



# Article Deep-Reinforcement-Learning-Based Low-Carbon Economic Dispatch for Community-Integrated Energy System under Multiple Uncertainties

Mingshan Mo<sup>1</sup>, Xinrui Xiong<sup>1</sup>, Yunlong Wu<sup>1</sup> and Zuyao Yu<sup>2,\*</sup>

- <sup>1</sup> School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; m202372495@hust.edu.cn (X.X.)
- <sup>2</sup> School of Naval Architecture and Ocean Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
- \* Correspondence: yuzuyao@hust.edu.cn

**Abstract:** A community-integrated energy system under a multiple-uncertainty low-carbon economic dispatch model based on the deep reinforcement learning method is developed to promote electricity low carbonization and complementary utilization of community-integrated energy. A demand response model based on users' willingness is proposed for the uncertainty of users' demand response behavior; a training scenario set of a reinforcement learning agent is generated with a Latin hypercube sampling method for the uncertainties of power, load, temperature, and electric vehicle trips. Based on the proposed demand response model, low-carbon economic dispatch of the community-integrated energy system under multiple uncertainties is achieved by training the agent to interact with the environment in the training scenario set and reach convergence after 250 training rounds. The simulation results show that the reinforcement learning agent achieves low-carbon economic dispatch under 5%, 10%, and 15% renewable energy/load fluctuation scenarios, temperature fluctuation scenarios, and uncertain scenarios of the number of trips, time periods, and mileage of electric vehicles, with good generalization performance under uncertain scenarios.

**Keywords:** demand response uncertainty; deep reinforcement learning; community-integrated energy system; low-carbon economic dispatch

# 1. Introduction

The increasing economic level and energy demand will lead to the problem of fossil energy depletion and ecological environment degradation, and the development of lowcarbon energy and the complementary use of energy has become a strategic choice for all countries in the world.

In the context of energy decarbonization, the use of renewable energy sources such as wind and solar is of great significance. However, renewable energy is characterized by randomness, so the integrated energy system containing renewable energy generation needs to have the ability to cope with the randomness of renewable energy.

In the community-integrated energy system (CIES) that includes renewable energy power generation, in addition to the demand for electricity, there are often demands for natural gas and cooling supply [1,2]. In a community-integrated energy system with electricity–gas–cooling coupling, the complementary characteristics of the three energies can be fully utilized to promote the consumption of renewable energy. However, the diversity of various energy devices, the complexity of joint control of equipment, and various uncertainties in the energy system bring challenges to the dispatch of a communityintegrated energy system.

In existing work, optimal dispatch of energy systems can be achieved with a variety of methods. Yang Li, Yuanyuan Zhang, and Shenbo Yang et al. [3–5] proposed a hierarchical



Citation: Mo, M.; Xiong, X.; Wu, Y.; Yu, Z. Deep-Reinforcement-Learning-Based Low-Carbon Economic Dispatch for Community-Integrated Energy System under Multiple Uncertainties. *Energies* **2023**, *16*, 7669. https://doi.org/10.3390/en16227669

Academic Editor: Mariano Giuseppe Ippolito

Received: 8 October 2023 Revised: 11 November 2023 Accepted: 13 November 2023 Published: 20 November 2023



**Copyright:** © 2023 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/). stochastic dispatch method for an integrated energy system based on Stackelberg game theory. For non-convex problems in an integrated energy system, Han Gao et al. [6] proposed a new optimization method based on Benders decomposition, whose sub-problems can be solved in parallel to accelerate the computation further. To address the complexity of multiple supplies and demands in an integrated energy system, X.J. Luo et al. [7] proposed a multi-energy system management strategy that includes three core algorithms for demand-side rolling optimization, supply-side rolling optimization, and feedback correction. To address the uncertainty of an integrated energy system, Peng Li, Guang Liu, Rujing Yan, and Xiaoqing Li et al. [8–11] applied robust optimization methods to solve the uncertainty in an integrated energy system based on multi-energy load coupling and proposed a stochastic robust optimal operation strategy for an integrated energy system.

With the development of artificial intelligence, deep reinforcement learning technology has had some applications in the optimal dispatch of energy systems. It can realize the joint optimal dispatch of various energy devices within an integrated energy system through the continuous interaction between the agent and the environment, with good adaptive learning capability.

Salman Sadiq Shuvo et al. [12] proposed a discrete action deep reinforcement learning method for managing smart devices based on the A2C (Advantage Actor-Critic) algorithm to optimize power costs. The method manages flexible loads as a discrete power staging control. Renzhi Lu et al. [13] considered uncontrollable, shiftable, and curtailable loads in the system and approximated the optimal policy using a discrete-action DQN (Deep Q-Network) method. Mifeng Ren et al. [14] built on the DQN method with a modelfree discrete-action Dueling-double deep Q-learning neural network (Dueling-DDQN) algorithm for joint dispatch of air conditioners, electric vehicles, and energy storage devices in a home energy management system model. Bo-Chen Lai et al. [15] proposed a multiagent reinforcement-learning-based community energy management system model in which the appliances are classified into three categories: uncontrollable appliances, shiftable appliances, and power-curtailable appliances, and a discrete-action Multi-agent Q-Learning algorithm is used for optimal dispatch based on the DQN approach. The control actions of the energy units are designed as a hierarchical regulation. To achieve continuous control of energy units in residential energy systems, Yujian Ye et al. [16] proposed a new real-time management strategy for residential energy systems based on the continuous-action Deep Deterministic Policy Gradient (DDPG) deep reinforcement learning approach to achieve multi-dimensional continuous state control of multiple energy units in energy systems to minimize energy costs for users. Hongyuan Ding et al. [17] classified smart home loads into HVAC, shiftable, uncontrollable, and thermal loads and proposed a continuous-action PD-DDPG (Primal-Dual Deterministic Policy Gradient) deep reinforcement learning method to optimize the control of home energy system devices based on the DDPG method. Lin Xue et al. [18] proposed a model-data-event-based low-carbon economic scheduling framework for the community-integrated energy system, and used an improved DDPG algorithm that takes into account generation and load uncertainty for real-time scheduling. Yue Qiu et al. [19] proposed a mathematical model of the local integrated energy system that takes into account supply- and load-side flexible resources and used an improved twin delayed Deep Deterministic Policy Gradient (TD3) algorithm to achieve operational optimization under renewable energy generation, electrical load, and thermal load uncertainty. Seong-Hyun Hong et al. [20] propose an energy management system (EMS) algorithm based on secure reinforcement learning to achieve more robust energy management considering generation and load uncertainties.

In summary, the application of reinforcement learning methods to a variety of demand response management has been realized in existing work. However, the uncertainty of demand response is usually not considered. In a community-integrated energy system, the user's demand response behavior is the result of the user's trade-offs and should be subject to uncertainty. Inspired by existing work, this paper develops a demand response model that takes into account user behavioral uncertainty and proposes a deep-reinforcementlearning-based low-carbon and economic dispatch method for community-integrated energy systems under multiple uncertainties.

The main contributions of this paper are as follows.

(1) A community-integrated energy system simulation model is developed to simulate the energy flow within the energy system. The simulation model includes energy units such as gas turbines, electric vehicles, and air conditioning systems, as well as demand response resources. The demand response resources are categorized into two categories in the simulation model, including curtailable load and shiftable load.

(2) To address the uncertainty of user participation in demand response behavior and consider real-time energy prices, a demand response model based on the degree of users' willingness is proposed to take into account the uncertainties of curtailable load and shiftable load.

(3) A Soft Actor–Critic deep-reinforcement-learning-based approach is proposed for the dispatch of community-integrated energy systems, which achieves a better dispatch scheme through the stronger exploratory capability of the Soft Actor–Critic algorithm, and adapts to the uncertainties of renewable energy generation, outdoor temperature, and electric vehicle trips through the training of the agent to improve the applicability of this method under multiple uncertainties.

### 2. Plant Model

The community-integrated energy system model based on deep reinforcement learning includes an environment model and an agent model. The environment model is the context of reinforcement learning in low-carbon economic dispatch, which consists of various component models such as renewable energy generation, energy demand, etc., and the energy market and carbon trading market. The agent is the decision-making subject of reinforcement learning, which learns the optimal dispatch strategy by continuously interacting with the environment, observing the environment, taking actions, and obtaining rewards. The general framework of the model is shown in Figure 1.



Figure 1. Model framework of community-integrated energy system based on deep reinforcement learning.

### 2.1. Environmental Model

The CIES environmental model integrates the electricity–gas coupling model on the energy supply side, the electricity–cooling coupling model on the energy supply side, the demand response model on the energy demand side, the energy storage device model on the energy storage side, etc. The internal components of CIES, together with the electricity market, the natural gas market, and the carbon trading market, constitute the environmental model.

- 2.1.1. Electricity-Gas Coupling Model on the Energy Supply Side
- (1) Gas turbine model

The gas turbine [21] converts natural gas into electricity as follows:

$$H_t = a_g (P_t^{\text{GT}})^2 + b_g P_t^{\text{GT}} + c_g$$
<sup>(1)</sup>

$$G_t = H_t / G^{HV} \tag{2}$$

where  $H_t$  means the heat consumption of the gas turbine;  $a_g$ ,  $b_g$ , and  $c_g$  are the heat consumption coefficients;  $G_t$  means the natural gas consumed using the gas turbine;  $G^{HV}$  means the thermal value.

The gas turbine is operated to satisfy the constraints as shown in Equations (3) and (4).

$$p_{\min}^{\rm GT} < p_t^{\rm GT} < p_{\max}^{\rm GT} \tag{3}$$

$$-\Delta P^{\text{GTmax}} \le P_t^{\text{GT}} - P_{t-1}^{\text{GT}} \le \Delta P^{\text{GTmax}}$$
(4)

where  $p_{\text{max}}^{\text{GT}}$  and  $p_{\text{min}}^{\text{GT}}$  are the maximum and minimum power output of the gas turbine, respectively;  $\Delta P^{\text{GTmax}}$  is the maximum climbing power of the gas turbine, respectively.

# (2) P2G equipment model

The P2G equipment [22] converts electrical energy to natural gas as follows:

$$G_t^{P2G} = P_t^{P2G} \eta^{P2G} / G^{HV}$$
<sup>(5)</sup>

where  $G_t^{P2G}$  means the natural gas generated;  $P_t^{P2G}$  means the electric power consumed;  $\eta^{P2G}$  means the conversion efficiency of the P2G equipment.

# 2.1.2. Electric-Cooling Coupling Model on the Energy Supply Side

The electricity–cooling coupling on the energy supply side is realized through the cooling storage air conditioner, which needs to meet the constraints as follows [23]:

$$0 \le H_t^{\text{ACc}} \le H_{\text{max}}^{\text{ACc}} \tag{6}$$

where  $H_t^{ACc}$  means the cooling created by the chiller at time *t*;  $H_{max}^{ACc}$  means the maximum cooling creation of the chiller at time *t*.

Indoor temperature changes can be described in Equations (7) and (8).

$$T_t^{\text{in}} = \varepsilon T_{t-1}^{\text{in}} + (1 - \varepsilon)(T_t^{\text{out}} - (H_t^{\text{ACd}} - Q_t)/A)$$
(7)

$$H_t^{\text{ACd}} = H_t^{\text{ACc}} + H_t^{\text{ACr}} - H_t^{\text{ACs}} \tag{8}$$

where  $H_t^{ACd}$  is the cooling provided to the room using the air conditioning at time t;  $Q_t$  is the heat gained by the building through solar radiation, indoor installations, etc., in addition to the heat transfer from the temperature difference at time t;  $\varepsilon$  is the air inertia coefficient, set to 0.95;  $H_t^{ACs}$  and  $H_t^{ACr}$  mean the cooling charging and discharging volume of the storage tank at time t, respectively.

The indoor temperature constraints to be met using air conditioning dispatch are shown in Equations (9) and (10).

$$\Gamma_0^{\rm in} = T^{\rm origin} \tag{9}$$

$$T_{\min}^{\text{in}} \le T_t^{\text{in}} \le T_{\max}^{\text{in}} \tag{10}$$

where  $T^{\text{origin}}$  is the initial temperature;  $T_t^{\text{in}}$  means the indoor temperature at time *t*;  $T_{\text{max}}^{\text{in}}$  and  $T_{\text{min}}^{\text{in}}$  mean the maximum and minimum appropriate temperature, respectively.

The electric power of the air conditioner at time *t* can be expressed as follows:

$$P_t^{AC} = \frac{H_t^{ACc}}{\mu^{ACc}} + \mu^{ACs} H_t^{ACs} + \mu^{ACr} H_t^{ACr}$$
(11)

where  $\mu^{ACc}$  is the energy efficiency ratio of the chiller;  $\mu^{ACs}$  and  $\mu^{ACr}$  are the energy efficiency ratios of the cooling charging and discharging, respectively.

# 2.1.3. Energy Storage Device Model on the Energy Storage Side

The energy storage device model involves constraints on the battery, the gas storage tank, and the cooling storage tank [11,24] as shown in Equations (12)–(15).

$$E_{\min}^{\rm ES} \le E_t^{\rm ES} \le E_{\max}^{\rm ES} \tag{12}$$

$$0 \le P_t^{\text{ESch}} \le P_{\max}^{\text{ESch}} \tag{13}$$

$$0 \le P_t^{\text{ESdis}} \le P_{\text{max}}^{\text{ESdis}} \tag{14}$$

$$E_t^{ES} = E_{t-1}^{ES} + \Delta t P_t^{ESch} \eta^{ESch} - \frac{\Delta t P_t^{ESdis}}{\eta^{ESdis}}$$
(15)

where  $E_t^{\text{ES}}$  is the residual capacity at time t;  $E_{\text{max}}^{\text{ES}}$  and  $E_{\text{min}}^{\text{ES}}$  are the maximum and minimum storage capacity, respectively;  $P_t^{\text{ESch}}$  and  $P_t^{\text{ESdis}}$  are the charging and discharging energy at time t;  $P_{\text{max}}^{\text{ESch}}$  and  $P_{\text{max}}^{\text{ESdis}}$  are the maximum charge and discharge energy per unit time, respectively;  $\eta^{\text{ESch}}$  and  $\eta^{\text{ESdis}}$  are the charging and discharging efficiency, respectively.

# 2.1.4. Demand Response Model Based on User's Willingness on the Energy Demand Side

In this paper, we consider the uncertainty of user participation in demand response and establish the demand response model based on the user's willingness (DRUW). The dispatch decisions for CIES are implemented with the community-integrated energy management system (CIEMS). After CIEMS issues the incentive signal, users consider whether to execute a demand response from their own interests according to the incentive signal. Since the response behavior of users involves their own interests, the demand response behavior of users is inseparable from the real-time energy price and demand response incentive compensation mechanism.

Define the willingness degree of a user to participate in load curtailment and load shifting demand response as follows.

$$\begin{cases} \theta_t^{\text{RL}} = \theta_t^{\text{RLB}} \beta_t^{\text{RL}} = \frac{c^{\text{RL}} \beta_t^{\text{RL}}}{\epsilon^{\text{dam}} \rho_t^{\text{Pu}}} \\ \theta_t^{\text{SL}} = \theta_t^{\text{SLB}} \beta_t^{\text{SL}} = \frac{c^{\text{SL}} \beta_t^{\text{SL}}}{\epsilon^{\text{dam}} (1/\rho_t^{\text{Pu}})} \end{cases}$$
(16)

where  $\theta_t^{\text{RL}}$  and  $\theta_t^{\text{SL}}$  are the willingness degree of a user to participate in load curtailment at time *t* and load shifting to time *t*, respectively;  $\theta_t^{\text{RLB}}$  and  $\theta_t^{\text{SLB}}$  are the benchmark willingness degrees of a user to participate in load curtailment at time *t* and load shifting to time *t*, respectively;  $\beta_t^{\text{RL}}$  and  $\beta_t^{\text{SL}}$  are the demand response incentive factors for load curtailment at time *t* and load shifting to time *t*, respectively, and take the values of [1,2].  $\varepsilon^{\text{dam}}$  is the response damping coefficient for a user's demand response to electric/gas energy prices, which depends on the user and on the energy source;  $\rho_t^{\text{pu}}$  is the normalized energy price at time *t*;  $c^{\text{RL}}$  is the unit curtailed electric/gas load compensation price;  $c^{\text{SL}}$  is the unit shifted electric/gas load compensation price. Based on the user's willingness degree, define their demand response probability as shown in Equation (17) and Figure 2.

$$\omega_{t} = \begin{pmatrix} 0 & \theta_{t} < \theta_{\min} \\ (\theta_{t} - \theta_{\min}) / (\theta_{\max} - \theta_{\min}) & \theta_{\min} < \theta_{t} < \theta_{\max} \\ 1 & \theta_{t} > \theta_{\max} \end{cases}$$
(17)

where  $\omega_t$  means the probability of a user's participation in the response at time *t*;  $\theta_{\min}$  means the lower limit of the response uncertainty interval of the user;  $\theta_{\max}$  means the upper limit of the response uncertainty interval of the user.



Figure 2. Relationship between user's response probability and response willingness.

As shown in Figure 2, the horizontal coordinate is the user's willingness degree, and the vertical coordinate is the probability of the user's participation in the demand response. When the user's willingness degree is lower than the lower limit of the response uncertainty interval at a certain moment, the user does not participate in the response, and the response probability is 0. When the user's willingness degree is higher than the upper limit of the response uncertainty interval, the user's response probability is 1. When the user's willingness degree is between the upper and lower limits of the response uncertainty interval, the user's response probability is proportional to the willingness degree and takes values between 0 and 1.

(1) Curtailable electric/gas load model in DRUW

$$P_t^{\rm RL} = \delta_t^{\rm RL} P_t^{\rm RL0} \tag{18}$$

where  $\delta_t^{\text{RL}}$  means the binary state variable of the electric/gas load curtailment response at time *t*;  $P_t^{\text{RL}}$  means the actual power curtailment of electric/gas load at time *t*;  $P_t^{\text{RL0}}$  means the expected power curtailment of electric/gas at time *t*.

The dispatch of curtailable electric/gas load needs to satisfy the continuous curtailment time constraint and the total curtailment times' constraint as shown in Equations (19) and (20).

$$T_{\min}^{\text{RL}} \le \sum_{t=\tau}^{\tau+T_{\max}^{\text{RL}}-1} \delta_t^{\text{RL}} \le T_{\max}^{\text{RL}}$$
(19)

$$\sum_{t=1}^{T} \delta_t^{\text{RL}} \le N_{\max}^{\text{RL}}$$
(20)

where  $T_{\text{max}}^{\text{RL}}$  and  $T_{\text{min}}^{\text{RL}}$  mean the upper and lower limits of the continuous curtailment time, respectively;  $N_{\text{max}}^{\text{RL}}$  means the upper limit of the total curtailment times.

(2) shiftable electric/gas load model in DRUW

$$P_{ts'}^{\rm SL} = \delta_{ts'}^{\rm SL} P_{ts}^{\rm SL0} \tag{21}$$

where *ts* means the load onset moment before the shiftable electric/gas load participates in the dispatch; *ts'* means the load onset moment after the shiftable electric/gas load participates in the dispatch;  $\delta_{ts'}^{SL}$  means the binary state variable of the shiftable electric/gas load response at time *ts'*;  $P_{ts'}^{SL}$  means the load shifted to period *ts'*;  $P_{ts}^{SL0}$  means the shifted electric/gas load at time *ts*.

The power distribution vector of the shiftable load before it participates in the dispatch is as follows:

$$L_{\text{before}}^{\text{SL}} = (0, \cdots, P_{ts}^{\text{SL}}, P_{ts+1}^{\text{SL}}, \cdots, P_{ts+td}^{\text{SL}}, \cdots, 0)$$
(22)

where *td* means the dispatch time of the shiftable electric/gas load;  $P_{ts}^{SL}$  means the shiftable electric/gas load in the period *ts* before it participates in the dispatch.

The power distribution vector after the participation of shiftable load in the dispatch is as follows:

$$L_{\text{after}}^{\text{SL}} = (0, \cdots, P_{ts'}^{\text{SL}}, P_{ts'+1}^{\text{SL}}, \cdots, P_{ts'+td}^{\text{SL}}, \cdots, 0)$$
(23)

Shiftable electrical/gas loads need to meet dispatch interval constraints as follows:

$$ts^{\min} < ts' < ts' + td < te^{\max} \tag{24}$$

where  $ts^{\min}$  and  $te^{\max}$  mean the lower and upper limits of the allowable dispatch interval for shiftable load, respectively.

### 2.1.5. Community Electric Vehicle Model on the Energy Demand Side

1

Electric vehicles rely on onboard batteries to participate in power dispatch, and while participating in dispatch, they have to meet the trip power demand of users, so they need to meet the constraints as shown in Equations (25) and (26) on top of the battery-related constraints.

$$\delta_t^{\text{EVch}} + \delta_t^{\text{EVdis}} \le \delta_t^{\text{V2G}} \tag{25}$$

$$E_{tg-1}^{\rm EV} - E_{\min}^{\rm EV} \ge s_{tg}^{\rm EV} \zeta^{\rm EV}$$
<sup>(26)</sup>

where  $\delta_t^{V2G}$  is a binary state variable, which indicates the connection state of the electric vehicle to the grid, and the constraint is to restrict the electric vehicle to charge and discharge only when it is connected to the grid, and cannot be in the charge and discharge state at the same time;  $\delta_t^{EVch}$  and  $\delta_t^{EVdis}$  are the charging and discharging states of the electric vehicle at time *t*, respectively; *tg* means the electric vehicle trip moment;  $E_{tg-1}^{EV}$  means the electric vehicle pre-trip storage;  $E_{min}^{EV}$  indicates the minimum capacity of an electric vehicle;  $s_{tg}^{EV}$  means the electric vehicle trip mileage at the period tg;  $\zeta^{EV}$  means the electric vehicle power consumption per unit mileage.

### 2.2. Agent Model

# 2.2.1. Markov Decision Process for CIES Dispatch

The dispatch process of CIES can be represented with a Markov Decision Process (MDP) [25–27], and the MDP can be described with a five-tuple (*S*, *A*, *P*,  $\gamma$ , *R*): where *S* means the observation space; *A* means the agent action space; *P* is the state transfer probability, i.e., the probability of executing action  $a_1$  in state  $s_1$  and the state transforming to  $s_2$ ; *R* means the reward given by the environment after the agent makes the action; and  $\gamma$  means the discount factor, which means the degree of influence of the reward obtained in future periods on the cumulative reward.

The community-integrated energy system dispatch cycle is 24 h, and the agent needs to make dispatch actions from the first period after observing the environmental state until the last period of the dispatch cycle, making a total of 24 decisions of dispatch actions. During the dispatch cycle, the state shifts once for each dispatch action made by the agent [28], as shown in Figure 3.



Figure 3. Schematic diagram of MDP state transition in CIES dispatch cycle.

Taking into account the discount factor  $\gamma$ , the return obtained by the agent at time *t* of the dispatch cycle can be described as follows [29]:

$$U_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \dots + \gamma^{23-t} R_{24}$$
(27)

The training goal of an agent is to learn an optimal policy that maximizes the return in the dispatch cycle.

### 2.2.2. SAC Deep Reinforcement Learning Algorithm

The SAC (Soft Actor–Critic) algorithm is a deep reinforcement learning algorithm based on the maximum entropy theory. Compared with the deterministic policy algorithm that selects the action with the largest action value in each step of the policy, the SAC algorithm introduces an entropy term in the policy objective and selects the action with the largest sum of action value and entropy term in each step of the policy, which has the characteristics of policy randomization, stronger generalization, robustness, and exploration ability compared with general deep reinforcement learning algorithms, and can avoid premature convergence to local optima [30,31]. The optimal strategy after introducing the entropy term is as follows:

$$\pi_{\text{soft}}^* = \operatorname{argmax}_{\pi} \operatorname{E}_{(s_t, a_t) \sim \rho_{\pi}} \left[ \sum_{t=0}^T \gamma^t R(s_t, a_t) + \alpha H(\pi(\bullet|s_t)) \right]$$
(28)

where  $\alpha$  is the temperature parameter, which is the weighting factor of the entropy term. The entropy term in Equation (28) is expressed as

$$H(\pi(\bullet|s_t)) = -E_\pi \log \pi(a'|s_t) \tag{29}$$

In addition to the policy function, the SAC algorithm also introduces entropy terms in the action value function and state value function, as shown in Equations (30) and (31), respectively.

$$Q_{\text{soft}}(s_t, a_t) = E_{s_{t+1}, a_{t+1}}[R(s_t, a_t) + \gamma Q(s_{t+1}, a_{t+1}) - \alpha \log(\pi(a_{t+1}|s_{t+1}))]$$
(30)

$$V_{\text{soft}}(s_t) = E_{a_t}[Q(s_t, a_t) - \alpha \log \pi(a_t|s_t)]$$
(31)

2.2.3. Agent Observation Space

The observation space is the CIES state information needed by the agent in the decisionmaking process, which can be expressed as follows:

$$S = [P_t^{\text{Ren}}, L_t^e, L_t^g, P_{t-1}^{\text{GT}}, \rho_t^e, \rho_t^g, T_t^{\text{in}}, E_t^{\text{AC}}, E_t^{\text{ES}}, E_t^{\text{CS}}, E_t^{\text{V2G}}, \delta_t^{\text{e,RL,aseq}}, \delta_t^{\text{g,RL,aseq}}, \delta_t^{\text{e,RL,aall}}, \delta_t^{\text{g,RL,aall}}]$$
(32)

where  $P_t^{\text{Ren}}$  means renewable energy power output;  $L_t^{\text{e}}$  and  $L_t^{\text{g}}$  mean electric load and natural gas load;  $\rho_t^{\text{e}}$  and  $\rho_t^{\text{g}}$  mean real-time electricity price and natural gas price;  $E_t^{\text{AC}}$ ,  $E_t^{\text{ES}}$ ,  $E_t^{\text{GS}}$ , and  $E_t^{\text{EV}}$  mean air conditioner, battery, gas tank, and electric vehicle capacity;  $\delta_t^{\text{V2G}}$  is the binary variable of whether the electric vehicle is connected to the grid or not;  $\delta_t^{\text{e,RL,aseq}}$  and  $\delta_t^{\text{g,RL,aseq}}$  mean the period of time for which the electric/gas load has been continuously curtailed, respectively;  $\delta_t^{e,RL,aall}$  and  $\delta_t^{g,RL,aall}$  mean the total number of times the electric/gas load has been curtailed, respectively.

## 2.2.4. Agent Action Space

The action space of an agent is the control variable that needs to be optimized to achieve CIES dispatch, which can be expressed as follows:

$$A = [P_t^{dc,ES}, P_t^{dc,GS}, P_t^{dc,EV}, P_t^{GT}, P_t^{P2G}, H_t^{ACc}, H_t^{ACrs}, \beta_t^{e,RL}, \beta_t^{g,RL}, \beta_t^{e,SL}, \beta_t^{g,SL}, P_t^{e,RL}, P_t^{g,RL}]$$
(33)

where  $P_t^{\text{dc,ES}}$ ,  $P_t^{\text{dc,GS}}$ , and  $P_t^{\text{dc,EV}}$  means the discharge/charge power of the battery, gas storage tank, and electric vehicle, when taking a positive value for discharging, and vice versa for charging;  $H_t^{\text{ACrs}}$  means the discharge/charge volume of the cooling storage tank;  $\beta_t^{\text{e,RL}}$ ,  $\beta_t^{\text{g,RL}}$ ,  $\beta_t^{\text{e,SL}}$ , and  $\beta_t^{\text{g,SL}}$  mean the incentive factors for curtailable electric/gas load and shiftable electric/gas load, respectively;  $P_t^{\text{e,RL}}$  and  $P_t^{\text{g,RL}}$  mean the expected curtailments in curtailable electric load and gas load, respectively.

# 2.2.5. Agent Reward Function

Low-carbon economic dispatch refers to dispatch with the objective function of maximizing net benefits when considering the equivalent economic costs associated with carbon emissions and minimizing carbon dioxide emissions during dispatch operations. The objective function of CIES optimal dispatch is to maximize the CIES net revenue. Based on the objective function of CIES dispatch, the reward function of the agent can be expressed as follows:

$$REW_t = I_t^{\text{sell,e}} + I_t^{\text{sell,g}} - C_t^{\text{CO2}} - C_t^{\text{GTpol}} - C_t^{\text{buy}} - C_t^{\text{RL}} - C_t^{\text{SL}} - CF_t$$
(34)

where  $I_t^{\text{sell},e}$  and  $I_t^{\text{sell},g}$  mean CIES revenue from electricity/gas sales, respectively;  $C_t^{\text{CO2}}$  means CIES carbon trading costs;  $C_t^{\text{GTpol}}$  means gas turbine pollution emission costs;  $C_t^{\text{buy}}$  means CIES purchased electricity/gas costs;  $C_t^{\text{RL}}$  means dispatch costs for load curtailment;  $C_t^{\text{SL}}$  means dispatch costs for load shifting;  $CF_t$  means action out-of-limit penalty costs.

(1) CIES revenue from electricity sales

$$I_t^{\text{sell,e}} = P_t^{\text{sell,in,e}} \rho_t^{\text{sell,in,e}} \Delta t + P_t^{\text{sell,out,e}} \rho_t^{\text{sell,out,e}} \Delta t$$
(35)

where  $P_t^{\text{sell,in,e}}$  and  $P_t^{\text{sell,out,e}}$  mean the power sold using CIES to the users and energy market at time *t*, respectively;  $\rho_t^{\text{sell,in,e}}$  and  $\rho_t^{\text{sell,out,e}}$  mean the price of electricity sold using CIES to the users and energy market at time *t*, respectively.

(2) CIES revenue from gas sales

$$I_t^{\text{sell,g}} = P_t^{\text{sell,g}} \rho_t^{\text{sell,g}} \Delta t \tag{36}$$

where  $P_t^{\text{sell,g}}$  means the volume of gas sold using CIES to users at time *t*;  $\rho_t^{\text{sell,g}}$  means the price of gas sold using CIES at time *t*.

(3) Cost of gas turbine pollution emissions

The pollution emission of gas turbines is mainly considered in this paper for sulfur oxides,  $SO_X$ , and nitrogen oxides,  $NO_X$ , as follows:

$$C_t^{\text{GTpol}} = P_t^{\text{GT}} m^{\text{GTSO}_{\chi}} c^{\text{SO}_{\chi}} \Delta t + P_t^{\text{GT}} m^{\text{GTNO}_{\chi}} c^{\text{NO}_{\chi}} \Delta t$$
(37)

where  $m^{\text{GTSO}_X}$  and  $m^{\text{GTNO}_X}$  are the pollutant emission coefficients;  $c^{\text{SO}_X}$  and  $c^{\text{NO}_X}$  are the pollutant emission unit cost coefficients of SO<sub>X</sub> and NO<sub>X</sub>, respectively.

(4) Cost of Carbon trading

The carbon allowances that CIES needs to purchase are as follows:

$$E_t^{\rm I} = \left(\sum_{t=1}^T P_t^{\rm GT} m^{\rm GTCO2} + \sum_{t=1}^T P_t^{\rm buy} m^{\rm gridCO2} - \sum_{t=1}^{24} k^{\rm P2G} P_t^{\rm P2G}\right) - \left(e^{\rm GT} \sum_{t=1}^T P_t^{\rm GT} + e^{\rm grid} \sum_{t=1}^T P_t^{\rm buy}\right)$$
(38)

where  $m^{\text{GTCO2}}$  and  $m^{\text{gridCO2}}$  are the carbon emission coefficients of the gas turbine and grid, respectively;  $k^{\text{P2G}}$  is the CO<sub>2</sub> absorption coefficient of the P2G equipment;  $e^{\text{GT}}$  and  $e^{\text{grid}}$  are the unit carbon emission allowances for the gas turbine and grid, respectively.

The cost of purchasing CO2 allowances [32] can be expressed as follows:

$$C_{t}^{\text{CO2}} = \begin{cases} \rho E_{t}^{\text{I}} & E_{t}^{\text{I}} \leq l \\ \rho (1+v)(E_{t}^{\text{I}}-l) + \rho l & l \leq E_{t}^{\text{I}} \leq 2l \\ \rho (1+2v)(E_{t}^{\text{I}}-2l) + \rho (2+v)l & 2l \leq E_{t}^{\text{I}} \leq 3l \\ \rho (1+3v)(E_{t}^{\text{I}}-3l) + \rho (3+3v)l & 3l \leq E_{t}^{\text{I}} \leq 4l \\ \rho (1+4v)(E_{t}^{\text{I}}-4l) + \rho (4+6v)l & E_{t}^{\text{I}} \geq 4l \end{cases}$$
(39)

where  $\rho$  is the carbon trading base price; v is the price growth rate; and l is the stepped interval of carbon price growth.

(5) Cost of purchasing electricity/gas

$$C_t^{\text{buy}} = P_t^{\text{buy}} \rho_t^{\text{buy}} \Delta t \tag{40}$$

where  $P_t^{\text{buy}}$  means the amount of electricity/gas purchased using CIES;  $\rho_t^{\text{buy}}$  means the price of electricity/gas purchased using CIES.

(6) Cost of dispatch curtailable electric/gas load

$$C_t^{\rm RL} = c^{\rm RL} \beta_t^{\rm RL} \delta_t^{\rm RL} P_t^{\rm RL} \Delta t \tag{41}$$

(7) Cost of dispatch shiftable electric/gas load

$$C_t^{\rm SL} = c^{\rm SL} \beta_t^{\rm SL} \delta_t^{\rm SL} P_t^{\rm SL} \Delta t \tag{42}$$

# 3. Model Training

### 3.1. Construction of the Training Scenario Set

To enhance the generalization performance of the agent in the presence of uncertainties in source, load, weather, and electric vehicle trips, the training scenario set of the agent was generated using a Latin hypercube sampling method based on the idea of stratification to generate 300 scenario sets [33,34], each including renewable energy power output, electric/gas load, outdoor temperature, and electric vehicle trip plan.

Wind power output depends mainly on natural wind speed and is described with Weibull distribution [35]; PV output and outdoor temperature depend mainly on solar radiation and are described with Beta distribution [36]; electric/gas load and electric vehicle trip plan are described with normal distribution [37,38].

## 3.2. Construction of the Training Scenario Set

On the basis of the described reinforcement learning model, the agent was trained using a comprehensive set of training scenarios. To analyze the effect of the maximum entropy strategy on the model, the models with temperature parameters of 0.05 and 0 were trained, and the return values and carbon trading costs at 2000 episodes are shown in Figure 4.



Figure 4. Training curve under different temperature parameters: (a) return value; (b) carbon trading costs.

As can be seen from Figure 4, the model with a temperature parameter of 0 converges faster because it does not consider the entropy term; the model with a temperature parameter of 0.05 considers the entropy term and begins to converge only after 250 episodes, and the training curve at the convergence stage fluctuates more, but the model is more capable of finding the optimal solution and converges to a higher return value. The model with a temperature parameter of 0.05 achieved a higher low-carbon goal.

# 4. Case Studies

### 4.1. Experimental Setup

The parameters of each power component of CIES are shown in Table 1. The real-time electric/gas prices are shown in Table 2. The parameters of the agent are shown in Table 3. The experimental data are shown in Figure 5.

Parameter	Value	Parameter	Value	Parameter	Value
$p_{\text{max}}^{\text{GT}}(\text{kW})$	100	$E_{\max}^{\text{ES,AC}}$ (kWh)	50	$\eta^{\mathrm{ESch,BA}}$	95%
$p_{\min}^{\text{GT}}(\text{kW})$	10	$E_{\min}^{\text{ES,AC}}$ (kWh)	0	$\eta^{\mathrm{ESdis,BA}}$	95%
$\Delta P^{\text{GTmax}}(\text{kW/h})$	70	$E_{\max}^{\text{ES,EV}}$ (kWh)	20	$\eta^{\text{ESch,GS}}$	95%
$m^{\text{GTSO}_X}$	0.0098	$E_{\min}^{\text{ES,EV}}$ (kWh)	6	$\eta^{\text{ESdis,GS}}$	95%
$m^{\text{GTNO}_X}$	0.543	$P_{\max}^{\text{ESch,BA}}(\text{kW})$	30	$\eta^{\text{ESch,EV}}$	95%
ag	0.11	$P_{\max}^{\text{ESdis,BA}}(\text{kW})$	30	$\eta^{\mathrm{ESdis,EV}}$	95%
$b_{g}$	2	$P_{\rm max}^{\rm ESch,GS}({\rm m}^3)$	50	$\mu^{ACc}$	2.6
Cg	0	$P_{\rm max}^{\rm ESdis,GS}({ m m}^3)$	50	$\mu^{ACs}$	0.0045
$E_{\rm max}^{\rm ES,BA}$ (kWh)	100	$P_{\max}^{\text{ESch,AC}}(\text{kW})$	20	$\mu^{ACr}$	0.0038
$E_{\min}^{\text{ES,BA}}(\text{kWh})$	10	$P_{\max}^{\text{ESdis,AC}}(\text{kW})$	20	$H_{\max}^{ACc}$	25
$E_{\text{max}}^{\text{ES,GS}}(\text{m}^3)$	150	$P_{\max}^{\text{ESch,EV}}(\text{kW})$	8	$\zeta^{ m EV}$	0.241
$E_{\rm min}^{\rm ES,GS}({ m m}^3)$	10	$P_{\max}^{\text{ESdis,EV}}(\text{kW})$	8		

Table 1. Parameters of each power component of CIES.

Table 2. Parameters of real-time electricity/gas price.

Time Period	Electricity Price (USD/kWh)	Natural Gas Prices (USD/m <sup>3</sup> )
Peak section	0.143	0.043
Flat section	0.114	0.036
Valley section	0.086	0.029



Table 3. Parameters of SAC agent.

Figure 5. Experimental data: (a) renewable energy and electrical load; (b) gas load.

In the DRUW, let the users' response damping coefficient be 1.0; then, the benchmark willingness curve of users' 24 h response to electric/gas demand is shown in Figure 6.



Figure 6. Demand response benchmark willingness curve.

To analyze the impact of the DRUW parameters proposed in this paper on the CIES dispatch, four experimental cases are established as shown in Table 4.

Table 4. Experimental cases of DRUW.

Case Index	Realization DRUW	Response Damping Coefficient	Demand Response Uncertainty Interval
1	×	_	_
2	$\checkmark$	1.0	[0.5,0.6]
3		1.3	[0.5,0.6]
4		1.0	[0.5–0.7]

# 4.2. Simulation Results

The dispatch results for the four cases are shown in Table 5.

Case Index	Electric Load Curtailment (kWh)	Gas Load Curtailment (m <sup>3</sup> )	Demand Response Compensation Costs (USD)	Cost of Electricity/Gas Purchase (USD)	CIES Net Revenue (USD)
1	0	0	0	183.34	1287.00
2	101.00	39.66	7.87	170.83	1368.32
3	2.29	3.35	2.13	187.55	1303.95
4	28.18	1.34	3.17	180.49	1316.06

Table 5. CIES dispatch results under DRUW cases.

As can be seen from Table 5, the CIES net revenues for case 2–case 4, where DRUW is implemented, are all higher than case 1, where DRUW is not implemented, due to the fact that the user's demand response participation allows for load curtailment during peak load periods and load shifting during low periods.

Comparing the results of case 3 and case 2, it can be seen that the user's participation in the electric/gas demand response decreases due to an increase in its response damping coefficient, resulting in a decrease in both the amount of electric/gas load curtailment and demand response compensation costs. In addition, the CIES needs to purchase energy from the market to meet the system energy supply/demand balance, the cost of purchased electric/gas increases by USD 16.72, and the net revenue decreases.

Comparing the results of case 4 and case 2, it can be seen that the uncertainty of user participation in the electric/gas demand response increases due to the increase in the user's demand response uncertainty interval and the increase in the upper limit of the interval, resulting in a decrease in both the amount of electric/gas load curtailment and demand response compensation costs. In addition, the increase in CIES's need to purchase energy from the market to meet the system energy supply/demand balance increases the cost of purchased electricity/gas by USD 9.66 and decreases the net revenue.

It can be seen that an increase in both the response damping coefficient and the uncertainty interval of the demand response leads to a decrease in demand response participation, which in turn leads to a decrease in CIES's net revenue.

### 4.3. Generalization Performance Analysis under Source and Load Uncertain Scenarios

The Monte Carlo method is used to generate source and load uncertain scenario sets with fluctuation rates of 5%, 10%, and 15%, respectively, and the scenario set includes renewable energy power output and electric/gas load, as shown in Figure 7.



Figure 7. Uncertain scenario set of renewable energy and electric/gas load.

The three uncertain scenarios in Figure 7 are optimally dispatched using the trained agent, and the dispatch results are shown in Figure 8, and the economic indicators related to the dispatch results are shown in Table 6. As can be seen from Figure 8 and Table 6, when there is uncertainty in the source and load, the agent can make corresponding decisions for



the uncertain environment, i.e., the optimal dispatch of CIES is achieved considering the uncertainty in the source and load.

Figure 8. Dispatch results for uncertain scenario.

<b>Table 6.</b> Economic indicators of CIES dispatch results under the source and 1	load uncertain scenarios.
---	---------------------------

Source and Load Fluctuation Rate	Cost of Electricity Purchase (USD)	Cost of Gas Purchase (USD)	Carbon Trading Costs (USD)	Demand Response Compensation Costs (USD)	CIES Net Revenue (USD)
5%	65.92	101.41	3.41	8.06	1403.97
10%	54.89	101.14	2.82	7.69	1392.19
15%	62.76	102.08	3.37	7.76	1392.87

# 4.4. Generalization Performance Analysis under Outdoor Temperature Uncertain Scenarios

The uncertain scenario sets of outdoor temperature with 5%, 10%, and 15% fluctuation rates were generated with the Monte Carlo method, as shown in Figure 9a. The dispatch results of indoor temperature are shown in Figure 10. As can be seen from Figure 9b, the indoor temperature in all three uncertain scenarios is limited to the required 25.5–27.5 °C during the 24 h dispatch cycle, satisfying the indoor temperature constraint.



**Figure 9.** Experimental data: (**a**) uncertain scenario set of outdoor temperature; (**b**) dispatch results of indoor temperature.





### 4.5. Generalization Performance Analysis under Uncertain Scenarios of Electric Vehicle Trips

Uncertainty about when electric vehicles will connect to and leave the grid due to the uncertainty of community users' daily trip plans, the agent needs to realize the optimal dispatch of EVs under the uncertain trip scenario. Let the number of EVs involved in the dispatch be 20, and consider the uncertainty of trip number, trip time, and trip distance. Establish three EV trip uncertain scenarios, as shown in Table 7. The results of EV dispatch under three uncertain scenarios are shown in Figure 10 and Table 8.

Table 7. Electric vehicle trip uncertain scenario.

Scenario Index	Number of Trips	Trip Time	Trip Mileage
1	1	7:00-16:00	24
2	1	9:00-18:00	20
2	2	7:00-13:00	12
3	Z	16:00-18:00	16

Table 8. Dispatch results under the uncertain scenarios of electric vehicle trips.

Scenario Index	Net Charging Volume (kWh)	Surplus Power Storage (kWh)	CIES Net Revenue (USD)
1	6.93	14.22	1368.92
2	5.00	15.18	1360.35
3	6.84	16.15	1355.24

As can be seen from Figure 10, in the three trip uncertain scenarios, the charging and discharging operations of the agent to the EVs only occur when electric vehicles are connected to the grid, and the charging hours in all three scenarios are concentrated in the low load hours from 0:00 a.m. to 1:00 a.m., which can reduce the charging cost; meanwhile, the EV discharging hours are concentrated in the peak load hours from 21:00 p.m. to 22:00 p.m., which can release the stored surplus power and improve the net revenue.

As can be seen from Table 8, the actual storage capacity of EVs in all three scenarios before their respective trip time can meet the trip power demand, reflecting the good generalization performance of the agent to the uncertain scenarios of EV trips.

# 5. Conclusions

The uncertainty of the demand response is rarely considered in existing applied research using reinforcement learning methods for energy system dispatch. However, in the community-integrated energy system, the user's demand response behavior should be subject to uncertainty. In this paper, we develop a demand response model that takes into account the uncertainty of user behavior, and a multiple-uncertainty community-energysystem low-carbon economic dispatch model based on a deep reinforcement learning method is proposed. The proposed model considers the uncertainties of various factors such as renewable energy, electric/gas load, temperature, and electric vehicle trip, and proposes a demand response model based on the user's willingness to the uncertainty of the user's demand response behavior, which is combined with the SAC reinforcement learning method to realize the low-carbon economic dispatch of a community-integrated energy system under multiple uncertainties. The simulation results show the following:

(1) In the DRUW, the increase in both the response damping coefficient and demand response uncertainty interval leads to the decrease in demand response participation, resulting in the decrease in operating net revenue of the community-integrated energy system.

(2) The trained agent has good adaptability to multiple uncertainties in the communityintegrated energy system and has good generalization performance in the scenarios with uncertainty in the user's demand response behavior as well as uncertainty in source, load, outdoor temperature, and electric vehicle trips.

This paper's model of demand response uncertainty and reinforcement-learning-based low-carbon economic dispatch in CIES with demand response uncertainty may positively influence future related research, but at the same time, there are some limitations and it is worthy of further improvement. For the user demand response uncertainty results, we mainly consider the two states of response and non-response, and in the future, a continuous response modeling the response can also be used for the demand response. A simple linear approximation is used for the description of the demand response uncertainty curve in this paper, which can be combined with Monte Carlo or other methods in the future to provide a more accurate modeling of the user's demand response uncertainty.

**Author Contributions:** Conceptualization, M.M.; Data curation, X.X.; Funding acquisition, Z.Y.; Methodology, M.M.; Resources, X.X.; Software, M.M.; Supervision, Z.Y.; Validation, M.M. and X.X.; Visualization, Y.W.; Writing—original draft, M.M.; Writing—review and editing, Y.W. 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 State Grid, HUST-State Grid Future of Grid Institute, grant number: 52130421N00B.

**Data Availability Statement:** The numerical data used to support the findings of this study are included within the article.

**Acknowledgments:** The authors would like to thank the reviewers for their valuable comments on this research.

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

# References

- 1. Wang, C.S.; Lv, C.X.; Li, P.; Song, G.Y.; Li, S.Q.; Xu, X.D.; Wu, J.Z. Modeling and optimal operation of community integrated energy systems: A case study from China. *Appl. Energy* **2018**, *230*, 1242–1254. [CrossRef]
- Zhou, Y.Z.; Wei, Z.N.; Sun, G.Q.; Cheung, K.W.; Zang, H.X.; Chen, S. A robust optimization approach for integrated community energy system in energy and ancillary service markets. *Energy* 2018, 148, 1–15. [CrossRef]
- Li, Y.; Wang, B.; Yang, Z.; Li, J.Z.; Chen, C. Hierarchical stochastic scheduling of multi-community integrated energy systems in uncertain environments via Stackelberg game. *Appl. Energy* 2022, 308, 118392. [CrossRef]
- Zhang, Y.Y.; Zhao, H.R.; Li, B.K.; Wang, X.J. Research on dynamic pricing and operation optimization strategy of integrated energy system based on Stackelberg game. *Int. J. Electr. Power Energy Syst.* 2022, 143, 108446. [CrossRef]
- Yang, S.B.; Tan, Z.F.; Zhou, J.H.; Xue, F.; Gao, H.D.; Lin, H.Y.; Zhou, F.A. A two-level game optimal dispatching model for the park integrated energy system considering Stackelberg and cooperative games. *Int. J. Electr. Power Energy Syst.* 2021, 130, 106959. [CrossRef]
- Gao, H.; Li, Z.S. A Benders Decomposition Based Algorithm for Steady-State Dispatch Problem in an Integrated Electricity-Gas System. *IEEE Trans. Power Syst.* 2021, 36, 3817–3820. [CrossRef]
- Luo, X.J.; Fong, K.F. Development of integrated demand and supply side management strategy of multi-energy system for residential building application. *Appl. Energy* 2019, 242, 570–587. [CrossRef]
- Li, P.; Wang, Z.X.; Wang, N.; Yang, W.H.; Li, M.Z.; Zhou, X.C.; Yin, Y.X.; Wang, J.H.; Guo, T.Y. Stochastic robust optimal operation of community integrated energy system based on integrated demand response. *Int. J. Electr. Power Energy Syst.* 2021, 128, 106735. [CrossRef]

- Liu, G.; Qin, Z.F.; Diao, T.Y.; Wang, X.W.; Wang, P.M.; Bai, X.Q. Low carbon economic dispatch of biogas-wind-solar renewable energy system based on robust stochastic optimization. *Int. J. Electr. Power Energy Syst.* 2022, 139, 108069. [CrossRef]
- Yan, R.J.; Wang, J.J.; Wang, J.H.; Tian, L.; Tang, S.Q.; Wang, Y.W.; Zhang, J.; Cheng, Y.L.; Li, Y. A two-stage stochastic-robust optimization for a hybrid renewable energy CCHP system considering multiple scenario-interval uncertainties. *Energy* 2022, 247, 123498. [CrossRef]
- Li, X.Q.; Zhang, L.Z.; Wang, R.Q.; Sun, B.; Xie, W.J. Two-Stage Robust Optimization Model for Capacity Configuration of Biogas-Solar-Wind Integrated Energy System. *IEEE Trans. Ind. Appl.* 2023, 59, 662–675. [CrossRef]
- 12. Shuvo, S.S.; Yilmaz, Y. Home Energy Recommendation System (HERS): A Deep Reinforcement Learning Method Based on Residents' Feedback and Activity. *IEEE Trans. Smart Grid* **2022**, *13*, 2812–2821. [CrossRef]
- 13. Lu, R.Z.; Bai, R.C.; Luo, Z.; Jiang, J.H.; Sun, M.Y.; Zhang, H.T. Deep reinforcement learning-based demand response for smart facilities energy management. *IEEE Trans. Ind. Electron.* 2021, *69*, 8554–8565. [CrossRef]
- Ren, M.F.; Liu, X.F.; Yang, Z.L.; Zhang, J.H.; Guo, Y.J.; Jia, Y.B. A novel forecasting based scheduling method for household energy management system based on deep reinforcement learning. *Sustain. Cities Soc.* 2022, 76, 103207. [CrossRef]
- Lai, B.C.; Chiu, W.Y.; Tsai, Y.P. Multiagent Reinforcement Learning for Community Energy Management to Mitigate Peak Rebounds Under Renewable Energy Uncertainty. *IEEE Trans. Emerg. Top. Comput. Intell.* 2022, 6, 568–579. [CrossRef]
- Ye, Y.J.; Qiu, D.W.; Wu, X.D.; Strbac, G.; Ward, J. Model-Free Real-Time Autonomous Control for a Residential Multi-Energy System Using Deep Reinforcement Learning. *IEEE Trans. Smart Grid.* 2020, 11, 3068–3082. [CrossRef]
- Ding, H.Y.; Xu, Y.; Hao, B.C.S.; Li, Q.Q.; Lentzakis, A. A safe reinforcement learning approach for multi-energy management of smart home. *Electr. Power Syst. Res.* 2022, 210, 108120. [CrossRef]
- 18. Xue, X.; Wang, J.X.; Zhang, Y.; Yong, W.Z.; Qi, J.; Li, H.T. Model-data-event based community integrated energy system low-carbon economic scheduling. *Renew. Sustain. Energy Rev.* 2023, *182*, 113379. [CrossRef]
- Qiu, Y.; Zhou, S.Y.; Xia, D.; Gu, W.; Sun, K.Y.; Han, G.Y.; Zhang, K.; Lv, H.K. Local integrated energy system operational optimization considering multi-type uncertainties: A reinforcement learning approach based on improved TD3 algorithm. *IET Renew. Power Gener.* 2023, 17, 2236–2256. [CrossRef]
- Hong, S.H.; Lee, H.S. Robust Energy Management System with Safe Reinforcement Learning Using Short-Horizon Forecasts. IEEE Trans. Smart Grid. 2023, 14, 2485–2488. [CrossRef]
- 21. Liu, Y.; Liu, T.Y. Research on System Planning of Gas-Power Integrated System Based on Improved Two-Stage Robust Optimization and Non-Cooperative Game Method. *IEEE Access* 2021, *9*, 79169–79181. [CrossRef]
- 22. Li, G.Q.; Zhang, R.F.; Jiang, T.; Chen, H.H.; Bai, L.Q.; Li, X.J. Security-constrained bi-level economic dispatch model for integrated natural gas and electricity systems considering wind power and power-to-gas process. *Appl. Energy* 2017, 194, 696–704. [CrossRef]
- 23. Sun, G.Q.; Qian, W.H.; Huang, W.J.; Xu, Z.; Fu, Z.X.; Wei, Z.N.; Chen, S. Stochastic Adaptive Robust Dispatch for Virtual Power Plants Using the Binding Scenario Identification Approach. *Energies* **2019**, *12*, 1918. [CrossRef]
- 24. Li, Y.; Zou, Y.; Tan, Y.; Cao, Y.J.; Liu, X.D.; Shahidehpour, M.; Tian, S.M.; Bu, F.P. Optimal Stochastic Operation of Integrated Low-Carbon Electric Power, Natural Gas, and Heat Delivery System. *IEEE Trans. Sustain. Energy* **2018**, *9*, 273–283. [CrossRef]
- Zhang, B.; Hu, W.H.; Cao, D.; Huang, Q.; Chen, Z.; Blaabjerg, F. Deep reinforcement learning-based approach for optimizing energy conversion in integrated electrical and heating system with renewable energy. *Energy Convers. Manag.* 2019, 202, 112199. [CrossRef]
- 26. Dong, J.; Wang, H.X.; Yang, J.Y.; Lu, X.Y.; Gao, L.; Zhou, X.R. Optimal Scheduling Framework of Electricity-Gas-Heat Integrated Energy System Based on Asynchronous Advantage Actor-Critic Algorithm. *IEEE Access* **2021**, *9*, 139685–139696. [CrossRef]
- 27. Zhang, B.; Hu, W.H.; Cao, D.; Huang, Q.; Chen, Z.; Blaabjerg, F. Economical operation strategy of an integrated energy system with wind power and power to gas technology—A DRL-based approach. *IET Renew. Power Gener.* 2020, 14, 3292–3299. [CrossRef]
- Boutilier, C.; Dean, T.; Hanks, S. Decision-theoretic planning: Structural assumptions and computational leverage. J. Artif. Intell. Res. 1999, 11, 1–94. [CrossRef]
- 29. Kober, J.; Bagnell, J.A.; Peters, J. Reinforcement learning in robotics: A survey. Int. J. Rob. Res. 2013, 32, 1238–1274. [CrossRef]
- 30. Han, X.Y.; Mu, C.X.; Yan, J.; Niu, Z.Y. An autonomous control technology based on deep reinforcement learning for optimal active power dispatch. *Int. J. Electr. Power Energy Syst.* **2023**, *145*, 108686. [CrossRef]
- Xiao, B.Y.; Yang, W.W.; Wu, J.M.; Walker, P.D.; Zhang, N. Energy management strategy via maximum entropy reinforcement learning for an extended range logistics vehicle. *Energy* 2022, 253, 124105. [CrossRef]
- 32. Yang, X.H.; Zhang, Z.L.; Mei, L.H.; Wang, X.P.; Deng, Y.H.; Wei, S.; Liu, X.P. Optimal configuration of improved integrated energy system based on stepped carbon penalty response and improved power to gas. *Energy* **2023**, *263*, 125985. [CrossRef]
- 33. Mei, F.; Zhang, J.T.; Lu, J.X.; Lu, J.J.; Jiang, Y.H.; Gu, J.Q.; Yu, K.; Gan, L. Stochastic optimal operation model for a distributed integrated energy system based on multiple-scenario simulations. *Energy* **2021**, *219*, 119629. [CrossRef]
- Wang, J.J.; Huo, S.J.; Yan, R.J.; Cui, Z.H. Leveraging heat accumulation of district heating network to improve performances of integrated energy system under source-load uncertainties. *Energy* 2022, 252, 124002. [CrossRef]
- 35. Wais, P. A review of Weibull functions in wind sector. Renew. Sustain. Energy Rev. 2017, 70, 1099–1107. [CrossRef]
- Ettoumi, F.Y.; Mefti, A.; Adane, A.; Bouroubi, M.Y. Statistical analysis of solar measurements in Algeria using beta distributions. *Renew. Energy* 2002, 26, 47–67. [CrossRef]

- 37. Das, S.; Malakar, T. Estimating the impact of uncertainty on optimum capacitor placement in wind-integrated radial distribution system. *Int. Trans. Electr. Energy Syst.* 2020, 30, e12451. [CrossRef]
- 38. Liu, D.Q. Cluster Control for EVs Participating in Grid Frequency Regulation by Using Virtual Synchronous Machine with Optimized Parameters. *Appl. Sci.* 2019, *9*, 1924. [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.