

Article

Short-Term Forecasting of the Total Power Generation from Wind Farms and Solar Power Plants in the National Power System Using Advanced Ensemble Machine Learning Models

Paweł Piotrowski 

Electrical Power Engineering Institute, Warsaw University of Technology, Koszykowa 75 Street,
00-662 Warsaw, Poland; pawel.piotrowski@pw.edu.pl

Abstract

The introduction of the article presents the state of renewable energy development in Poland and statistical information on its dynamics in the context of sustainable development, highlighting both the positive aspects of this situation and the potential risks to the national power system. These risks stem from the inherent instability of renewable energy generation and the seasonal variability of production. The main part of the article provides a comprehensive statistical analysis of time series data (wind energy generation and solar energy generation) aimed at identifying the appropriate input variables for forecasting models. In addition to the two time series of electricity generation, other exogenous variables and feature engineering techniques were incorporated. In the forecasting section, short-term forecasts of energy generation in the national power system from wind farms and solar power plants were developed. The forecasts for both types of renewable energy sources (RESs) were conducted separately and then integrated into a single time series to assess which forecasting approach is more effective. A detailed analysis was carried out to determine the optimal hyperparameters for individual machine learning models. Subsequently, an ensemble model was developed, integrating multiple single models. The article concludes with final insights and practical recommendations regarding the selection of preferred models and input variables that ensure the highest forecast accuracy. Additionally, potential future developments of the models and further research in this field are discussed in the context of sustainable development.

Keywords: renewable energy; forecasting; sustainable development; wind farm; solar power plant; power generation; national power system; machine learning; artificial intelligence; ensemble model



Academic Editor: Lars Johanning

Received: 21 December 2025

Revised: 26 January 2026

Accepted: 7 February 2026

Published: 10 February 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

While a large number of studies have addressed forecasting of individual wind farms or photovoltaic (PV) plants, the forecasting of aggregated generation (wind and/or PV) at the national or system-wide scale is comparatively less common—but increasingly important as the share of variable renewables grows. At the system level, aggregated predictions support tasks such as reserve scheduling, intra-hour balancing, market clearing, and grid operation, hence improving accuracy and operational integration. The main task of this article is forecasting hourly wind and solar power generation with a +1 h horizon, which is essential for balancing the national power system. The motivation behind this research stems from the fact that the +1 h forecasting horizon for renewable energy sources

(RESs) generation holds the greatest significance for system operators, particularly in the context of real-time balancing and the activation of spinning reserves. It is important to note that most existing studies focus on forecasting horizons ranging from 3 to 24 h, or on forecasting generation at the level of individual wind farms and power plants.

Article [1] presents a +1 h forecast for the Spanish national power system (hourly aggregate). The study demonstrated that a Transformer-based model trained on wind data for the Spain European Network of Transmission System Operators for Electricity (ENTSO-E) 2020–2024 outperforms Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models. Beyond accuracy, attention maps revealed the Transformer effectively captures short-term fluctuations and long-range dependencies, offering a transparent mechanism for interpreting historical information [1].

Article [2] forecasts PV generation for Germany's national territory. A 43,824 h dataset (2015–2019) of solar power Load Profile (LP), irradiance, and ten meteorological variables was used with eight Machine Learning and Deep Learning (ML/DL) algorithms. Adaptive Boosting (ADA) and Random Forest (RF) outperformed convolutional neural networks. The high correlation (0.9778) between the LP and ground-level irradiance enables forecasting from a combination of irradiance and meteorological variables [2].

The research problem under consideration concerns the situation in which Numerical Weather Prediction (NWP) data are not available. In this case, wind farms and PV power plants are geographically dispersed across the entire country, which makes it difficult to assign precise point-based NWP data to the aggregated forecasted process. The scientific literature identifies several typical approaches to solving this forecasting problem.

The first approach relies on time series forecasting, where the model input consists of selected lagged values of the target variable, optionally complemented by additional engineered features (feature engineering). Such features may include indicators of the time of day or season, installed capacity, as well as transformed values of the time series (e.g., smoothed or filtered versions) to reduce stochastic noise in the process [3–6]. In these cases, commonly used models include Auto Regressive (AR), Auto Regressive Integrated Moving Average (ARIMA), Trigonometric Seasonal Box–Cox ARMA Time Series (TBATS) and deep autoregressive learning architectures such as LSTM, GRU, Convolutional Neural Network (CNN) and Transformer networks [1,7–9].

The second approach is based on a physico-statistical framework for forecasting power generation from PV plants. An example of such a method is presented in [10]. The authors successfully predict hourly solar power values for horizons of up to two hours ahead, without using any NWP inputs. The dataset consists of 15 min solar power measurements from 21 rooftop PV systems located in a small village in Denmark. The proposed method follows a two-stage procedure. In the first stage, a statistical normalization of solar power is performed using a clear-sky model. The clear-sky reference is estimated through statistical smoothing techniques. In the second stage, forecasts of the normalized solar power are generated using adaptive linear time series models. The results indicate that for forecast horizons up to two hours, the most relevant inputs are the available historical observations of solar power, whereas for longer horizons, NWP-based variables become the dominant predictors [10].

The third approach is based on satellite-based nowcasting of Global Horizontal Irradiance (GHI). GHI represents the total amount of solar radiation incident on a unit of horizontal surface. It consists of two main components: Direct Normal Irradiance (DNI)—the direct solar radiation received from the Sun without scattering—and Diffuse Horizontal Irradiance (DHI)—the diffuse radiation reaching the surface from the entire sky after being scattered in the atmosphere (e.g., by clouds or aerosols). GHI is the most important input parameter for PV power generation models, as it enables the calculation of

the total energy that PV modules, oriented at various tilt angles, can produce through the transposition of GHI onto the plane of the PV array. The authors of [11] shows a data-driven approach to short-term forecasting of GHI using geostationary satellite imagery. The authors apply various machine learning models, including convolutional and also ensemble methods to extract cloud motion and brightness features directly from satellite data. The results show that satellite-based nowcasting significantly improves forecast accuracy for horizons up to 2 h in comparison to traditional statistical models and offers a practical solution for real-time solar energy management.

Article [12] utilizes a Transformer-based model for short-term solar irradiance forecasting. The self-attention mechanism captures long-range temporal dependencies more effectively than RNNs like LSTM and GRU. Relying solely on historical data without NWP inputs, this model is ideal for real-time, scalable applications. Conversely, ref. [13] proposes a CNN-based approach for intra-hour forecasting using sky image sequences. This method extracts spatial-temporal cloud dynamics directly from images, eliminating NWP dependency. Experimental results demonstrate that this CNN approach significantly outperforms traditional statistical and persistence models, offering a practical solution for real-time solar nowcasting.

It should be emphasized that forecasting the generation of electricity from wind farms and solar power plants for horizons longer than several hours without the use of NWP data is challenging and characterized by increased prediction errors. In particular, this limitation applies to wind power generation forecasts, due to the high degree of randomness associated with the primary driving factor—wind speed. It is less pronounced in the case of solar power generation forecasts, owing to the relatively higher regularity and periodicity of meteorological conditions (solar irradiance), especially during the summer months. An interesting example of such an approach, combining NWP forecasts with ML models at the national scale, is presented in article [14], which addresses 24 h ahead forecasts of wind power generation in Poland. A numerical weather prediction model, operationally used at the Institute of Meteorology and Water Management—National Research Institute (IMGW-PIB) in Poland, together with hourly records of total wind power generation in the Polish power system, were employed to develop and validate the forecasting models. The best-performing model, based on Extreme Gradient Boosting (XGBoost) and trained on two years of data (2018–2019), achieved a Mean Absolute Percentage Error (MAPE) of 26.7% and a Root Mean Square Error (RMSE) of 4.5% for day-ahead hourly forecasts of national wind power generation in 2020 [14].

The problem of spatial forecasting for wind farms is discussed in article [15]. Spatial forecasting is understood here as simultaneous forecasting for both large (e.g., country) and small (e.g., district) areas. Such an approach is very useful to different stakeholders of the power industry, and especially to system operators. The study presents an analysis of forecasted meteorological data (NWP) aimed at improving the efficiency and accuracy of spatial wind power generation forecasts using the MultiLayer Perceptron (MLP) neural network type.

The primary objectives of this study are:

- (I) To compare the effectiveness of various ensemble models (variants of Random Forests and Gradient Boosting Decision Trees) in the task of 1 h ahead forecasting of power generation from wind farms, solar power plants, and the aggregated generation from both sources within the National Power System (covering the entire country). This comparison will utilize two data approaches: one using only lagged values of the target series and another incorporating additional input variables derived through feature engineering.

- (II) To propose novel, proprietary ensemble integrators designed to combine different base ensemble models (variants of Random Forests and Gradient Boosting Decision Trees) with the aim of improving forecasting quality.
- (III) To verify whether it is more beneficial to forecast the combined time series of wind and solar power generation directly, or if it is more advantageous to generate separate forecasts for wind power and solar power, and subsequently aggregate (sum) the two individual forecasts to obtain the total forecast for both RES.

The remainder of this paper is structured as follows. Section 1 (Introduction) provides an introduction to the topic. Section 2 discusses The issue of renewable energy development in Poland in the context of sustainable development. Section 3 (Data) includes the data analysis, while Section 4 (Methods) describes the proposed forecasting models. Section 5 (Experimental Results) presents the results of the conducted research, and Section 6 (Discussion) contains a comprehensive discussion of the findings. The paper concludes with a brief Section 7 (Summary).

2. The Issue of Renewable Energy Development in Poland in the Context of Sustainable Development

Over the past ten years, the development of wind energy in Poland has been largely determined by regulatory factors. The most significant constraint proved to be the so-called “10H rule” (where 10H stands for ten times the total height of the wind turbine). This regulation introduced a requirement that every new wind turbine must be located at a distance of at least ten times its total height from the nearest residential building or protected area. In practice, this led to the blocking of more than 95% of the country’s territory for new onshore wind investments. Between 2017 and 2021, an unclear and unstable auction system was also in force, which—although formally designed to support renewable energy—effectively limited the inflow of private capital due to frequent legislative changes and the lack of predictability in the support mechanism. An additional problem was the legal uncertainty concerning the tax classification of wind turbines: in 2017–2018, tax authorities interpreted that the entire turbine structure (including its technological components) constituted a building subject to property tax, which significantly increased operating costs for investors. It was not until 2022–2023 that the process of regulatory liberalization began. The amendment to the Act on Investments in Wind Power Plants [16] relaxed the 10H rule, allowing wind turbines to be located at a minimum distance of 700 m from residential buildings, provided that municipal approval is obtained and a full environmental impact assessment is conducted. The new legislation empowers municipalities to determine acceptable distances in local spatial development plans, restoring real opportunities for the development of new onshore wind projects. Despite this liberalization, the wind sector continues to face administrative constraints—including lengthy planning procedures, the lack of standardized noise assessment criteria, and uncertainty regarding future participation rules in support mechanisms (renewable energy auctions and contracts for difference). Furthermore, the absence of updated spatial development plans in many municipalities prevents the effective use of potentially available land for wind energy investments.

The dynamics of the development of solar power plants and wind farms in Poland from 2015 to 2025 are presented in Figure 1. Figure 1 clearly shows the effect of the liberalization of legal regulations (2022–2023), resulting in a significant change in the dynamics of installed capacity on onshore wind farms.

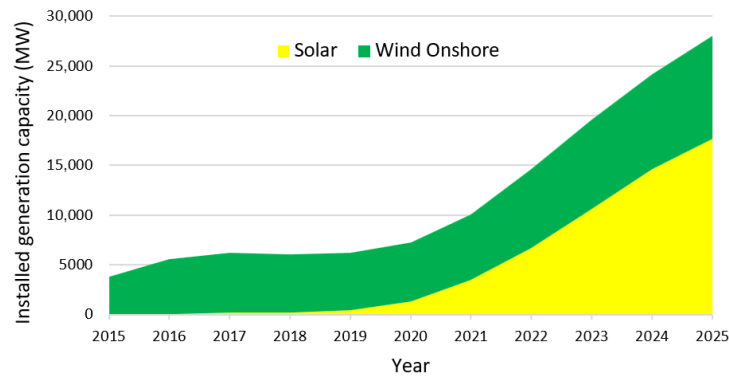


Figure 1. The dynamics of the development of solar power plants and wind farms in Poland from 2015 to 2025.

Between 2015 and 2025, there were also dynamic changes in the percentage share of installed capacity from onshore wind farms and solar power plants relative to the total installed capacity of all available and non-available power plants in Poland. These changes are presented in Figure 2. Over the course of ten years, the percentage share of installed capacity from onshore wind farms and solar power plants increased more than tenfold. This represents a significant achievement in Poland's pursuit of sustainable development through the use of renewable energy sources.

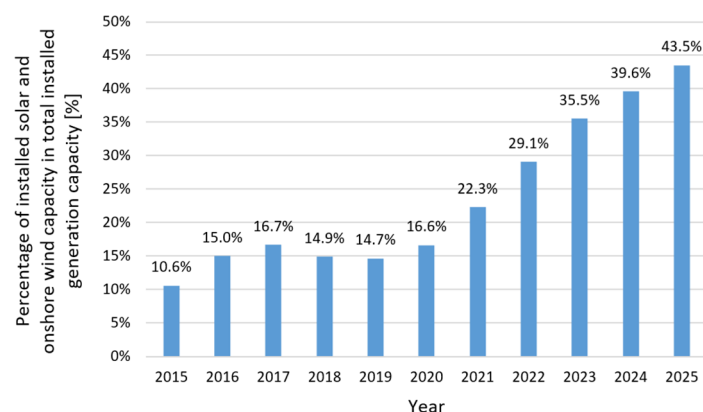


Figure 2. The dynamics of the changes in the percentage share of installed capacity from onshore wind farms and solar power plants relative to total installed capacity in Poland from 2015 to 2025.

The challenges and issues related to the development of wind and solar energy in Poland have been extensively discussed in numerous scientific articles. The study in article [17] provides a comprehensive overview of recent advancements and challenges in the development of renewable energy in Poland. The paper analyzes statistical data and policy documents to evaluate the growth dynamics of the wind and photovoltaic sectors between 2015 and 2023. The paper emphasizes the significant increase in installed PV capacity and the steady expansion of onshore wind power, driven mainly by the European Union (EU) climate policy and domestic support mechanisms. This research highlights barriers such as grid congestion, administrative delays, and also limited energy storage capacity. The conclusion is that further progress requires modernization of transmission infrastructure and greater integration of forecasting tools into national energy planning. However, article [18] investigates the regional potential for renewable energy development across Poland's sixteen voivodeships. Using a multi-criteria assessment framework, authors evaluate environmental, economic and infrastructural conditions influencing the deployment of various renewable technologies. The research reveals substantial differences between western and eastern regions, especially in access to transmission networks, invest-

ment attractiveness and local policy support. The article also identifies solar and biomass energy as the most promising technologies for decentralized generation in rural areas. The authors recommend regionally tailored strategies to ensure balanced renewable energy growth and alignment with Poland's sustainable development goals.

The paper [19] investigates the key determinants influencing the rapid expansion of photovoltaic energy in Poland. The authors describe legal regulations, environmental conditions, investment incentives and social acceptance as the main drivers of PV growth in period 2015–2023. The analysis shows the pivotal role of the “Mój Prąd” subsidy program, net metering mechanisms and also decreasing technology costs. However, it also identifies grid limitations and administrative procedures as very significant barriers. The research concludes that long-term policy stability and improvements in grid infrastructure are essential for maintaining dynamic PV growth in Poland.

However in article [20] there is detailed assessment of the development of the wind energy sector in Poland and focusing primarily on onshore installations. The authors analyze capacity expansion trends, geographical distribution of wind farms and also their impact on the national transmission and distribution grid. Simulation results reveal that high regional concentration of wind capacity in the northern part of Poland creates grid stability challenges and congestion during high-wind periods. The research also shows the potential integration of offshore wind projects into the Polish power system. The authors emphasize that coordinated grid expansion and improved forecasting systems are necessary to ensure system reliability alongside the continued growth of wind power in Poland.

3. Data

3.1. Statistical Data Analysis

The power generation data used from all onshore wind farms and all solar power plants in Poland originate from [21]. The time series cover the period from 10 April 2020 to 23 October 2025. For this time span, both analyzed time series were available.

The variability of power generation from all wind farms in Poland for hourly values from 10 April 2020 to 23 October 2025, is presented in Figure 3. Meanwhile, Figure 4 illustrates the variability of power generation from all solar power plants in Poland for the same time period.

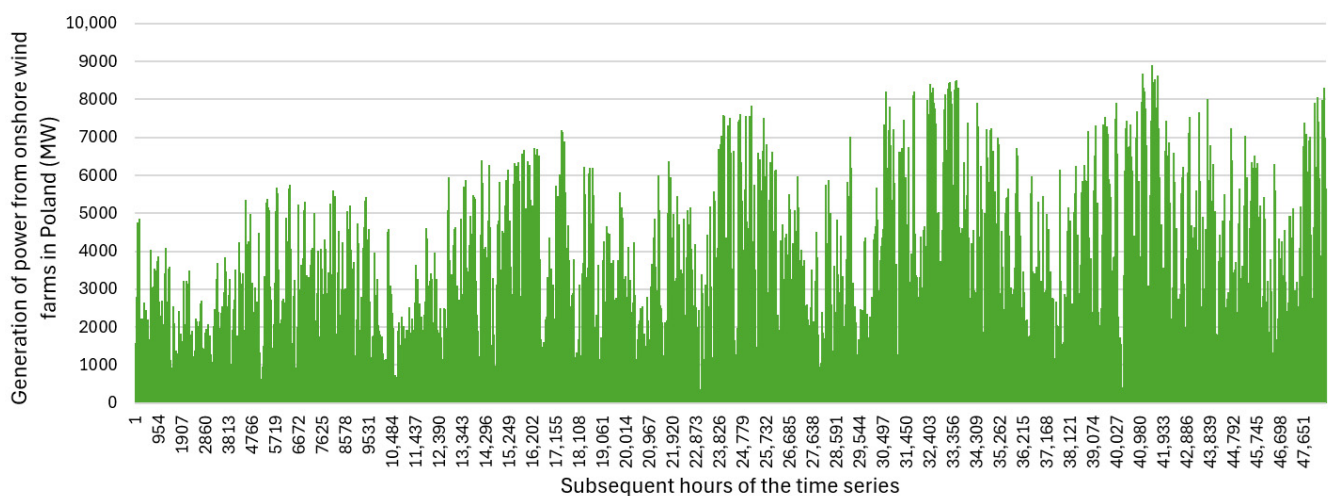


Figure 3. The variability of power generation from all wind farms in Poland for hourly values from 10 April 2020 to 23 October 2025.

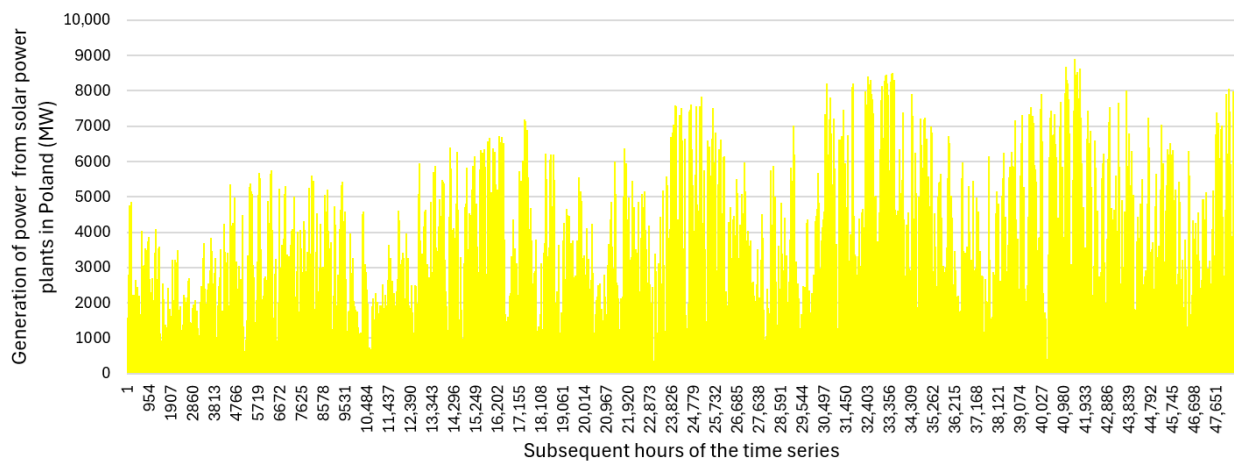


Figure 4. The variability of power generation from all solar power plants in Poland for hourly values from 10 April 2020 to 23 October 2025.

Both Figures 3 and 4 clearly show a significant increase in the process, which results from the growth of total installed capacity for both renewable energy sources between 2020 and 2025, as shown in Figure 1. Power generation from solar power plants exhibits strong seasonality, a phenomenon that is clearly visible in Figure 4 (with maximum generation occurring in the summer and near-summer months).

Figure 5 presents the average hourly power generation values in a given month from all wind farms and solar power plants in Poland, along with the total power generation from both renewable energy sources for the subsequent months of the time series from April 2020 to October 2025. The total power generation shows an upward trend over the analyzed period, with minor oscillations. Power generation from wind farms and solar power plants also exhibit an increasing trend, but in both cases, a strong seasonal pattern is clearly visible. This is further illustrated in Figure 6, which shows the calculated average hourly power generation values for both renewable energy sources for each month of the year separately, based on data from April 2020 to October 2025. In Figure 6, in addition to the pronounced seasonality of both processes, their differences are also evident—solar power generation peaks during the summer and late spring months, while wind power generation is at its lowest during the same period of time. The calculated Pearson's linear correlation coefficient (R) between the two analyzed processes is -0.919 , which suggests the inverse annual variability of power generation from wind farms and solar power plants. However, the variability in power generation across the months of the year is much greater for solar power plants than for wind farms. The author is aware that calculating R using monthly mean values significantly overestimates the strength of the correlation.

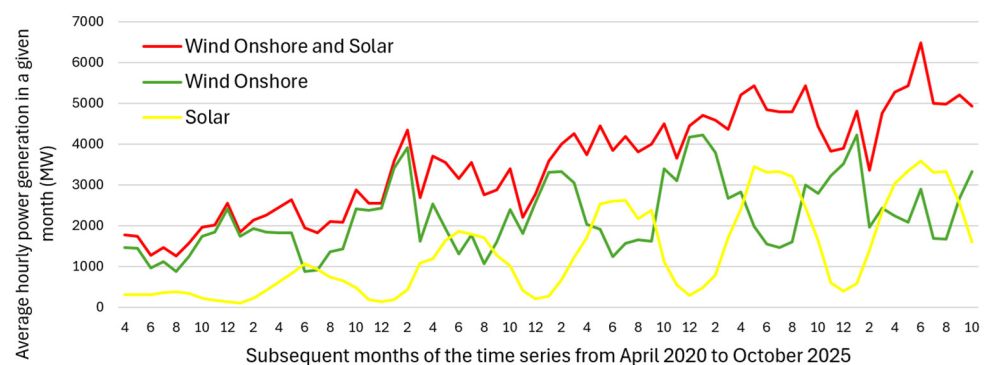


Figure 5. The average hourly power generation values in a given month from all wind farms and solar power plants in Poland from 10 April 2020 to 23 October 2025.

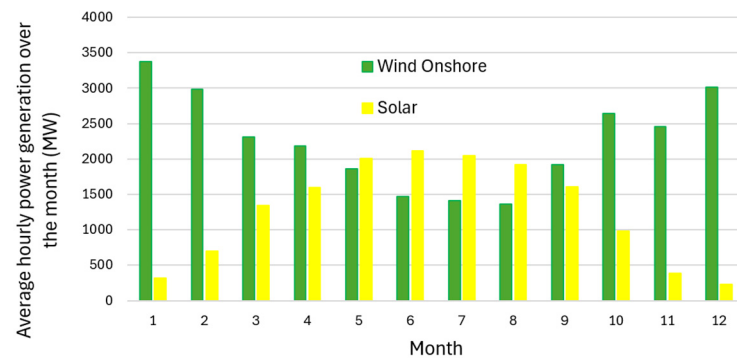


Figure 6. The average hourly power generation values for both renewable energy sources for each month of the year.

As the first step the analysis was conducted to verify which lagged values of the fore-casted time series of power generation (wind farms, solar power plants and total both renewable energy sources) could potentially serve as valuable input data for predictive models. For this purpose, R values for the power generation time series up to 4 days backward (96 lagged 1 h values) were calculated. Figure 7 shows the results of calculations of autocorrelation function (ACF) of power generation.

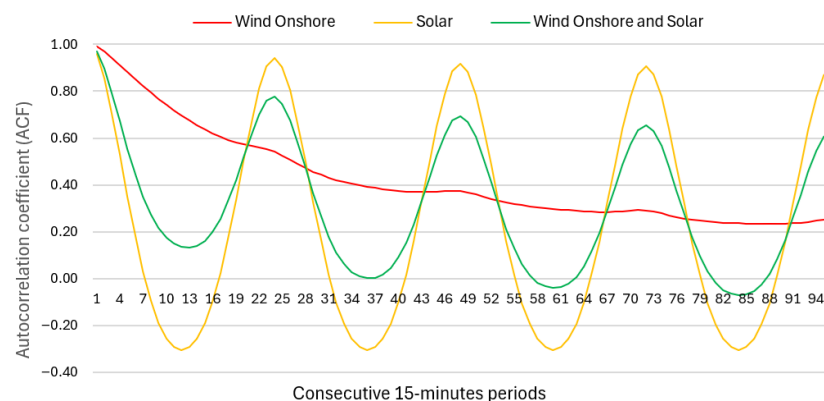


Figure 7. Autocorrelation function of the three time series of power generation up to 4 days back.

As shown in Figure 7, valuable input data for forecasting models consists of the lagged values of the time series by multiples of 24 h. This should be considered when selecting input data for forecasting models. In the case of power generation from wind farms, no such dependence exists—only the most recent lagged values of the time series may be relevant as input data for forecasting models. On the other hand, the time series representing the combined generation from both renewable energy sources exhibits cyclic increases in Autocorrelation Function (ACF) values on a daily basis. This is similar to the time series of power generation from solar plants, although the ACF values are slightly smaller.

3.2. Feature Engineering

In time series forecasting, a standard procedure involves using selected lagged values of the predicted process as input data, most often based on the Pearson correlation coefficient (R)—the larger the absolute value, the stronger the linear relationship. Enhancing the forecasting model with additional information is generally considered beneficial. Due to the seasonality of generation from both renewable energy sources and their diurnal variability, two additional input features are introduced: MONTH (the forecast month) and HOUR (the forecast hour). To reduce the “noise” present in the analyzed time series, a weighted averaging of the time series was applied. The smoothed time series values of

power generation were calculated using Equation (1) [22]. The weights (0.5, 0.3, 0.2) were selected based on the methodology proposed in [22] and empirical verification, assigning the highest importance to the most recent historical values to effectively dampen stochastic noise while retaining trend information. In the forecasting models, the smoothed value of power generation serves as an additional input variable from the T-1 period.

$$PG_{t-1}^{smoothed} = PG_{t-2} \cdot w_{t-2} + PG_{t-3} \cdot w_{t-3} + PG_{t-4} \cdot w_{t-4}, \quad (1)$$

where $PG_{t-1}^{smoothed}$ is the smoothed value of power generation for period $t - 1$, PG_{t-k} is the value of the power generation for period $t - k$, and w_k are weights ($w_{t-2} = 0.5$, $w_{t-3} = 0.3$ and $w_{t-4} = 0.2$).

The power generation process of solar power plants is limited, within the 24 h cycle, to the period between sunrise and sunset. Therefore, an additional marker—DAYLIGHT—is introduced into the forecasting model as an input variable, indicating whether power generation can occur (marker value = 1) or is not possible (marker value = 0).

3.3. Input Data Selection and Dataset Partitioning

In this study, forecasting models are proposed, and separate one-hour-ahead forecasts are generated for the following time series: wind power generation (WPG), solar power generation (SPG), and the combined wind-and-solar power generation (WSPG). The purpose of performing forecasts for the combined generation from both renewable energy sources is to assess the validity of decomposing the forecasting task, i.e., conducting separate forecasts for each individual renewable energy source. In reference [23], it was demonstrated that when the shared information between wind and PV generation is low, the uncertainty should be modeled separately and forecasted using independent models, whereas when the shared information is high, a multi-task (joint) forecasting model is more appropriate.

For the WPG, SPG, and WSPG series, six lagged values were selected for the models. Their statistical significance was verified ($p < 0.05$). The ACF analysis revealed that the first six lags of the time series were statistically significant ($p < 0.05$). Similarly, the partial autocorrelation function (PACF) was analyzed for the first six lags. In all cases, these values were found to be statistically significant ($p < 0.05$).

Figure 8 presents the complete set of input data used in the forecasting models for one-hour-ahead wind power generation, together with the calculated values of R between each input variable and the output variable, i.e., the wind power generation at time T (WPG_T). Despite the very low R values for the input variables MONTH and HOUR, both may still be useful since machine learning models are capable of capturing non-linear dependencies, whereas R assesses only linear relationships.

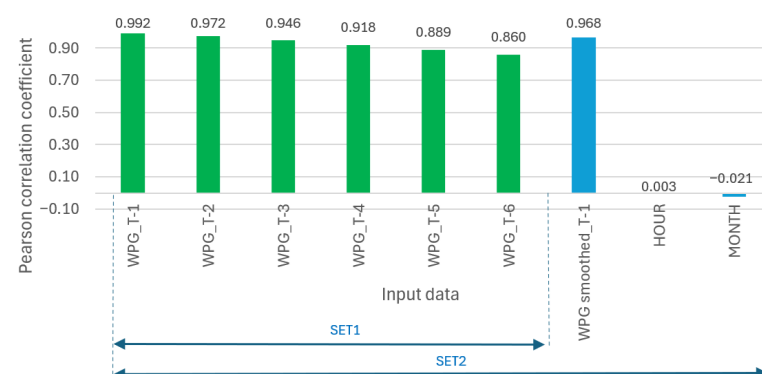


Figure 8. Pearson linear correlation coefficients for the complete set of input data used in wind power generation forecasting.

Figure 9 presents the complete set of input data used in the forecasting models for one-hour-ahead solar power generation, along with the calculated values of the Pearson correlation coefficient between each input variable and the output variable, i.e., the solar power generation at time T (SPG_T). In this case, a pronounced influence of the lagged values of the generation process, specifically those from 1, 2, and 3 days earlier, can be clearly observed.

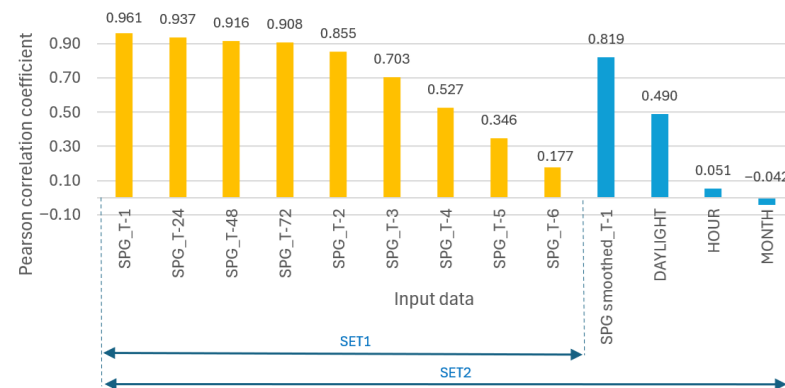


Figure 9. Pearson linear correlation coefficients for the complete set of input data used in solar power generation forecasting.

Figure 10 presents the complete set of input data used in the forecasting models for one-hour-ahead wind and solar power generation, together with the calculated values of the Pearson correlation coefficient between each input variable and the output variable, i.e., the combined wind and solar power generation at time T (WSPG_T). In this case, the lagged values of the process from 1, 2, and 3 h earlier, as well as the smoothed value from period T-1, show the strongest influence. The lagged values from 1, 2, and 3 days earlier exhibit a slightly weaker correlation, which results from the fact that the analyzed time series represents a combined generation process from two renewable energy sources.

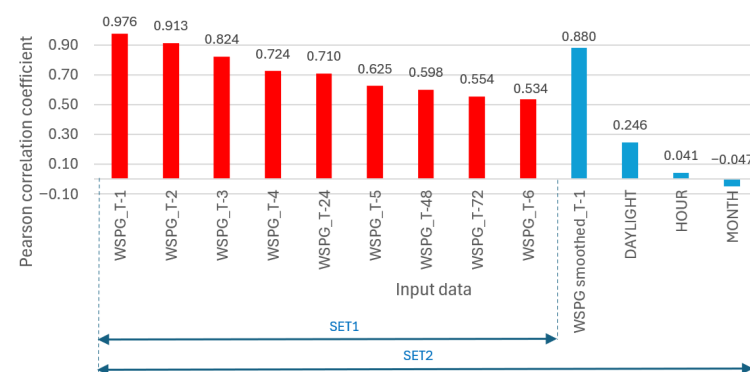


Figure 10. Pearson linear correlation coefficients for the complete set of input data used in wind and solar power generation forecasting.

The forecasts are performed in two variants for the WPG, SPG, and WSPG processes: one using all available information (SET2) and the other using only the lagged values of the predicted process, i.e., without additional input variables generated through feature engineering (SET1). The forecasting experiments conducted using input data SET1 and input data SET2 aim to verify the effectiveness and justification of incorporating additional features (feature engineering) beyond the lagged values of the predicted time series.

The three time series cover the period from 10 April 2020 to 23 October 2025, comprising a total of 48,651 samples. The available samples for each time series were divided into a training subset (approximately 70%), a validation subset (approximately 15%), and a

test subset (approximately 15%). The division was implemented using repeating 35-day cycles with an identical structure: 25 days of data for training, 5 days for validation, and 5 days for testing. Figure 11 illustrates the block-based partitioning of data into training, validation, and test sets. This approach ensures that each model is trained on data that are representative of the entire multi-year time series, capturing both the seasonality and the diurnal variability of the process. Moreover, this partitioning guarantees representative data for the validation subset (for proper hyperparameter tuning) as well as for the test subset (for the final assessment of model accuracy). Unlike conventional chronological splitting, which might bias the evaluation towards a specific season (e.g., testing only on autumn months), the applied block-based strategy ensures that the models are evaluated across diverse meteorological conditions representing the entire year. To strictly prevent data leakage, hyperparameter optimization processes were computed exclusively using the training subsets.

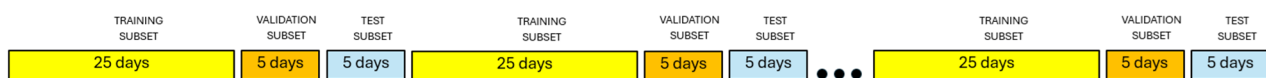


Figure 11. Block-based partitioning of data into training, validation, and test subsets.

4. Methods

4.1. Advanced Ensemble Machine Learning Models

Below is a short description of the proposed forecasting methods from the baseline model (reference model) to the most complex ensemble models. The study utilized two different Random Forest models (homogeneous ensembles), five different Gradient Boosting Decision Tree models (homogeneous ensembles), and five different ensemble models (heterogeneous ensembles) that integrated the forecasts obtained from the Random Forest and the Gradient Boosting Decision Tree models in various ways.

4.1.1. Naive Method

In the analyzed forecasting problem, the reference model against which the machine learning models are evaluated is the standard Naïve (naive method). Despite its simplicity, this method is widely recognized in the literature as a robust benchmark for short-term renewable energy forecasting, particularly for very short horizons where meteorological inertia is high. For wind power generation, solar power generation and total power generation from wind farms and photovoltaic plants, the naive model is defined by Equation (2). In this case, the forecasted value corresponds to the most recently observed power generation measurement.

$$\hat{y}_t = y_{t-1} \quad (2)$$

where $\hat{y}_t$ —forecast of power generation for the 1 h period t and y_{t-1} —power generation for period lagged by $t - 1$ from forecast period t .

4.1.2. Random Forest Regression

RF (Random Forest) is a homogeneous ensemble model where all constituent models are decision trees. In the regression task, the prediction in a single decision tree is the mean target value of all instances associated with the leaf node. The final prediction is the average value of all single decision trees. Random Forests utilize deep, low-bias trees; the ensemble process effectively reduces the overall variance of the model. The regularization hyperparameters generally include: the maximum depth of a single decision tree, the minimum number of data points required in a node before the node is split, the minimum number of data points allowed in a leaf node, and the maximum number of nodes. The selection of predictors for each of the single decision trees is made by randomly choosing

from all available predictors [24,25]. The problem of overfitting in this case is primarily mitigated by the decorrelation effect achieved through the ensemble process.

In the conducted research aimed at finding the appropriate RF hyperparameters, various values were tested for the following parameters:

- the number of trees,
- the number of features (predictors) in an individual tree,
- the minimum number of samples required to form a leaf node,
- the maximum depth (levels),
- the minimum number of samples required to split a node,
- the maximum number of nodes.

4.1.3. Extremely Randomized Trees

ET (Extremely Randomized Trees) model differs from the standard RF in that, besides randomly selecting a subset of features, it also randomly selects the split threshold for each node, instead of calculating the optimal threshold. This additional randomness often leads to faster training and can help reduce the model's variance. The tested hyperparameters were identical to those used for the RF model.

4.1.4. Gradient Boosting Decision Tree

GBDT (Gradient Boosting Decision Tree) is the oldest form of Gradient Boosting. GBDT is an ensemble technique that builds trees sequentially, where each new decision tree is trained to correct the prediction errors (residuals) of the preceding sequence of trees. It leverages the mathematical concept of gradient descent to minimize the loss function by iteratively moving in the direction of the steepest descent in the function space [26].

4.1.5. eXtreme Gradient Boosting

XGBoost (eXtreme Gradient Boosting) is the highly optimized and scalable implementation of the Gradient Boosting framework. It is known for its speed and performance due to parallel processing and efficient handling of sparse data, often dominating machine learning competitions. Model introduces advanced regularization techniques (L1 and L2) to prevent overfitting, which contributes to its robust predictive power compared to the original, oldest GBDT implementation [27,28].

4.1.6. Light Gradient Boosting Machine

LightGBM (Light Gradient Boosting Machine) is efficient implementation of the GBDT model. Its primary innovation is the use of a histogram-based decision tree algorithm and a leaf-wise (depth-limited) growth strategy, as opposed to the level-wise approach used by XGBoost or GBDT. These optimizations make LightGBM significantly faster and require much less memory, especially when dealing with large datasets, making it highly preferred for large-scale data applications [29,30].

4.1.7. Histogram-Based Gradient Boosting Regressor

HGBR (Histogram-based Gradient Boosting Regressor) is implementation of the Gradient Boosting framework, heavily inspired by LightGBM's model. It achieves efficiency by replacing the costly sorting of continuous feature values with a histogram-based algorithm, which bins features into a limited number of discrete values. This binning significantly speeds up the search for optimal split points, leading to much faster training times and better scalability compared to the original oldest GBDT [31].

4.1.8. Categorical Boosting

CatBoost (Categorical Boosting) is high-performance gradient boosting library developed by Yandex. Its key distinguishing feature, particularly valuable in regression tasks where feature types are mixed, is its native and efficient handling of categorical features using a novel algorithm called Ordered Target Encoding, which avoids prediction shift caused by traditional target encoding. Additionally, CatBoost utilizes a symmetric tree structure to minimize overfitting and employs a unique permutation-based strategy to avoid bias in gradient estimation making it a robust and reliable choice [32,33].

Summarizing the information about Gradient Boosting models, especially for forecasting problems, LightGBM is often the most advantageous model due to its exceptional speed and scalability, allowing for rapid experimentation and handling of big data [34]. Examples of model applications in forecasting energy generation from RES are described in [35,36]. However GBDT is the least advantageous choice, as its slow training and high resource consumption are outperformed by all modern histogram-based variants (LightGBM and HGBR). CatBoost is the best choice when the dataset includes many raw categorical features that do not require manual preprocessing.

In the conducted research aimed at finding the appropriate Gradient Boosting hyperparameters, various values were tested for the following parameters:

- Learning rate;
- The number of trees;
- The minimum number of samples required to form a leaf node;
- The maximum depth (levels);
- The minimum number of samples required to split a node;
- The maximum number of nodes.

4.1.9. Arithmetic Mean Forecasts Integrator

INT_AVE (Arithmetic Mean Forecasts Integrator) computes the final forecast for each period T as the arithmetic mean of the individual model forecasts. The ensemble exclusively considers the best-performing homogeneous model among the seven tested models, along with those models that achieved a nMAE (Normalized Mean Absolute Error) only slightly higher than the best homogeneous model. As the selection criterion for the INT_AVE ensemble, a threshold was adopted: models with an nMAE error up to 10% greater than the error of the best homogeneous model were included.

4.1.10. No Extreme Forecasts Integrator

INT_NO_EXTREMES (No Extreme Forecasts Integrator) computes the final forecast for each period T by excluding the single highest and single lowest forecast values from the selected homogeneous models (nMAE error up to 10% greater than the error of the best homogeneous model). This exclusion process is repeated for every period T , meaning that INT_NO_EXTREMES does not always utilize the same set of homogeneous models. Consequently, INT_NO_EXTREMES represents a unique ensemble model with a “switching” set of constituent models (selecting a set out of the seven available models). The final prediction result is calculated by Equation (3).

$$\hat{y}_i = \frac{1}{n-2} \cdot \left(\sum_{k=1}^s \hat{y}_i^k - \min\{\hat{y}_i^k\} - \max\{\hat{y}_i^k\} \right) \quad (3)$$

where i is the prediction point, $\hat{y}_i$ is the final predicted value, $\hat{y}_i^k$ is the predicted value by predictor number k , and n is the number of predictors in the ensemble before the elimination of the outputs of predictors yielding extreme forecasts from the set of results.

4.1.11. Expertly Weighted Mean Forecasts Integrator

INT_WEIGHT_AVE (Expertly Weighted Mean Forecasts Integrator) computes the final forecast for each period T as a special weighted mean of the selected homogeneous models (with a nMAE error up to 10% greater than the error of the best homogeneous model). The algorithm stipulates that, after model selection for the ensemble, the forecast with the lowest value and the forecast with the highest value in period T are assigned a weight of 0.05. The remaining forecasts are assigned equal weights. Consequently, INT_WEIGHT_AVE also represents a unique ensemble model with a “switching” set of constituent models (selecting a specific set out of the seven available models).

4.1.12. Median Forecasts Integrator

INT_MEDIAN (Median Forecasts Integrator) computes the final forecast for each period T as the median of the selected homogeneous models (with an Normalized Mean Absolute Error (nMAE) error up to 10% greater than the error of the best homogeneous model). The primary advantage of using the median over the arithmetic mean is its robustness against outliers. This feature makes INT_MEDIAN a valuable tool for creating stable and reliable forecasts, especially in noisy environments.

4.1.13. Average of Extremes Forecasts Integrator

INT_AVE_EXT (Average of Extremes Forecasts Integrator) calculates the final forecast for period T based on the selected homogeneous models (with an nMAE error up to 10% greater than the error of the best homogeneous model) by computing the mid-range (the mean value of the highest and lowest forecasts) for period T . This expert approach is validated through verification, as it essentially relies solely on the arithmetic mean calculated from the two most extreme forecasts. The sequential steps of the performed research are shown in Figure 12.

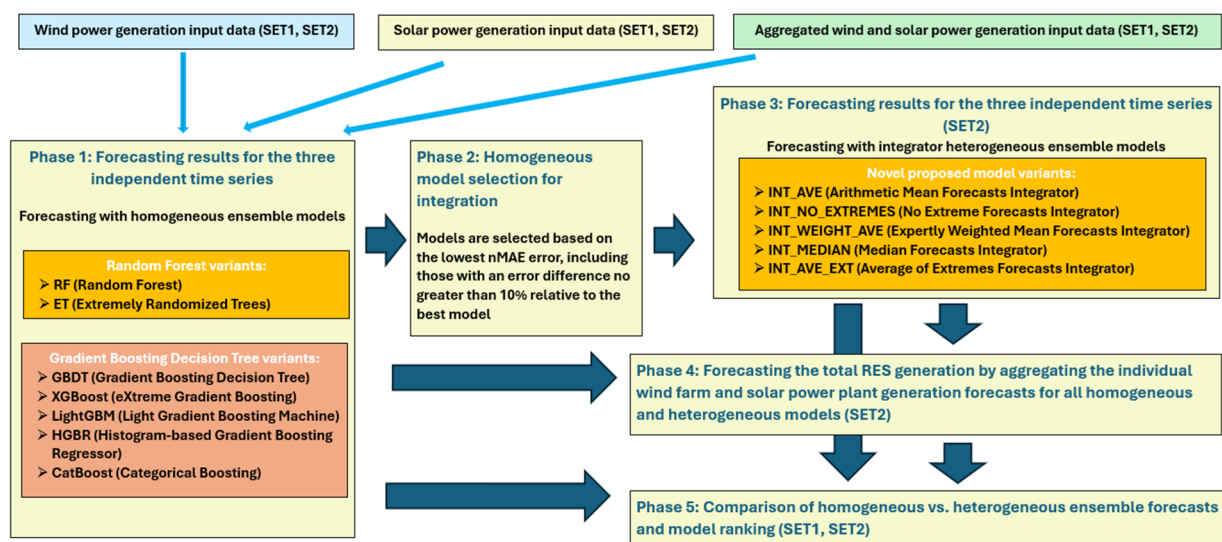


Figure 12. The workflow of the research stages.

4.2. Evaluation Metrics

Normalized Mean Absolute Error (nMAE) is determined by Equation (4). Due to the zero values which occur in the power generation time series, it is impossible to use the popular Mean Absolute Percentage Error (MAPE) error [37]. Given that the forecasted time series exhibit an increasing trend, the use of MAE without special normalization would be incorrect, as the Absolute Errors (AE) would systematically increase along with the rising generation levels. To address this issue, a time-varying normalizing factor $g_{norm}^{Q(s)year}$ was

applied. The normalizing factor is defined as the mean hourly power generation, computed using all hourly data points for all consecutive quarters (3-month periods) in each year. In total, twenty-three distinct normalizing factors with increasing values were applied, corresponding to each individual quarter.

$$nMAE = \frac{1}{n} \sum_{i=1}^n \frac{1}{g_{norm}^{Q(s)year}} |y_i - \hat{y}_i|, \quad (4)$$

where $g_{norm}^{Q(s)year}$ is the normalizing factor, s denotes the quarter index within a specific *year*, y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of prediction points.

The Relative Normalized Mean Absolute Error (RelnMAE) is defined as the ratio between the $nMAE$ of the model under evaluation and the $nMAE_{naive}$ of a benchmark method (naive model)—Equation (5). It facilitates the assessment of model performance against a fixed baseline. The $RelnMAE$ metric exhibits an inverse relationship with the model's quality. Specifically, a $RelnMAE$ value exceeding 1 signifies that the evaluated method performs worse than the reference baseline (e.g., the naive model).

$$RelnMAE = \frac{nMAE}{nMAE_{naive}}, \quad (5)$$

The Normalized Root Mean Square Error (nRMSE) is calculated by Equation (6). The nRMSE measure is typically used for power generation forecasts from RES [38,39].

$$nRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{g_{norm}^{Q(s)year}} \right)^2}, \quad (6)$$

The Normalized Mean Bias Error (nMBE) captures the average bias in the prediction and is defined by Equation (7). The nMBE indicates whether the forecasts are overestimated (nMBE < 0) or underestimated (nMBE > 0) [40].

$$nMBE = \sum_{i=1}^n \frac{1}{g_{norm}^{Q(s)year}} (y_i - \hat{y}_i) \quad (7)$$

To summarize this section, the nMAE error was adopted as the primary performance metric for the study for quality assessment purposes. The secondary error metric utilized is nRMSE. RelnMAE and nMBE, however, serve only as supporting (or auxiliary) error metrics.

5. Experimental Results

The research successfully obtained forecasting results for four distinct tasks. Table 1 presents the forecasts for the wind farm power generation time series, Table 2 details the forecasts for the solar power plant generation time series, and Table 3 shows the results for the time series of combined wind and solar power generation (forecasted jointly). Conversely, Table 4 displays the forecasts for the total combined wind and solar power generation, which were calculated as the sum of the individual wind and solar power generation forecasts (post-aggregation). The nMAE (%) error was selected as the primary error measure, and within the tables, the lowest error for each metric is highlighted in blue, while the highest error is highlighted in red. These results specifically reflect the performance of the optimized models after hyperparameter tuning and are presented for both input data sets: SET1 (utilizing only selected lagged values of the forecasted time series) and SET2 (employing the full set of available input data).

Table 1. Results of total wind farm power generation forecasts for all analyzed models, utilizing the SET1 and SET2 input data variants and tuned hyperparameters, sorted by the ascending nMAE error metric on the test subset.

Model Name	Input Dataset	nMAE (%)	nRMSE (%)	relnMAE	nMBE [%]	nMAE Improvement vs. Naive Method (%)
INT_MEDIAN	SET2	4.8353	6.7939	0.64876	0.04920	35.124
INT_NO_EXTREMES	SET2	4.8415	6.8170	0.64959	0.06020	35.041
INT_WEIGHT_AVE	SET2	4.8442	6.8230	0.64995	0.06060	35.005
INT_AVE	SET2	4.8555	6.8455	0.65147	0.06181	34.853
LightGBM	SET2	4.8734	6.8132	0.65386	0.05581	34.614
INT_AVE_EXT	SET2	4.8913	6.9115	0.65627	0.06422	34.373
XGBoost	SET2	4.9160	6.8824	0.65959	0.07375	34.041
HGBR	SET2	4.9256	6.9122	0.66087	0.07188	33.913
RF	SET2	5.1251	7.2641	0.68764	0.03698	31.236
LightGBM	SET1	5.1376	7.2482	0.68932	0.01065	31.068
HGBR	SET1	5.1649	7.2433	0.69298	0.02681	30.702
XGBoost	SET1	5.1685	7.2897	0.69347	0.01814	30.653
RF	SET1	5.1797	7.3466	0.69497	0.04016	30.503
ET	SET2	5.2463	7.4400	0.70391	0.05983	29.609
ET	SET1	5.2845	7.4986	0.70902	0.06093	29.098
GBDT	SET2	5.3860	7.5495	0.72265	0.06625	27.735
GBDT	SET1	5.4735	7.7149	0.73438	0.06122	26.562
CatBoost	SET1	6.0375	8.4848	0.81006	0.05183	18.994
CatBoost	SET2	6.0795	8.4948	0.81570	0.03447	18.430
Naive	-	7.4532	10.4959	1.00000	0.01309	-

Table 2. Statistical significance of forecast accuracy: Diebold–Mariano test results for total wind power generation.

Best Model (INT_MEDIAN)	DM Statistic (S)	p-Value	Significance
INT_NO_EXTREMES	−1.0116	0.3117	ns
LightGBM	−1.7407	0.0817	ns
XGBOOST	−3.4660	0.0005	***

Notes: *** $p < 0.001$, ns—not significant.

The results presented in Table 1 unequivocally indicate the superiority of the ensemble integrator models. The nMAE improvement versus the naive method reaches 35.1% for the best-performing model. Furthermore, the differences in the nMAE error are negligible among the top eight best models. The considerable nMAE error recorded by the CatBoost models is surprising. The Random Forest models achieved average results, with the RF model proving superior to the ET model.

To statistically validate the predictive superiority of the best model (INT_MEDIAN), the Diebold–Mariano (DM) test with a MAE loss function was conducted for selected models (including the runner-up and the best-performing non-ensemble model) on the test dataset. The results (Table 2) indicate that the DM statistic for the XGBOOST model is negative and statistically significant ($p < 0.001$). This allows for the rejection of the null hypothesis of equal forecast accuracy, confirming that the INT_MEDIAN model provides significantly more accurate forecasts than the XGBOOST model. For the INT_NO_EXTREMES and the LightGBM model, the difference in forecast accuracy was not statistically significant.

Table 3. Results of total solar power plants generation forecasts for all analyzed models, utilizing the SET1 and SET2 input data variants and tuned hyperparameters, sorted by the ascending nMAE error metric on the test subset.

Model Name	Input Dataset	nMAE (%)	nRMSE (%)	relnMAE	nMBE [%]	nMAE Improvement vs. Naive Method (%)
INT_WEIGHT_AVE	SET2	7.2631	14.2536	0.25807	−0.00406	74.193
INT_MEDIAN	SET2	7.2638	14.2519	0.25809	−0.00595	74.191
INT_NO_EXTREMES	SET2	7.2638	14.2519	0.25809	−0.00595	74.191
INT_AVE	SET2	7.2795	14.3056	0.25865	0.00352	74.135
INT_AVE_EXT	SET2	7.3570	14.4715	0.26140	0.01299	73.860
LightGBM	SET2	7.3962	14.4851	0.26280	−0.07566	73.720
XGBoost	SET2	7.4748	14.7507	0.26559	−0.02189	73.441
HGBR	SET2	7.6968	14.9700	0.27347	−0.04532	72.653
RF	SET2	8.1254	15.9606	0.28870	0.15695	71.130
LightGBM	SET1	8.3947	16.5555	0.29827	−0.23865	70.173
HGBR	SET1	8.4286	16.5470	0.29948	−0.33779	70.052
XGBoost	SET1	8.4985	16.6831	0.30196	−0.13568	69.804
ET	SET2	8.5427	16.9016	0.30353	0.95550	69.647
RF	SET1	8.6382	17.0083	0.30692	−0.12749	69.308
ET	SET1	9.3386	18.4200	0.33181	0.77824	66.819
GBDT	SET2	9.7191	18.3692	0.34533	0.51411	65.467
CatBoost	SET2	10.0769	19.1394	0.35804	0.11451	64.196
GBDT	SET1	10.7664	20.7286	0.38254	0.51601	61.746
CatBoost	SET1	10.9013	20.6889	0.38733	0.08260	61.267
naive	-	28.1444	46.8422	1.00000	−0.00002	-

Table 4. Statistical significance of forecast accuracy: Diebold–Mariano test results for total solar power plants generation.

Best Model (INT_WEIGHT_AVE)	DM Statistic (S)	p-Value	Significance
INT_MEDIAN	−0.2330	0.8158	ns
INT_AVE_EXT	−2.8157	0.0049	**
LightGBM	−3.3109	0.0009	***

Notes: *** $p < 0.001$, ** $p < 0.01$, ns—not significant.

The results presented in Table 3 indicate also the superiority of the ensemble integrator models. The nMAE improvement versus the naive method reaches 74.2% for the best-performing model. This suggests that the improvement (nMAE error) achieved by forecasting models for total solar power plant generation, relative to the naive model, is more than double that achieved for total wind farm power generation. Furthermore, the differences in the nMAE error are negligible among the top eight best models. The considerable nMAE error recorded by the CatBoost models is a surprising—very similar error to the GBDT models. The Random Forest models achieved average results, with the RF model proving superior to the ET model.

To statistically validate the predictive superiority of the best model (INT_WEIGHT_AVE), the Diebold–Mariano (DM) test with a MAE loss function was conducted for selected models (including the runner-up and the best-performing non-ensemble model) on the test dataset. The results (Table 4) indicate that the DM statistics for the INT_AVE_EXT and LightGBM models are negative and statistically significant ($p < 0.01$ and $p < 0.001$, respectively). This allows for the rejection of the null hypothesis of equal forecast accuracy, confirming that the INT_WEIGHT_AVE model provides significantly more accurate forecasts than the INT_AVE_EXT and LightGBM models. For the INT_MEDIAN model, the difference in forecast accuracy was not statistically significant.

The results presented in Table 5 also show the superiority of the ensemble integrator models. The nMAE improvement versus the naive method reaches 54.5% for the best-performing model. Furthermore, the differences in the nMAE error are negligible among the top five best models (all ensemble integrator models). The Random Forest models achieved average results, with the RF model proving superior to the ET model.

Table 5. Results of total power generation forecasts (integrated as a single time series) for wind and solar power plants across all analyzed models, utilizing SET1 and SET2 input data variants with tuned hyperparameters, sorted by ascending nMAE error on the test subset.

Model Name	Input Dataset	nMAE (%)	nRMSE (%)	relnMAE	nMBE [%]	nMAE Improvement vs. Naive Method (%)
INT_AVE	SET2	4.6326	6.3955	0.45527	0.00979	54.473
INT_WEIGHT_AVE	SET2	4.6489	6.4122	0.45688	0.00673	54.312
INT_AVE_EXT	SET2	4.6526	6.4243	0.45724	0.01363	54.276
INT_NO_EXTREMES	SET2	4.6570	6.4219	0.45767	0.00596	54.233
INT_MEDIAN	SET2	4.6570	6.4219	0.45767	0.00596	54.233
XGBoost	SET2	4.7367	6.5476	0.46550	0.06251	53.450
LightGBM	SET2	4.7781	6.5487	0.46957	0.06280	53.043
HGBR	SET2	4.7886	6.5804	0.47060	−0.02147	52.940
RF	SET2	5.1034	7.0071	0.50154	−0.06466	49.846
ET	SET2	5.2848	7.2306	0.51937	−0.10959	48.063
XGBoost	SET1	5.4314	7.5017	0.53378	−0.08582	46.622
GBDT	SET2	5.4483	7.4269	0.53544	0.12434	46.456
LightGBM	SET1	5.4588	7.5408	0.53647	−0.11361	46.353
RF	SET1	5.5268	7.6241	0.54315	−0.12735	45.685
HGBR	SET1	5.5325	7.6023	0.54371	−0.14937	45.629
CatBoost	SET2	5.7357	7.8303	0.56368	0.05845	43.632
ET	SET1	5.7999	7.9617	0.57000	−0.14918	43.000
GBDT	SET1	6.1797	8.4558	0.60731	−0.05385	39.269
CatBoost	SET1	6.6512	9.0402	0.65365	−0.06915	34.635
naive	-	10.1754	14.8599	1.00000	0.03356	-

To statistically validate the predictive superiority of the best model (INT_AVE), the Diebold–Mariano (DM) test with a MAE loss function was conducted for selected models (including the runner-up and the best-performing non-ensemble model) on the test dataset. The results (Table 6) indicate that the DM statistics for the INT_WEIGHT_AVE, INT_MEDIAN and XGBoost models are negative and statistically significant ($p < 0.05$, $p < 0.05$ and $p < 0.001$, respectively). This allows for the rejection of the null hypothesis of equal forecast accuracy, confirming that the INT_AVE model provides significantly more accurate forecasts than the INT_WEIGHT_AVE, INT_MEDIAN and XGBoost models.

Table 6. Statistical significance of forecast accuracy: Diebold–Mariano test results for total power generation forecasts (integrated as a single time series).

Best Model (INT_AVE)	DM Statistic (S)	p-Value	Significance
INT_WEIGHT_AVE	−2.0169	0.0437	*
INT_MEDIAN	−2.4507	0.0143	*
XGBoost	−4.4420	0.0000	***

Notes: *** $p < 0.001$, * $p < 0.05$, ns—not significant.

The results presented in Table 7 show also the superiority of the ensemble integrator models. The nMAE improvement versus the naive method reaches 58.2% for the best-performing model. Furthermore, the differences in the nMAE error are negligible among the top seven models.

Table 7. Time series forecast results for combined wind farm power and solar power plant generation, calculated as the sum of individual wind and solar forecasts, across all analyzed models. The results utilize the SET1 and SET2 input data variants and tuned hyperparameters, and are sorted by the ascending nMAE error metric on the test subset.

Model Name	Input Dataset	nMAE (%)	nRMSE (%)	relnMAE	nMBE [%]	nMAE Improvement vs. Naive Method (%)
INT_MEDIAN	SET2	4.2607	5.8880	0.41872	0.02438	58.128
INT_NO_EXTREMES	SET2	4.2663	5.8861	0.41927	0.01023	58.073
INT_WEIGHT_AVE	SET2	4.2677	5.8875	0.41941	0.00835	58.059
INT_AVE	SET2	4.2771	5.8980	0.42034	0.00269	57.966
INT_AVE_EXT	SET2	4.3154	5.9480	0.42410	−0.00861	57.590
LightGBM	SET2	4.3253	5.9530	0.42507	0.06211	57.493
XGBoost	SET2	4.3471	6.0112	0.42722	0.05332	57.278
HGBR	SET2	4.4072	6.0536	0.43312	0.03291	56.688
LightGBM	SET1	4.5289	6.2252	0.44508	0.00682	55.492
RF	SET2	4.5668	6.3094	0.44880	−0.00705	55.120
XGBoost	SET1	4.5668	6.2910	0.44881	−0.04171	55.119
HGBR	SET1	4.5836	6.3148	0.45046	0.01365	54.954
ET	SET2	4.6368	6.3878	0.45568	−0.12782	54.432
RF	SET1	4.6876	6.4897	0.46068	−0.01492	53.932
ET	SET1	4.7785	6.5618	0.46961	−0.14537	53.039
GBDT	SET2	4.8693	6.6359	0.47853	−0.05918	52.147
GBDT	SET1	5.1408	7.0807	0.50522	−0.17368	49.478
CatBoost	SET2	5.2863	7.1600	0.51951	−0.00132	48.049
CatBoost	SET1	5.4931	7.4880	0.53984	−0.08035	46.016
naive	-	10.1754	14.8599	1.00000	0.03356	-

The LightGBM model emerged as the superior base model, although the second-ranked XGBoost demonstrated only marginally lower performance. For the LightGBM model, the optimal hyperparameters identified through grid search tuning were as follows: number of estimators: 500; learning rate: 0.1; maximum number of leaves per tree: 31; tree depth: −1 (unconstrained); minimum samples in leaf: 25; feature fraction (column subsampling): 1 (disabled); subsample: 1 (disabled); L1 regularization: 0; L2 regularization: 1; and early stopping: disabled.

The Random Forest models achieved average results, with the RF model proving superior to the ET model. The worst model was CatBoost among all tested models (excluding the naive model).

To statistically validate the predictive superiority of the best model (INT_MEDIAN), the Diebold–Mariano (DM) test with a MAE loss function was conducted for selected models (including the runner-up and the best-performing non-ensemble model) on the test dataset. The results (Table 8) indicate that the DM statistics for the INT_NO_EXTREMES and LightGBM models are negative and statistically significant ($p < 0.01$). This allows for the rejection of the null hypothesis of equal forecast accuracy, confirming that the INT_MEDIAN model provides significantly more accurate forecasts than the INT_NO_EXTREMES and LightGBM models. For the INT_AVE_EXT model, the difference in forecast accuracy was not statistically significant.

Referring to Table 7, Figure 13 presents a comparison (visualization) of the aggregated RES power generation forecasts (calculated as the sum of the wind farm power generation forecast and the solar power plant power generation forecast) with the actual power generation values for the test range. Figure 13a is a scatter plot comparing the forecasts from the reference model (naive model) and the actual generation values. In contrast, Figure 13b is a scatter plot comparing the forecasts from the best ensemble model (INT_MEDIAN) and the actual generation values.

Table 8. Statistical significance of forecast accuracy: Diebold–Mariano test results for combined wind farm power and solar power plant generation, calculated as the sum of individual wind and solar forecasts.

Best Model (INT_MEDIAN)	DM Statistic (S)	<i>p</i> -Value	Significance
INT_NO_EXTREMES	−3.1779	0.0015	**
INT_AVE_EXT	−1.4116	0.1581	ns
LightGBM	−3.1312	0.0017	**

Notes: ** $p < 0.01$, ns—not significant.

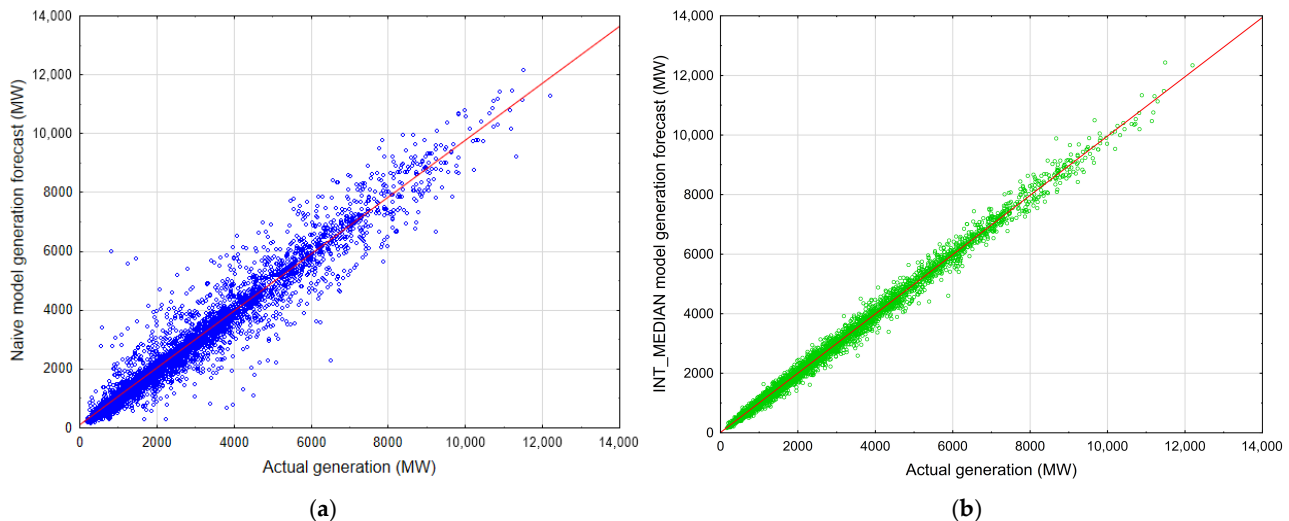


Figure 13. (a) Scatter plot comparing the forecasts from the reference model (naive model) and the actual generation values; (b) scatter plot comparing the forecasts from the best ensemble model (INT_MEDIAN) and the actual generation values.

6. Discussion

The research successfully accomplished all stated objectives. The author compared the effectiveness of various ensemble models (variants of Random Forests and Gradient Boosting Decision Trees) in the task of 1 h ahead forecasting of power generation from wind farms, solar power plants, and the aggregated generation from both sources within the National Power System (covering the entire country).

The general analysis, based on the results obtained for the individual forecasting tasks, showed that the most advantageous model was most frequently LightGBM, while the second most frequent best performer was the XGBoost model. It is worth noting that the least favorable model was CatBoost, and the second-to-last was the GBDT model. Regarding the relatively poor performance of the CatBoost model, it is important to note that this algorithm, by default, constructs symmetric trees, where the same split condition is applied to all nodes at a given level. Symmetric trees tend to be less flexible when fitting highly complex, non-linear relationships in continuous (quantitative) data, such as those found in RES time series. The author suggests that the use of symmetric trees is a probable reason for these results, though it is currently only a hypothesis.

In contrast, the LightGBM model—utilizing a leaf-wise growth strategy—can develop significantly deeper and more asymmetrical branches for specific, narrow variable ranges. This allows it to more effectively capture the intricate dynamics of power generation fluctuations from both wind farms and solar power plants. The comparative analysis of feature importance highlights inherent architectural differences. A sensitivity analysis was performed for the LightGBM and CatBoost models regarding solar power generation forecasting using the SET2 input data set. LightGBM's leaf-wise growth algorithm demonstrated higher sensitivity to strong short-term predictors, specifically the generation from

the previous hour, $SPG(T-1)$, and HOUR variables. CatBoost, constrained by its symmetric tree structure designed to reduce overfitting, distributed feature importance a little more evenly. In the context of short-term solar power forecasting, where recent generation levels are dominant predictors, this structural regularization likely prevented CatBoost from capturing rapid fluctuations as effectively as the ‘greedier’ leaf-wise approach.

Analyzing the results of the Random Forest models across all forecasted time series, it can be observed that the RF model consistently achieved lower nMAE errors than the ET model. The Random Forest (RF) model outperformed Extra Trees (ET) because it employs optimal split points for each feature, allowing for a more precise fit to the complex nonlinearities of RES data. In contrast, ET introduces additional randomness by selecting split points at random, which increased the model’s bias and resulted in slightly higher nMAE errors in this specific forecasting task.

Comparing the improvements of the best-performing models across all forecasted power generation time series relative to the baseline (naive model), the most significant reduction in nMAE was observed for solar power generation. In this instance, the top-performing model achieved an nMAE 74.2% lower than that of the naive method. The significant improvement in solar forecasting compared to the persistence model (74.2%) is consistent with the findings of Bacher et al. (2009) [10]. As established in their work, solar generation is governed by a strong, deterministic diurnal cycle. While the persistence model fails to account for the predictable movement of the sun, machine learning models effectively capture these solar geometry patterns, leading to a much higher skill score than in the case of more stochastic wind generation. Conversely, the smallest improvement in nMAE relative to the baseline was recorded for wind power generation forecasts, with a reduction of 35.1%. Wind power generation exhibits higher stochasticity, making it more difficult to achieve significant improvements over the baseline naive method.

A comparison of quality between the input data set SET1 (only lagged values) and the input data set SET2 (lagged values and additional information obtained using feature engineering) unequivocally demonstrated that the SET2 set achieved lower forecasting errors. By using SET2 instead of SET1, the nMAE error was reduced by as much as 12.8% in the case of generation forecasts from wind farms and solar power plants (single time series). However, the smallest percentage reduction in the nMAE error occurred for the combined generation forecasts from both RES, calculated as the sum of separate wind farm and solar power plant generation forecasts, where the nMAE error decreased by 4.5%.

The proposed novel, proprietary ensemble integrators, designed to combine different base ensemble models (variants of Random Forests and Gradient Boosting Decision Trees) with the aim of improving forecasting quality, proved to be highly effective. They outperformed the base ensemble models across all forecasted time series. Among the ensemble integrators, the INT_MEDIAN and INT_NO_EXTREMES models stand out. Conversely, the INT_AVE_EXT model proved to be the least advantageous. The expert concept of the INT_AVE_EXT ensemble integrator, which relies solely on the two most distant forecasts calculated for each period T , proved to be unsuccessful. Other forecasts, which were closer to the actual power generation values, were typically not located near the midpoint between the extreme forecast values, thus preventing the INT_AVE_EXT model from achieving an optimal prediction. Nevertheless, it yielded better results than the base ensemble models, with only one exception.

Regarding wind power generation forecasting, the best-performing ensemble integrator (INT_MEDIAN) achieved an nMAE value 0.8% lower than that of the leading base ensemble model (LightGBM).

In contrast, for solar power generation forecasts, the best-performing ensemble integrator (INT_WEIGHT_AVE) achieved an nMAE value 1.8% lower than that of the leading base ensemble model (LightGBM).

Furthermore, in the case of wind and solar power generation forecasted as a single time series, the best-performing ensemble integrator (INT_AVE) achieved an nMAE value 2.2% lower than that of the leading base ensemble model (XGBoost).

In the case of problem decomposition—where the total generation from both RESs is calculated as the sum of their individual forecasts—the best-performing ensemble integrator (INT_MEDIAN) achieved an nMAE value 1.5% lower than that of the leading base ensemble model (LightGBM).

Furthermore, as one of the most critical research objectives, we verified whether it is more beneficial to forecast the combined time series of wind and solar power generation directly, or if it is more advantageous to generate separate forecasts for wind power and solar power, and subsequently aggregate (sum) the two individual forecasts to obtain the total forecast for both RES.

For the case of the National Power System in Poland, utilizing both input data sets SET2 and SET1, the combined RES generation forecasts calculated from separately obtained forecasts for wind farm power generation and solar power plant power generation proved to be more advantageous. For the SET1 set, the nMAE error was reduced by 15.9% (XGBoost). The ensemble integrator models were tested using only the SET2 dataset, as this variant proved to be more advantageous for the base ensemble models, yielding lower nMAE errors, so for the SET2 set, the nMAE error was reduced by 8.0% by the INT_MEDIAN (the sum of separate forecasts for each RES) model in comparison to the INT_AVE (aggregated RES generation) model. Notably, the nMAE errors for various models were consistently lower when using the SET2 set—thus, the application of the SET2 set is justified for forecasting RES power generation.

Figure 14 compares forecasts (SET2 input data) derived from a single combined time series (using the best INT_AVE model) and forecasts resulting from the sum of separate forecasts for each RES generation (using the best INT_MEDIAN model). Figure 14 demonstrates that the INT_MEDIAN forecast curve exhibits a closer alignment with the actual generation curve than the INT_AVE forecast curve.

Figure 15 presents a comparison of the performance improvement (percentage decrease in nMAE error) for each of the tested models using the SET2 dataset, achieved through problem decomposition. In this approach, total RES generation is forecasted as the sum of individual forecasts rather than as a single time series representing the combined generation of both sources. It should be emphasized that the comparison in Figure 15 does not indicate which model achieved the lowest overall nMAE error through decomposition; rather, it highlights the extent to which decomposition contributed to error reduction for each specific model (overall, the INT_MEDIAN model remains the top-ranked model when utilizing decomposition). The most significant improvement due to decomposition was observed for the ET model (12.26% nMAE reduction); however, the best-performing model overall, INT_MEDIAN, saw a smaller reduction of 8.51%.

The strong negative correlation (-0.919) described in Section 3, between wind and solar power generation is a key statistical driver behind the superior performance of the decomposition approach. Summing individual forecasts for each RES source yielded better results than applying a single time series forecasting model to the combined generation. A model trained on an aggregated time series (WSPG) perceives a ‘smoothed’ signal, representing the sum of two distinct time series. Consequently, the extremes of one RES source are ‘masked’ by the other. Paradoxically, this may hinder the model’s ability to learn specific patterns characteristic of each individual RES source. This explains why the models

trained on the aggregated series (WSPG) yielded higher nMAE forecast errors compared to the sum of individual forecasts from both RESs.

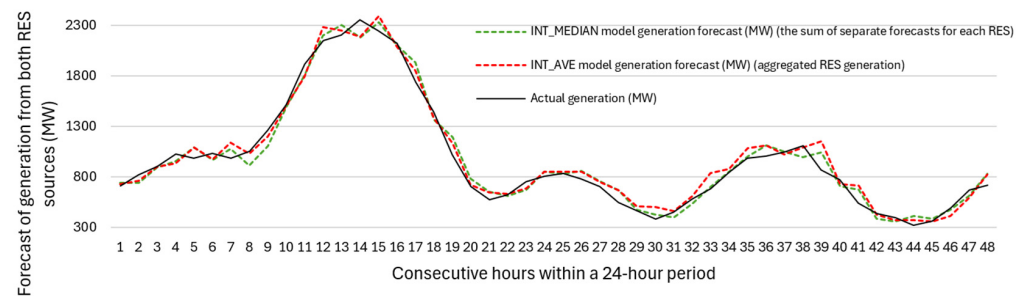


Figure 14. Comparison of forecasting results for the INT_MEDIAN and INT_AVE models (input data SET2) for two consecutive days in March 2022.

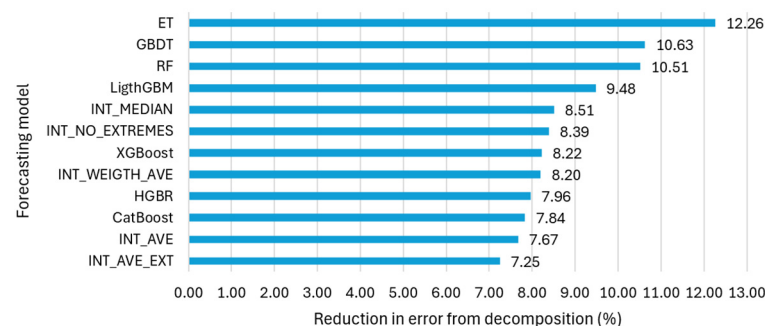


Figure 15. Comparison of the performance improvement (percentage decrease in nMAE error) for each of the tested models using the SET2 dataset, achieved through problem decomposition (forecasting total RES generation as the sum of individual forecasts versus a single aggregated time series).

It is also worth noting that forecasts employing the WSPG approach result in a loss of dynamic information, as aggregating two RES time series prior to the training process partially ‘blurs’ the signals of rapid fluctuations. For instance, a sudden drop in wind generation may be masked by a stable increase in PV generation. Consequently, the integrated model loses the ability to precisely identify critical turning points for specific RESs.

Furthermore, the complexity of the relationship between the two sources must be considered. The relationship between meteorological features and the sum of two negatively correlated RESs is significantly more non-linear and challenging for a single ‘base ensemble’ model to map than the relationships for each source individually. These factors further explain why problem decomposition yielded superior results—characterized by lower forecast errors—compared to the single time series forecasting of combined RES generations.

Practical recommendations are summarized in Table 9, which serves as a Model Selection Guide based on Scenario.

Table 9. Model Selection Guide based on Scenario.

Scenario	Recommended
High volatility/complex non-linear data	LightGBM/XGBoost
Need for high robustness/avoiding outliers	Ensemble Integrators (Median/No Extremes)
Limited computing resources	LightGBM

The recommended procedure for national power system operators is the separate forecasting (decomposition) of renewable energy generation.

In future research plans, prognostic studies are scheduled to be performed using other forecasting models, with an emphasis on models belonging to the category of artificial neural networks (MLP, LSTM, GRU, and Transformer) and the construction of hybrid models.

7. Conclusions

This study successfully compared the effectiveness of various ensemble models (Random Forest and GBDT variants) for 1 h ahead power generation forecasting from wind farms and solar power plants within the National Power System.

Model Performance and Input Data—the LightGBM model was the most advantageous among homogeneous ensembles, consistently outperforming XGBoost, GBDT, and CatBoost. Critically, using the SET2 input data (lagged values plus feature engineering) significantly reduced the nMAE error compared to SET1 (lagged values only), achieving a reduction of up to 12.8% for single-source generation forecasts. The application of SET2 is therefore justified.

Ensemble Integration—Novel, proprietary ensemble integrators (specifically INT_MEDIAN and INT_NO_EXTREMES) proved highly effective, consistently outperforming the base ensemble models across all time series.

Optimal Forecasting Strategy—it was verified that the most beneficial strategy is to generate separate forecasts for wind and solar power and then aggregate (sum) the results to obtain the total combined forecast. This Forecast-then-Aggregate approach reduced the nMAE error of the combined RES forecast by 8% when utilizing the SET2 data, confirming its superiority over direct combined forecasting.

Funding: This research received no external funding.

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: The study did not involve humans.

Data Availability Statement: The data presented in this study are openly available in [Transparency Platform: <https://transparency.entsoe.eu>].

Acknowledgments: The author thanks the Gemini 3 Pro Large Language Model (Google) for assistance with the translation from Polish to English, and for improving the clarity and grammatical structure of the manuscript. All scientific conclusions and findings remain the sole responsibility of the author.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACF	Autocorrelation Function
ADA	Adaptive Boosting
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
CatBoost	Categorical Boosting
CNN	Convolutional Neural Network
DHI	Diffuse Horizontal Irradiance
DL	Deep learning
DNI	Direct Normal Irradiance
DS	Diebold–Mariano test
ENTSO-E	European Network of Transmission System Operators for Electricity
ET	Extremely Randomized Trees

EU	European Union
GBDT	Gradient-Boosted Decision Tree
GHI	Global Horizontal Irradiance
GRU	Gated Recurrent Unit
HGBR	Histogram-based Gradient Boosting Regressor
IMGW-PIB	Institute of Meteorology and Water Management—National Research Institute
LightGBM	Light Gradient Boosting Machine
LP	Load Profile
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	MultiLayer Perceptron
MSE	Mean Squared Error
nMAE	Normalized Mean Absolute Error
nMBE	Normalized Mean Bias Error
nRMSE	Normalized Root Mean Square Error
NWP	Numerical Weather Prediction
PACF	Partial Autocorrelation Function
PV	Photovoltaic
R	Pearson's linear correlation coefficient
RelnMAE	Relative Normalized Mean Absolute Error
RES	Renewable Energy Sources
RF	Random Forest
RMSE	Root Mean Square Error
TBATS	Trigonometric Seasonal Box–Cox ARMA Time Series
XGBoost	Extreme Gradient Boosting

References

- Oriol, T.; Cifuentes, J.; Marulanda, G. Enhancing Wind Power Forecasting in the Spanish Market through Transformer Neural Networks and Temporal Optimization. *Sustainability* **2025**, *17*, 8655. [\[CrossRef\]](#)
- Wood, A.D. Country-wide solar power load profile for Germany 2015 to 2019: The impact of system curtailments on prediction models. *Energy Convers. Manag.* **2022**, *269*, 116096. [\[CrossRef\]](#)
- Aieb, A.; Liotta, A.; Jacob, A.; Yaqub, M.A. Short-Term Forecasting of Non-Stationary Time Series. *Eng. Proc.* **2024**, *68*, 34. [\[CrossRef\]](#)
- Maitra, S.; Politis, D.N. Prepivoted Augmented Dickey-Fuller Test with Bootstrap-Assisted Lag Length Selection. *Stats* **2024**, *7*, 1226–1243. [\[CrossRef\]](#)
- Surakhi, O.; Zaidan, M.A.; Fung, P.L.; Hossein Motlagh, N.; Serhan, S.; AlKhanafseh, M.; Ghoniem, R.M.; Hussein, T. Time-Lag Selection for Time-Series Forecasting Using Neural Network and Heuristic Algorithm. *Electronics* **2021**, *10*, 2518. [\[CrossRef\]](#)
- Leites, J.; Cerqueira, V.; Soares, C. Lag Selection for Univariate Time Series Forecasting Using Deep Learning: An Empirical Study. In *Progress in Artificial Intelligence*; Santos, M.F., Machado, J., Novais, P., Cortez, P., Moreira, P.M., Eds.; Springer: Cham, Switzerland, 2025; pp. 321–332. [\[CrossRef\]](#)
- Tuncar, E.; Saglam, S.; Oral, B. A review of short-term wind power generation forecasting methods in recent technological trends. *Energy Rep.* **2024**, *12*, 197–209. [\[CrossRef\]](#)
- Konstantinou, M.; Peratikou, S.; Charalambides, A. Solar Photovoltaic Forecasting of Power Output Using LSTM Networks. *Atmosphere* **2021**, *12*, 124. [\[CrossRef\]](#)
- Dhaked, D.; Dadhich, S.; Birla, D. Power output forecasting of solar photovoltaic plant using LSTM. *Green Energy Intell. Transp.* **2023**, *2*, 100113. [\[CrossRef\]](#)
- Bacher, P.; Madsen, H.; Nielsen, H.A. Online Short-Term Solar Power Forecasting. *Sol. Energy* **2009**, *83*, 1772–1783. [\[CrossRef\]](#)
- Cui, Y.; Wang, P.; Meirink, J.F.; Ntantis, N.; Wijnands, J.S. Solar Radiation Nowcasting Based on Geostationary Satellite Images and Deep Learning Models. *Sol. Energy* **2024**, *282*, 112866. [\[CrossRef\]](#)
- Pospíchal, J.; Kubovčík, M.; Dirgová Luptáková, I. Solar Irradiance Forecasting with Transformer Model. *Appl. Sci.* **2022**, *12*, 8852. [\[CrossRef\]](#)
- Feng, C.; Zhang, J.; Zhang, W.; Hodge, B.-M. Convolutional Neural Networks for Intra-Hour Solar Forecasting Based on Sky Image Sequences. *Appl. Energy* **2022**, *310*, 118438. [\[CrossRef\]](#)

14. Bochenek, B.; Jurasz, J.; Jaczewski, A.; Stachura, G.; Sekuła, G.; Strzyżewski, G.; Wdowikowski, M.; Figurski, M. Day-Ahead Wind Power Forecasting in Poland Based on Numerical Weather Prediction. *Energies* **2021**, *14*, 2164. [\[CrossRef\]](#)
15. Piotrowski, P.; Baczyński, D.; Kopyt, M.; Szafranek, K.; Helt, P.; Gulczyński, T. Analysis of forecasted meteorological data (NWP) for efficient spatial forecasting of wind power generation. *Electr. Power Syst. Res.* **2019**, *175*, 105891. [\[CrossRef\]](#)
16. PSEW—Polskie Stowarzyszenie Energetyki Wiatrowej. New Rules for Onshore Wind Farms—Liberalisation of the 10H Rule. 2023. Available online: <https://www.psew.pl/en/new-rules-for-onshore/> (accessed on 11 December 2025).
17. Aydın, B.; Stecuła, K.; Olczak, P.; Kulpa, J.; Stecuła, B. Exploring the Green Horizon: Recent Research on Renewable Energy in Poland—A Review. *Energies* **2025**, *18*, 1695. [\[CrossRef\]](#)
18. Wasiuta, A. Assessing Renewable Energy Development Potential in Polish Voivodeships. *Sustainability* **2024**, *16*, 11261. [\[CrossRef\]](#)
19. Błaszczak, A.; Matuszak-Flejszman, A.; Nawrocki, K. Determinants of the Development of Photovoltaics in Poland. *Renew. Energy* **2024**, *233*, 121161. [\[CrossRef\]](#)
20. Robak, S.; Raczkowski, R.; Piekarz, M. Development of the Wind Generation Sector and Its Effect on the Grid in Poland. *Energies* **2023**, *16*, 6805. [\[CrossRef\]](#)
21. Transparency Platform. Available online: <https://transparency.entsoe.eu> (accessed on 24 October 2025).
22. Piotrowski, P.; Kopyt, M. Short-term forecasts of energy generation in a solar power plant using various machine learning models, along with ensemble and hybrid methods. *Energies* **2024**, *17*, 4234. [\[CrossRef\]](#)
23. Yang, X.; Dong, Y.; Yang, L.; Wu, T. A Hybrid Wind–PV Power Generation Forecasting Model Based on Uncertainty Quantification. *Appl. Soft Comput.* **2025**, *185*, 113965. [\[CrossRef\]](#)
24. Géron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2019.
25. Parol, M.; Piotrowski, P.; Kapler, P.; Piotrowski, M. Forecasting of 10-Second Power Demand of Highly Variable Loads for Microgrid Operation Control. *Energies* **2021**, *14*, 1290. [\[CrossRef\]](#)
26. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
27. 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 (KDD), San Francisco, CA, USA, 13–17 August 2016.
28. eXtreme Gradient Boosting. Available online: <https://github.com/dmlc/xgboost> (accessed on 11 December 2025).
29. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 3149–3157.
30. Light Gradient Boosting Machine. Available online: <https://github.com/Microsoft/LightGBM> (accessed on 11 December 2025).
31. Deep Dive into scikit-learn's HistGradientBoosting Classifier and Regressor. Available online: <https://github.com/thomasjpfan/pydata-2019-histgradientboosting> (accessed on 11 December 2025).
32. Yandex Catboost. Available online: <https://github.com/catboost/catboost> (accessed on 11 December 2025).
33. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 3–8 December 2018; pp. 6638–6648.
34. Keels, J. XGBoost, Light GBM and CatBoost. A Comparison of Decision Tree Algorithms and Applications to a Regression Problem. Available online: <https://medium.com/octave-john-keells-group/xgboost-light-gbm-and-catboost-a-comparison-of-decision-tree-algorithms-and-applications-to-a-f1d2d376d89c> (accessed on 12 December 2025).
35. Gao, Q. Multi-Temporal Scale Wind Power Forecasting Based on Lasso-CNN-LSTM-LightGBM. *EAI Endorsed Trans. Energy Web* **2024**, *11*, 1. [\[CrossRef\]](#)
36. Liao, S.; Tian, X.; Liu, B.; Liu, T.; Su, H.; Zhou, B. Short-Term Wind Power Prediction Based on LightGBM and Meteorological Reanalysis. *Energies* **2022**, *15*, 6287. [\[CrossRef\]](#)
37. Yildiz, C.; Acikgoz, H.; Korkmaz, D.; Budak, U. An improved residual-based convolutional neural network for very short-term wind power forecasting. *Energy Convers. Manag.* **2021**, *228*, 113731. [\[CrossRef\]](#)
38. Hu, H.; Wang, L.; Lv, S.X. Forecasting energy consumption and wind power generation using deep echo state network. *Renew. Energy* **2020**, *154*, 598–613. [\[CrossRef\]](#)
39. Du, P.; Wang, J.; Yang, W.; Niu, T. A novel hybrid model for short-term wind power forecasting. *Appl. Soft Comput.* **2019**, *80*, 93–106. [\[CrossRef\]](#)
40. Amila, T.; Peiris, J.J.; Rathnayake, U. Forecasting Wind Power Generation Using Artificial Neural Network: “Pawan Danawi” A Case Study from Sri Lanka. *J. Electr. Comput. Eng.* **2021**, *2021*, 5577547. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.