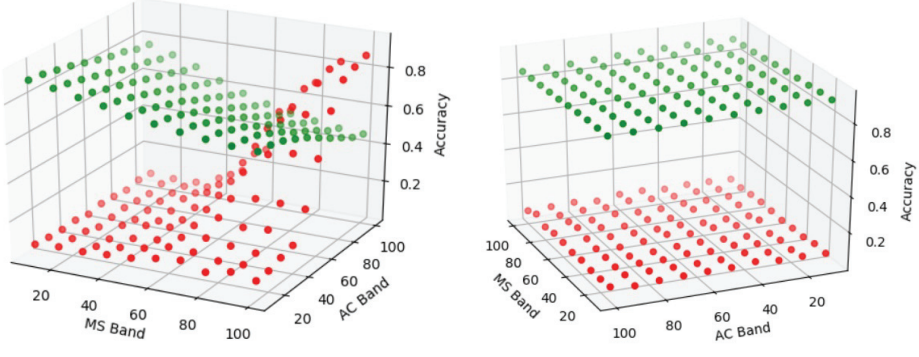
$$\max_{z \neq y}\{\bar{g}_{y,z}(x) - \tau_{y,z}\} \geq 0.$$

Otherwise, the input will be considered benign. In case the test recognizes the image as malicious one, the "corrected" class label z is defined as

$$\max_{z}\{\bar{g}_{y,z}(x) - \tau_{y,z}\}.$$

Implementation details: The original source code for the Odds defense [17] on CIFAR-10 and ImageNet was provided by the authors: https://github.com/yk/icml19_public (accessed on 1 May 2020). We use their code as a guideline for our own defense implementation. We develop the defense for the CIFAR-10 and Fashion-MNIST and datasets. For each dataset, we apply the untargeted 10-iteration PGD attack on the vanilla classifier that will be used in the defense. Note this is a white-box attack. The parameters for the PGD attack are $\epsilon = 0.005$ for CIFAR-10 and $\epsilon = 0.015$ for Fashion-MNIST respectively. By applying the white-box PGD attack we can create the adversarial datasets for the defense. We choose these attack parameters because they yield adversarial examples with small noise. In [17], the authors assume that the adversarial examples are created by adding small noise. Hence, they are not robust against adding the white noises. For a given image, it is normalized first to be in the range $[-0.5, 0.5]$. For each pixel, we generate a noise from $\mathcal{N}(0, 0.05)$ and add it to the pixel.

For CIFAR-10, we create 50,000 adversarial examples. For Fashion-MNIST, we create 60,000 adversarial examples. We calculate μ, σ and τ for each data set for FPR = 1%, 10%, 20%, 30%, 40%, 50% and 80% as described in the mathematical background. For each image, we evaluate it 256 times to compute $\bar{g}_{y,z}(x)$. Table A16 shows the prediction accuracy of the defense for the clean (non-adversarial) dataset for CIFAR-10 and Fashion-MNIST. To compute the clean prediction accuracy, we use 1000 samples from the test dataset of CIFAR-10 and Fashion-MNIST.



(**a**) CIFAR-10 (**b**) Fashion-MNIST

Figure A1. Feature distillation experiments to determine the hyperparameters for the defense. The x and y axis of the grid correspond to the specific hyperparameters for the defense. The Accuracy Sensitive band (denoted as AC in the figure) is the same as QS_1. The Malicious Defense band (denoted as MS in the figure) is the same as QS_2. On the z-axis the accuracy is measured. For every point in this grid two accuracy measurements are taken. The green dot corresponds to the clean accuracy using the QS values specified by the x-y coordinates. The red dot corresponds to the defense accuracy using the QS values specified by the x-y coordinates.

Table A4. CIFAR-10 pure black-box attack. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	CW-T	EAD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-U	EAD-U	Acc
ADP	0.003	0.016	0.004	0.011	0.003	0.003	0.044	0.022	0.001	0.03	0.009	0.013	0.0152
BaRT-1	0.007	0.026	0.027	0.032	−0.005	−0.005	0.151	0.135	0.089	0.153	−0.07	−0.066	−0.0707
BaRT-10	0.001	−0.001	0.045	−0.012	−0.052	−0.053	−0.039	−0.086	−0.019	−0.041	−0.457	−0.456	−0.4409
BaRT-4	0.006	0.035	0.024	0.009	−0.005	−0.021	0.098	0.099	0.061	0.101	−0.186	−0.175	−0.1765
BaRT-7	0.009	0.014	0.037	−0.001	−0.032	−0.036	0.036	0.025	−0.009	0.021	−0.337	−0.353	−0.3164
BUZz-2	0.053	0.057	0.099	0.05	0.011	0.011	0.352	0.306	0.315	0.338	0.047	0.049	−0.0771
BUZz-8	0.083	0.076	0.131	0.07	0.014	0.014	0.504	0.479	0.508	0.473	0.075	0.078	−0.1713
ComDef	−0.005	0.022	0.004	0.013	−0.014	−0.015	0.078	0.033	−0.02	0.043	−0.054	−0.059	−0.043
DistC	0.009	0.027	0.014	0.024	−0.008	−0.014	0.022	0.041	0.016	0.051	−0.104	−0.11	−0.0955
ECOC	−0.006	0.047	−0.007	0.042	0.004	0.001	0.131	0.103	−0.003	0.099	−0.029	−0.033	−0.0369
FD	−0.018	0.001	0.018	0	−0.025	−0.035	−0.017	0.032	0.01	0.014	−0.248	−0.252	−0.2147
k-WTA	0.003	0.02	0.001	0.021	−0.002	−0.003	0.07	0.019	0.001	0.028	−0.07	−0.08	−0.0529
Odds	0.054	0.041	0.071	0.041	0.003	0.002	0.314	0.207	0.233	0.22	0.011	0.011	−0.2137
Vanilla	0.902	0.917	0.853	0.924	0.984	0.984	0.443	0.453	0.384	0.455	0.923	0.919	0.9278

Table A5. CIFAR-10 adaptive black-box attack 1%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	−0.015	−0.001	−0.017	−0.013	−0.055	0.002	−0.062	−0.007	0.008	−0.001	0.002	−0.003	0.0152
BaRT-1	0.011	0.017	0.013	0.013	0.125	0.128	0.103	0.121	0.009	−0.061	0.009	−0.061	−0.0695
BaRT-10	−0.05	−0.044	−0.071	−0.042	−0.287	−0.277	−0.325	−0.28	−0.058	−0.439	−0.053	−0.394	−0.4408
BaRT-4	−0.003	0.001	−0.014	−0.011	−0.019	0.002	−0.06	−0.016	−0.008	−0.213	−0.01	−0.185	−0.1834
BaRT-7	−0.035	−0.023	−0.016	−0.017	−0.151	−0.125	−0.208	−0.149	−0.026	−0.307	−0.024	−0.284	−0.319
BUZz-2	0.002	0.017	0.012	0.015	0.148	0.149	0.103	0.148	0.015	0.005	0.016	0.004	−0.0771
BUZz-8	0.026	0.027	0.024	0.024	0.234	0.228	0.2	0.227	0.017	0.005	0.017	0.006	−0.1713
ComDef	0.014	0.016	0.012	0.016	0.13	0.137	0.109	0.131	0.01	−0.007	0.004	−0.004	−0.0424
DistC	−0.003	0.003	0.001	0.01	0.043	0.067	0.007	0.076	0.004	−0.033	0.004	−0.029	−0.0933
ECOC	0.017	0.022	0.014	0.022	0.186	0.192	0.14	0.194	0.008	0.002	0.005	0.001	−0.0369
FD	−0.014	0.006	−0.009	0.007	−0.026	0.012	−0.036	0.001	0.003	−0.012	−0.006	−0.01	−0.2147
k-WTA	−0.022	−0.004	−0.019	−0.006	−0.023	0.04	−0.018	0.042	0.003	−0.008	−0.01	−0.011	−0.0529
Odds	0.009	0.014	0.005	0.004	0.135	0.124	0.104	0.125	0.014	−0.002	0.013	0.001	−0.214
Vanilla	0.973	0.973	0.976	0.974	0.751	0.766	0.793	0.764	0.983	0.995	0.982	0.994	0.9278

Table A6. CIFAR-10 adaptive black-box attack 25%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	−0.024	−0.049	−0.087	−0.047	−0.074	−0.12	−0.185	−0.131	−0.012	−0.007	−0.008	−0.006	0.0152
BaRT-1	0.015	0.032	0.018	0.029	0.184	0.099	0.089	0.11	−0.012	−0.057	0.001	−0.04	−0.0724
BaRT-10	−0.025	0.006	−0.012	0.015	−0.142	−0.179	−0.208	−0.182	−0.046	−0.398	−0.053	−0.425	−0.4384
BaRT-4	0.003	0.003	0	0.019	0.004	−0.015	−0.072	−0.047	−0.036	−0.196	−0.026	−0.187	−0.1764
BaRT-7	−0.014	−0.013	−0.022	−0.007	−0.106	−0.125	−0.201	−0.15	−0.055	−0.302	−0.04	−0.316	−0.3089
BUZz-2	0.032	0.053	0.051	0.05	0.274	0.228	0.231	0.232	0.003	0.007	0.003	0.011	−0.0771
BUZz-8	0.069	0.07	0.084	0.078	0.419	0.336	0.374	0.335	0.003	0.009	0.005	0.015	−0.1713
ComDef	0.031	0.041	0.029	0.039	0.137	0.126	0.078	0.111	−0.004	−0.009	−0.005	−0.004	−0.0421
DistC	−0.044	−0.007	−0.049	−0.011	−0.022	−0.019	−0.112	−0.02	−0.01	−0.036	−0.011	−0.032	−0.0944
ECOC	−0.044	−0.056	−0.119	−0.041	0.004	−0.073	−0.183	−0.091	−0.004	−0.006	−0.011	−0.009	−0.0369
FD	−0.045	−0.023	−0.045	−0.014	−0.062	−0.05	−0.146	−0.055	−0.011	−0.035	−0.014	−0.031	−0.2147
k-WTA	−0.052	−0.068	−0.112	−0.066	−0.074	−0.174	−0.258	−0.2	−0.008	−0.021	−0.009	−0.019	−0.0529
Odds	0.045	0.047	0.048	0.051	0.237	0.182	0.161	0.16	−0.003	−0.001	−0.003	0	−0.2132
Vanilla	0.924	0.924	0.91	0.921	0.551	0.638	0.597	0.644	0.997	0.991	0.994	0.985	0.9278

Table A7. CIFAR-10 adaptive black-box attack 50%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	−0.036	−0.116	−0.137	−0.107	−0.077	−0.21	−0.199	−0.229	−0.002	0.001	0.006	0.001	0.0152
BaRT-1	0.034	0.011	0.021	0.028	0.148	0.071	0.051	0.071	−0.009	−0.062	−0.013	−0.064	−0.0753
BaRT-10	0.036	0.046	0.086	0.037	−0.044	−0.092	−0.008	−0.104	−0.043	−0.414	−0.034	−0.433	−0.4399
BaRT-4	0.036	0.02	0.055	0.058	0.075	0.043	0.036	0.02	−0.024	−0.173	−0.039	−0.183	−0.1772
BaRT-7	0.03	0.016	0.055	0.048	0.026	−0.034	0	−0.025	−0.045	−0.297	−0.046	−0.306	−0.3181
BUZz-2	0.088	0.08	0.11	0.093	0.367	0.289	0.316	0.293	0.007	0.012	0.01	0.011	−0.0771
BUZz-8	0.124	0.106	0.162	0.12	0.542	0.428	0.521	0.434	0.007	0.013	0.01	0.014	−0.1713
ComDef	0.01	−0.033	−0.039	−0.014	0.03	−0.015	−0.067	−0.036	−0.005	−0.012	−0.002	−0.016	−0.0411
DistC	−0.021	−0.036	−0.059	−0.014	−0.041	−0.065	−0.117	−0.059	−0.014	−0.042	−0.012	−0.045	−0.0922
ECOC	−0.025	−0.045	−0.11	−0.035	0.02	−0.079	−0.145	−0.09	0.001	−0.004	0.001	−0.019	−0.0369
FD	0.013	0.002	0.008	0.029	0.018	0.021	−0.01	0.015	−0.014	−0.035	−0.014	−0.038	−0.2147
k-WTA	−0.002	−0.139	−0.171	−0.131	−0.064	−0.226	−0.241	−0.248	−0.002	−0.022	−0.005	−0.029	−0.0529
Odds	0.073	0.064	0.098	0.074	0.283	0.181	0.185	0.181	−0.002	−0.002	−0.006	−0.005	−0.2133
Vanilla	0.87	0.886	0.826	0.872	0.423	0.529	0.426	0.531	0.993	0.987	0.99	0.986	0.9278

Table A8. CIFAR-10 adaptive black-box attack 75%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	−0.038	−0.167	−0.201	−0.167	−0.092	−0.231	−0.216	−0.234	−0.009	−0.006	−0.001	0	0.0152
BaRT-1	0.002	0.027	0.007	0.015	0.117	0.072	0.029	0.069	−0.013	−0.067	−0.006	−0.069	−0.0706
BaRT-10	0.018	0.052	0.061	0.03	−0.055	−0.036	−0.001	−0.03	−0.065	−0.428	−0.058	−0.417	−0.4349
BaRT-4	0.014	0.034	0.045	0.038	0.083	0.088	0.065	0.066	−0.035	−0.2	−0.031	−0.197	−0.1829
BaRT-7	0.016	0.057	0.072	0.05	0.048	0.03	0.001	0.014	−0.035	−0.3	−0.036	−0.334	−0.308
BUZz-2	0.074	0.094	0.104	0.086	0.332	0.328	0.344	0.324	0.007	0.011	0.007	0.014	−0.0771
BUZz-8	0.105	0.126	0.159	0.114	0.541	0.484	0.55	0.464	0.007	0.011	0.008	0.015	−0.1713
ComDef	−0.013	0.003	−0.034	−0.008	0.014	−0.013	−0.063	−0.007	−0.001	−0.019	−0.001	−0.017	−0.0434
DistC	−0.051	−0.042	−0.083	−0.042	−0.078	−0.073	−0.122	−0.077	−0.012	−0.055	−0.016	−0.059	−0.0913
ECOC	−0.06	−0.049	−0.143	−0.054	−0.008	−0.086	−0.163	−0.099	0.004	−0.009	0.002	−0.008	−0.0369
FD	−0.013	0.055	0.004	0.024	0.006	0.097	0.007	0.048	−0.01	−0.032	−0.007	−0.02	−0.2147
k-WTA	−0.036	−0.157	−0.254	−0.162	−0.094	−0.252	−0.243	−0.283	−0.007	−0.031	−0.014	−0.044	−0.0529
Odds	0.05	0.07	0.088	0.05	0.246	0.19	0.196	0.179	−0.002	−0.009	−0.003	−0.014	−0.2133
Vanilla	0.887	0.864	0.822	0.875	0.425	0.478	0.392	0.496	0.993	0.989	0.992	0.984	0.9278

Table A9. CIFAR-10 adaptive black-box attack 100%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	−0.023	−0.163	−0.172	−0.136	−0.002	−0.163	−0.112	−0.148	0.004	0.006	−0.004	−0.002	0.0152
BaRT-1	0.02	0.011	0.022	0.027	0.173	0.166	0.139	0.169	−0.016	−0.069	−0.018	−0.054	−0.0707
BaRT-10	0.044	0.05	0.126	0.073	0.057	0.078	0.123	0.118	−0.063	−0.405	−0.047	−0.404	−0.4409
BaRT-4	0.044	0.053	0.089	0.053	0.148	0.167	0.184	0.203	−0.02	−0.183	−0.017	−0.199	−0.1765
BaRT-7	0.038	0.06	0.1	0.069	0.113	0.147	0.167	0.161	−0.028	−0.282	−0.045	−0.309	−0.3164
BUZz-2	0.103	0.11	0.168	0.123	0.473	0.426	0.493	0.451	0.009	0.014	0.008	0.01	−0.0771
BUZz-8	0.127	0.13	0.203	0.145	0.628	0.568	0.667	0.576	0.009	0.014	0.009	0.013	−0.1713
ComDef	−0.008	0.005	−0.01	0.007	0.133	0.097	0.084	0.102	−0.003	−0.019	−0.007	−0.022	−0.043
DistC	−0.011	−0.017	−0.041	−0.002	0.005	0.018	−0.01	0.026	0.004	−0.025	0.004	−0.031	−0.0955
ECOC	−0.04	−0.056	−0.105	−0.054	0.091	0.012	−0.03	0.033	0.008	0.002	0.003	−0.008	−0.0369
FD	0.007	0.048	0.062	0.074	0.105	0.181	0.138	0.194	0.002	−0.02	−0.004	−0.029	−0.2147
k-WTA	−0.019	−0.136	−0.174	−0.129	−0.015	−0.157	−0.124	−0.138	−0.003	−0.029	−0.009	−0.034	−0.0529
Odds	0.075	0.077	0.134	0.082	0.312	0.277	0.319	0.299	0.006	−0.012	0	−0.01	−0.2137
Vanilla	0.866	0.861	0.777	0.848	0.334	0.387	0.259	0.374	0.991	0.986	0.991	0.987	0.9278

Table A10. Fashion-MNIST pure black-box attack. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	CW-T	EAD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-U	EAD-U	Acc
ADP	0.01	−0.05	−0.035	−0.018	−0.003	−0.002	−0.031	−0.006	−0.03	−0.024	0.012	0.018	0.013
BaRT-1	0.041	0.055	0.089	0.062	−0.002	0	0.119	0.173	0.137	0.159	−0.021	−0.017	−0.0317
BaRT-4	0.042	0.05	0.079	0.057	−0.019	−0.015	0.118	0.172	0.149	0.15	−0.132	−0.106	−0.1062
BaRT-6	0.049	0.038	0.084	0.052	−0.029	−0.038	0.091	0.139	0.118	0.133	−0.174	−0.18	−0.1539
BaRT-8	0.055	0.018	0.083	0.059	−0.036	−0.047	0.041	0.09	0.108	0.092	−0.239	−0.222	−0.2212
BUZz-2	0.108	0.1	0.161	0.109	0.003	0.006	0.362	0.482	0.447	0.469	0.04	0.051	−0.0819
BUZz-8	0.128	0.109	0.176	0.12	0.005	0.008	0.47	0.566	0.552	0.563	0.069	0.078	−0.1577
ComDef	0.008	0.053	0.048	0.066	−0.003	0.001	−0.005	0.089	0.06	0.09	0.001	0.011	−0.0053
DistC	0.007	0.018	0.027	0.031	−0.004	−0.003	0.005	0.038	0.026	0.027	−0.008	−0.001	−0.0093
ECOC	0.012	0.056	0.079	0.073	0	0.003	0.043	0.113	0.085	0.108	0.001	0.006	−0.0141
FD	−0.002	0.006	0.022	0.014	−0.017	−0.019	−0.046	0.057	0.011	0.035	−0.094	−0.098	−0.0823
k-WTA	−0.006	0.002	0.013	0.014	−0.001	0	−0.064	0.044	0.012	0.028	0.001	−0.001	−0.0053
Odds	0	0.002	0.004	0.003	0.003	0.001	0.001	0.028	0.015	0.023	0.026	0.021	−0.1809
Vanilla	0.865	0.889	0.817	0.879	0.995	0.992	0.429	0.363	0.351	0.374	0.914	0.905	0.9356

Table A11. Fashion-MNIST adaptive black-box attack 1%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	0.03	−0.018	−0.009	−0.004	0.051	−0.029	0.008	−0.027	−0.018	0.023	0.005	0.031	0.013
BaRT-1	0.083	0.063	0.05	0.061	0.229	0.137	0.175	0.171	0.041	−0.022	0.009	−0.026	−0.0308
BaRT-4	0.069	0.04	0.034	0.049	0.153	0.056	0.067	0.055	−0.035	−0.182	−0.032	−0.165	−0.0999
BaRT-6	0.046	0.033	−0.006	0.013	0.113	0.008	0.026	0.036	−0.081	−0.274	−0.062	−0.215	−0.1615
BaRT-8	0.046	0.012	−0.028	0.018	0.048	−0.027	−0.03	−0.053	−0.098	−0.326	−0.111	−0.296	−0.2258
BUZz-2	0.155	0.122	0.111	0.117	0.529	0.425	0.436	0.436	0.061	0.079	0.022	0.039	−0.0819
BUZz-8	0.187	0.136	0.123	0.126	0.679	0.488	0.515	0.504	0.064	0.086	0.026	0.05	−0.1577
ComDef	0.032	0.055	0.02	0.025	0.086	0.114	0.123	0.13	0.038	0.042	0.011	0.013	−0.0058
DistC	−0.021	−0.033	−0.029	−0.039	−0.024	−0.057	−0.025	−0.029	0.007	0.029	−0.002	0.008	−0.0093
ECOC	−0.01	0.03	0.008	0.019	0.061	0.038	0.042	0.051	−0.033	−0.1	−0.026	−0.08	−0.0141
FD	−0.073	−0.08	−0.099	−0.06	−0.099	−0.168	−0.15	−0.136	−0.043	−0.1	−0.028	−0.097	−0.0823
k-WTA	0.035	0.036	0.027	0.044	0.072	0.044	0.049	0.068	0.02	0.05	0.016	0.035	−0.0053
Odds	0.031	0.043	0.019	0.038	0.064	0.051	0.065	0.085	0.021	0.033	0.006	0.017	−0.1833
Vanilla	0.807	0.864	0.876	0.873	0.29	0.503	0.475	0.486	0.935	0.91	0.972	0.947	0.9356

Table A12. Fashion-MNIST adaptive black-box attack 25%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	0.089	0.025	0.036	0.004	0.035	−0.038	−0.008	−0.038	−0.04	−0.092	−0.013	−0.032	0.013
BaRT-1	0.11	0.238	0.195	0.217	0.191	0.165	0.159	0.145	−0.038	−0.134	−0.019	−0.097	−0.0314
BaRT-4	0.141	0.293	0.268	0.256	0.215	0.246	0.224	0.256	−0.067	−0.216	−0.041	−0.218	−0.1018
BaRT-6	0.113	0.285	0.261	0.273	0.209	0.224	0.208	0.206	−0.065	−0.28	−0.056	−0.217	−0.1627
BaRT-8	0.133	0.29	0.294	0.285	0.198	0.195	0.21	0.197	−0.091	−0.341	−0.055	−0.278	−0.221
BUZz-2	0.224	0.42	0.411	0.415	0.542	0.603	0.572	0.601	0.033	0.067	0.034	0.073	−0.0819
BUZz-8	0.288	0.465	0.452	0.454	0.783	0.818	0.808	0.815	0.034	0.073	0.035	0.083	−0.1577
ComDef	0.003	0.17	0.08	0.159	0.043	0.112	0.089	0.105	0.022	0.016	0.015	−0.004	−0.0048
DistC	−0.052	−0.034	−0.062	−0.044	0.013	−0.043	−0.037	−0.054	−0.006	−0.058	−0.001	−0.037	−0.0096
ECOC	0.047	0.188	0.169	0.175	0.014	0.063	0.059	0.06	−0.07	−0.282	−0.067	−0.242	−0.0141
FD	−0.086	−0.012	−0.037	−0.025	−0.048	−0.05	−0.066	−0.072	−0.01	−0.063	−0.036	−0.088	−0.0823
k-WTA	0.012	0.017	−0.001	−0.014	−0.043	−0.029	−0.026	−0.031	−0.279	−0.411	−0.437	−0.402	−0.8516
Odds	−0.064	0.02	0.017	0.024	−0.007	0.042	0.025	0.025	−0.017	−0.037	−0.022	−0.022	−0.1807
Vanilla	0.696	0.53	0.538	0.539	0.108	0.141	0.135	0.149	0.966	0.927	0.965	0.917	0.9356

Table A13. Fashion-MNIST adaptive black-box attack 50%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	0.115	−0.015	0.021	−0.024	0.01	−0.022	−0.013	−0.027	−0.019	−0.065	−0.011	−0.044	0.013
BaRT-1	0.143	0.227	0.263	0.24	0.183	0.205	0.197	0.195	−0.047	−0.114	−0.02	−0.1	−0.0312
BaRT-4	0.179	0.325	0.327	0.314	0.241	0.246	0.258	0.224	−0.059	−0.216	−0.024	−0.184	−0.1
BaRT-6	0.175	0.331	0.357	0.336	0.251	0.278	0.256	0.268	−0.045	−0.248	−0.03	−0.243	−0.1563
BaRT-8	0.188	0.324	0.342	0.325	0.201	0.24	0.264	0.235	−0.064	−0.296	−0.046	−0.258	−0.2174
BUZz-2	0.264	0.446	0.473	0.444	0.534	0.627	0.611	0.625	0.017	0.057	0.019	0.06	−0.0819
BUZz-8	0.321	0.482	0.514	0.482	0.766	0.835	0.823	0.826	0.018	0.061	0.02	0.067	−0.1577
ComDef	0.044	0.143	0.123	0.158	0.016	0.083	0.08	0.084	0.003	−0.01	−0.004	−0.015	−0.0067
DistC	0.029	−0.024	−0.009	−0.029	0.038	−0.019	−0.007	−0.035	−0.006	−0.038	−0.012	−0.054	−0.0094
ECOC	0.097	0.23	0.238	0.235	0.013	0.075	0.091	0.072	−0.049	−0.133	−0.05	−0.129	−0.0141
FD	−0.019	0	0.009	0	−0.055	−0.019	−0.02	−0.043	−0.02	−0.049	−0.024	−0.069	−0.0823
k-WTA	0.057	−0.006	−0.018	−0.013	−0.037	−0.042	−0.035	−0.058	−0.012	−0.028	−0.032	−0.049	−0.0053
Odds	−0.012	0.027	0	0.016	−0.024	0.013	0.005	0.006	−0.005	0.011	−0.011	0.004	−0.1828
Vanilla	0.666	0.516	0.479	0.516	0.132	0.127	0.107	0.132	0.982	0.939	0.98	0.933	0.9356

Table A14. Fashion-MNIST adaptive black-box attack 75%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	0.089	0.024	0.103	0.018	−0.033	−0.028	−0.026	−0.016	−0.016	−0.054	−0.006	−0.054	0.013
BaRT-1	0.144	0.299	0.336	0.297	0.151	0.239	0.195	0.254	−0.027	−0.112	−0.002	−0.078	−0.0316
BaRT-4	0.173	0.347	0.41	0.345	0.196	0.304	0.269	0.36	−0.046	−0.17	−0.022	−0.167	−0.107
BaRT-6	0.175	0.372	0.437	0.354	0.202	0.309	0.29	0.327	−0.043	−0.23	−0.027	−0.183	−0.1503
BaRT-8	0.148	0.368	0.422	0.35	0.159	0.303	0.267	0.297	−0.063	−0.309	−0.035	−0.281	−0.2154
BUZz-2	0.232	0.471	0.522	0.478	0.5	0.626	0.594	0.636	0.01	0.05	0.02	0.055	−0.0819
BUZz-8	0.281	0.501	0.563	0.504	0.715	0.838	0.809	0.857	0.01	0.051	0.021	0.061	−0.1577
ComDef	0.029	0.226	0.192	0.221	−0.044	0.127	0.076	0.145	0.002	−0.006	0.009	−0.015	−0.0052
DistC	−0.01	−0.049	−0.007	−0.03	−0.004	−0.025	−0.013	−0.002	−0.016	−0.043	−0.013	−0.056	−0.0096
ECOC	0.04	0.218	0.275	0.232	−0.033	0.075	0.075	0.099	−0.063	−0.156	−0.043	−0.151	−0.0141
FD	−0.087	0.003	0.026	0.004	−0.099	−0.03	−0.039	−0.01	−0.025	−0.039	−0.013	−0.054	−0.0823
k-WTA	0.009	−0.042	−0.007	−0.036	−0.126	−0.035	−0.056	−0.018	−0.002	−0.011	−0.006	−0.024	−0.0053
Odds	−0.043	0.063	0.064	0.049	−0.054	0.049	−0.003	0.068	−0.002	0	0.004	−0.01	−0.1807
Vanilla	0.698	0.494	0.423	0.49	0.195	0.116	0.114	0.096	0.99	0.949	0.979	0.939	0.9356

Table A15. Fashion-MNIST adaptive black-box attack 100%. Note the defense numbers in the table are the defense accuracy minus the vanilla defense accuracy. This means they are relative accuracies. The very last row is the actual defense accuracy of the vanilla network.

	FGSM-T	IFGSM-T	MIM-T	PGD-T	FGSM-U	IFGSM-U	MIM-U	PGD-U	CW-T	CW-U	EAD-T	EAD-U	Acc
ADP	0.086	−0.014	0.039	−0.012	−0.093	−0.038	−0.007	−0.029	−0.006	−0.027	−0.005	−0.03	0.013
BaRT-1	0.129	0.278	0.304	0.274	0.125	0.26	0.228	0.258	−0.015	−0.1	−0.015	−0.062	−0.0317
BaRT-4	0.165	0.319	0.383	0.317	0.176	0.276	0.273	0.288	−0.052	−0.182	−0.032	−0.148	−0.1062
BaRT-6	0.175	0.347	0.397	0.346	0.136	0.314	0.306	0.293	−0.058	−0.237	−0.044	−0.213	−0.1539
BaRT-8	0.159	0.344	0.389	0.327	0.166	0.287	0.254	0.274	−0.051	−0.255	−0.033	−0.243	−0.2212
BUZz-2	0.227	0.432	0.489	0.43	0.462	0.657	0.62	0.653	0.006	0.037	0.007	0.057	−0.0819
BUZz-8	0.279	0.469	0.535	0.466	0.672	0.818	0.809	0.835	0.007	0.039	0.009	0.061	−0.1577
ComDef	0.014	0.131	0.103	0.136	−0.029	0.088	0.074	0.093	−0.003	−0.018	−0.006	−0.012	−0.0053
DistC	0.014	−0.015	−0.001	−0.012	−0.047	−0.035	−0.011	−0.029	−0.011	−0.034	−0.021	−0.026	−0.0093
ECOC	0.057	0.193	0.233	0.176	−0.081	0.068	0.074	0.088	−0.026	−0.083	−0.026	−0.076	−0.0141
FD	−0.094	−0.038	−0.006	−0.041	−0.158	−0.037	−0.043	−0.037	−0.026	−0.071	−0.032	−0.064	−0.0823
k-WTA	0.047	−0.032	−0.024	−0.013	−0.138	−0.045	−0.048	−0.04	−0.008	−0.018	−0.026	−0.041	−0.0053
Odds	−0.036	0.011	0.009	0.017	−0.01	0.03	0.024	0.036	−0.002	−0.017	−0.004	−0.002	−0.1809
Vanilla	0.707	0.529	0.46	0.531	0.234	0.123	0.111	0.118	0.993	0.961	0.991	0.939	0.9356

Table A16. Clean prediction accuracy of the Odds defense on Fashion-MNIST and CIFAR-10 with different FPRs.

FPR	1%	10%	20%	30%	40%	50%	80%
FashionMNIST	78.6%	79.6%	78.5%	79.5%	78.6%	78.8%	79.1%
CIFAR-10	0.3%	27.8%	43.2%	61.1%	75.2%	86.2%	99.3%

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.
2. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition . *arXiv* **2014**, arXiv:1409.1556. Available online: https://arxiv.org/abs/1409.1556 (accessed on 16 September 2021).

3. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. *arXiv* **2014**, arXiv:1412.6572. Available online: https://arxiv.org/abs/1412.6572 (accessed on 16 September 2021).
4. Papernot, N.; McDaniel, P.D.; Goodfellow, I.J.; Jha, S.; Celik, Z.B.; Swami, A. Practical Black-Box Attacks against Machine Learning. *ACM Asia CCS* **2017**, *2017*, 506–519.
5. Chen, P.Y.; Zhang, H.; Sharma, Y.; Yi, J.; Hsieh, C.J. Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, Dallas, TX, USA, 3 November 2017; pp. 15–26.
6. Chen, J.; Jordan, M.I. Boundary Attack++: Query-Efficient Decision-Based Adversarial Attack. *arXiv* **2014**, arXiv:1904.02144v1. Available online: https://gaokeji.info/abs/1904.02144v1 (accessed on 16 September 2021).
7. Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.J.; Fergus, R. Intriguing properties of neural networks. *arXiv* **2013**, arXiv:1312.6199. Available online: https://arxiv.org/abs/1312.6199 (accessed on 16 September 2021).
8. Papernot, N.; McDaniel, P.; Goodfellow, I. Transferability in machine learning: from phenomena to black-box attacks using adversarial samples. *arXiv* **2016**, arXiv:1605.07277. Available online: https://arxiv.org/abs/1605.07277 (accessed on 16 September 2021).
9. Athalye, A.; Carlini, N.; Wagner, D. Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 274–283.
10. Liu, Y.; Chen, X.; Liu, C.; Song, D. Delving into Transferable Adversarial Examples and Black-box Attacks. *arXiv* **2017**, arXiv:1611.02770. Available online: https://arxiv.org/abs/1611.02770 (accessed on 16 September 2021).
11. Pang, T.; Xu, K.; Du, C.; Chen, N.; Zhu, J. Improving Adversarial Robustness via Promoting Ensemble Diversity. *Int. Conf. Mach. Learn.* **2019**, *97*, 4970–4979 .
12. Verma, G.; Swami, A. Error Correcting Output Codes Improve Probability Estimation and Adversarial Robustness of Deep Neural Networks. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.
13. Jia, X.; Wei, X.; Cao, X.; Foroosh, H. Comdefend: An efficient image compression model to defend adversarial examples. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6084–6092.
14. Raff, E.; Sylvester, J.; Forsyth, S.; McLean, M. Barrage of random transforms for adversarially robust defense. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 6528–6537.
15. Xiao, C.; Zhong, P.; Zheng, C. Enhancing Adversarial Defense by k-Winners-Take-All. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
16. Kou, C.; Lee, H.K.; Chang, E.C.; Ng, T.K. Enhancing transformation-based defenses against adversarial attacks with a distribution classifier. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
17. Roth, K.; Kilcher, Y.; Hofmann, T. The Odds are Odd: A Statistical Test for Detecting Adversarial Examples. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 5498–5507.
18. Liu, Z.; Liu, Q.; Liu, T.; Xu, N.; Lin, X.; Wang, Y.; Wen, W. Feature distillation: Dnn-oriented jpeg compression against adversarial examples. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 10–15 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 860–868.
19. Carlini, N.; Athalye, A.; Papernot, N.; Brendel, W.; Rauber, J.; Tsipras, D.; Goodfellow, I.; Madry, A.; Kurakin, A. On evaluating adversarial robustness. *arXiv* **2019**, arXiv:1902.06705. Available online: https://arxiv.org/abs/1902.06705 (accessed on 16 September 2021).
20. Carlini, N.; Wagner, D. Adversarial examples are not easily detected: Bypassing ten detection methods. In Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, Dallas, TX, USA, 3 November 2017; pp. 3–14.
21. He, W.; Wei, J.; Chen, X.; Carlini, N.; Song, D. Adversarial example defense: Ensembles of weak defenses are not strong. In Proceedings of the 11th {USENIX} Workshop on Offensive Technologies ({WOOT} 17), Vancouver, BC, Canada, 14–15 August 2017.
22. Tramer, F.; Carlini, N.; Brendel, W.; Madry, A. On adaptive attacks to adversarial example defenses. *arXiv* **2020**, arXiv:2002.08347. Available online: https://arxiv.org/abs/2002.08347 (accessed on 16 September 2021).
23. Dong, Y.; Fu, Q.A.; Yang, X.; Pang, T.; Su, H.; Xiao, Z.; Zhu, J. Benchmarking Adversarial Robustness on Image Classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
24. Mahmood, K.; Nguyen, P.H.; Nguyen, L.M.; Nguyen, T.; van Dijk, M. BUZz: BUffer Zones for defending adversarial examples in image classification. *arXiv* **2019**, arXiv:1910.02785. Available online: https://arxiv.org/abs/1910.02785 (accessed on 16 September 2021).
25. Yuan, X.; He, P.; Zhu, Q.; Li, X. Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2805–2824. [CrossRef] [PubMed]
26. Brendel, W.; Rauber, J.; Bethge, M. Decision-Based Adversarial Attacks: Reliable Attacks Against Black-Box Machine Learning Models. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April 30–3 May 2018.

27. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards deep learning models resistant to adversarial attacks. *arXiv* **2021**. arXiv:1706.06083. Available online: https://arxiv.org/abs/1706.06083 (accessed on 16 September 2021).
28. Carlini, N.; Wagner, D. Towards evaluating the robustness of neural networks. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (sp), San Jose, CA, USA, 22–26 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 39–57.
29. Guo, C.; Gardner, J.R.; You, Y.; Wilson, A.G.; Weinberger, K.Q. Simple black-box adversarial attacks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
30. Chen, J.; Gu, Q. Rays: A ray searching method for hard-label adversarial attack. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, 6–10 July 2020; pp. 1739–1747.
31. Kurakin, A.; Goodfellow, I.; Bengio, S. Adversarial examples in the physical world. *arXiv* **2017**. arXiv:1607.02533.
32. Dong, Y.; Liao, F.; Pang, T.; Su, H.; Zhu, J.; Hu, X.; Li, J. Boosting adversarial attacks with momentum. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 9185–9193.
33. Chen, P.Y.; Sharma, Y.; Zhang, H.; Yi, J.; Hsieh, C.J. Ead: Elastic-net attacks to deep neural networks via adversarial examples. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
34. Chen, J.; Jordan, M.I.; Wainwright, M.J. Hopskipjumpattack: A query-efficient decision-based attack. In Proceedings of the 2020 IEEE Symposium on Security and Privacy (sp), San Francisco, CA, USA, 17–21 May 2020; IEEE: Piscataway, NJ, USA, 2020, pp. 1277–1294.
35. Byun, J.; Go, H.; Kim, C. Small Input Noise is Enough to Defend Against Query-based Black-box Attacks. *arXiv* **2021**, arXiv:2101.04829. Available online: https://arxiv.org/abs/2101.04829 (accessed on 16 September 2021)
36. Athalye, A.; Engstrom, L.; Ilyas, A.; Kwok, K. Synthesizing robust adversarial examples. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 284–293.
37. Carlini, N.; Wagner, D. MagNet and "Efficient Defenses against Adversarial Attacks" Are Not Robust to Adversarial Examples. *arXiv* **2017**, arXiv:cs.LG/1711.08478. Available online: https://arxiv.org/abs/1711.08478 (accessed on 16 September 2021).
38. Xie, C.; Wang, J.; Zhang, Z.; Ren, Z.; Yuille, A. Mitigating adversarial effects through randomization. *arXiv* **2018**, arXiv:1711.01991. Available online: https://arxiv.org/abs/1711.01991 (accessed on 16 September 2021).
39. Krizhevsky, A.; Nair, V.; Hinton, G. Learning Multiple Layers of Features from Tiny Images. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.9220&rep=rep1&type=pdf (accessed on 16 September 2021)
40. Xiao, H.; Rasul, K.; Vollgraf, R. Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv* **2017**. arXiv:1708.07747. Available online: https://arxiv.org/abs/1708.07747 (accessed on 16 September 2021).
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
42. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
43. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980. Available online: https://arxiv.org/abs/1412.6980 (accessed on 16 September 2021).
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. *Lect. Notes Comput. Sci.* **2016**, 630–645. [CrossRef]
45. Kou, C.K.L.; Lee, H.K.; Ng, T.K. A compact network learning model for distribution regression. *Neural Netw.* **2019**, *110*, 199–212. [CrossRef] [PubMed]

entropy

MDPI

Article

Feature Selection for Regression Based on Gamma Test Nested Monte Carlo Tree Search

Ying Li [1], Guohe Li [1,*] and Lingun Guo [1,2]

[1] Beijing Key Lab of Petroleum Data Mining, Department of Geophysics, China University of Petroleum, Beijing 102249, China; 2016315014@student.cup.edu.cn (Y.L.); 2019310406@student.cup.edu.cn (L.G.)
[2] College of Software, Henan Normal University, Xinxiang 453007, China
* Correspondence: lgh102200@sina.com

Abstract: This paper investigates the nested Monte Carlo tree search (NMCTS) for feature selection on regression tasks. NMCTS starts out with an empty subset and uses search results of lower nesting level simulation. Level 0 is based on random moves until the path reaches the leaf node. In order to accomplish feature selection on the regression task, the Gamma test is introduced to play the role of the reward function at the end of the simulation. The concept Vratio of the Gamma test is also combined with the original UCT-tuned1 and the design of stopping conditions in the selection and simulation phases. The proposed GNMCTS method was tested on seven numeric datasets and compared with six other feature selection methods. It shows better performance than the vanilla MCTS framework and maintains the relevant information in the original feature space. The experimental results demonstrate that GNMCTS is a robust and effective tool for feature selection. It can accomplish the task well in a reasonable computation budget.

Keywords: feature selection; regression; nested monte carlo tree search (NMCTS); filter; gamma test; GNMCTS

Citation: Li, Y.; Li, G.; Guo, L. Feature Selection for Regression Based on Gamma Test Nested Monte Carlo Tree Search. *Entropy* **2021**, *23*, 1331. https://doi.org/10.3390/e23101331

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 31 August 2021
Accepted: 7 October 2021
Published: 12 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Feature selection is a commonly used procedure in data pre-processing. It is further categorized into the filter, wrapper and embedded methods. The filter method generates an optimal feature subset according to a certain evaluation function; it is independent of a succeeded classifier or regressor. Therefore, it can obtain the final result faster. On the contrary, the wrapper method evaluates feature subset according to classifier or regressor result. Thus, it can achieve better performance on the classifier or regressor, but it takes a longer time for the whole process. The embedded method integrates feature selection and model training together. It utilizes learned hypotheses to accomplish feature selection during model-optimized training. In order to achieve a more flexible model combination, the filter method is a good choice.

The Monte Carlo Tree Search (MCTS) method has achieved many states of art performances in the game domain, such as Go [1,2], LOA, Bubble Breaker, SameGame, etc. [3]. These games can be viewed as a large-scale Markov decision process. From this perspective, it can also deal with online planning, route scheduling and combinatorial optimization problems. The success of AlphaGo has had a profound influence on artificial intelligence (AI) approaches. Many reinforcement learning methods were adapted in feature selection problems and achieved satisfactory results. Typically, MCTS for feature selection has developed many fine frameworks [4–6]. It can be categorized into the filter or wrapper method depending on the specific framework design. On the one hand, the classifier or regressor results can be directly returned as a reward. On the other hand, evaluation value calculated from certain criteria such as information gain, Fisher's score, etc., can be used as a reward during iteration. The process can then be considered as a filter method. To be specific, the tree search combines selective strategy and simulation strategy called rollout to obtain the

optimal solution. It has a trade-off of exploration versus exploitation, which is also well known as the e–e dilemma. The UCT technique is the most popular way to control the growth of the search tree. The UCB-tuned1 was proposed soon after; this technique adapted well in single-player games, so in this paper, its basic form was also used for feature subset selection. Typically, this paper mainly focused on the regression task. Gamma test was introduced to play the role of the evaluation function. Since MCTS is based on selective sampling and simulation, the result is backpropagated until the episode ends; node values are only updated until then. Speed of convergence and efficient calculation becomes a key. The Gamma test [7–10] is a non-parametric tool to measure the non-linear relationship between inputs and outputs. It has time complexity O(Mlog M), where M is the number of data points. One run of the Gamma test for thousands of data points usually takes a few seconds. Therefore, the Gamma test can fit this task well. Usually, MCTS takes random moves or follows a simple heuristic strategy during simulation. Nested MCTS (NMCTS) has a stronger performance compared to regular MCTS [11]. NMCTS has beaten MCTS in the deterministic Markov decision process domains such as SameGame, Clickomania. It is natural to expect NMCTS could achieve better performance in feature selection compared to MCTS. NMCTS of higher nesting level uses best search result of lower nesting level as simulation result. A level 1 search corresponds to regular MCTS. Based on the MCTS feature selection method, we proposed the Gamma test nested MCTS method for feature selection in this paper. The main contributions of the study are listed as below:

- The novel method GNMCTS is proposed to solve feature selection on regression tasks, which is less explored in recent researches;
- GNMCTS uses the Gamma test as a reward function, which is easy to implement and takes only a few seconds on a dataset with tens or hundreds of feature dimensions;
- GNMCTS searches the feature space more efficiently through nesting; the two hyper-parameters, nesting level and iteration numbers, are flexible to tune, which can be set to different values on different nesting levels;
- GNMCTS is tested on seven real-world datasets, and the results are compared with the other six feature selection methods based on reinforcement learning. The result shows the superiority of GNMCTS.

The paper is organized as follows: Section 2 briefly reviews the related work. In Section 3, the background methodology on the basic MCTS framework of feature selection is briefly introduced. Section 4 focuses on the GNMCTS method. Given the background of MCTS application in the feature selection domain in Section 4.1, NMCTS was extended to solve the problem in Section 4.2. A revised reward function based on the Gamma test is introduced in Section 4.3. Section 5 mainly compared GNMCTS with other feature selection methods on UCI and WEKA datasets. Conclusions and future work are stated in Section 6.

2. Related Work

Feature selection is widely used during data pre-processing. It aims to reduce the data dimensions without losing valuable information and accelerate the succeeded tasks while retaining high accuracy.

The wrapper methods are dependent on the specific classification or regression algorithms. The result of the classifier or regressor acts as an evaluation standard for the candidate feature subsets. Huang [12] proposed a method called FCSVM-RFE for gene detection, where representative genes are ranked by SVM-RFE after gene clustering. Masood [13] proposed to use an incremental search strategy combined with an extreme learning machine classifier. The research of these wrapper methods focused on alleviating time complexity. However, the inherent property of an expensive computation budget is not easy to conquer. Filter methods employ certain measurements such as information gain [14] to evaluate subsets. The main focus lay on improving accuracy, but most researchers pay attention to classification tasks that are not appropriate for regression.

Hybrid methods take advantage of both categories. These methods have independent metrics and specific learning algorithms to measure the subsets.

From the perspective of searching strategy, feature selection methods can be categorized into exhaustive, heuristic, meta-search. Exhaustive search is basically impossible to implement on real-world datasets. This leaves the researchers two directions [15] to explore search space: guiding the search process under specific heuristics or using greedy hill-climbing methods. The latter is often simple to implement, such as sequential forward or backward selection (SFS, SBS), the best first search. These methods follow a monotonic behavior of feature selection. The popular heuristics include genetic algorithm (GA), ant colony optimization (ACO) and particle swarm optimization (PSO). Nguyen [16] presented a comprehensive survey on the state-of-the-art works applying swarm intelligence to achieve feature selection in classification, with a focus on the representation and search mechanisms. Sharma [17] conducted a systematic review methodology for synthesis and analysis of one hundred and seventy-six articles. The parameters related to these nature-inspired methods are complex to control and needed to be tuned with great effort. While feature selection based on reinforcement learning method was recently developed with the success of AlphaGo. Fan W. [18] proposed an Interactive Reinforced Feature Selection (IRFS) framework that guides agents by not just self-exploration experience but also diverse external skilled trainers to accelerate learning for feature exploration. The hyper-parameters in these methods are relatively easy to control, and fewer parameters require to be tuned.

The stopping criteria have a direct influence on the size of the candidate feature subset. It indicates when the search procedure should be stopped. The commonly used criteria include (1) pre-defined number of iterations, (2) pre-defined number of features, (3) difference or improvements between successive iteration steps and (4) judgment by specific evaluation functions. The above criteria do not couple with different methods flexible enough. Automatic stopping criteria should be customized depending on the specific learning algorithms.

In summary, to overcome the problems stated above, the proposed method in this paper focused on the design of the filter feature selection method for the regression task. In order to evaluate the candidate subsets, the Gamma test was used, and NMCTS in game theory was introduced with the merits of easily controllable hyperparameters. The automatic stopping criteria were designed considering the structure characteristic of the search tree and the property of the Gamma test.

3. Background Methodology

3.1. Basic Procedure of Monte Carlo Tree Search (MCTS) for Feature Selection

Feature selection can be regarded as a sequential decision problem. It has many common points with a single-player game that has no opponent. To be specific, the action space and state space are finite and discrete. Given a set of features $F_{All} = \{X_1, X_2, \ldots, X_M\}$, MCTS algorithm will finally return the best action set as the best feature subset F_{best}. A brief introduction of MCTS for the feature selection problem is represented in Figure 1. The algorithm can be summarized into the following four basic steps, which are:

(1) Selection: Let N_{root} define the root node where the feature subset is empty (i.e., $F_{root\_sub} \in \varnothing$), starting from N_{root}, use some tree policy to gradually descend inside the tree until the path reaches a non-terminal state leaf node N_i. Choosing an action corresponds to adding a selected feature to the candidate feature subset $F_{sub} = F_{sub} \cup \{N_i\}$, F_{sub} is also used as the state of N_i;

(2) Expansion: Expand N_i until it has no more legal actions that correspond to the case where the remaining feature set is empty (i.e., $F_{All} \backslash F_{sub} = \varnothing$) or pre-conditioned number of expanded children is reached. Then, add expanded children node N_j to N_i. Initialize N_j with new node state as $F_{sub} = F_{sub} \cup \{N_j\}$, record its parent N_i. The features already appeared in F_{sub} will no longer be in the legal actions;

(3) Simulation: This procedure is also called a rollout or a playout. In general, starting from the leaf node N_i, the successive nodes are chosen step by step by some simulation policy until it reaches a terminal state or pre-conditioned computation budget;
(4) Backpropagation: The simulation result is backpropagated through the nodes during the selection phase on the path, and their statistics are updated. The statistics include the visit number of nodes and their values.

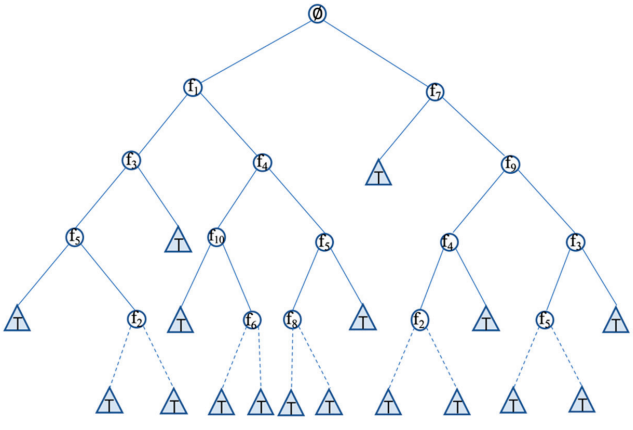


Figure 1. Monte Carlo Tree Search (MCTS).

The tree search strategy includes two policies. The two policies involved in the selection phase and simulation phase, respectively, are:

(1) Tree policy: It is a strategy to select features. Furthermore, it can be split into two aspects. One is selected during the tree build-up period, and another is the final selection of picking up the best feature sequence F_{best}. The former has many variations [19]; the most popular version proposed by Auer et al. is called UCB1, represented by Equation (2), the policy indicates to execute an action with promising potentials which can maximize value in Equation(2),

$$\overline{\mu_j} = \frac{Q(s,a)}{n(s,a)} \tag{1}$$

$$UCB1 = \overline{\mu_j} + \sqrt{\frac{C_e \cdot \ln n}{n_j}} \tag{2}$$

where $\overline{\mu_j}$ defines average gain of the selected feature, s is the current state which represents F_{sub} in the feature selection problem, a represents the currently selected action that corresponds to adding a new feature to the current subset. $Q(s,a)$ is an instant reward after adding the new feature to the current subset. $n(s,a)$ defines the number of visits of the current node n_i, n_j defines the number of visits of its children nodes. With the increasing visited number of uncertainty nodes, the asymmetrical growing search tree gradually prefer those nodes that gain a higher exploitation score $\overline{\mu_j}$. The confidence interval shrinks with repeated visits.

To a large degree, how much exploration part accounts for evaluation result relies on the exploration constant C_e. Aiming at the choice of this parameter, Oleksandr I. Marchenko proposed the MCTS-TSC (tree shape control) method, which used the original depth–width criteria [20]. For the feature selection problem, there is no fixed shape such as depth dominant or width dominant for the search tree. It is implicit in constraining the growing direction of the tree. Considering the complexity of the algorithm and computation budget, C_e chosen by trails is a better and easier idea, for those who do not care about the cost may combine the newest technique on pruning.

For the final feature subset decision, the target is to achieve the highest classification accuracy or minimum regression error, so the tree should choose nodes with the best score record that have been seen so far rather than the average score.

Default policy: It is a strategy to implement a rollout. There are two ways to perform this: either by a uniform random selection policy or by some simple heuristic based on prior domain knowledge. The enhancements on the rollout policy can be found in Cameron B. Browne [21].

The pseudocode for MCTS is listed in Algorithm 1 as follows:

Algorithm 1 MCTS(time_limit,iteration_limit,explorationRate)
//explorationRate defines the degree of exploration

```
root = treeNode(initialState, None)
While (time<time_limit & count<iteration_limit) do
  randomPolicy(state):
      while not state.isTerminal():
          try:
              action = random.choice(state.getPossibleActions())
          except IndexError:
              raise Exception("Non-terminal state has no possible actions: " + str(state))
          state = state.takeAction(action)
      return state.getReward()
       def selectNode(self, node):
          while not node.isTerminal:
               if node.isFullyExpanded:
                    node = self.getBestChild(node, self.explorationConstant)
               else:
                    return self.expand(node)
          return node
       def expand(self, node):
          actions = node.state.getPossibleActions(node)
          for action in actions:
                    newNode = treeNode(node.state.takeAction(action), node)
                    node.children[action] = newNode
                    if len(actions) == len(node.children):
                      node.isFullyExpanded = True
                    return newNode
       def backpropogate(self, node, reward):
            while node is not None:
                    node.numVisits += 1
                    node.totalReward += reward
                    node = node. Parent
```

3.2. Gamma Test

The Gamma test is a non-linear modeling and analysis tool to test the relationship between input and output variables on the numerical dataset. It fits the job of feature subset selection fast enough; the time complexity of the Gamma test is $O(MlogM)$, where M is the number of input samples. One single run of the Gamma test takes roughly only a few seconds on a dataset that consists of thousands of instances with hundred features. The Gamma test has already been applied in many industrial and natural resource problems [22–25]. In the section, a brief introduction of the calculation steps and theory are organized.

The relationship between input X_i and output y_i can be represented by a smooth function in the following form:

$$y_i = f(X_i) + r \tag{3}$$

where $f(X)$ is the assumed regression model, r is a noise that cannot be explained by $f(X)$. When there is no noise, r is zero.

Define $X_{N[i,k]}$ as a list of k nearest neighbors of the ith point X_i in the input space $\{X_1, X_2, X_3, \ldots, X_M\}$ found by KD tree. p is defined as the number of the nearest neighbors used to calculate statistic Γ. Based on many researches and experiments [26], it is shown that $p = 10$ can obtain better results in a reasonable time.

Define $y_{N[i,k]}$ as the list of the target value corresponding to the nearest neighbor sequence $X_{N[i,k]}$. It should be noticed that they are not the list of kth nearest neighbors to the ith point y_i. Calculate the Euclidean distance between the nearest neighbors and the query point in the input and output space,

$$\delta_M(k) = \frac{1}{M}\sum_{i=1}^{M}\left|X_{N[i,k]} - X_i\right|^2 \tag{4}$$

$$\gamma_M(k) = \frac{1}{2M}\sum_{i=1}^{M}\left|y_{N[i,k]} - y_i\right|^2 \tag{5}$$

By Equation (3), and the continuity of unknown function $f(X)$, the probability of $\gamma_M(k) \to var(\gamma)$ as $\delta_M(k) \to 0$. However, it is impossible for $\delta_M(k)$ to reach zero infinitely. Therefore, the limit value $\gamma_M(k)$ that infinitely approximates $var(\gamma)$ cannot be directly calculated. Finally, by Equation (5),

$$\gamma_M(k) \to var(\gamma)\ as\ \delta_M(k) \to 0 \tag{6}$$

the Gamma test assumes that the relationship between the k-neighbor pairs $\delta_M(k)$, $\gamma_M(k)$ are approximately linear, and the **slope** is a constant A,

$$\gamma_M(k) = A\delta_M(k) + var(\gamma) + o(\delta_M(k)),\ as\ \delta_M(k) \to 0 \tag{7}$$

Based on the above assumptions, the least-squares linear fit is performed on $\{(\delta_M(k), \gamma_M t(k)), 1 < k \leq p\}$. Equation (7) can be written as

$$\gamma_M(k) = A\delta_M(k) + \Gamma \tag{8}$$

The intercept Γ is the estimated noise variance. The evidence of linear progression can be found in the research by Evans [9]. In some cases, Γ value is negative. The first reason is that number of samples is too small, such as under a hundred points, there are no sufficient data points to obtain an accurate outcome. Another reason is the regression model is so smooth that data points can be fully explained. When $\Gamma \leq 0$, it is replaced by $|\Gamma|$. Similarly, the case that $\Gamma > \mathrm{var(y)}$ may occur. When this case is true, some pre-process on data, such as abnormal point detection, should be performed. Since the Gamma test can only examine the non-linear relationship between inputs and output, linear regression should also be considered.

4. GNMCTS for Feature Selection

4.1. Nested Monte Carlo Tree Search Subsection

The nested Monte Carlo tree search (NMCTS) was proposed by Hendrik Baier [11]; it was an enhancement work on Nested Monte Carlo Search (NMCS) [21]. The method was tested on many single-player games such as Solitare, SameGame, Bubble Breaker, etc. [27–30]. It was compared with basic NMCS on different nest levels. NMCTS outperformed regular MCTS on those single-player games, and it can also deal with large Markov decision processes. Therefore, it should adopt the feature selection problem well. NMCTS is different from MCTS in the simulation phase. Selection, expansion and backpropagation phases still remained the same as described in Section 3.1. The NMCTS combined MCTS on a lower base level, leaving itself called recursively on higher nest levels. The techniques of MCTS, such as UCB-tuned1, can also be used in NMCTS. While MCTS uses random feature selection beginning with a given state until reaching a terminal state during rollout, NMCTS uses a heuristic that for every feature selection starting from the given state, and

level n search calls level $n-1$ search result. Then, select the feature with the highest score from level $n-1$ search. As illustrated in Figure 2, curve lines represent for level 0 search. It is a normal random simulation. Then, level 1 search calls the result of level 0 search and selects the action with the best score. Level 2 search calls level 1 search and selects the feature with the best score.

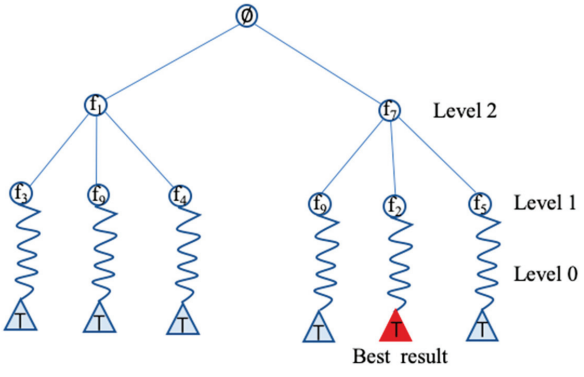


Figure 2. A level 2 NMCTS illustration.

The best feature sequence is recorded every iteration and compared in case the performance is not improved by adding the new feature. After the computation budget runs out, the best score and sequence are returned. The pseudocode of NMCTS is shown below in Algorithm 2.

Algorithm 2 NMCTS (startNode, Seq, max_iter, level)
//Seq defines best F_{sub} the tree has found so far

```
best_reward = inf.
best_seq = ()
Current_node = startNode
For iteration number in the called level:
    While Current_node is not terminal and not fully expanded:
Current_node = selection(Current_node)
Seq = Current_node.feature_subset
If level=1:
    While Current_node is not terminal:
      Reward, Seq = Random_rollout(Current_node)
Else:
    Reward, Seq = NMCTS(Current_Node, Seq, max_iter, level-1)
Back_propogation(Current_Node, reward)
If Reward < best_reward:
    best_reward = reward
    best_seq = Seq
```

4.2. Gamma Test as Evaluation Function for Regression Task

Next, a simple example was illustrated to show that the Gamma test could be used in feature selection.

The butterfly dataset [31,32] consists of two relevant features, three redundant features and three irrelevant features, which correspond to X1, X2, J3, J4, J5, I6, I7 and I8. In this trial, we generated 10,000 data points with eight features above. Figure 3 illustrates a 3d projection of relevant feature values X1 and X2 on the Y-axis. In Figure 4, an irrelevant feature I6 was added, which was considered as noise. The exhaustive search must traverse 2^8-1 combinations. As it took only a few seconds, we computed the gamma value for all the possible combinations, and the minimum gamma value should indicate the best

relevant feature combination. The combination of the first two features obtained the minimum gamma value of 0.00043 among all cases, which is close to zero, as shown in Figure 5. This validated Gamma test estimated the best feature subset correctly.

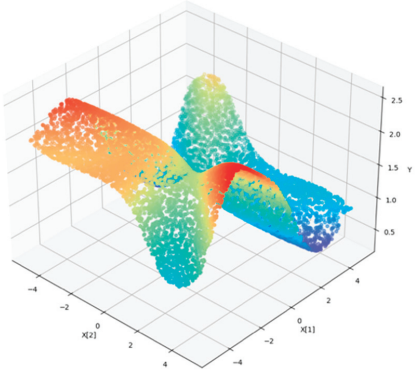


Figure 3. Butterfly 3d projection with X1, X2 and Y.

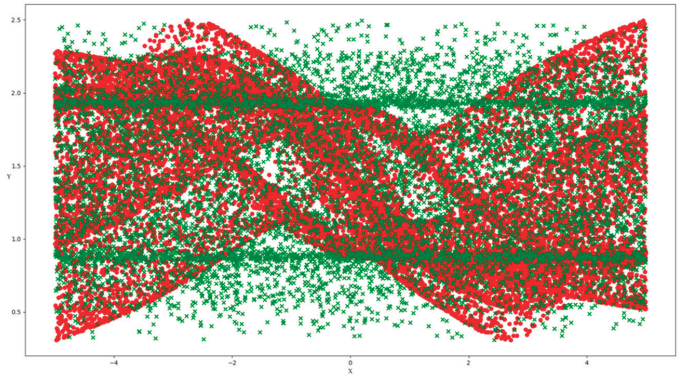


Figure 4. Butterfly scatter plot with X1, X2, I6 and Y.

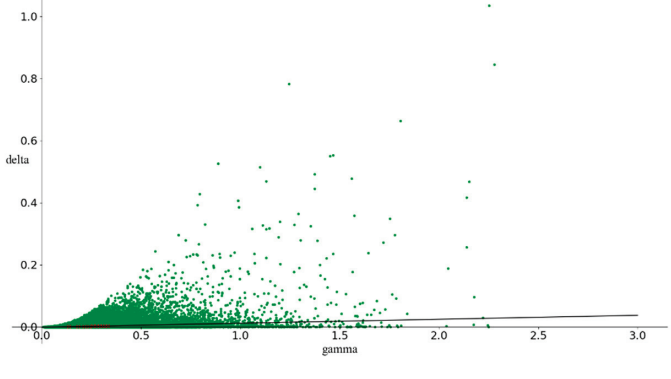


Figure 5. Gamma scatter plot for the smooth function.

4.3. Gamma Test Modified Node Selection Policy

In two-player games, the reward is often denoted with {−1,0,1}, representing loss, draw or win. The reward interval of a node falls within [−1,1]. The value of Γ has a large range of variations in different feature subsets. According to Maarten P.D. Schadd [3], there are two solutions; one is scaling the reward back into the interval [−1,1], and the other solution is adding a constant to calculate the reward that would fit the application domain. In the feature selection problem, although the exact maximum Γ value is not known, according to Equation (7), it can be evaluated by the real variance of the target data var(y). For feature selection, a modified UCT version is used. The target is to maximize Equation (9),

$$\mu_j + \mathrm{Ce}\cdot\sqrt{\frac{ln\ n}{n_j}} + \sqrt{\frac{\Sigma\Gamma^2 - n_j * \mu_j^2 + \mathrm{D}}{n_j}} \tag{9}$$

The left two terms of Equations (9) are the same in Equation (2), the third term contains the sum of squared rollout reward $\Sigma\Gamma^2$ represents a possible deviation of the child node, it is corrected by the expected results $n_j * \mu_j^2$. Ce and D are constants discussed above aiming at exploring rarely visited nodes. In our experiment, D is set with the value of var(y). Finally, the best feature subset can be found by best policy π^*, which minimizes the Γ value; this can be written in the form of Equation (10).

$$\pi^* = \underset{\pi}{\operatorname{argmin}}\ \Gamma \tag{10}$$

An indicator variable defined:

$$\mathrm{Vratio} = \frac{\Gamma}{\mathrm{var(y)}} \tag{11}$$

Vratio provides a scale-invariant measure; normally, the value is in the range [0, 1]. If the Vratio value is close to zero, then it means the input variable has a strong non-linear relationship with the target. If the Vratio value is close to one, then the prediction target can hardly be explained by input variables; the performance of the regressor is more likely to be a random walk.

To be noted, the filter feature selection method has to generate a subset with a certain number of features. Moreover, the final number of selected features has a direct influence on the result and succeeding computation cost. Romaric [5] proposed to add a stopping feature in the default policy. A stopping feature is chosen with probability $\mathrm{rand}(0,1) > 1 - q^d$, where d is the depth of the current node in the simulation and q is a constant, where $q < 1$. With the growth of the tree, d becomes larger, the probability of the stopping feature being selected also becomes bigger. In this paper, the Vratio is considered to replace q, and the modified stopping condition becomes:

$$\mathrm{rand}(0,1) > (1 - \mathrm{V_{ratio}})^d \tag{12}$$

The intuition for the inequality is to achieve a satisfactory regression result with a small number of features. Since Vratio can show the goodness of fitting by the current feature subset, the smaller Vratio is, the smaller the probability of selecting the stopping feature. Then, the tree can further explore the potential path. Otherwise, the larger Vratio is, the sooner the simulation phase ends. The deeper the search tree grows, the bigger probability for the stopping feature to be selected. Another stopping condition takes consideration of the original feature set size of F. For a high dimension feature set, the

timing for stopping should be delayed in case feature space is not explored enough. The stopping feature will work if any case in Equation (12) or Equation (13) happens.

$$\text{rand}(0,1) < \frac{\text{node.depth}}{\text{size}(F)} \tag{13}$$

5. Experimental Results

This section demonstrates the performance of the NMCTS gamma algorithm on selecting the best feature combination, and the experiments were conducted on seven benchmark datasets. All the experiments were implemented in Python with environment 48 Intel(R) Xeon(R) Silver 4214 CPU 2.20 GHz and 125 GB of RAM.

5.1. Datasets

Seven datasets were used for comparison and performance validation. Datasets were taken from two publicly available repositories [33,34], UCI and WEKA. Specific information is shown in Table 1. The feature dimensions and the number of instances varied to gain diversity in characteristics. Both the features and labels are numeric. If datasets contained some ID information, then that column was deleted. The range of labels was listed in the fifth column of Table 1. The Parkinsons_Updrs dataset is composed of a range of biomedical voice measurements from 42 people with early-stage Parkinson's disease. There are two prediction targets, motor Updrs and total Updrs. To be convenient for comparison, we only considered the total Updrs as a target in the experiments. However, one can calculate the scores, respectively, using the proposed algorithm on multi-output datasets. The Puma32h dataset was synthetically generated from a realistic simulation of the dynamics of a Unimation Puma 560 robot arm. The task is to predict the angular acceleration of one of the robot arm's links. The Bank32nh was synthetically generated from a simulation of how bank customers choose their banks. Tasks are based on predicting the fraction of bank customers who leave the bank because of full queues. Ailerons addresses a control problem, namely flying an F16 aircraft. The attributes describe the status of the airplane, while the goal is to predict the control action on the ailerons of the aircraft. Pol describes a telecommunication problem in a commercial application. Triazines predicts the activity from the descriptive, structural attributes. Residential building includes construction cost, sale prices, project variables, and economic variables corresponding to real estate single-family residential apartments in Tehran, Iran, and the goal is to predict sale prices.

Table 1. Benchmark datasets.

No.	Dataset	Instances	Features	Label Range
1	Parkinsons_Updrs	5875	19	[5.0377,39.511]
2	Puma32h	4123	33	[−0.0847,0.0898]
3	Bank32nh	8192	33	[0,0.8197]
4	Ailerons	13,750	41	[−0.0036,0]
5	Pol	15,000	49	[0,100]
6	Triazines	186	61	[0.1,0.9]
7	Residential building	372	109	[50,6800]

5.2. Experimental Settings

We conducted five-fold cross-validation for all the comparison experiments. The iteration number limit was set to 1000. The corresponding dimension reduction effect and computation time were compared on six datasets of different sizes. The experiment was repeated 20 times then took average values as results. For comparison purposes, the best feature subsets of each feature selection method in Table 2 were tested on the same gradient boosting regressor from the scikit-learn module. Specific parameters of this regressor were: The number of estimators was set to 25, max depth was 4, min samples split was 2, the learning rate was 0.2, the loss was the least square. Before inputting the algorithm, standard

normalization was performed for all the datasets. Features with 0 variances that show no contribution to the prediction model were deleted at first.

Table 2. Experimental methods.

Method	Description
PSO	Particle Swarm Optimization based method [35]
QBSO	Q-learning based Bee Swarm optimization method [36,37]
MCTS RreliefF	Improved relief feature selection algorithm based on MCTS [38]
MCTS RAVE	Feature selection as a One-Player Game [39]
FSTD	Feature selections using Temporal Difference [40]
GRNN	General Regression Neural Network [41]

5.3. Comparison Methods and Metrics

We compared the NMCTS gamma algorithm with six state-of-the-art feature selection methods for the regression task listed in Table 2. We mainly focused on feature selection methods using reinforcement algorithms which included temporal difference learning, Q-learning and enhanced MCTS methods.

A brief introduction of parameter settings related to methods in Table 2 are listed below:

- The objective function of particle swarm optimization (PSO) consists of customized evaluation function results and the feature number reduction ratio. For comparison purposes, the evaluation function's part in it was substituted by the Gamma test;
- QBSO integrated the Bee Swarm Optimization algorithm with Q learning for solving feature selection tasks. The original algorithm was designed for classification. In the regression case, the fitness of BSO was substituted from the accuracy of the KNN classifier to the mean square error of the KNN regressor. The reward function of Q Learning only differed in minor sign modification from its original paper;
- For MCTS_RreliefF, as the ReliefF algorithm was used to implement classification on multiclass outputs feature selection problem, we changed it into RreliefF algorithm; the other framework in the paper remained the same, including most parameter settings in [38];
- For MCTS with global rave and local rave (MCTS_RAVE), the reward function of MCTS was originally AUC. It was also substituted by the Gamma test;
- For the Temporal Difference learning method, the reward function was also changed into the Gamma test. Learning rate alpha was 0.5, epsilon in the ε-greedy strategy was 0.5. Epsilon decay rate and alpha decay rate were set to 0.995, and the discount parameter was 0.3, parameter b in heuristic was 0.6, stop condition parameter was 3;
- GRNN used the Radical basis function as the kernel. The kernel bandwidth was decided by Silverman's rule of thumb. Type of the gradient search solver was chosen L-BFGS-B;
- GNMCTS used level 2 nesting search. The iteration number of nesting was set to 10 for level 2 and 100 for level 1. The UCT exploration constant Ce was 0.3. The expansion width of each node was 10. The rest parameters were the same with the MCTS_RAVE method.

The final results were evaluated on seven metrics, including the mean squared error (MSE), mean absolute error (MAE), R-square (R2), explained variance score (EV), dimension reduction (DR) effect, confidence interval and computation time. The expressions of these measurements are as follows:

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{14}$$

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{15}$$

$$\mathrm{R2} = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \tag{16}$$

$$\mathrm{EV} = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 - \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|}{\sum_{i=1}^{m}(y_i - \bar{y}_i)^2} \tag{17}$$

$$P(L_m < \hat{y}_{i_i} < U_m) = \gamma \tag{18}$$

The smaller MSE and MAE are, the more accurate predictions are. On the contrary, the larger R2 and EV are, the more powerful of model predictions are. When the value is close to 1, it indicates the model can perfectly predict all data correctly. When the value is close to 0, it indicates the model performance essentially acts as a baseline model. When the value drops below 0, it indicates the model is worse than the baseline model. This could be the reason why there is no linear or non-linear relationship between inputs and outputs. The difference between R2 and EV lies in the mean value of the residual, i.e., whether $\frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|$ is 0 or not. In Equation (18), γ is a number between 0 and 1, and it was set 0.95 in this paper. L_m, U_m are lower and upper confidence bound of variable y_i.

The dimension reduction ability is represented by Equation (19). The numerator and denominator are the number of selected features and total feature subset, respectively.

$$\mathrm{DR} = 1 - \frac{\#\text{selected features}}{\text{total features}} \tag{19}$$

5.4. Results and Comparisons

According to the aforementioned parameter settings, experiments were conducted as previously described.

As shown in Table 3a,b, GNMCTS obtained minimum MSE and MAE on Bank32nh and Parkinson's datasets. On the rest dataset, the results were very close to the best results obtained by GRNN and PSO. GRNN obtained the four best records on triazines, Puma32h, Pol, ailerons and residential building. This could explain why the GRNN method was the wrapper feature selection method. It adjusted neural weights of the hidden layer according to the MSE of regression. Therefore, it has inherent lower MSE and MAE than filter methods, but it cannot deal with a high dimension dataset when the feature number and instance number are large. Additionally, it took a much longer computation time compared with other methods. GRNN failed when calculating the triazines dataset. These were the main problems with GRNN. PSO obtained the smallest MAE and MSE on the triazines dataset but did not perform well in other datasets. GNMCTS was robust and easy to implement. The GNMCTS method obtained better results than MCTS_Rrelieff, PSO, QBSO, MCTS_rave and TD_learning within the same time control. Specifically, GNMCTS outperformed MCTS as expected on four datasets and achieved similar results on Puma32h, Pol and Residential building datasets. This would improve if more iterations were allowed on level 1 or 2 nest level. As the iteration limit was 1000 for both GNMCTS and MCTS, this limited iteration number of GNMCTS on level 1 multiplied by that of level 2 must equal 1000. This would weaken exploration ability on lower-level search space. With the increase in iterations, GNMCTS would finally outperform MCTS. The results of GNMCTS compared with the original dataset without feature selection had slightly improved or maintained the same.

Table 3. GNMCTS results compared with other methods on seven datasets.

(a) MSE:

	MCTS_ RrelieF	GRNN_ isotropic	PSO	QBSO	MCTS_rave	TD_ learning	GNMCTS	Original
triazines	0.2283	–	0.0172	0.2246	0.0183	0.0235	0.0182	0.0169
puma32h	6.60×10^{-5}	6.50×10^{-5}	6.90×10^{-5}	9.07×10^{-4}	6.60×10^{-5}	9.22×10^{-4}	6.70×10^{-5}	6.90×10^{-5}
pol	1330.2471	83.5928	1613.6182	1233.6613	99.4774	722.0634	96.1439	84.5149
bank32nh	0.0108	0.0074	0.0072	0.0133	0.0071	0.0151	0.0071	0.0071
ailerons	7.79×10^{-8}	2.81×10^{-8}	4.99×10^{-8}	9.68×10^{-8}	6.80×10^{-5}	7.63×10^{-8}	3.73×10^{-8}	2.78×10^{-8}
residential	94,879.2274	51,048.8556	1,107,680	1,098,121	51,887.8991	238,679.153	54,071.2573	54,071.2573
parkisons	18.2702	13.7788	64.3326	64.3017	14.2117	55.4174	13.6362	13.6377

(b) MAE:

Gradient Boost	MCTS_ Rrelieff	GRNN_ isotropic	PSO	QBSO	MCTS _rave	TD_ learning	GNMCTS	Original
triazines	0.1122	–	0.0928	0.1013	0.0951	0.0123	0.0984	0.0906
puma32h	0.0065	0.0064	0.0066	0.0234	0.0065	0.0235	0.0065	0.0066
pol	29.2173	5.3454	34.6598	27.6021	5.9007	18.0749	5.7809	5.4873
bank32nh	0.0732	0.0564	0.0557	0.0828	0.0556	0.0906	0.0552	0.0554
ailerons	1.99×10^{-4}	1.22×10^{-4}	1.69×10^{-4}	2.40×10^{-4}	6.59×10^{-3}	2.11×10^{-4}	1.44×10^{-4}	1.21×10^{-4}
residetial	153.4199	109.2683	723.8313	718.452	98.3058	321.5966	117.245	104.9476
parkisons	3.3436	2.8993	6.7796	6.8061	3.0077	6.1205	2.9299	2.9301

(c) R2:

Gradient Boost	MCTS_ Rrelieff	GRNN_ isotropic	PSO	QBSO	MCTS_rave	TD_ learning	GNMCTS	Original
triazines	0.0692	–	0.3046	0.0481	0.2249	0.0399	0.2479	0.3012
puma32h	0.9261	0.9267	0.9229	−0.0187	0.9256	−0.0353	0.925	0.9227
pol	0.2358	0.9519	0.073	0.2913	0.9428	0.5853	0.9449	0.9514
bank32nh	0.2699	0.4962	0.5136	0.103	0.513	−0.0156	0.5111	0.519
ailerons	0.5309	0.8309	0.6997	0.4171	0.9237	0.5411	0.7755	0.833
residetial	0.9343	0.964	0.2285	0.2354	0.9631	0.8338	0.962	0.9574
parkisons	0.7234	0.7912	0.0259	0.0264	0.7846	0.1608	0.7934	0.7934

(d) EV:

Gradient Boost	MCTS_ Rrelieff	GRNN_ isotropic	PSO	QBSO	MCTS_rave	TD_ learning	GNMCTS	Original
triazines	0.075	–	0.3192	0.0831	0.2346	0.0507	0.2646	0.3189
puma32h	0.9262	0.9268	0.923	−0.0165	0.9257	−0.0334	0.925	0.9227
pol	0.236	0.952	0.0732	0.2914	0.9428	0.5854	0.9447	0.9514
bank32nh	0.2701	0.4962	0.5137	0.1043	0.514	−0.0148	0.5111	0.5191
ailerons	0.5312	0.831	0.7001	0.4175	0.9238	0.5415	0.7756	0.833
residential	0.9358	0.9648	0.2346	0.2419	0.9638	0.8362	0.9625	0.9581
parkinsons	0.7237	0.7914	0.0266	0.027	0.7848	0.1611	0.7936	0.7936

(e) Confidence bound

Gradient Boost	GNMCTS	Original
triazines	[0.6295,0.6691]	[0.6209,0.6968]
puma32h	[−0.0010,0.0028]	[−0.0010,0.0028]
pol	[27.5489,30.3289]	[27.5455,30.3560]
bank32nh	[0.0794,0.0875]	[0.0795,0.0876]
ailerons	$[-8.8193 \times 10^{-4}, -8.6153 \times 10^{-4}]$	$[-8.8513 \times 10^{-4}, -8.5827 \times 10^{-4}]$
residential	[1114.2649,1660.1809]	[1114.0922,1649.7011]
parkinsons	[20.9721,21.6521]	[20.9665,21.6493]

In Table 3c,d, GNMCTS obtained satisfactory results. Compared with the original dataset without feature selection, it slightly improved on three datasets and held the line on triazines, Pol, Bank32nh, Ailerons. R2 and EV of QBSO and TD learning methods on Puma32h were negative, and the TD learning method also obtained a negative value on Bank32h. These results indicated the models were worse than the baseline model. The baseline model took advantage of mean prediction values, so it was like a conserved guess about the prediction result. This could be due to that the two methods had chosen irrelevant features. GNMCTS, GRNN and MCTS rave methods especially outperform other methods on the Pol dataset. In Table 3e, 95% confidence intervals of the mean value of prediction on

seven datasets are presented. As shown in the table, the confidence interval slightly shrunk or remained the same after feature selection compared to the original full feature set. The interval between low and high confidence bound is within a reasonable value.

In order to demonstrate the ability of dimension reduction, the number of selected features in Table 3 was compared with the original dataset. The DR result of GNMCTS is shown in Figure 6. GNMCTS could effectively reduce the feature dimension on most datasets. The Parkinson updrs original dataset only contains 19 columns, so GNMCTS did not need too many iterations to find the optimal solution, but for comparison purpose, we set the iteration number to 1000 which enforce GNMCTS return a relative redundant solution.

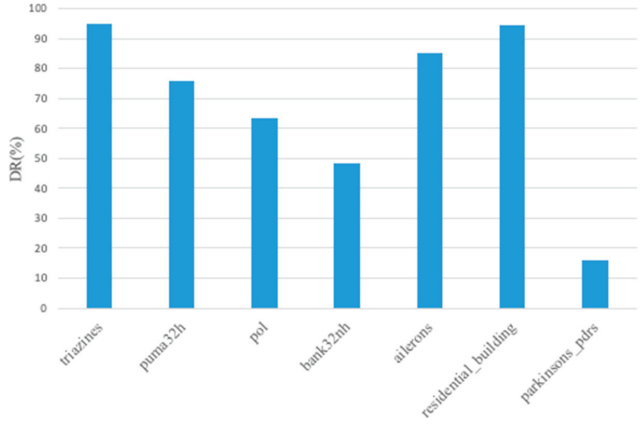


Figure 6. Graphical representation of dimension reduction (DR) achieved by GNMCTS on all datasets.

The computation times for each method were recorded, as shown in Figure 7. As GRNN failed to predict triazines, the results of this dataset were not shown. With the same iteration number, we can see QBSO was the most time-consuming method. The second most time-consuming method was MCTS_Rrelieff, followed by PSO. The cost of the TD learning method was closed to MCTS RAVE and GNMCTS but was less time-consuming than GRNN.

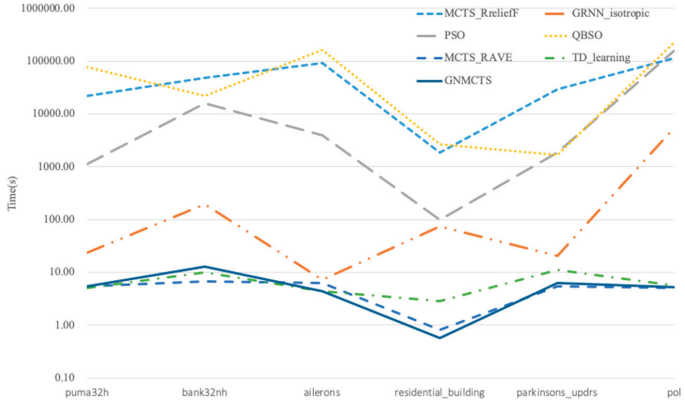


Figure 7. Calculation time comparison illustration of different methods.

We also performed the Friedman test on MAE in Table 3b. The Friedman test was used further to compare the generalization of learning methods on different datasets. The p-value was 1.8834×10^{-7}, which was close to 0 and far smaller than 0.05. This means the performances of methods apparently differed from one another.

6. Conclusions

The Monte Carlo Tree Search (MCTS) is a method for searching optimal decisions in a given deterministic environment. It generates an asymmetrical growing tree because of the searching strategy. It combines selectivity and randomness in the search process. The merit of this kind of method is strong learning power without any domain knowledge. This characteristic makes the reinforcement learning method a perfect inspiring player and teacher. It can show some unique ways of solving problems where other methods failed. The proposed method GNMCTS inherits the merits of MCTS and can obtain a better robust result by nesting. Through experimental analysis, GNMCTS obtained satisfactory results compared to other feature methods. It can effectively reduce the feature dimension with a reasonable computation budget. GNMCTS can fit feature selection for regression tasks for data with various dimensions. The Gamma test could indicate how many data points it takes to converge, called the M-test; this could accelerate MCTS greatly. Future work may focus on the revised UCT formulation combined with this M-test and develop an algorithm-based parallelization of NMCTS.

Author Contributions: Writing, Y.L.; Validation, Y.L.; Investigation, L.G.; Methodology, Y.L.; Supervision, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partly supported by the National Natural Science Foundation of China (60473125,61701213), science and technology planning projects of Karamay (2020CGZH0009), Scientific Research Foundation of Karamay Campus of China University of Petroleum (Beijing)(RCYJ2016B-03-001), the Natural Science Foundation of Fujian Province (Nos.2021J01473 and 2021J01475), and the Research Fund for Educational Department of Fujian Province (No. JAT190392).

Data Availability Statement: Data can be found at https://archive.ics.uci.edu/ml/datasets.php (accessed on 9 June 2021) or https://www.openml.org/home (accessed on 9 June 2021) Codes for methods mentioned in Section 5 can be found at https://github.com/ring0o0o/nmcts.git (accessed on 9 June 2021).

Acknowledgments: We would like to thank Zheng Yifeng for his valuable suggestions and help to improve this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fu, M.C. AlphaGo and Monte Carlo tree search: The simulation optimization perspective. In Proceedings of the 2016 Winter Simulation Conference (WSC), Arlington, VA, USA, 11–14 December 2016; pp. 659–670.
2. Gelly, S.; Silver, D. Combining online and offline knowledge in UCT. In Proceedings of the International Conference of Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 273–280.
3. Schadd, M.P.; Winands, M.H.; Tak, M.J.; Uiterwijk, J.W. Single-player Monte-Carlo tree search for SameGame. *Knowl. Based Syst.* **2012**, *34*, 3–11. [CrossRef]
4. Chaudhry, M.U.; Lee, J.-H. MOTiFS: Monte Carlo tree search based feature selection. *Entropy* **2018**, *20*, 385. [CrossRef]
5. Gaudel, R.; Sebag, M. Feature selection as a one-player game. In Proceedings of the International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 359–366.
6. Fard, S.M.H.; Hamzeh, A.; Hashemi, S. A game theoretic framework for feature selection. In Proceedings of the 9th International Conference on Fuzzy Systems and Knowledge Discovery, Chongqing, China, 29–31 May 2012; pp. 845–850.
7. Jones, A.J. New tools in non-linear modelling and prediction. *Comput. Manag. Sci.* **2004**, *1*, 109–149. [CrossRef]
8. Kemp, S.E.; Wilson, I.D.; Ware, J.A. A tutorial on the gamma test. *Int. J. Simul.* **2004**, *6*, 67–75.
9. Evans, A.D. A proof of the Gamma test. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2002**, *458*, 2759–2799. [CrossRef]
10. Evans, A.D.; Jones, A.J.; Schmidt, W.M. Asymptotic moments of near–neighbour distance distributions. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2002**, *458*, 2839–2849. [CrossRef]
11. Baier, H.; Winands, M. Nested Monte-Carlo tree search for online planning in large MDPs. In Proceedings of the 20th European Conference on Artificial Intelligence, Montpellier, France, 27–31 August 2012; Volume 242, pp. 109–114.

12. Huang, X.; Zhang, L.; Wang, B. Feature clustering based support vector machine recursive feature elimination for gene selection. *Appl. Intell.* **2018**, *48*, 594–607. [CrossRef]
13. Masood, M.K.; Soh, Y.C.; Jiang, C. Occupancy estimation from environmental parameters using wrapper and hybrid feature selection. *Appl. Soft Comput.* **2017**, *60*, 482–494. [CrossRef]
14. Bommert, A.; Sun, X.; Bischl, B.; Rahnenführer, J.; Lang, M. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data Anal.* **2020**, *143*, 106839. [CrossRef]
15. Venkatesh, B.; Anuradha, J. A Review of feature selection and its methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. [CrossRef]
16. Nguyen, B.H.; Xue, B.; Zhang, M. A survey on swarm intelligence approaches to feature selection in data mining. *Swarm Evol. Comput.* **2020**, *54*, 100663. [CrossRef]
17. Sharma, M.; Kaur, P. A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. *Arch. Comput. Methods Eng.* **2020**, *28*, 1103–1127. [CrossRef]
18. Fan, W.; Liu, K.; Liu, H.; Wang, P.; Ge, Y.; Fu, Y. AutoFS: Automated Feature selection via diversity-aware interactive reinforcement learning. In Proceedings of the IEEE International Conference on Data Mining (ICDM), Istanbul, Turkey, 30–31 July 2020; pp. 1008–1013.
19. Rimmel, A. Improvements and Evaluation of the Monte Carlo Tree Search Algorithm. Ph.D. Thesis, Université Paris Sud, Le Kremlin-Bicêtre, France, October 2009.
20. Marchenko, O.I.; Marchenko, O.O. Monte-Carlo tree search with tree shape control. In Proceedings of the IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kyiv, Ukraine, 29 May–2 June 2017; pp. 812–817.
21. Browne, C.B.; Powley, E.; Whitehouse, D. A survey of Monte Carlo tree search methods. *Trans. Comput. Intell. AI Games* **2012**, *4*, 1–43. [CrossRef]
22. Hogan, S.; Jarvis, P.; Wilson, I. Using the gamma test in the analysis of classification models for time-series events in urodynamics investigation. In Proceedings of the International Conference on Innovative Techniques and Applications of Artificial Intelligence, Cambridge, UK, 13–15 December 2011; pp. 299–310.
23. Narges, S.; Mohammad, A.F.; Ahmad, S.; Mohammad, H.M.A. Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. *J. Nat. Gas Sci. Eng.* **2013**, *14*, 238–249.
24. Iturrarán-Viveros, U. Smooth regression to estimate effective porosity using seismic attributes. *J. Appl. Geophys.* **2012**, *76*, 1–12. [CrossRef]
25. Noori, R.; Karbassi, A.; Sabahi, M.S. Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction. *J. Environ. Manag.* **2010**, *91*, 767–771. [CrossRef] [PubMed]
26. Jaafar, W.; Han, D. Variable Selection using the gamma test forward and backward selections. *J. Hydrol. Eng.* **2012**, *17*, 182–190. [CrossRef]
27. Akiyama, H.; Komiya, K.; Kotani, Y. Nested Monte-Carlo search with AMAF heuristic. In Proceedings of the International Conference on Technologies and Applications of Artificial Intelligence, Hsinchu, Taiwan, 18–20 November 2010; pp. 172–176.
28. Sironi, C.F.; Liu, J.; Winands, M. Self-adaptive monte-carlo tree search in general game playing. *Trans. Games* **2018**, *1*, 132–144. [CrossRef]
29. Rimmel, A.; Teytaud, F.; Cazenave, T. Optimization of the nested Monte-Carlo algorithm on the traveling salesman problem with time windows. In Proceedings of the International Conference on Applications of Evolutionary Computation, Torino, Italy, 27–29 April 2011; pp. 501–510.
30. Mehat, J.; Cazenave, T. Combining UCT and nested Monte-Carlo search for single-player general game playing. *Trans. Comput. Intell. AI Games* **2011**, *2*, 271–277. [CrossRef]
31. Golay, J.; Leuenberger, M.; Kanevski, M. Feature selection for regression problems based on the morisita estimator of intrinsic dimension. *Pattern Recognit.* **2017**, *70*, 126–138. [CrossRef]
32. Golay, J.; Kanevski, M. A new estimator of intrinsic dimension based on the multipoint morisita index. *Pattern Recognit.* **2015**, *48*, 4070–4081. [CrossRef]
33. Carmona, L.; Pedro, S. Filter-type variable selection based on information measures for regression tasks. *Entropy* **2012**, *14*, 323–343. [CrossRef]
34. Tsanas, A.; Little, M.; McSharry, P.; Ramig, L. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *Trans. Biomed. Eng.* **2009**, *57*, 884–893. [CrossRef] [PubMed]
35. Zhang, Y.; Wang, S.; Phillips, P. Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowl. Based Syst.* **2014**, *64*, 22–31. [CrossRef]
36. Sadeg, S.; Hamdad, L.; Remache, A.R. QBSO-FS: A Reinforcement learning based bee swarm optimization metaheuristic for feature selection. In Proceedings of the International Work-Conference on Artificial Neural Networks Proceedings Part 2, Gran Canaria, Spain, 12–14 June 2019; pp. 785–796.
37. Sadeg, S.; Hamdad, L.; Benatchba, K. BSO-FS: Bee Swarm optimization for feature selection in classification. In Proceedings of the International Work-Conference on Artificial Neural Networks Proceedings Part 1, Palma de Mallorca, Spain, 10–12 June 2015; pp. 387–399.
38. Zheng, J.; Zhu, H.; Chang, F. An improved relief feature selection algorithm based on Monte-Carlo tree search. *Syst. Sci. Control. Eng.* **2019**, *7*, 304–310. [CrossRef]

39. Fard, S.; Hamzeh, A.; Hashemi, S. Using reinforcement learning to find an optimal set of features. *Comput. Math. Appl.* **2013**, *66*, 1892–1904. [CrossRef]
40. Sali, R.; Sodiq, A.; Akakpo, A. Feature selection using reinforcement learning. *arXiv* **2021**, arXiv:2101.09460.
41. Amato, F.; Guignard, F.; Jacquet, P. On Feature selection using anisotropic general regression neural network. *arXiv* **2020**, arXiv:2010.05744.

entropy

MDPI

Article

Adaptive Event-Triggered Synchronization of Uncertain Fractional Order Neural Networks with Double Deception Attacks and Time-Varying Delay

Zhuan Shen, Fan Yang, Jing Chen, Jingxiang Zhang, Aihua Hu and Manfeng Hu *

School of Science, Jiangnan University, Wuxi 214122, China; 6191204005@stu.jiangnan.edu.cn (Z.S.); 6191204018@stu.jiangnan.edu.cn (F.Y.); 8201703038@jiangnan.edu.cn (J.C.); zhangjingxiang@jiangnan.edu.cn (J.Z.); aihuahu@jiangnan.edu.cn (A.H.)
* Correspondence: humanfeng@jiangnan.edu.cn; Tel.: +86-510-8591-0233

Abstract: This paper investigates the problem of adaptive event-triggered synchronization for uncertain FNNs subject to double deception attacks and time-varying delay. During network transmission, a practical deception attack phenomenon in FNNs should be considered; that is, we investigated the situation in which the attack occurs via both communication channels, from S-C and from C-A simultaneously, rather than considering only one, as in many papers; and the double attacks are described by high-level Markov processes rather than simple random variables. To further reduce network load, an advanced AETS with an adaptive threshold coefficient was first used in FNNs to deal with deception attacks. Moreover, given the engineering background, uncertain parameters and time-varying delay were also considered, and a feedback control scheme was adopted. Based on the above, a unique closed-loop synchronization error system was constructed. Sufficient conditions that guarantee the stability of the closed-loop system are ensured by the Lyapunov-Krasovskii functional method. Finally, a numerical example is presented to verify the effectiveness of the proposed method.

Keywords: uncertain fractional order neural network; adaptive event-triggered scheme; double deception attacks; time-varying delay

Citation: Shen, Z.; Yang, F.; Chen, J.; Zhang, J.; Hu, A.; Hu, M. Adaptive Event-Triggered Synchronization of Uncertain Fractional Order Neural Networks with Double Deception Attacks and Time-Varying Delay. *Entropy* **2021**, *23*, 1291. https://doi.org/10.3390/e23101291

Academic Editor: Luis Hernández-Callejo

Received: 7 September 2021
Accepted: 25 September 2021
Published: 30 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Neural networks, which bridge the micro-world of communications with the physical world for processing information as mathematical models, widely exist in a broad range of areas, such as intelligent control, secure communication, and pattern recognition [1–4]. Due to the complexity of the dynamic characteristics of some physical systems, a traditional integer-order neural network model cannot accurately represent their dynamic behaviors. Fractional order calculus is not only a generalized form of the traditional integer-order calculus; it also has some irreplaceable properties of integral order calculus, such as the special feature of time memory [4–7]. Based on these features, the fractional order differential equation has been used to model neural networks [8–12]. Synchronization, among several phenomena arising from the complex nonlinear dynamics of neural networks, has gained lots of attention and has been applied in many integer-order neural networks [13–17]. However, there are few studies about the synchronization problem of FNNs, which was the first motivation of this paper.

The event-triggered scheme (ETS) depends on a predefined event-triggered condition to determine whether the sampled data should be transmitted to the next control unit rather than a fixed period; therefore, replacing the time-triggered scheme (TTS) to save network communication resources and guarantee the system's performance simultaneously was suggested in [16,18–23]. Although ETS was adopted in the latest three studies of different fractional order, real-valued systems [21–23], there was still a common disadvantage: the threshold coefficients of traditional ETS are all constants and cannot be timely adjusted

to fit a system's evolution. However, the adaptive event-triggered scheme (AETS), as a combination of adaptive control and traditional ETS, can overcome the conservativeness to make good use of communication resources dynamically. Therefore, designing an AETS with an adaptive threshold coefficient for FNNs to further improve the utilization of communication resources was the second motivation of the current work.

On the other hand, a security problem, due to advanced modern communication technology, has recently emerged as a hot topic in the engineering applications [24,25], especially in autonomous vehicle platooning [26,27]. Since the control components such as sensors, controllers, and actuators are connected by the shared communication networks to achieve remote control, compromise by malicious adversaries is extremely risky [22,28,29]. As a typical representative of malicious attacks, a deception attack can replace the original data with false data to destroy the system [22,28–31]. To the best of the authors' knowledge, the synchronization problem of FNNs regarding deception attacks has been investigated in the literature [22], although the deception attacks were only allowed to occur in the controller to actuator (C-A) channel, governed by a Bernoulli variable. However, in communication networks, attacks may occur in the sensor to controller (S-C) channel and C-A channel simultaneously. Moreover, it is well known that a Bernoulli process is a special kind of the Markov process. Therefore, inspired by the aforementioned discussion, investigating double deception attacks governed by Markov processes in the synchronization of FNNs under AETS was the third motivation. Given the actual environmental conditions, neural networks inevitably suffer from noise and limitations of equipment, so uncertainties in parameters and time-varying delay have also been taken into account. The main contributions are outlined below.

(1) The synchronization problem of FNNs under network attacks is firstly proposed with an AETS to further save network bandwidth resources. The AETS has an adaptive law for adjusting its threshold coefficient such that the controller can timely access system information to stabilize the error system.
(2) A generalized deception attack for FNNs is investigated; that is, the deception attack may occur in S-C and C-A channels simultaneously. Moreover, the attack behaviors are governed by independent Markov processes that are more extensive than the Bernoulli processes in other studies.
(3) Parameters' uncertainties and time-varying delay are also investigated in light of the synchronization problem of FNNs and a double deception attack in the AETS. That is more practicable to some extent.

The remainder of this paper is organized as follows. In Section 2, some preliminaries are introduced and the model is formulated. The main results, including theorems, are shown in Section 3. In Section 4, a simulation which verified the main results is presented. Finally, the discussion and conclusions are presented in Section 5.

Notation: In this paper, R^n and $\|\cdot\|$ denote the n-dimensional Euclidean vector space and the Euclidean norm for vectors, respectively. $R^{n\times n}$ is the set of all $n \times n$ real matrices. T denotes the transposition of the vectors or matrices. I represents the identity matrix with appropriate dimensions, and $He[A] = A + A^T$. The symbol N represents the sets of all natural numbers and $N_0 = N \cup \{0\}$. The signal "$*$" denotes the symmetric block of matrix. $col(\dots)$ and $diag(\dots)$ represent a column vector and a diagonal matrix, respectively.

Remark 1. *Network attacks may occur in both S-C and C-A channels during network transmission, as shown in Figure 1. We only found a few studies investigating relevant network attacks, and they only used single-channel attacks: the C-A channel [22]; the S-C channel [32–34]. In addition, in prior studies the behaviors of network attacks were governed by Bernoulli variables, usually. To the authors' knowledge, there is no literature simultaneously considering network attacks in S-C and C-A channels in FNNs. Moreover, in this paper, the double network attacks governed by two independent Markov processes are more general than Bernoulli processes.*

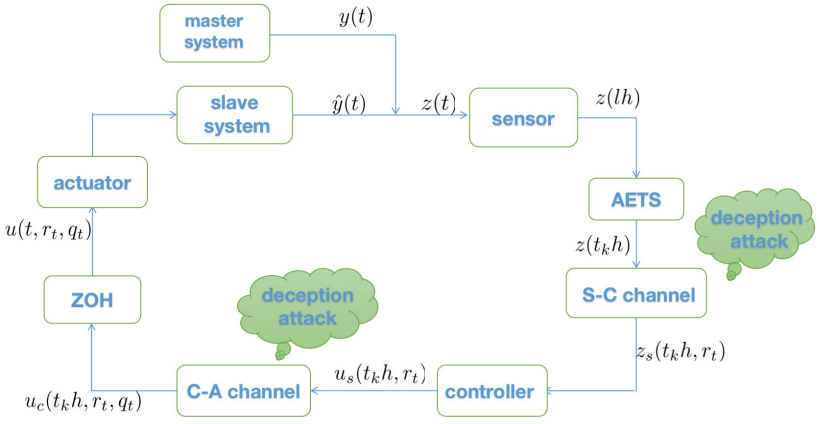


Figure 1. The framework of the closed-loop synchronization error system.

2. Preliminaries and Model Formulation

In this section, the basic definitions and relations about fractional calculus are introduced; then a closed-loop synchronization error system is constructed.

2.1. Fractional Order Calculations

Definition 1. *The fractional integral of order r for an integrable function* $f(x) : [t_0, +\infty] \to R$ *is defined as [19]:*

$$ {}_{t_0}I_t^r f(t) = \frac{1}{\Gamma(r)} \int_{t_0}^{t} \frac{f(\beta)}{(t-\beta)^{1-r}} \, \mathrm{d}\beta, $$

where $0 < r < 1$*, and* $\Gamma(\cdot)$ *is the Gamma function.*

Definition 2. *The Caputo fractional derivative of order* $r > 0$ *for a function* $f(t) \in C^n([t_0, +\infty), R)$ *is defined as [22]:*

$$ {}_{t_0}D_t^r f(t) = \frac{1}{\Gamma(n-r)} \int_{t_0}^{t} \frac{f^{(n)}(\beta)}{(t-\beta)^{r-n+1}} \, \mathrm{d}\beta, $$

where $t \geq t_0$ *and n is an integer such that* $0 < n-1 < r < n$*. Moreover, when* $0 < r < 1$*,*

$$ {}_{t_0}D_t^r f(t) = \frac{1}{\Gamma(1-r)} \int_{t_0}^{t} \frac{f^{'}(\beta)}{(t-\beta)^{r}} \, \mathrm{d}\beta. $$

From the ldefinitions 1 and 2, it is clear that the Caputo fractional derivative satisfies the following properties:

(1) ${}_{t_0}D_t^r {}_{t_0}I_t^s f(t) = {}_{t_0}D_t^r {}_{t_0}D_t^{-s} f(t) = {}_{t_0}D_t^{r-s} f(t)$, where $r \geq s \geq 0$.
(2) ${}_{t_0}D_t^r C = 0$, where C is a constant.
(3) ${}_{t_0}D_t^r(v_1 f(t) + v_2 g(t)) = v_1 {}_{t_0}D_t^r f(t) + v_2 {}_{t_0}D_t^r g(t)$, where v_1 and v_2 are any constants.

Lemma 1 ([22])**.** *For a differentiable function vector* $x(t) \in R^n$*, an equality with the following form is true:*

$$ {}_{t_0}D_t^r(x^T(t)Px(t)) \leq 2x^T(t)P {}_{t_0}D_t^r x(t), $$

where r and $P \in R^{n \times n}$ *satisfy* $0 < r < 1$ *and* $P > 0$*, respectively.*

Lemma 2 ([35]). *For a given positive definite matrix $\mathcal{R} \in R^{n\times n}$, given scalars a, b satisfying $a < b$, the following inequality holds for any continuously differentiable function $e(x)$ in $[a, b] \to R^n$:*

$$(b-a)\int_a^b e^T(s)\mathcal{R}e(s)\,\mathrm{d}s \geq \left(\int_a^b e(s)\,\mathrm{d}s\right)^T \mathcal{R}\left(\int_a^b e(s)\,\mathrm{d}s\right).$$

Lemma 3 ([36]). *For $\eta(t) \in [0, \eta]$ and any matrices $R, S \in R^{n\times n}$ satisfying $\begin{bmatrix} R & S \\ * & R \end{bmatrix} \geq 0$, the following inequality holds:*

$$-\eta\int_{t-\eta}^{t} \dot{e}^T(s)R\dot{e}(s)\,\mathrm{d}s \leq \xi^T(t)\Theta\xi(t),$$

where $\xi(t) = col\{e(t), e(t-\eta(t)), e(t-\eta)\}$ and

$$\Theta = \begin{bmatrix} -R & R-S & S \\ * & -2R+He[S] & R-S \\ * & * & -R \end{bmatrix}.$$

Lemma 4 ([32]). *For given matrix $S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$, where $S_{12} = S_{21}^T$, the following conditions are equivalent.*

$$(1) S < 0;$$
$$(2) S_{22} < 0, S_{11} - S_{21}S_{22}^{-1}S_{12} < 0.$$

2.2. Model Formulation

Consider the following uncertain FNN model as the master system:

$$\begin{aligned} {}_{t_0}D_t^r x(t) &= -(A+\Delta A(t))x(t) + (B+\Delta B(t))\hat{f}(x(t)) \\ &\quad + (D+\Delta D(t))\hat{f}(x(t-\eta(t))) + I(t), \\ y(t) &= Cx(t), \\ x(t_0) &= \phi_1(t_0), t_0 \in [-\eta, 0], \end{aligned} \tag{1}$$

where $0 < r < 1$ denotes the order of fractional order derivative. $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T \in R^n$ is the state vector of the neuron. $y(t)$ is the measurable output vector. $\eta(t)$ satisfies $0 \leq \eta(t) \leq \eta$, and $\dot{\eta}(t) \leq \bar{\eta}$ denotes the time-varying coupling delay. $\hat{f}(x(t)) = (\hat{f}_1(x_1(t)), \hat{f}_2(x_2(t)), \ldots, \hat{f}_n(x_n(t)))$ and $\hat{f}(x(t-\eta(t))) = (\hat{f}_1(x_1(t-\eta(t))), \hat{f}_2(x(t-\eta(t))), \ldots, \hat{f}_n(x(t-\eta(t)))) \in R^n$ are the activation functions. $I(t)$ is an external input vector. $A = diag(a_1, a_2, \ldots, a_n) \in R^{n\times n}$, are the self-feedback connection weight matrices. $B = (b_{ij})_{n\times n} \in R^{n\times n}, D = (d_{ij})_{n\times n} \in R^{n\times n}$ are the connection weight matrices. Furthermore, $\Delta A(t), \Delta B(t), \Delta D(t)$ are the matrices with time-varying parameters, which are norm bounded and satisfy

$$[\Delta A(t), \Delta B(t), \Delta D(t)] = GS(t)[E_a, E_b, E_d],$$

where G, E_a, E_b, E_d are known constant matrices, $S(t)$ is an unknown time-varying matrix function satisfying $S^T(t)S(t) \leq I$. Assume that master system (1) have a unique solution

with initial value $\phi_1(t_0)$ and that it is continuously differential on $t_0 \in [-\eta, 0]$ [37].

Next, consider the corresponding slave system as follows:

$$\begin{aligned} {}_{t_0}D_t^r\hat{x}(t) &= -(A+\Delta A(t))\hat{x}(t) + (B+\Delta B(t))\hat{f}(\hat{x}(t)) \\ &\quad + (D+\Delta D(t))\hat{f}(\hat{x}(t-\eta(t))) + I(t) + u(t), \\ \hat{y}(t) &= C\hat{x}(t), \\ \hat{x}(t_0) &= \phi_2(t_0), t_0 \in [-\eta, 0], \end{aligned} \tag{2}$$

where $\hat{x}(t) = (\hat{x}_1(t), y_2(t), \ldots, \hat{x}_n(t))^T$ is the state vector. Similarly, assume slave system (2) also has a unique solution with initial value $\phi_2(t_0)$, which is continuously differential on $t_0 \in [-\eta, 0]$, and $u(t)$ is the control input, and the others are same as the master system.

In order to realize the synchronization between systems (1) and (2), define the synchronization error $z(t) = C(\hat{x}(t) - x(t))$, and the parameter uncertainty of each part is treated as a whole. The following error system can be obtained:

$$\begin{aligned} {}_{t_0}D_t^r e(t) &= -Ae(t) + Bf(e(t)) + Df(e(t-\eta(t))) + Gm(t) + u(t), \\ m(t) &= S(t)(-E_a e(t) + E_b f(e(t)) + E_d f(e(t-\eta(t)))), \\ z(t) &= Ce(t), \\ e(t_0) &= \phi(t_0), t_0 \in [-\eta, 0], \end{aligned} \tag{3}$$

where $f(e(t)) = \hat{f}(\hat{x}(t)) - \hat{f}(x(t)), f(e(t-\eta(t))) = \hat{f}(\hat{x}(t-\eta(t))) - \hat{f}(x(t-\eta(t)))$. The initial value of error system (3) is $\phi(t_0) = \phi_2(t_0) - \phi_1(t_0), t_0 \in [-\eta, 0]$. It is well known that system (3) has a unique solution [38].

Remark 2. *The model considered in this paper can be regarded as a generalization of [22]. Such an attack has only been considered in the C-A channel and governed by a Bernoulli process in FNNs [22], in which the event-triggered threshold coefficient is a constant and cannot fit a system's evolution dynamically. The FNNs studied in this paper not only adopt AETS to further improve the utilization of communication resources, but parameters' uncertainties and double deception attacks are also investigated.*

The following assumption will be used later on.

Assumption 1. *The neuron activation function $f(e(t))$ is continuous and bounded, and satisfies the following conditions:*

$$0 \le \frac{f_i(e_1(t)) - f_i(e_2(t))}{e_1(t) - e_2(t)} \le \phi_i, \tag{4}$$

for $i = 1, 2, \ldots, n$, where ϕ_i are known positive constants.

Let the two adversary network attacks during the communication be characterized by two independent right-continuous Markov processes r_t, q_t on the probability space taking values in the finite state space $M = \{1, 2, \ldots, s\}$ with generator $\pi = (\pi_{ij})_{s\times s}, \rho = (\rho_{ij})_{s\times s}$ given by

$$Pr\{r_{t+k} = j | r_t = i\} = \begin{cases} \pi_{ij}k + o(k) & i \ne j, \\ 1 + \pi_{ii}k + o(k) & i = j. \end{cases}$$

$$Pr\{q_{t+k} = n | q_t = m\} = \begin{cases} \rho_{mn}k + o(k) & m \ne n, \\ 1 + \rho_{mn}k + o(k) & m = n. \end{cases}$$

where $k > 0, \lim_{k\to 0}\frac{o(k)}{k} = 0, \pi_{ij} \ge 0, i \ne j, \rho_{mn} \ge 0, m \ne n$, and for every $i, m \in M, \pi_{ii} = -\sum_{j\ne i}\pi_{ij}, \rho_{mm} = -\sum_{n\ne m}\rho_{mn}$.

To save on network bandwidth as much as possible, an AETS was adopted in this study. The sensor with sampling period h was time-driven, and the output error $z(t)$ was measured by the sensor at the sampling instant $lh, l \in N_0$. Let $t_k h$ denote the triggered

instant; then the next triggered instant is denoted by $t_{k+1}h$. $t_k + ih, i \in N$ denotes the current sampling time. Whether or not the sampled data $z(t_k + ih)$ should be transmitted is determined by the adaptive event-triggered condition:

$$\tilde{z}_k^T(t)\Omega\tilde{z}_k(t) - d(t)z^T(t_k + ih)\Omega z(t_k + ih) \leq 0, \tag{5}$$

where $\tilde{z}_k(t) = z(t_k h) - z(t_k + ih), z(t_k h)$ denotes the latest transmitted data, $\Omega > 0$ is a weighting matrix to be designed, and the adaptive threshold coefficient $d(t)$ satisfies the following adaptive law:

$$\dot{d}(t) = (\frac{1}{d(t)^2} - \frac{\bar{w}}{d(t)})\tilde{z}_k^T(t)\Omega\tilde{z}_k(t), \tag{6}$$

where $\bar{w} \geq 1$ can adjust the monotonicity of $d(t)$ [32], and the next triggered instant can be denoted as follows:

$$t_{k+1}h = t_k h + min\{ih|\tilde{z}_k^T\Omega\tilde{z}_k > d(t)z^T(t_k + ih)\Omega z(t_k + ih), i \in N\}.$$

Based on the reality of the network communication, the delay s_k is considered at the instant $t_k h$. Assume that $0 \leq s_k \leq \bar{s}$, where $\bar{s} = max\{s_k\}$. The sampling date $z(t_k h)$ will be transmitted at the instant $t_k + s_k$. Then the time interval $[t_k h + s_k, t_{k+1}h + s_{k+1})$ can be divided $I_0 = [t_k h + s_k, t_k h + h + \bar{s}), I_i = [t_k h + ih + \bar{s}, t_k h + ih + h + \bar{s}), i = 1, 2, \ldots, \delta - 1$, and $\delta = t_{k+1} - t_k - 1$, $I_\delta = [t_k h + \delta h + \bar{s}, t_{k+1} + d_{k+1})$. Then $\tilde{z}_k(t) = z(t_k h) - z(t_k h + ih)$ is equivalent to:

$$\tilde{z}_k(t) = \begin{cases} z(t_k h) - z(t_k h), & t \in I_0, \\ z(t_k h) - z(t_k h + ih), & t \in I_i, \\ z(t_k h) - z(t_k h + \delta h), & t \in I_\delta \end{cases} \tag{7}$$

which can be written as

$$\tilde{z}_k(t) = z(t_k h) - z(t - \tau(t)), t \in [t_k h + s_k, t_{k+1}h + s_{k+1}) \tag{8}$$

in which

$$\tau(t) = \begin{cases} t - t_k h, & t \in I_0, \\ t - t_k h - ih, & t \in I_i, \\ t - t_k h - \delta h, & t \in I_\delta. \end{cases} \tag{9}$$

According to Equation (9), it is easy to get

$$0 \leq \tau(t) \leq h + \bar{s}, t \in [t_k h + s_k, t_{k+1}h + s_{k+1}).$$

Remark 3. *From the adaptive event-triggered condition (5), it is easy to know the minimum event-triggered interval is a constant, which means that there is no Zeno behavior.*

As shown in Figure 1, deception attacks may occur on the S-C communication channel, and the integrity of normal transmission data will be damaged by malicious attacks. To depict the stochastic occurrence modeling of deception attacks, Markov processes are adopted in this paper. Then the control input in time interval $[t_k h + s_k, t_{k+1}h + s_{k+1}), k = 1, 2, \ldots,$ can be denoted as

$$\begin{aligned} z_s(t_k h, r_t) &= b^s(r_t)z(t_k h) + \bar{b}^s(r_t)g_s(z(t_k h)), \\ &= b^s(r_t)\Big(z(t - \tau(t)) + \tilde{z}_k(t)\Big) + \bar{b}^s(r_t)g_s(z(t_k h)). \end{aligned} \tag{10}$$

where $b^s(1) = 1, b^s(2) = 0, \bar{b}^s(r_t) = 1 - b^s(r_t)$, and $g_s : R^n \to R^n$ is the energy bounded deception signal in the S-C communication channel satisfying

$$\|g_s(x(t))\| \leq \|G_s x(t)\|. \tag{11}$$

where $G_s \in R^{n\times n}$ is a known constant matrix satisfying $G_s > 0$. If $r_t = 1$, the data will be transmitted normally without any attack. Conversely, $r_t = 2$ means that malicious attack signals occur in the S-C channel.

The main purpose of this study was to synchronize uncertain FNNs under AETS, subject to double deception attacks and time-varying delay. Construct the state feedback controller:

$$\begin{aligned} u(t) &= u_s(t_k h, r_t), \\ &= Kz_s(t_k h, r_t), t \in [t_k h + s_k, t_{k+1} h + s_{k+1}), \end{aligned} \tag{12}$$

where the feedback gain matrix K needs to be determined.

In a similar routine to that of the S-C communication channel, when the released data $u_s(t_k h, r_t)$ are transmitted through the C-A communication channel, the channel may be attacked again. Therefore, the control output signal can be denoted as

$$\begin{aligned} u(t) &= u_c(t_k h, r_t, q_t), \\ &= b^c(q_t)u_s(t_k h, r_t) + \bar{b}^c(q_t)g_c(u_s(t_k h, r_t)), \\ &= b^c(q_t)b^s(r_t)KCe(t-\tau(t)) + b^c(q_t)b^s(r_t)K\tilde{z}(t) + b^c(q_t)\bar{b}^s(r_t)Kg_s(\bar{z}(t)) \\ &\quad + \bar{b}^c(q_t)g_c(u_s(t_k h, r_t)), \quad t \in [t_k h + s_k, t_{k+1} h + s_{k+1}), \end{aligned} \tag{13}$$

where $\bar{z}(t) = \tilde{z}(t) + z(t - \tau(t))$, $b^c(1) = 1, b^c(2) = 0, \bar{b}^c(q_t) = 1 - b^c(q_t)$, and $g_c : R^n \to R^n$ is the energy bounded deception signal in the C-A communication channel satisfying

$$\|g_c(x(t))\| \leq \|G_c x(t)\|. \tag{14}$$

where $G_c \in R^{n\times n}$ is a known constant matrix satisfying $G_c > 0$. For simplicity, for every $i, m \in M, r_t = i, q_t = m$, $b^s(r_t), b^c(q_t)$ are denoted in this paper by b_i^s and b_m^c, respectively. Similarly, for a matrix $P_1(r_t, q_t)$, it is denoted by P_1^{im}. In addition, for a matrix P_1^{im}, there is the following definition:

$$\bar{P}_1^{im} = \sum_{j\in M} \pi_{ij} P_1^{jm} + \sum_{n\in M} \rho_{mn} P_1^{in}. \tag{15}$$

Then, it is easy to obtain the error system

$$\begin{aligned} {}_{t_0}D_t^r e(t) &= -Ae(t) + Bf(e(t)) + Df(e(t-\eta(t))) + Gm(t) + b_m^c b_i^s K\tilde{z}(t) \\ &\quad + b_m^c b_i^s KCe(t-\tau(t)) + b_m^c \bar{b}_i^s Kg_s(\bar{z}(t)) + \bar{b}_m^c g_c(u_s(t_k h, r_t)), \\ m(t) &= S(t)(-E_a e(t) + E_b f(e(t)) + E_d f(e(t-\eta(t)))), \\ z(t) &= Ce(t), \\ e(t_0) &= \phi(t_0), t_0 \in [-max\{\eta, h\}, 0]. \end{aligned} \tag{16}$$

The following two definitions will be used in the proof of Theorem 1.

Definition 3 ([39]). *Let $V(t, e(t), r_t = i, q_t = m)$ be the positive Lyapunov–Krasovskii functional and $\mathcal{L}(\cdot)$ be a weak infinitesimal operator. Then*

$$\mathcal{E}\Big\{\int_0^t \mathcal{L}V(s, e(s), i, m)\,\mathrm{d}s\Big\} = \mathcal{E}V(t, e(t), i, m) - \mathcal{E}V(0, \phi(t_0), r_0, q_0),$$

where $\mathcal{E}$ denotes the expectation.

Definition 4 ([40,41]). *The synchronization error system (16) is said to be globally, stochastically, asymptotically stable in the mean square sense, if for any initial conditions $\phi(t_0)$ defined on $[-max\{\eta, h\}, 0]$ and $r_0, q_0 \in M$ the following condition is satisfied:*

$$\lim_{t\to\infty} \mathcal{E}\Big\{\int_0^t e^T(s)e(s)\,\mathrm{d}s \mid \phi(t_0), r_0, q_0\Big\} < \infty.$$

So far, a closed-loop synchronization error system (16) has been constructed. In the following, in order to realize the synchronization between systems (1) and (2), the stability of error system (16) will be proven.

3. Results

Two theorems are developed in this section. Firstly, the synchronization criterion for systems (1) and (2) is presented in Theorem 1. Then, on the basis of Theorem 1, the criterion for feedback controller design is developed by Theorem 2.

Theorem 1. *Suppose Assumption 1 holds. The FNNs (1) and (2) are globally, stochastically, asymptotically synchronized under the feedback control scheme (12) in the mean square sense, for the given scalars r and control gain matrix K, if there exist positive definite matrices $P, \Omega, P_1^{im}, P_3^{im}, N_1$, $N_3, R_1^{im}, R_2^{im}, M_1^{im}, M_2^{im}, L_1, L_2, J_1, J_2, Q_1, Q_2$; positive definite diagonal matrices Δ_1, Δ_2; and matrices $P_2^{im}, N_2, S^{im}, T^{im}$; and positive scalars $\varepsilon, \lambda_1, \lambda_2$,such that the following LMIs for every i, m hold:*

$$\begin{bmatrix} \Pi_{1,1} & \Pi_{1,2} & \Pi_{1,3} & \Pi_{1,4} & \Pi_{1,5} & \Pi_{1,6} & \Pi_{1,7} \\ * & \Pi_{2,2} & \Pi_{2,3} & 0 & \Pi_{2,5} & 0 & 0 \\ * & * & \Pi_{3,3} & 0 & 0 & 0 & \Pi_{3,7} \\ * & * & * & \Pi_{4,4} & 0 & \Pi_{4,6} & \Pi_{4,7} \\ * & * & * & * & \Pi_{5,5} & 0 & \Pi_{5,7} \\ * & * & * & * & * & \Pi_{6,6} & 0 \\ * & * & * & * & * & * & \Pi_{7,7} \end{bmatrix} < 0, \tag{17}$$

$$\bar{R}_1^{im} < L_1, \bar{R}_2^{im} < L_2, \bar{M}_1^{im} < J_1, \bar{M}_2^{im} < J_1, \tag{18}$$

$$\begin{bmatrix} R_2^{im} & S^{im} \\ * & R_2^{im} \end{bmatrix} \geq 0, \begin{bmatrix} M_2^{im} & T^{im} \\ * & M_2^{im} \end{bmatrix} \geq 0, \tag{19}$$

$$\begin{bmatrix} P_1^{im} & P_2^{im} \\ * & P_3^{im} \end{bmatrix} > 0, \begin{bmatrix} N_1 & N_2 \\ * & N_3 \end{bmatrix} > 0, \tag{20}$$

where

$$\Pi_{1,1} = \begin{bmatrix} \Xi_{1,1} & \Xi_{1,2} \\ * & \Xi_{2,2} \end{bmatrix}, \Pi_{1,2} = \begin{bmatrix} S^{im} & M_2^{im} - T^{im} & T^{im} \\ R_2^{im} - S^{im} & 0 & 0 \end{bmatrix},$$

$$\Pi_{1,3} = \begin{bmatrix} \Xi_{1,6} & -\varepsilon E_a E_d + PD \\ 0 & 0 \end{bmatrix}, \Pi_{6,6} = -(\lambda_2 G_c^T G_c)^{-1},$$

$$\Pi_{1,4} = \begin{bmatrix} b_m^c \bar{b}_i^s PK & \bar{b}_m^c P & b_m^c b_i^s PK \\ 0 & 0 & \lambda_1 C^T G_s^T G_s \end{bmatrix}, \Pi_{1,5} = \begin{bmatrix} P_1^{im} + (P_2^{im})^T & P_2^{im} + P_3^{im} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\Pi_{1,6} = \begin{bmatrix} 0 \\ b_i^s C^T K^T \end{bmatrix}, \Pi_{1,7} = \Psi \otimes \begin{bmatrix} -A^T \\ b_i^s b_m^c C^T K^T \end{bmatrix}, \Pi_{2,3} = \begin{bmatrix} 0 & 0 \\ 0 & \Phi\Delta_2 - (1-\bar{\eta})N_2 \\ 0 & 0 \end{bmatrix},$$

$$\Pi_{2,2}=\begin{bmatrix}-Q_1-R_2^{im} & 0 & 0\\ * & \Xi_{4,4} & M_2^{im}-T^{im}\\ * & * & -Q_2-M_2^{im}\end{bmatrix},\Pi_{2,5}=\begin{bmatrix}-P_1^{im} & -P_2^{im} & 0\\ 0 & 0 & 0\\ -(P_2^{im})^T & -P_3^{im} & 0\end{bmatrix},$$

$$\Pi_{3,3}=\begin{bmatrix}\Xi_{6,6} & \varepsilon E_aE_d\\ * & \Xi_{7,7}\end{bmatrix},\Pi_{3,7}=\Psi\otimes\begin{bmatrix}B^T\\ D^T\end{bmatrix},\Pi_{4,4}=\begin{bmatrix}-\lambda_1 I & 0 & 0\\ * & -\lambda_2 I & 0\\ * & * & \Xi_{10,10}\end{bmatrix},$$

$$\Pi_{4,6}=\begin{bmatrix}\bar{b}_i^sK^T\\ 0\\ b_i^sK^T\end{bmatrix},\Pi_{4,7}=\Psi\otimes\begin{bmatrix}\bar{b}_i^sb_m^cK^T\\ \bar{b}_m^c\\ b_i^sb_m^cK^T\end{bmatrix},\Pi_{5,5}=\begin{bmatrix}\bar{P}_1^{im}-R_1^{im} & \bar{P}_2^{im} & 0\\ * & \bar{P}_3^{im}-M_1^{im} & 0\\ * & * & -\varepsilon I\end{bmatrix},$$

$$\Pi_{5,7}=\Psi\otimes\begin{bmatrix}0\\ 0\\ G^T\end{bmatrix},\Pi_{7,7}=\begin{bmatrix}-(R_2^{im})^{-1} & 0 & 0 & 0\\ * & -(M_2^{im})^{-1} & 0 & 0\\ * & * & -2h(L_2)^{-1} & 0\\ * & * & * & -2\eta(J_2)^{-1}\end{bmatrix},$$

$$\Xi_{1,1}=-2PA+Q_1+Q_2+N_1+h^2R_1^{im}+\eta^2M_1^{im}+\frac{h^3}{2}L_1+\frac{\eta^3}{2}J_1-R_2^{im}-M_2^{im}+\varepsilon E_a^2,$$

$$\Xi_{1,2}=b_m^cb_i^sPKC+R_2^{im}-S^{im},\Psi=\begin{bmatrix}h & \eta & h^2 & \eta^2\end{bmatrix},$$

$$\Xi_{2,2}=-2R_2^{im}+He[S^{im}]+C^T\Omega C+\lambda_1C^TG_s^TG_sC+\Omega,$$

$$\Xi_{4,4}=-(1-\bar{\eta})N_1-2M_2^{im}+He[T^{im}],\Xi_{6,6}=N_3-2\Delta_1+\varepsilon E_b^2,$$

$$\Xi_{7,7}=-(1-\bar{\eta})N_3-2\Delta_2+\varepsilon E_d^2,\quad \Xi_{10,10}=\lambda_1G_s^TG_s-\varpi\Omega,$$

$$\Xi_{1,6}=PB+N_2+\Phi\Delta_1-\varepsilon E_aE_b.$$

Proof. Consider the following fractional order Lyapunov–Krasovskii functional:

$$V(t,e(t))=\sum_{k=1}^{9}V_k(t,e(t),r_t,q_t),$$

where

$$V_1(t,e(t),r_t,q_t)={}_{t0}D_t^{r-1}e^T(t)Pe(t),$$

$$V_2(t,e(t),r_t,q_t)=\frac{1}{2}d^T(t)d(t),$$

$$V_3(t,e(t),r_t,q_t)=\begin{bmatrix}\int_{t-h}^t e(s)\,\mathrm{d}s\\ \int_{t-\eta}^t e(s)\,\mathrm{d}s\end{bmatrix}^T\begin{bmatrix}P_1^{im} & P_2^{im}\\ * & P_3^{im}\end{bmatrix}\begin{bmatrix}\int_{t-h}^t e(s)\,\mathrm{d}s\\ \int_{t-\eta}^t e(s)\,\mathrm{d}s\end{bmatrix},$$

$$V_4(t,e(t),r_t,q_t)=\int_{t-h}^t e^T(s)Q_1e(s)\,\mathrm{d}s+\int_{t-\eta}^t e^T(s)Q_2e(s)\,\mathrm{d}s,$$

$$V_5(t,e(t),r_t,q_t)=\int_{t-\eta(t)}^t\begin{bmatrix}e(s)\\ f(e(s))\end{bmatrix}^T\begin{bmatrix}N_1 & N_2\\ * & N_3\end{bmatrix}\begin{bmatrix}e(s)\\ f(e(s))\end{bmatrix}\mathrm{d}s,$$

$$V_6(t,e(t),r_t,q_t)=h\int_{-h}^0\int_{t+\theta}^t\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}^T\begin{bmatrix}R_1^{im} & 0\\ 0 & R_2^{im}\end{bmatrix}\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}\mathrm{d}s\mathrm{d}\theta,$$

$$V_7(t,e(t),r_t,q_t)=\eta\int_{-\eta}^0\int_{t+\theta}^t\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}^T\begin{bmatrix}M_1^{im} & 0\\ 0 & M_2^{im}\end{bmatrix}\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}\mathrm{d}s\mathrm{d}\theta,$$

$$V_8(t,e(t),r_t,q_t)=h\int_{-h}^0\int_{\theta}^0\int_{t+\beta}^t\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}^T\begin{bmatrix}L_1 & 0\\ 0 & L_2\end{bmatrix}\begin{bmatrix}e(s)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\mathrm{d}\beta\mathrm{d}\theta,$$

$$V_9(t,e(t),r_t,q_t)=\eta\int_{-\eta}^0\int_{\theta}^0\int_{t+\beta}^t\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}^T\begin{bmatrix}J_1 & 0\\ 0 & J_2\end{bmatrix}\begin{bmatrix}e(s)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\mathrm{d}\beta\mathrm{d}\theta.$$

For simplicity, $V_i=V_i(t,e(t),r_t,q_t),i=1,2,\ldots,9$.

The weak infinitesimal operator $\mathcal{L}$ is defined as follows:

$$\mathcal{L}V(t,e(t),r_t,q_t) = \frac{\partial V(t,e(t),r_t,q_t)}{\partial t} + \dot{e}^T(t)\frac{\partial V(t,e(t),r_t,q_t)}{\partial e(t)}\bigg|_{r_t=i,q_t=m} + \sum_{j=1}^{2}\pi_{ij}V(e(t),j,m) + \sum_{n=1}^{2}\rho_{mn}V(e(t),i,n).$$

By calculating the weak infinitesimal derivatives of $V(t,e(t),r_t,q_t)$ along with the error system (16), one has

$$\mathcal{L}V_1 \le 2e^T(t)PD_t^{\gamma}e(t), \quad \mathcal{L}V_2 = d(t)\dot{d}(t), \tag{21}$$

$$\begin{aligned}\mathcal{L}V_3 = {} & 2\begin{bmatrix}\int_{t-h}^{t}e(s)\,\mathrm{d}s\\ \int_{t-\eta}^{t}e(s)\,\mathrm{d}s\end{bmatrix}^T\begin{bmatrix}P_1^{im} & P_2^{im}\\ * & P_3^{im}\end{bmatrix}\begin{bmatrix}e(t)-e(t-h)\\ e(t)-e(t-\eta)\end{bmatrix} + \begin{bmatrix}\int_{t-h}^{t}e(s)\,\mathrm{d}s\\ \int_{t-\eta}^{t}e(s)\,\mathrm{d}s\end{bmatrix}^T\\ & \times\begin{bmatrix}\bar{P}_1^{im} & \bar{P}_2^{im}\\ * & \bar{P}_3^{im}\end{bmatrix}\begin{bmatrix}\int_{t-h}^{t}e(s)\,\mathrm{d}s\\ \int_{t-\eta}^{t}e(s)\,\mathrm{d}s\end{bmatrix},\end{aligned} \tag{22}$$

$$\mathcal{L}V_4 = e^T(t)(Q_1+Q_2)e(t) - e^T(t-h)Q_1e(t-h) - e^T(t-\eta)Q_2e(t-\eta), \tag{23}$$

$$\begin{aligned}\mathcal{L}V_5 = {} & \begin{bmatrix}e(t)\\ f(e(t))\end{bmatrix}^T\begin{bmatrix}N_1 & N_2\\ * & N_3\end{bmatrix}\begin{bmatrix}e(t)\\ f(e(t))\end{bmatrix} - (1-\bar{\eta})\begin{bmatrix}e(t-\eta(t))\\ f(e(t-\eta(t)))\end{bmatrix}^T\\ & \times\begin{bmatrix}N_1 & N_2\\ * & N_3\end{bmatrix}\begin{bmatrix}e(t-\eta(t))\\ f(e(t-\eta(t)))\end{bmatrix},\end{aligned} \tag{24}$$

$$\begin{aligned}\mathcal{L}V_6 = {} & h^2\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}^T\begin{bmatrix}R_1^{im} & 0\\ 0 & R_2^{im}\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix} - h\int_{t-h}^{t}\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}\begin{bmatrix}R_1^{im} & 0\\ 0 & R_2^{im}\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\\ & + h\int_{-h}^{0}\int_{t+\theta}^{t}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\begin{bmatrix}\bar{R}_1^{im} & 0\\ 0 & \bar{R}_2^{im}\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\mathrm{d}\theta,\end{aligned} \tag{25}$$

$$\begin{aligned}\mathcal{L}V_7 = {} & \eta^2\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}^T\begin{bmatrix}M_1^{im} & 0\\ 0 & M_2^{im}\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix} - \eta\int_{t-\eta}^{t}\begin{bmatrix}e(s)\\ \dot{e}(s)\end{bmatrix}\begin{bmatrix}M_1^{im} & 0\\ 0 & M_2^{im}\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\\ & + \eta\int_{-\eta}^{0}\int_{t+\theta}^{t}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\begin{bmatrix}\bar{M}_1^{im} & 0\\ 0 & \bar{M}_2^{im}\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\mathrm{d}\theta,\end{aligned} \tag{26}$$

$$\mathcal{L}V_8 = \frac{h^3}{2}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}^T\begin{bmatrix}L_1 & 0\\ 0 & L_2\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix} - h\int_{-h}^{0}\int_{t+\theta}^{t}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}^T\begin{bmatrix}L_1 & 0\\ 0 & L_2\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\mathrm{d}\theta, \tag{27}$$

$$\mathcal{L}V_9 = \frac{\eta^3}{2}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}^T\begin{bmatrix}J_1 & 0\\ 0 & J_2\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix} - \eta\int_{-\eta}^{0}\int_{t+\theta}^{t}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}^T\begin{bmatrix}J_1 & 0\\ 0 & J_2\end{bmatrix}\begin{bmatrix}e(t)\\ \dot{e}(t)\end{bmatrix}\mathrm{d}s\mathrm{d}\theta. \tag{28}$$

By using Lemmas 1 and 2, it follows that

$$-h\int_{t-h}^{t}\dot{e}^T(s)R_2^{im}\dot{e}(s)\,\mathrm{d}s \le \xi_1^T(t)\Theta_1\xi_1(t), \tag{29}$$

$$-\eta\int_{t-\eta}^{t}\dot{e}^T(s)M_2^{im}\dot{e}(s)\,\mathrm{d}s \le \xi_2^T(t)\Theta_2\xi_2(t), \tag{30}$$

$$-h\int_{t-h}^{t}e^T(s)R_1^{im}e(s)\,\mathrm{d}s \le -\Big(\int_{t-h}^{t}e(s)\,\mathrm{d}s\Big)^T R_1^{im}\int_{t-h}^{t}e(s)\,\mathrm{d}s, \tag{31}$$

$$-\eta\int_{t-\eta}^{t}e^{T}(s)M_1^{im}e(s)\,\mathrm{d}s\le-\Big(\int_{t-\eta}^{t}e(s)\,\mathrm{d}s\Big)^{T}M_1^{im}\int_{t-\eta}^{t}e(s)\,\mathrm{d}s, \tag{32}$$

where $\xi_1(t)=col\{e(t),e(t-\tau(t)),e(t-h)\},\xi_2(t)=col\{e(t),e(t-\eta(t)),e(t-\eta)\}$, and

$$\Theta_1=\begin{bmatrix}-R_2^{im} & R_2^{im}-S^{im} & S^{im}\\ * & -2R_2^{im}+He[S^{im}] & R_2^{im}-S^{im}\\ * & * & -R_2^{im}\end{bmatrix},$$

$$\Theta_2=\begin{bmatrix}-M_2^{im} & M_2^{im}-T^{im} & T^{im}\\ * & -2M_2^{im}+He[T^{im}] & M_2^{im}-T^{im}\\ * & * & -M_2^{im}\end{bmatrix}.$$

It can be obtained from $m(t)$ that

$$\begin{aligned}&\varepsilon e^{T}(t)E_a^2e(t)-2\varepsilon e^{T}(t)E_aE_bf(e(t))-2\varepsilon e^{T}(t)E_aE_df(e(t-\eta(t)))\\&+\varepsilon f^{T}(e(t))E_b^2f(e(t))+2\varepsilon f^{T}(e(t))E_bE_df(e(t-\eta(t)))\\&+\varepsilon f^{T}(e(t-\eta(t)))E_d^2f(e(t-\eta(t)))-\varepsilon m^{T}(t)m(t)\ge0.\end{aligned} \tag{33}$$

Moreover, from the adaptive event-triggered condition, activation function, (11) and (14), it follows that

$$d(t)\dot{d}(t)\le z^{T}(T-\tau(t))\Omega z(T-\tau(t))-\bar{w}\tilde{z}^{T}(t)\Omega\tilde{z}(t), \tag{34}$$

$$-2f^{T}(e(t))\Delta_1f(e(t))+2e^{T}(t)\Phi\Delta_1f(e(t))\ge0, \tag{35}$$

$$-2f^{T}(e(t-\eta(t)))\Delta_2f(e(t-\eta(t)))+2e^{T}(t-\eta(t))\Phi\Delta_2f(e(t-\eta(t)))\ge0, \tag{36}$$

$$\lambda_1\bar{z}^{T}(t)G_s^{T}G_s\bar{z}(t)-\lambda_1g_s^{T}(\bar{z}(t))g_s(\bar{z}(t))\ge0, \tag{37}$$

$$\lambda_2u_c^{T}G_c^{T}G_cu_c-\lambda_2g_c^{T}(u_c)g_c(u_c)\ge0. \tag{38}$$

Let

$$\begin{aligned}\zeta(t)=col\Big\{&e(t),e(t-\tau(t)),e(t-h),e(t-\eta(t)),e(t-\eta),f(e(t)),f(e(t-\eta(t))),\\&g_s(\bar{z}),g_c(u_c),\tilde{z}(t),\int_{t-h}^{t}e^{T}(s)\,\mathrm{d}s,\int_{t-\eta}^{t}e^{T}(s)\,\mathrm{d}s,m(t)\Big\},\end{aligned}$$

together with (21)–(38). Then, the following can be obtained.

$$\mathcal{L}V(t,e(t),r_t,q_t)\le\zeta^{T}(t)\Xi\zeta(t).$$

From the aforementioned part, we know that matrix inequality (17) guarantees $\Xi<0$ holds. That further guarantees that $\mathcal{L}V(t,e(t),r_t,q_t)<0$ holds for every $i,m\in M$.

Let $\lambda_0=\lambda_{min}(-\Xi)$; then $\lambda_0>0$. For any $t>0$, we have:

$$\mathcal{L}V(t,e(t),r_t,q_t)\le-\lambda_0\zeta^{T}(t)\zeta(t)\le-\lambda_0e^{T}(t)e(t).$$

By Definition 3, one can obtain:

$$\mathcal{E}V(t,e(t),i,m)-\mathcal{E}V(0,\phi(t_0),r_0,q_0)\le-\lambda_0\mathcal{E}\Big\{\int_0^{t}e^{T}(t)e(t)\,\mathrm{d}s\Big\},$$

hence, for $t \geq 0$:

$$\mathcal{E}\Big\{\int_0^t e^T(t)e(t)\,\mathrm{d}s\Big\} \leq \frac{1}{\lambda_0}\mathcal{E}V(0,\phi(t_0),r_0,q_0),$$

based on Definition 4, which implies that error system (16) is globally, stochastically, asymptotically stable in the mean square sense. That means systems (1) and (2) get globally, stochastically, asymptotically synchronized in the mean square sense. The proof is completed. □

Notice that Theorem 1 only gives sufficient conditions for the synchronization of systems (1) and (2), and fails to solve the design problem of the controller (12). Therefore, the design method of the control gain K is constructed in Theorem 2.

Theorem 2. *Suppose Assumption 1 holds. The FNNs (1) and (2) are globally, stochastically, asymptotically synchronized in the mean square sense, for the given scalars r, if there exist positive definite matrices* $P, \Omega, P_1^{im}, P_3^{im}, N_1, N_3, R_1^{im}, R_2^{im}, M_1^{im}, M_2^{im}, L_1, L_2, J_1, J_2, Q_1, Q_2$*; positive definite diagonal matrices* Δ_1, Δ_2*; and matrices* $P_2^{im}, N_2, S^{im}, T^{im}, Y$*; and positive scalars* $\varepsilon, \lambda_1, \lambda_2$*,such that the following LMIs for every i, m hold:*

$$\begin{bmatrix} \tilde{\Pi}_{1,1} & \Pi_{1,2} & \Pi_{1,3} & \tilde{\Pi}_{1,4} & \Pi_{1,5} & \tilde{\Pi}_{1,6} & \tilde{\Pi}_{1,7} \\ * & \Pi_{2,2} & \Pi_{2,3} & 0 & \Pi_{2,5} & 0 & 0 \\ * & * & \Pi_{3,3} & 0 & 0 & 0 & \tilde{\Pi}_{3,7} \\ * & * & * & \Pi_{4,4} & 0 & \tilde{\Pi}_{4,6} & \tilde{\Pi}_{4,7} \\ * & * & * & * & \Pi_{5,5} & 0 & \tilde{\Pi}_{5,7} \\ * & * & * & * & * & \tilde{\Pi}_{6,6} & 0 \\ * & * & * & * & * & * & \tilde{\Pi}_{7,7} \end{bmatrix} < 0, \tag{39}$$

$$\bar{R}_1^{im} < L_1, \bar{R}_2^{im} < L_2, \bar{M}_1^{im} < J_1, \bar{M}_2^{im} < J_1, \tag{40}$$

$$\begin{bmatrix} R_2^{im} & S^{im} \\ * & R_2^{im} \end{bmatrix} \geq 0, \begin{bmatrix} M_2^{im} & T^{im} \\ * & M_2^{im} \end{bmatrix} \geq 0, \tag{41}$$

$$\begin{bmatrix} P_1^{im} & P_2^{im} \\ * & P_3^{im} \end{bmatrix} > 0, \begin{bmatrix} N_1 & N_2 \\ * & N_3 \end{bmatrix} > 0, \tag{42}$$

where

$$\tilde{\Pi}_{1,1} = \begin{bmatrix} \Xi_{1,1} & \tilde{\Xi}_{1,2} \\ * & \Xi_{2,2} \end{bmatrix}, \tilde{\Pi}_{1,4} = \begin{bmatrix} b_m^c \bar{b}_i^s Y & \bar{b}_m^c P & b_m^c b_i^s Y \\ 0 & 0 & \lambda_1 C^T G_s^T G_s \end{bmatrix}, \tilde{\Pi}_{1,6} = \begin{bmatrix} 0 \\ b_i^s C^T Y^T \end{bmatrix},$$

$$\tilde{\Pi}_{1,7} = \Psi \otimes \begin{bmatrix} -A^T P \\ b_i^s b_m^c C^T Y^T \end{bmatrix}, \tilde{\Pi}_{3,7} = \Psi \otimes \begin{bmatrix} B^T P \\ D^T P \end{bmatrix}, \tilde{\Pi}_{4,6} = \begin{bmatrix} \bar{b}_i^s Y^T \\ 0 \\ b_i^s Y^T \end{bmatrix},$$

$$\tilde{\Pi}_{4,7} = \Psi \otimes \begin{bmatrix} \bar{b}_i^s b_m^c Y^T \\ \bar{b}_m^c P \\ b_i^s b_m^c Y^T \end{bmatrix}, \tilde{\Pi}_{5,7} = \Psi \otimes \begin{bmatrix} 0 \\ 0 \\ G^T P \end{bmatrix}, \tilde{\Pi}_{6,6} = -2\alpha_1 P + \alpha_1^2 \lambda_2 G_c^T G_c,$$

$$\tilde{\Pi}_{7,7} = diag\{-2\alpha_2 P + \alpha_2^2 R_2^{im}, -2\alpha_3 P + \alpha_3^2 M_2^{im}, -4h\alpha_4 P + 2h\alpha_4^2 L_2, -4h\alpha_5 P + 2h\alpha_5^2 J_2\},$$

$$\tilde{\Xi}_{1,2} = b_m^c b_i^s YC + R_2^{im} - S^{im},$$

and the other parameters are the same as in Theorem 1, among them the feedback gain matrix is defined with $K = P^{-1}Y$.

Proof. For any scalar $\alpha > 0$, the following inequality holds:

$$(\alpha\Omega - P)\Omega^{-1}(\alpha\Omega - P) \geq 0.$$

Based on the inequality, it can be obtained that:

$$-P\Omega^{-1}P \leq -2\alpha P + \alpha^2\Omega.$$

By defining $\chi = diag\{\overbrace{I,\ldots,I}^{13}, P, P, P, P, P\}$, multipling (17) by χ on the left side and the right side, respectively, and replacing the term in $\Pi_{6,6}$ with $-2\alpha_1 P + \alpha_1^2\lambda_4 G_c^T G_c$, $\tilde{\Pi}_{6,6}$ can be obtained. In the same way, $\tilde{\Pi}_{7,7}$ replaces $\Pi_{7,7}$. In addition, $Y = KP$ is also replaced. Then linear matrix inequality (39) can be obtained. That completes the proof. □

4. Numerical Simulations

In this section, a simulation is presented to demonstrate the effectiveness of the proposed approach. Consider the FNNs which are described by Equation (1) and (2) with the following parameters:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1.8 & -0.1 \\ -2 & 0.4 \end{bmatrix}, D = \begin{bmatrix} -1.7 & -0.6 \\ 0.5 & -2.5 \end{bmatrix},$$

$$E_a = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}, E_b = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}, E_d = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix},$$

$$G = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.02 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The nonlinear function was selected as $\hat{f}(x) = \tanh(x)$, so it can be calculated that $\Phi = I$. Due to the time-varying delay $\eta(t) = \frac{0.1e^t}{1+e^t}$, $\eta = 0.1, \bar{\eta} = 0.025$ can be obtained, respectively. The functions of deception signals are were chosen to be $g_s(x) = \tanh(x), g_c(x) = \tanh(x)$; therefore, one can get $G_s = I, G_c = I$. In this numerical example, we set the sampling period to $h = 0.05$, $\gamma = 0.98$, the initial value of the adaptive event-triggered parameter d_0 to 0.8, the external input vector $I(t)$ to 0, $\epsilon_1 = 0.1, \epsilon_2 = 0.1, \epsilon_3 = 0.1, \epsilon_4 = 0.1, \epsilon_5 = 0.1$. Additionally, the generators of Markov process r_t, q_t were

$$\pi_{ij} = \begin{bmatrix} -0.4 & 0.4 \\ 0.5 & -0.5 \end{bmatrix}, \rho_{ij} = \begin{bmatrix} -0.4 & 0.4 \\ 0.65 & -0.65 \end{bmatrix}.$$

Based on the proposed method, by solving the LMIs in Theorem 2, one can obtain the desired controller gain and the adaptive event-triggered weighting matrix as follows:

$$K = \begin{bmatrix} -0.0178 & 0.0026 \\ -0.0021 & -0.0270 \end{bmatrix}, \Omega = \begin{bmatrix} 0.0007 & 0.0007 \\ 0.0007 & 0.0011 \end{bmatrix}. \tag{43}$$

We chose the initial values $\phi_1(t_0) = (0.5; -0.1)$, $\phi_2(t_0) = (0.1; 0.2)$. Figure 2 shows the state trajectories of synchronization errors without control input. As can be seen from Figure 2, if there is no control input, the error system itself is unstable, which means that the systems cannot be synchronized. Using the feedback controller (12), the simulation results were obtained, as shown in Figures 3–7. Figure 3 shows the state trajectories of synchronization errors with control input, and one can see that synchronization errors finally converged to zero under the designed control protocol, which shows that the systems can achieve synchronization. Figures 4 and 5 depict the states of double deception attacks, whose states caused the oscillations of the synchronization error and the control input. Figure 6 depicts the trajectories of control input, from which one can see that the control input gradually tended to 0; that is, when the system achieves synchronization, external control is no longer required. Figure 7 shows the evolution of adaptive threshold coefficient $d(t)$ in AETS. From the adaptive law (6), the adaptive threshold coefficient can be timely adjusted according to the synchronization error. Therefore, when the error system is stable, that is, when synchronization is achieved, the parameter will no longer be adjusted and

will tend toward a constant. From the above simulation results, it can easily be seen that the proposed synchronization problem in this paper was effectively solved.

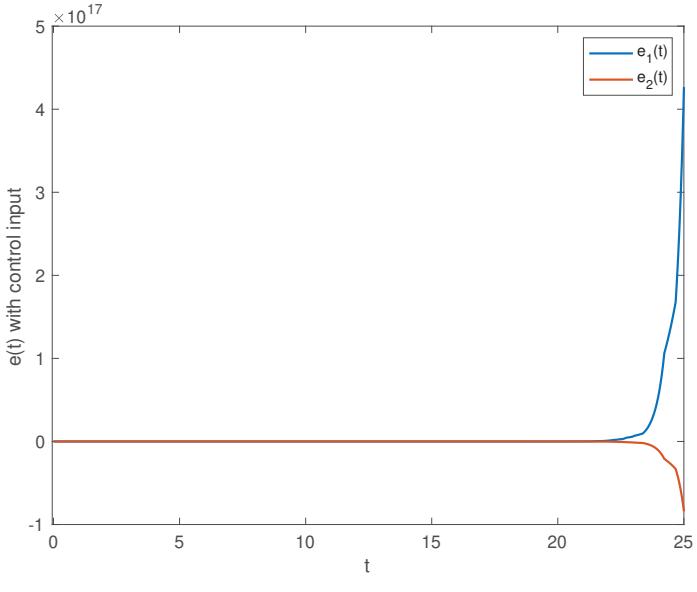


Figure 2. Synchronization error $e_i(t)(i = 1, 2)$ without control input.

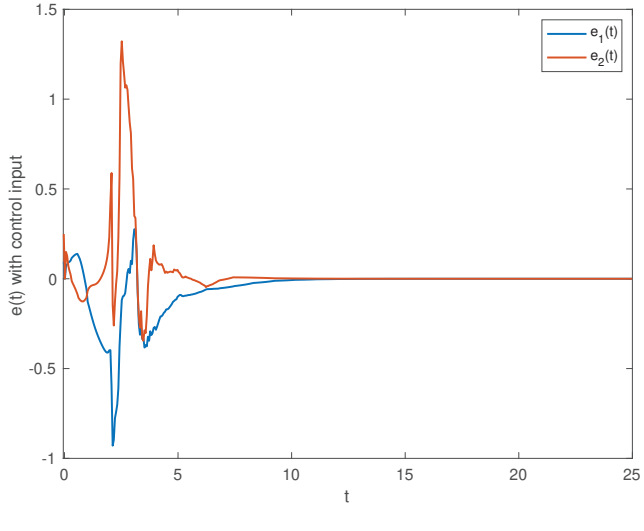


Figure 3. Synchronization error $e_i(t)(i = 1, 2)$ with control input $u_i(t)(i = 1, 2)$.

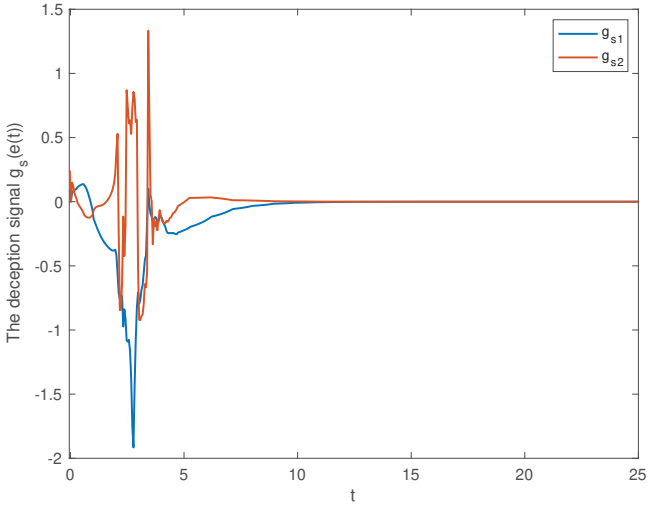


Figure 4. The state of the deception signal in the S-C channel.

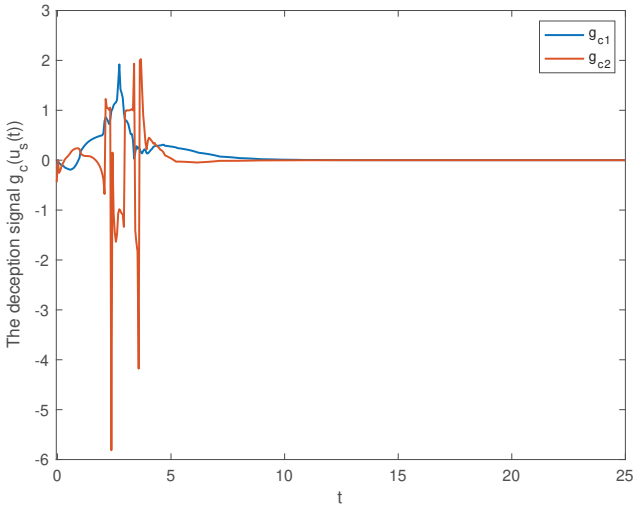


Figure 5. The state of the deception signal in the C-A channel.

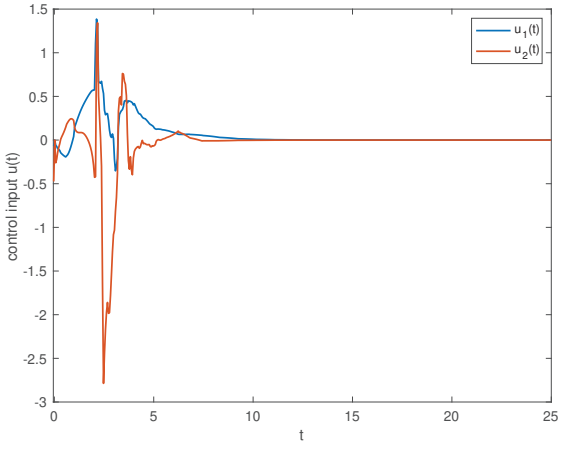


Figure 6. The trajectories of control input $u_i(t)(i = 1, 2)$.

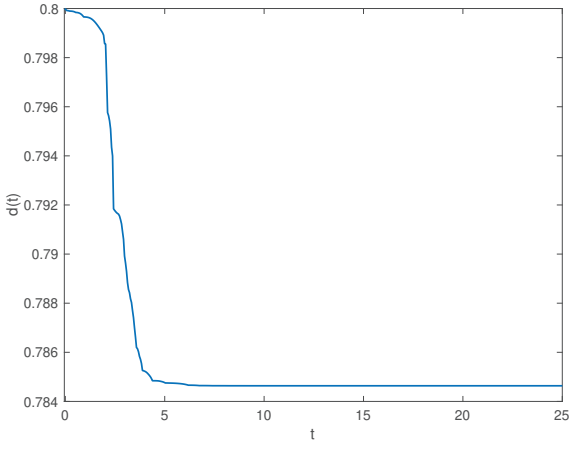


Figure 7. The trajectory of event-triggered parameter $d(t)$.

5. Discussion and Conclusions

The adaptive event-triggered synchronization problem of uncertain FNNs with double deception attacks and time-varying delay has been investigated in this paper. Noteworthy is that, regarding fractional order systems receiving deception attacks using traditional event-triggered methods given in the literature [22], we believe that the literature has not been comprehensive enough. Not only the traditional ETS technology, but also the attack phenomena were governed by Bernoulli processes, and attacks only occurred in the C-A channel. Thus, in this study, the AETS was adopted to determine the signals the needed to be transmitted. The deception attacks in communication channels from the sensor to controller and from controller to actuator are governed by two independent Markov processes. Considering the AETS, double deception attacks, and parameter uncertainties, a time-varying closed-loop fractional order synchronization error system was constructed. Sufficient conditions were formulated to guarantee the considered system is stochastically stable by employing the Lyapunov–Krasovskii functional method. Finally, a numerical example was presented to verify its effectiveness and the feasibility of the proposed

method. Thereby, we showed that our approach is more meaningful and comprehensive. It should be mentioned that besides deception attacks, denial of service (DoS) attacks is another interesting issue for FNNs and deserves further exploration. In addition, solving the problem of multiple communication channels for FNNs will be part of our future research efforts.

Author Contributions: Z.S. was in charge of the construction of model and writing. F.Y. was in charge of the simulation. J.C. and J.Z. mainly contributed to the synchronization analysis. A.H. mainly contributed via the supervision of program. M.H. was in charge of the review and editing of the whole paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work is jointly supported by the National Natural Science Foundation of China under grant 61973137 and the Natural Science Foundation of Jiangsu Province under grant BK20181342.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this paper:

FNNs	Fractional order neural networks
ETS	Event-triggered scheme
TTS	Time-triggered scheme
AETS	Adaptive event-triggered scheme
S-C Channel	Sensor to controller channel
C-A Channel	Controller to actuator channel

References

1. Wang, Y.; Guo, J.; Liu, G.B.; Lu, J.W. Finite-time sampled-data synchronization for uncertain neutral-type semi-Markovian jump neural networks with mixed time-varying delays. *Appl. Math. Comput.* **2021**, *403*, 126197.
2. Ding, S.B.; Wang, Z.S.; Rong, N.N. Intermittent control for quasi synchronization of delayed discrete-time neural networks. *IEEE Trans. Cybern.* **2021**, *51*, 862–873. [CrossRef] [PubMed]
3. Tian, Y.F.; Wang, Z.S. A new result on $H\infty$ performance state estimation for static neural networks with time-varying delays. *Appl. Math. Comput.* **2021**, *388*, 125556. [CrossRef]
4. Yuan, M.M.; Wang, W.P.; Wang, Z. Exponential synchronization of delayed memristor-based uncertain complex-valued neural networks for image protection. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *99*, 1–15. [CrossRef] [PubMed]
5. Xi, H.L.; Li, Y.X.; Huang, X. Generation and nonlinear dynamical analyses of fractional-order memristor-based lorenz systems. *Entropy* **2014**, *16*, 6240–6253. [CrossRef]
6. Sun, H.G.; Yong, Z.; Baleanu, D.; Chen, Y.Q. A new collection of real world applications of fractional calculus in science and engineering. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *64*, 213. [CrossRef]
7. Vladimir, Z.; Ilya, K. Best approximation of the fractional semi-derivative operator by exponential series. *Mathematics* **2018**, *6*, 12.
8. Chen, L.P.; Qu, J.F.; Chai, Y. Synchronization of a class of fractional-order chaotic neural networks. *Entropy* **2013**, *15*, 3265–3276. [CrossRef]
9. Cao, J.D.; Stamov, G.; Stamova, I. Almost periodicity in impulsive fractional-order reaction-diffusion neural networks with time-varying delays. *IEEE Trans. Cybern.* **2020**, *51*, 151–161. [CrossRef]
10. Jia, J.; Huang, X.; Li, Y.X.; Cao, J.D. Global stabilization of fractional-order memristor-based neural networks with time delay. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 997–1009. [CrossRef]
11. Xiao, J.Y.; Cao, J.D.; Cheng, J. Novel methods to finite-time Mittag–Leffler synchronization problem of fractional-order quaternion-valued neural networks. *Inf. Sci.* **2020**, *526*, 221–244. [CrossRef]
12. Hu, H.P.; Wang, J.K.; Xie, F.L. Dynamics analysis of a new fractional-order hopfield neural network with delay and its generalized projective synchronization. *Entropy* **2019**, *21*, 1. [CrossRef]
13. Wu, Y.B.; Zhu, J.L.; Li, W.X. Intermittent discrete observation control for synchronization of stochastic neural networks. *IEEE Trans. Cybern.* **2020**, *50*, 2414–2424. [CrossRef]
14. Cao, Y.T.; Wang, S.B.; Guo, Z.Y.; Huang, T.W. Synchronization of memristive neural networks with leakage delay and parameters mismatch via event-triggered control. *Neural Netw.* **2019**, *119*, 178–189. [CrossRef]

15. Rakkiyappan, R.; Dharani, S.; Cao, J.D. Synchronization of neural networks with control packet loss and time-varying delay via stochastic sampled-data controller. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 3215–3226. [CrossRef]
16. Dai, M.C.; Xia, J.W.; Xia, H.; Shen, H. Event-triggered passive synchronization for Markov jump neural networks subject to randomly occurring gain variations. *Neurocomputing* **2019**, *33*, 403–411. [CrossRef]
17. Shanmugam, L.; Mani, P.; Rajan, R. Adaptive synchronization of reaction-diffusion neural networks and its application to secure communication. *IEEE Trans. Cybern.* **2020**, *50*, 911–922. [CrossRef]
18. Hu, T.T.; He, Z.; Zhang, X.J.; Zhong, S.M. Leader-following consensus of fractional-order multi-agent systems based on event-triggered control. *Nonlinear Dyn.* **2020**, *99*, 2219–2232. [CrossRef]
19. Cheng, Y.L.; Hu, T.T.; Li, Y.H.; Zhong, S.M. Consensus of fractional-order multi-agent systems with uncertain topological structure: A Takagi-Sugeno fuzzy event-triggered control strategy. *Fuzzy Sets Syst.* **2021**, *416*, 64–85. [CrossRef]
20. Wei, M.; Li, Y.X.; Tong, S. Event-triggered adaptive neural control of fractional-order nonlinear systems with full-state constraints. *Neurocomputing* **2020**, *412*, 320–326. [CrossRef]
21. Li, Q.P.; Liu, S.Y.; Chen, Y.G. Combination event-triggered adaptive networked synchronization communication for nonlinear uncertain fractional-order chaotic systems. *Appl. Math. Comput.* **2018**, *333*, 521–535. [CrossRef]
22. Xiong, M.H.; Tan, Y.S.; Zhang, B.Y.; Fei, S.M. Observer-based event-triggered output feedback control for fractional-order cyber–physical systems subject to stochastic network attacks. *ISA Trans.* **2020**, *104*, 15–25. [CrossRef] [PubMed]
23. Yu, N.X.; Zhu, W. Event-triggered impulsive chaotic synchronization of fractional-order differential systems. *Appl. Math. Comput.* **2021**, *388*, 125554. [CrossRef]
24. Zouad, F.; Kemih, K.; Hamiche, H. A new secure communication scheme using fractional order delayed chaotic system: Design and electronics circuit simulation. *Analog Integr. Circ. Signal Process.* **2019**, *99*, 619–632. [CrossRef]
25. Bettayeb, M.; Said, D. Single channel secure communication scheme based on synchronization of fractional-order chaotic chua's systems. *Trans. Inst. Meas. Control* **2018**, *40*, 3651–3664. [CrossRef]
26. Zhao, C.C.; Duan, X.M.; Cai, L.; Cheng, P. Vehicle platooning with non-ideal communication networks. *IEEE Trans. Veh. Technol.* **2021**, *70*, 18–32. [CrossRef]
27. Petrillo, A.; Pescape, A.; Santini, S. A collaborative approach for improving the security of vehicular scenarios: The case of platooning. *Comput. Commun.* **2018**, *122*, 59–75. [CrossRef]
28. Wang, Z.B.; Song, M.K.; Zheng, S.Y. Invisible adversarial attack against deep neural networks: An adaptive penalization approach. *IEEE Trans. Dependable Secur. Comput.* **2021**, *18*, 1474–1488. [CrossRef]
29. Rahman, R.; Tomar, D. Threats of price scraping on e-commerce websites: Attack model and its detection using neural network. *J. Comput. Virol. Hacking Tech.* **2021**, *17*, 75–89. [CrossRef]
30. Deng, Z.; Lun, X.; Yu, R. Data security transmission mechanism in industrial neural control systems against deception attack. *Int. J. Secur. Appl.* **2016**, *10*, 391–404.
31. Shen, B.; Wang, Z.D.; Wang, D.; Li, Q. State-saturated recursive filter design for stochastic time-varying nonlinear complex networks under deception attacks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 3788–3800. [CrossRef]
32. Pan, R.Q.; Tan, Y.S.; Du, D.S. Adaptive event-triggered synchronization control complex networks with quantization and cyber-attacks. *Neurocomputing* **2020**, *382*, 249–258. [CrossRef]
33. Liu, J.L.; Xia, J.L.; Cao, J.; Tian, E.G. Quantized state estimation for neural networks with cyber attacks and hybrid triggered communication scheme. *Neurocomputing* **2018**, *291*, 35–49. [CrossRef]
34. Sun, Y.M.; Yu, J.Y.; Yu, X.H.; Gao, H.J. Decentralized adaptive event-triggered control for a class of uncertain systems with deception attacks and its application to electronic circuits. *IEEE Trans. Circ. Syst.* **2020**, *67*, 12. [CrossRef]
35. Song, Q.K.; Shu, H.Q.; Zhao, Z.J.; Liu, Y.R. Lagrange stability analysis for complex-valued neural networks with leakage delay and mixed time-varying delays. *Neurocomputing* **2017**, *244*, 33–41. [CrossRef]
36. Peng, C.; Han, Q.L.; Yue, D. Communication delay distribution dependent decentralized control for large-scale systems with IP-based communication networks. *IEEE Trans. Control Syst. Technol.* **2013**, *21*, 820–830. [CrossRef]
37. Chen, G.R.; Zhou, J.; Liu, Z.R. Global synchronization of couple delayed neural networks and applications to chaotic cnn models. *Int. J. Bifurc. Chaos* **2004**, *14*, 2229–2240. [CrossRef]
38. Yu, W.W.; Cao, J.D. Synchronization control of stochastic delayed neural networks. *Phys. A* **2007**, *373*, 252–260. [CrossRef]
39. Vijay Aravind, R.; Balasubramaniam, P. Stochastic stability of fractional-order Markovian jumping complex-valued neural networks with time-varying delays. *Neurocomputing* **2021**, *439*, 122–133. [CrossRef]
40. Tian, J.K.; Li, Y.M.; Zhao, J.; Zhong, S. Delay-dependent stochastic stability criteria for Markovian jumping neural networks with mode-dependent time-varying delays and partially known transition rates. *Appl. Math. Comput.* **2012**, *218*, 5769–5781. [CrossRef]
41. Balasubramaniam, P.; Rakkiyappan, R. Delay-dependent robust stability analysis for Markovian jumping stochastic Cohen–Grossberg neural networks with discrete interval and distributed time-varying delays. *Nonlinear Anal. Hybrid Syst.* **2009**, *3*, 207–214. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

MDPI Books Editorial Office
E-mail: books@mdpi.com
www.mdpi.com/journal/books

Academic Open
Access Publishing

www.mdpi.com

ISBN 978-3-0365-7649-7