energies

MDPI

Article

Incentive-Based Integrated Demand Response and Distributionally Robust Optimization for Enabling Economic Dispatch of Integrated Energy Systems

Rongjun Chen, Yang Liu * and Huaqiang Li

College of Electrical Engineering, Sichuan University, Chengdu 610065, China;
chenrongjun@stu.scu.edu.cn (R.C.); lihuaqiang@scu.edu.cn (H.L.)
* Correspondence: yang.liu@scu.edu.cn; Tel.: +86-18683570439

Abstract

Source-load fluctuations have significant impacts on the economic performance of integrated energy system (IES) dispatch. Existing studies mainly focus on alleviating the supply-demand matching pressure using system-side operation optimization and resource configuration. Therefore, two critical issues arise: the first is that the regulation potential of demand-side resources has not been fully exploited, and the second is that coordinated regulation between the system side and the demand side remains insufficient. To address these issues, this paper presents a two-stage optimization framework for IESs considering an incentive-based integrated demand response (IBIDR) model. Firstly, a two-stage optimization framework consisting of day-ahead and intra-day stages is established, which integrates the demand-side IBIDR and the system-side economic dispatch on a global scale. Secondly, in the day-ahead stage, the IBIDR model—considering multi-energy demand coupling and the bidirectional response characteristics of consumers—is presented. A multi-agent proximal policy optimization (MAPPO) algorithm is employed by the model to fully explore the regulation potential of the demand-side resources. Then, the model enables the generation of incentive strategies for the demand-side resources. Finally, in the intra-day stage, a distributionally robust optimization (DRO) model is further presented to address the source-load uncertainties, in which a wind power generation (WPG) ambiguity set constructed using imprecise Dirichlet theory (IDT) is adopted to represent the WPG uncertainty. The experimental results demonstrate that the presented framework can enhance the regulation capability and economic performance of the IES under source-load fluctuations. The IBIDR model can accurately characterize the multi-energy response characteristics of consumers. The MAPPO algorithm generates effective incentive strategies to encourage consumer participation in the IBIDR. The IDT ambiguity set improves WPG uncertainty characterization and reduces dispatch conservativeness.

Keywords: integrated energy system; integrated demand response; proximal policy optimization; distributionally robust optimization; uncertainty

Check for updates

Academic Editor: Mihaela Simionescu

Received: 28 August 2026
Revised: 9 October 2026
Accepted: 13 October 2026
Published: 17 October 2026

Copyright: © 2026 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.

1. Introduction

With the global advancement in low-carbon energy transition, the wind power penetration level in the power grid has been continuously increasing. As a result, wind power has gradually become an important component of modern power grids [1–3]. However, affected by complicated meteorological conditions, wind power generation (WPG) exhibits

significant fluctuation characteristics. Meanwhile, consumer energy consumption patterns have become increasingly diverse, leading to more complex variations in multi-energy load demands. Therefore, dynamic variations on both the source and load sides increase the difficulty of prediction, which may affect the security and stability of power system operations [4,5]. As an important entity carrying the interactions between source and load, an integrated energy system (IES) enables the conversion and utilization of electricity, heat, and gas. The advantages of multi-energy complementarity and flexible regulation facilitate renewable energy accommodation and enhance the regulation capability to handle the load variations to some extent [6–9]. However, as the amount of source–load fluctuations increases, the regulation demand may exceed the regulation capacity and even propagate through multi-energy coupling relationships to the entire IES, further increasing the difficulty of supply-demand coordination within the IES. Such issues significantly affect the economic dispatch performance of IESs. Therefore, how to further enhance the regulation capability of IESs to handle source-load fluctuations requires urgent research. In this context, demand-side resources can actively reshape load behaviors and provide additional regulation flexibility. However, their regulation potential needs to be effectively exploited and coordinated with system-side resources. Moreover, WPG uncertainty and prediction deviations in multi-energy load demands cannot be completely eliminated through demand-side regulation. These remaining deviations may result in a mismatch between the day-ahead dispatch schedule and actual intra-day operation. Accordingly, improving the regulation capability of the IES requires the coordination of demand-side active regulation and system-side uncertainty management.

1.1. Literature Review

In terms of improving the capability of regulation demand, solutions can be explored from the perspectives of system planning, operation, and regulation mechanisms. From the system-side perspective, existing studies mainly focus on operation optimization and resource configuration. Liu et al. [10] and Wu et al. [11] established optimal dispatch models to exploit the regulation capability of flexible resources within the system. Zhang et al. [12] presented a differentiated energy storage configuration optimization model for both the source and load sides of microgrids to alleviate renewable energy and load fluctuations. Although these models can improve the regulation capability of the system to some extent, they mainly rely on system-side resources to passively handle existing source-load fluctuations. Demand-side resources which are able to actively shape load behaviors remain insufficiently exploited. Therefore, researchers have presented a number of novel regulation mechanisms from the perspective of the demand side. Caliano et al. [13] developed a microgrid energy management system to achieve active control over flexible loads, enhancing demand-side regulation capability. However, this model mainly takes the economic operation of the microgrid as its optimization objective. Insufficient attention is paid to the voluntary participation of consumers, which may affect their sustained participation. Salim et al. [14] presented a peer-to-peer transaction-based coordination method for community energy sharing, which promotes energy trading among consumers through market pricing mechanisms and improves economic benefits for consumers. However, the price signals are mainly determined by market supply and demand. They lack guidance from system regulation demands, which limits the exploitation of the regulation potential of demand-side resources. Therefore, it is of great significance to develop a demand-side regulation mechanism that can improve system regulation capability while ensuring consumer benefits.

As an important mechanism for active electrical demand-side regulation, demand response (DR) can guide consumers to change electricity consumption behaviors through

regulation signals, which facilitate the matching of demand-side resources with system operation requirements [15]. With the development of IESs, the traditional DR has gradually evolved into an integrated demand response (IDR), which can mainly be classified into a price-based integrated demand response (PBIDR) and an incentive-based integrated demand response (IBIDR) [16]. The former guides consumers to respond through price signals, making it difficult to fully satisfy the differentiated regulation requirements of consumers, and the latter can design flexible incentive strategies according to consumer response characteristics, enabling more targeted response regulation [17,18]. Consumers are also more inclined to accept direct economic incentives rather than indirect rewards reflected through price variations [19]. Therefore, an IBIDR can more effectively unlock the regulation potential of demand-side resources. However, an IBIDR is affected by multi-energy coupling relationships. It also requires an accurate characterization of differentiated response behaviors of consumers under incentive prices. To address these issues, Hu et al. [20] developed a consumer response model considering the price elasticity of demand and multi-energy coupling relationships, effectively characterizing the multi-energy response behaviors of consumers under incentive prices. Liu et al. [21] introduced data centers into IES operation optimization. They also constructed a multi-energy flow coupling network to characterize the differentiated response characteristics of consumers. However, their study is limited to unidirectional load reduction responses; bidirectional response characteristics involving both load increases and reductions remain insufficiently considered. This point not only prevents accurate characterization of the actual response process of consumers but may also restrict the effectiveness of the IBIDR model for improving system operation performance. Therefore, the IBIDR model still requires further refinement.

Meanwhile, the interactive decision between operators and consumers in IBIDRs significantly increases the difficulty of optimizing incentive strategies. Existing studies have attempted to address this issue using mathematical programming, game theory, and heuristic algorithms. Zheng et al. [22] modeled the IBIDR process as a bi-level stochastic programming model and derived optimal expectation decisions using an equivalent transformation. Gao et al. [23] presented a residential demand response analysis method based on evolutionary game theory, where incentive prices corresponding to equilibrium points were determined through game interactions. Li et al. [24] utilized the particle swarm optimization algorithm to optimize incentive strategies in IBIDRs. However, these methods generally determine incentive strategies through model-based optimization, equilibrium analysis, or iterative search under given operating conditions. Variations in system operation states may require the decision process to be recalculated or updated, which increases the difficulty of rapidly adjusting incentive strategies. In recent years, deep reinforcement learning (DRL) has become an important approach for serving incentive strategy optimization in IBIDRs due to its remarkable learning capability [25]. Zhang et al. [26] and Ahmed et al. [27] employed deep Q-network (DQN) to approximate action-value functions for strategy optimization. However, DQN is mainly applicable to discrete action spaces, while incentive strategies in the IBIDR usually involve continuous action spaces. To address this issue, Li et al. [28] and Yazar et al. [29] employed a deep deterministic policy gradient (DDPG) to optimize continuous incentive strategies. However, a DDPG algorithm may suffer from unstable convergence and policy oscillations in complicated high-dimensional environments. In contrast, the proximal policy optimization (PPO) algorithm exhibits better performance in terms of training stability and convergence efficiency [30]. Nevertheless, the IBIDR process in IESs usually involves cooperative decision among multiple agents. The above methods have difficulty simultaneously balancing local decisions and global coordination.

Additionally, WPG output is affected by natural conditions, which cannot be controlled through demand-side regulation. Furthermore, load fluctuations still remain. Therefore, the uncertainties caused by source-load fluctuations need to be further addressed at the system operation level. Stochastic optimization (SO) and robust optimization (RO) are commonly used approaches for addressing uncertainty issues [31,32]. Among them, SO generally relies on the probability distributions of uncertain variables, which are often difficult to obtain accurately [33]. In contrast, RO characterizes uncertain variables by constructing uncertainty sets, which avoids precise probability distributions [34]. Nevertheless, RO usually determines optimization decisions based on worst-case scenarios, which may lead to overly conservative results. To address this issue, Wu et al. [35] presented an adjustable robust optimization (ARO) model for implementing the optimized dispatch. By introducing adjustable parameters, the conservativeness of dispatch strategies can be flexibly controlled to satisfy practical dispatch requirements. However, the optimization performance of ARO is highly dependent on parameter settings. Inappropriate parameter selection may degrade the economic performance of dispatch. To further balance the economic performance and robustness of dispatch results, distributionally robust optimization (DRO) has been adopted. DRO constructs ambiguity sets based on historical data to characterize uncertain distributions. Dispatch performance using DRO highly depends on the construction of ambiguity sets. Alismail et al. [36] constructed an ambiguity set using multiple-order moment constraints to constrain the set of feasible probability distributions. However, in their study, moment constraints only capture statistical characteristics. The actual distribution shape of samples is ignored. This results in overly broad ambiguity sets, making it difficult to represent true distribution characteristics. Cao et al. [37] presented a Wasserstein distance-based ambiguity set to quantify the risk of energy supply shortages caused by renewable energy intermittency. Although this ambiguity set offers data-driven advantages, the Wasserstein ball tends to expand when the sample size is limited. This significantly increases the conservativeness of the optimization model. Therefore, constructing an ambiguity set that represents WPG uncertainty under limited sample conditions remains a challenge.

1.2. Contributions of the Work

To clearly highlight the specific contributions of this study, the main differences between this study and previous studies are summarized in Table 1. The following research gaps in previous studies remain to be addressed:

(i) An integrated framework that jointly considers system-side operation and demand-side response remains lacking.
(ii) Bidirectional response characteristics remain insufficiently characterized in IBIDR models. Solution methods for incentive strategy optimization have difficulty balancing local decisions with global coordination, limiting the full exploitation of demand-side regulation potential.
(iii) Representing WPG uncertainty through ambiguity sets under limited sample conditions without introducing excessive conservatism remains difficult.

Motivated by the above analysis, this study investigates whether coordinated regulation between the demand side and system side can improve IES operation under source-load fluctuations. It also examines whether demand-side regulation potential can be more effectively exploited and whether the conservativeness of WPG uncertainty modeling can be reduced under limited sample conditions. To address these issues, this paper presents an MAPPO-IDTDRO-based two-stage optimization framework. The main contributions of this paper are summarized as follows.

(1) A two-stage optimization framework consisting of day-ahead and intra-day stages is developed. In the day-ahead stage, incentive strategies in IBIDRs are optimized to encourage consumers to actively participate in demand response. In the intra-day stage, economic dispatch is performed to address source-load uncertainties. The coordination between the two stages enhances the regulation capability of IESs against supply-demand imbalances caused by source-load fluctuations. It also improves the economic performance of IES operation.

(2) An IBIDR model considering multi-energy coupling and bidirectional response characteristics is constructed to characterize the differentiated response behaviors of consumers. The MAPPO algorithm is employed to learn coordinated decision-making through multi-agent interactions, fully exploring the regulation potential of demand-side resources. Based on learning results, the IBIDR model generates incentive strategies to encourage active consumer participation and improve demand response effectiveness.

(3) Based on historical data samples, IDT is employed to construct an ambiguity set for WPG. By fully utilizing the statistical information contained in the samples, the ambiguity set is progressively refined as the sample size increases. This allows the uncertainty bounds to be adjusted according to the sample information available. The IDT ambiguity set is further incorporated into the DRO model to reduce conservativeness and improve the economic performance of IES dispatch.

Table 1. Comparison between this study and previous studies.

Reference	Optimization Scope	Demand-Side Regulation		System-Side Uncertainty Optimization	
		IBIDR Modeling	Incentive Strategy Optimization	Optimization Method	Uncertainty Modeling
Zhang et al. [12]	System side	×	×	×	×
Caliano et al. [13]; Salim et al. [14]	Demand side	×	×	×	×
Hu et al. [20]; Liu et al. [21]	Demand side	Multi-energy coupling; unidirectional response	×	×	×
Zheng et al. [22]; Gao et al. [23]; Li et al. [24]	Demand side	Multi-energy coupling	Model-based	×	×
Zhang et al. [26]; Ahmed et al. [27]; Li et al. [28]	Demand side	Multi-energy coupling	Learning-based; local decision	×	×
Wu et al. [35]	System side	×	×	ARO	Uncertainty set
Alismail et al. [36]; Cao et al. [37]	System side	×	×	DRO	Multiple-order moment/ Wasserstein ambiguity set
This study	Demand-side and system-side coordination	Multi-energy coupling; bidirectional response	MAPPO-based multi-agent coordination	IDTDRO	IDT ambiguity set

"×" indicates that the corresponding aspect was not considered in the referenced study.

The rest of the paper is organized as follows. Section 2 introduces the modeling of the IES and the details of the two-stage optimization framework. Section 3 presents the modeling of the consumer and operator in the IBIDR, followed by the theory and design of the MAPPO-based interactive solution method for incentive strategies in IBIDRs. Section 4 introduces the IDT ambiguity set construction method and the DRO model for optimal IES dispatch. Section 5 presents and analyzes the experimental results. Section 6 discusses the results. Section 7 concludes the paper.

2. Two-Stage Optimization Framework for the IES

The two-stage optimization framework is shown in Figure 1. The IES consists of electricity, heat, and gas devices, which provide corresponding energy services to consumers. Implementation of the IBIDR involves three participants: a multi-energy trading market, an integrated energy system operator (IESO), and an integrated energy consumer (IEC). The multi-energy trading market is only responsible for supplying electricity, heat, and gas with corresponding energy prices to the IESO. The market does not directly participate in the optimization process. The IESO is the actual operator and manager of the IES. It is responsible for the economic dispatch of various energy devices within the system and the determination of incentive strategies in the IBIDR. The IEC is responsible for providing demand-side flexibility resources.

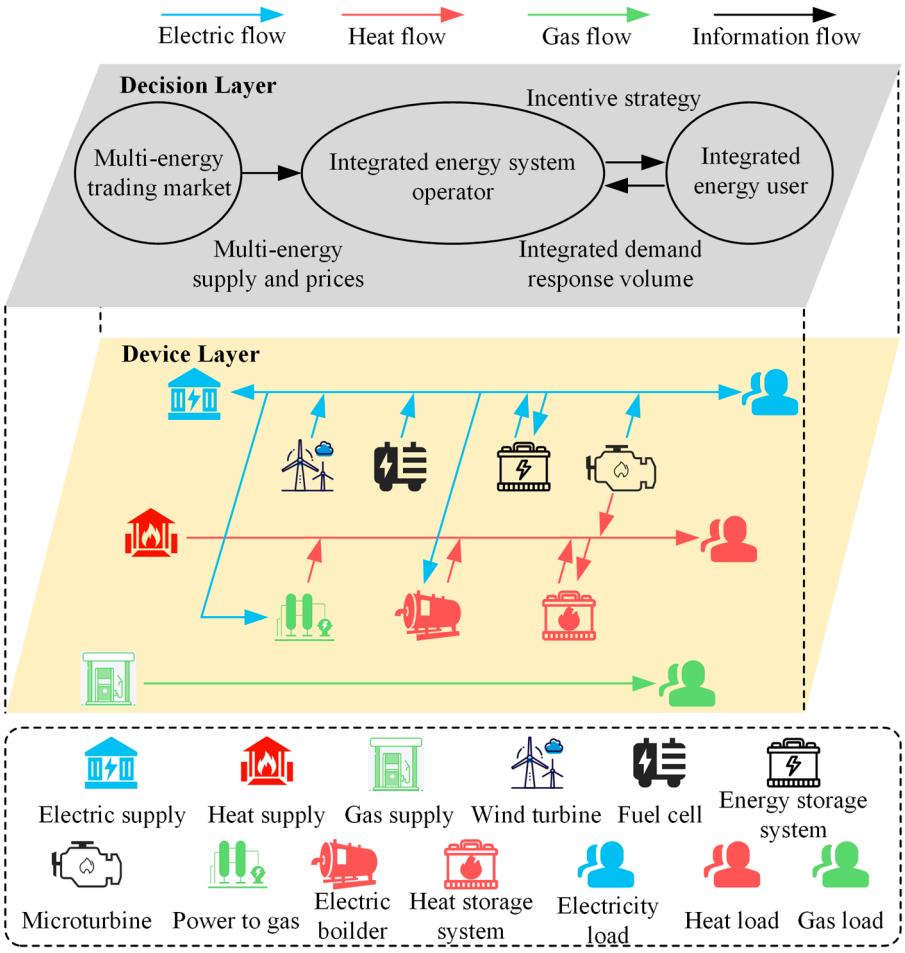


Figure 1. Two-stage optimization framework.

The day-ahead stage focuses on the optimization of incentive strategies in the IBIDR. However, prediction deviations in WPG and multi-energy load demands may lead to a mismatch between the actual intra-day operation and the day-ahead dispatch schedule. Therefore, the outputs of energy devices are adjusted upward or downward in the intra-day stage to maintain the energy balance of the IES.

3. Optimization of Incentive Strategy Considering Operator-Consumer Interactions in the IBIDR

To obtain the optimal incentive strategy in the operator-consumer interaction process, this section constructs an IBIDR model solved using MAPPO. In the model, the IESO serves

as the incentive strategy decisionmaker that is responsible for issuing incentive prices. The IEC serves as the incentive strategy responder that adjusts its energy consumption according to the received incentive prices.

3.1. Modeling of the IBIDR

3.1.1. IEC Model

The objective of the IEC is to maximize its benefit after receiving the incentive prices provided by the IESO. According to Equation (1), the overall benefit to the IEC P_{con} consists of the incentive benefit P_{inc} obtained from participating in the IBIDR process and the discomfort cost C_{dis} caused by adjusting its energy consumption demand. P_{inc} and C_{dis} can be calculated by Equations (2) and (3). The demand response amount vector $\boldsymbol{R}(t)$, which consists of $R_{\mathrm{e}}(t)$, $R_{\mathrm{h}}(t)$, and $R_{\mathrm{g}}(t)$, is calculated using Equation (4). The incentive price vector $\boldsymbol{P}(t)$ is shown in Equation (5).

$$P_{\mathrm{con}} = P_{\mathrm{inc}} - C_{\mathrm{dis}}, \tag{1}$$

$$P_{\mathrm{inc}} = \sum_{t=1}^{T} \left[\boldsymbol{P}(t)^{\mathrm{T}} \boldsymbol{R}(t) \right] \Delta t, \tag{2}$$

$$C_{\mathrm{dis}} = \sum_{t=1}^{T} \left[\frac{1}{2} (\boldsymbol{E}(t)\boldsymbol{R}(t))^{\mathrm{T}} \boldsymbol{\nu}(t) \boldsymbol{E}(t) \boldsymbol{R}(t) + \boldsymbol{J}(t) \boldsymbol{E}(t) \boldsymbol{R}(t) \right] \Delta t, \tag{3}$$

$$\boldsymbol{R}(t) = \boldsymbol{S}(t) \Delta \boldsymbol{R}(t), \tag{4}$$

$$\boldsymbol{P}(t) = \left[p_{\mathrm{e}}(t), p_{\mathrm{h}}(t), p_{\mathrm{g}}(t) \right]^{\mathrm{T}}, \tag{5}$$

where $p_{\mathrm{e}}(t)$, $p_{\mathrm{h}}(t)$, and $p_{\mathrm{g}}(t)$ denote the incentive prices of electricity, heat, and gas loads, and $R_{\mathrm{e}}(t)$, $R_{\mathrm{h}}(t)$, and $R_{\mathrm{g}}(t)$ denote the load response amount of electricity, heat, and gas loads. $\Delta \boldsymbol{R}(t)$ and $\boldsymbol{S}(t)$ denote the vectors of demand response adjustments and corresponding states for electricity, heat, and gas loads. For a load reduction response, $\boldsymbol{S}(t) > 0$, $\Delta \boldsymbol{R}(t) > 0$. For a load increase response, $\boldsymbol{S}(t) < 0$, $\Delta \boldsymbol{R}(t) < 0$. This setting ensures that $\boldsymbol{R}(t)$ remains positive. The load increase response is mainly employed to absorb surplus wind power. Therefore, electricity loads can participate in both load reduction and load increase responses. Heat and gas loads can only participate in load reduction response. The parameters $\boldsymbol{\nu}(t)$, $\boldsymbol{J}(t)$, and $\mathbf{E}(t)$ are employed to characterize the statistical characteristics of IECs, as described in Equations (6)–(8).

$$\boldsymbol{\nu}(t) = \mathrm{diag}[\nu_{\mathrm{e}}(t), \nu_{\mathrm{h}}(t), \nu_{\mathrm{g}}(t)], \tag{6}$$

$$\boldsymbol{J}(t) = [\vartheta_{\mathrm{e}}(t), \vartheta_{\mathrm{h}}(t), \vartheta_{\mathrm{g}}(t)], \tag{7}$$

$$\boldsymbol{E}(t) = \begin{bmatrix} e_{\mathrm{e}}(t) & e_{\mathrm{o}}(t) & e_{\mathrm{o}}(t) \\ e_{\mathrm{o}}(t) & e_{\mathrm{h}}(t) & e_{\mathrm{o}}(t) \\ e_{\mathrm{o}}(t) & e_{\mathrm{o}}(t) & e_{\mathrm{g}}(t) \end{bmatrix}, \tag{8}$$

where $\nu_{\mathrm{e}}(t)$, $\nu_{\mathrm{h}}(t)$, and $\nu_{\mathrm{g}}(t)$ denote the response willingness of IECs for electricity, heat, and gas loads; $\vartheta_{\mathrm{e}}(t)$, $\vartheta_{\mathrm{h}}(t)$, and $\vartheta_{\mathrm{g}}(t)$ denote the minimum incentive prices of IECs for electricity, heat, and gas loads; $e_{\mathrm{e}}(t)$, $e_{\mathrm{h}}(t)$, and $e_{\mathrm{g}}(t)$ denote the effects of changes in electricity, heat, and gas demands on their own comfort levels; and $e_{\mathrm{o}}(t)$ denotes the effect between different energy demands.

The IEC determines its load response amount based on P_{con}. Since P_{con} is strictly concave, its demand response optimization problem has a unique optimal solution [38]. The first-order derivative of P_{con} with respect to $\boldsymbol{R}$ is then set to zero to obtain the optimal response strategy $\boldsymbol{R}^{*}$, as shown in Equations (9) and (10):

$$\frac{\partial P_{\text{con}}}{\partial \boldsymbol{R}} = \boldsymbol{P} - \boldsymbol{E}^{\text{T}}\boldsymbol{\nu}\boldsymbol{E}\boldsymbol{R} - \boldsymbol{E}^{\text{T}}\boldsymbol{J}^{\text{T}}, \tag{9}$$

$$\boldsymbol{R}^{*} = \left(\boldsymbol{E}^{\text{T}}\boldsymbol{\nu}\boldsymbol{E}\right)^{-1}\left(\boldsymbol{P} - \boldsymbol{E}^{\text{T}}\boldsymbol{J}^{\text{T}}\right). \tag{10}$$

To satisfy practical operation conditions, $R_{\text{e}}(t)$, $R_{\text{h}}(t)$, and $R_{\text{g}}(t)$ should satisfy the constraints given in Equation (11):

$$\begin{cases} R_{\text{e}}^{\min} \le R_{\text{e}}(t) \le R_{\text{e}}^{\max} \\ R_{\text{h}}^{\min} \le R_{\text{h}}(t) \le R_{\text{h}}^{\max} \\ R_{\text{g}}^{\min} \le R_{\text{g}}(t) \le R_{\text{g}}^{\max} \end{cases}, \tag{11}$$

where $R_{\text{e}}^{\min}$, $R_{\text{h}}^{\min}$, and $R_{\text{g}}^{\min}$ denote the minimum load response amounts of electricity, heat, and gas loads, and $R_{\text{e}}^{\max}$, $R_{\text{h}}^{\max}$, and $R_{\text{g}}^{\max}$ denote the corresponding maximum load response amounts.

3.1.2. IESO Model

According to Equation (12), the overall benefit of the IESO P_{IESO} consists of the benefits from the electricity P_{es}, heat P_{hs}, and gas P_{gs} subsystems. Each energy subsystem benefit is calculated as the difference between the operation cost savings achieved through the IBIDR and the corresponding incentive costs, as shown in Equations (13)–(15).

$$P_{\text{IESO}} = P_{\text{es}} + P_{\text{hs}} + P_{\text{gs}}, \tag{12}$$

$$P_{\text{es}} = (C_{\text{MT}} + C_{\text{FC}} + C_{\text{Grid}} + C_{\text{WPG}}) - (C_{\text{DR,MT}} + C_{\text{DR,FC}} + C_{\text{DR,Grid}} + C_{\text{DR,WPG}} + C_{\text{inc,e}}), \tag{13}$$

$$P_{\text{hs}} = C_{\text{Heat}} - (C_{\text{DR,Heat}} + C_{\text{inc,h}}), \tag{14}$$

$$P_{\text{gs}} = C_{\text{Gas}} - (C_{\text{DR,Gas}} + C_{\text{inc,g}}), \tag{15}$$

where C_{MT}, C_{FC}, C_{Grid}, and C_{WPG} denote the microturbine (MT) operation cost, fuel cell (FC) operation cost, electricity interaction cost, and wind power curtailment penalty cost before the IBIDR; $C_{\text{DR,MT}}$, $C_{\text{DR,FC}}$, $C_{\text{DR,Grid}}$, and $C_{\text{DR,WPG}}$ denote the corresponding costs after the IBIDR; C_{Heat} and C_{Gas} denote the heat and gas supply cost before the IBIDR; $C_{\text{DR,Heat}}$ and $C_{\text{DR,Gas}}$ denote the corresponding costs after the IBIDR; and $C_{\text{inc,e}}$, $C_{\text{inc,h}}$, and $C_{\text{inc,g}}$ denote the incentive costs in the electricity, heat, and gas subsystems.

3.2. MAPPO-Based Incentive Strategy Optimization Method

The MAPPO algorithm can implicitly learn the state-action mapping relationship during the IBIDR process and make rapid decisions according to new states after training. In addition, the centralized training and decentralized execution mechanism in MAPPO can fully exploit global information while ensuring the independence of individual agent decisions. Therefore, the MAPPO algorithm is adopted to solve the incentive strategy optimization problem in the IBIDR.

3.2.1. State-Action Space and Reward Function

DRL learns optimal strategies through continuous interactions between agents and the environment, which can be described using a Markov decision process (MDP). An MDP mainly consists of the state space, action space, and reward function. In this paper, each energy subsystem is modeled as an independent agent. The local state spaces of the electricity $\boldsymbol{o}_{\text{e},t}$, heat $\boldsymbol{o}_{\text{h},t}$, and gas $\boldsymbol{o}_{\text{g},t}$ agents are shown in Equations (16)–(18).

$$\boldsymbol{o}_{\text{e},t} = \left\{t, P_{\text{WPG}}^{\text{pre}}(t), P_{\text{EL}}^{\text{pre}}(t), \lambda_{\text{buy,e}}(t), \lambda_{\text{sell,e}}(t), \text{SOC}_{\text{e}}(t), R_{\text{e}}(t-1), s_{\text{e}}(t-1), p_{\text{e}}(t-1)\right\}, \tag{16}$$

$$\boldsymbol{o}_{\mathrm{h},t} = \left\{t, P_{\mathrm{WPG}}^{\mathrm{pre}}(t), Q_{\mathrm{HL}}^{\mathrm{pre}}(t), \lambda_{\mathrm{buy,h}}(t), \mathrm{SOC}_{\mathrm{h}}(t), R_{\mathrm{h}}(t-1), s_{\mathrm{h}}(t-1), p_{\mathrm{h}}(t-1)\right\}, \quad (17)$$

$$\boldsymbol{o}_{\mathrm{g},t} = \left\{t, P_{\mathrm{WPG}}^{\mathrm{pre}}(t), G_{\mathrm{GL}}^{\mathrm{pre}}(t), \lambda_{\mathrm{buy,g}}(t), R_{\mathrm{g}}(t-1), s_{\mathrm{g}}(t-1), p_{\mathrm{g}}(t-1)\right\}, \quad (18)$$

where t denotes the scheduling time; $P_{\mathrm{WPG}}^{\mathrm{pre}}(t)$, $P_{\mathrm{EL}}^{\mathrm{pre}}(t)$, $Q_{\mathrm{HL}}^{\mathrm{pre}}(t)$, and $G_{\mathrm{GL}}^{\mathrm{pre}}(t)$ denote the predicted WPG output and load demands of electricity, heat, and gas; $\lambda_{\mathrm{buy,e}}(t)$ and $\lambda_{\mathrm{sell,e}}(t)$ denote the electricity purchasing and selling prices; $\lambda_{\mathrm{buy,h}}(t)$ and $\lambda_{\mathrm{buy,g}}(t)$ denote the heat and gas purchasing prices; $\mathrm{SOC}_{\mathrm{e}}(t)$ and $\mathrm{SOC}_{\mathrm{h}}(t)$ denote the states of charge of the energy storage system (ESS) and heat storage system (HSS); $R_{\mathrm{e}}(t-1)$, $R_{\mathrm{h}}(t-1)$, and $R_{\mathrm{g}}(t-1)$ denote the load response amounts of electricity, heat, and gas loads at the previous time step; $s_{\mathrm{e}}(t-1)$, $s_{\mathrm{h}}(t-1)$, and $s_{\mathrm{g}}(t-1)$ denote the corresponding load response states at the previous time step; and $p_{\mathrm{e}}(t-1)$, $p_{\mathrm{h}}(t-1)$, and $p_{\mathrm{g}}(t-1)$ denote the corresponding incentive prices at the previous time step. The global state space $\boldsymbol{s}_t$ is shown in Equation (19):

$$\boldsymbol{s}_t = \{\boldsymbol{o}_{\mathrm{e},t}, \boldsymbol{o}_{\mathrm{h},t}, \boldsymbol{o}_{\mathrm{g},t}\}. \quad (19)$$

The electricity, heat, and gas agents output the normalized actions $a_{\mathrm{e},t}$, $a_{\mathrm{h},t}$, and $a_{\mathrm{g},t}$, respectively. These actions constitute the action space a_t, as shown in Equation (20). $a_{\mathrm{e},t}$, $a_{\mathrm{h},t}$, and $a_{\mathrm{g},t}$ are mapped to the corresponding incentive prices $p_{\mathrm{e}}(t)$, $p_{\mathrm{h}}(t)$, and $p_{\mathrm{g}}(t)$ according to Equation (21).

$$\begin{cases} a_t = \{a_{\mathrm{e},t}, a_{\mathrm{h},t}, a_{\mathrm{g},t}\} \\ a_{\mathrm{e},t} \in [-1,1], a_{\mathrm{h},t} \in [-1,1], a_{\mathrm{g},t} \in [-1,1] \end{cases}, \quad (20)$$

$$\begin{cases} p_{\mathrm{e}}(t) = p_{\mathrm{e}}^{\min} + |a_{\mathrm{e},t}|(p_{\mathrm{e}}^{\max} - p_{\mathrm{e}}^{\min}) \\ p_{\mathrm{h}}(t) = p_{\mathrm{h}}^{\min} + \frac{1}{2}(a_{\mathrm{h},t}+1)(p_{\mathrm{h}}^{\max} - p_{\mathrm{h}}^{\min}) \\ p_{\mathrm{g}}(t) = p_{\mathrm{g}}^{\min} + \frac{1}{2}(a_{\mathrm{g},t}+1)(p_{\mathrm{g}}^{\max} - p_{\mathrm{g}}^{\min}) \end{cases}, \quad (21)$$

where $p_{\mathrm{e}}^{\min}$, $p_{\mathrm{h}}^{\min}$, and $p_{\mathrm{g}}^{\min}$ denote the minimum incentive prices for electricity, heat, and gas loads, and $p_{\mathrm{e}}^{\max}$, $p_{\mathrm{h}}^{\max}$, and $p_{\mathrm{g}}^{\max}$ denote the corresponding maximum incentive prices. For the electricity agent, a positive $a_{\mathrm{e},t}$ indicates that $p_{\mathrm{e}}(t)$ obtained from Equation (21) is applied to load reduction. A negative $a_{\mathrm{e},t}$ indicates $p_{\mathrm{e}}(t)$ that is applied to load increase.

The reward function guides agents to learn incentive strategies. Therefore, the benefits of the electricity P_{es}, heat P_{hs}, and gas P_{gs} subsystems defined in Equations (13)–(15) are adopted as the reward functions $r_{\mathrm{e},t}$, $r_{\mathrm{h},t}$, and $r_{\mathrm{g},t}$ of the corresponding agents, as shown in Equations (22)–(24). Although defined separately, these rewards are coupled through the multi-energy response. Each critic uses the global state to estimate state value under multi-agent interactions. The overall reward function $r_{\mathrm{oa},t}$ in Equation (25) is the sum of the three rewards. Thus, an improvement in any local reward that does not compromise the others also improves $r_{\mathrm{oa},t}$, aligning local optimization with global coordination.

$$r_{\mathrm{e},t} = P_{\mathrm{es}}, \quad (22)$$

$$r_{\mathrm{h},t} = P_{\mathrm{hs}}, \quad (23)$$

$$r_{\mathrm{g},t} = P_{\mathrm{gs}}, \quad (24)$$

$$r_{\mathrm{oa},t} = r_{\mathrm{e},t} + r_{\mathrm{h},t} + r_{\mathrm{g},t}. \quad (25)$$

3.2.2. MAPPO Algorithm Architecture and Execution Process

The MAPPO algorithm adopts an actor–critic framework [30], as shown in Figure 2. Each agent consists of an old actor network, a new actor network, and a critic network. Both the actor and critic networks employ multilayer perceptrons (MLPs) for feature extraction. The network structure is shown in Figure 3. Each of three hidden layers contains

128 neurons. Each hidden layer consists of a fully connected layer, ReLU activation, and layer normalization. The forward propagation of the networks is described as follows.

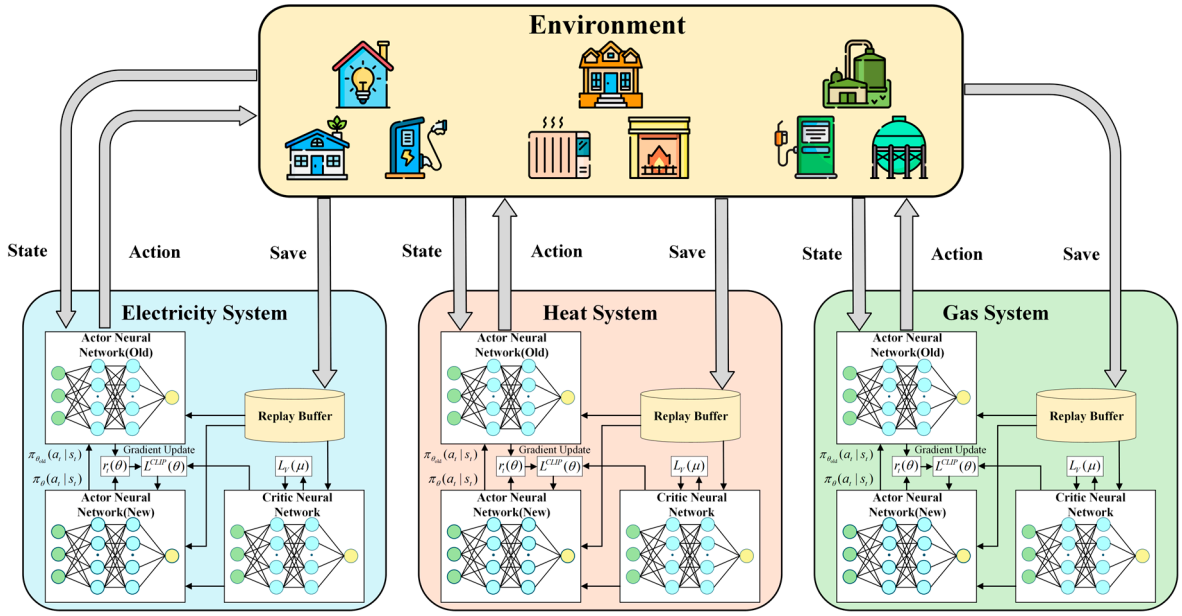


Figure 2. Overall architecture of the MAPPO algorithm.

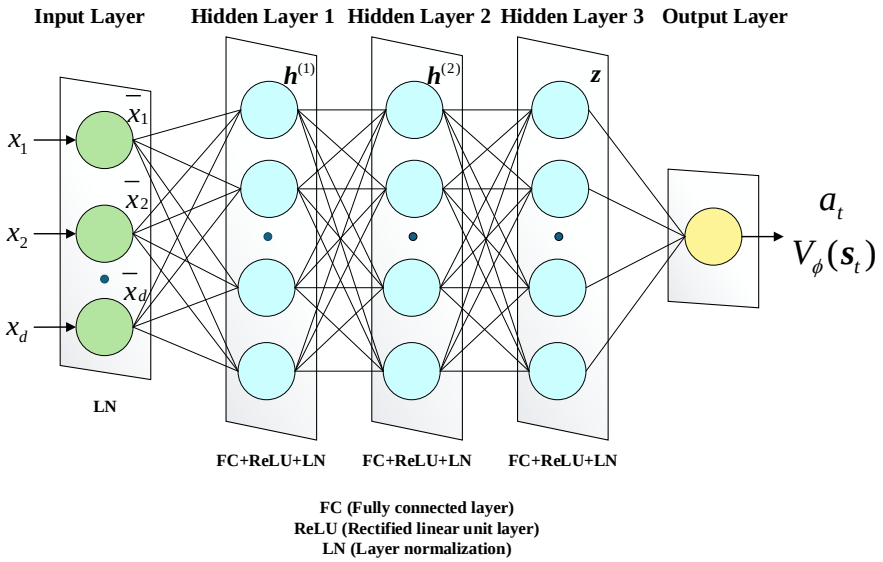


Figure 3. MLP network structure.

Firstly, the local observation vectors $\boldsymbol{o}_{\mathrm{e},t}$, $\boldsymbol{o}_{\mathrm{h},t}$, and $\boldsymbol{o}_{\mathrm{g},t}$ are fed into their respective actor networks. The global state vector $\boldsymbol{s}_t$ is fed into the critic networks. The input vectors are processed through layer normalization, as described in Equations (26)–(28):

$$\overline{x}_i = \gamma_i \frac{x_i - \mu_x}{\sqrt{{\sigma_x}^2 + \varepsilon}} + \beta_i, \quad i = 1, \ldots, d, \tag{26}$$

$$\mu_x = \frac{1}{d} \sum_{i=1}^{d} x_i, \tag{27}$$

$$\sigma_x{}^2 = \frac{1}{d} \sum_{i=1}^{d} (x_i - \mu_x)^2, \tag{28}$$

where x_i denotes the ith feature of the input vector; d denotes the feature dimension; μ_x and ${\sigma_x}^2$ denote the mean and variance of the input vector; $\overline{x}_i$ denotes the ith feature after layer normalization; ε is a small constant; and γ_i and β_i denote the trainable scaling factor and bias.

Subsequently, the normalized input vector $\overline{\boldsymbol{x}} = [\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_d]$ is processed through hidden layers for feature extraction. Finally, the feature encoding $\boldsymbol{z}$ is obtained, as shown in Equation (29):

$$\begin{cases} \boldsymbol{h}^{(0)} = \overline{\boldsymbol{x}} \\ \boldsymbol{h}^{(l)} = LayerNorm\left[ReLU\left(\boldsymbol{W}^{(l)}\boldsymbol{h}^{(l-1)} + \boldsymbol{b}^{(l)}\right)\right], \quad l = 1,2,3 \\ \boldsymbol{z} = \boldsymbol{h}^{(3)} \end{cases}, \tag{29}$$

where l denotes the hidden-layer index; $\boldsymbol{h}^{(l)}$ denotes the output feature of hidden layer l; $\boldsymbol{W}^{(l)}$ and $\boldsymbol{b}^{(l)}$ denote the weight matrix and bias vector of hidden layer l; $ReLU(\cdot)$ denotes the activation function; and $LayerNorm(\cdot)$ denotes the layer normalization calculated according to Equations (26)–(28).

The MAPPO-based solution process consists of two stages, including offline training and online application. For each agent during offline training, the old actor network maps its feature encoding $\boldsymbol{z}^a$ to the action mean μ^a through the action output layer. The action standard deviation σ^a is obtained from the trainable log-standard-deviation parameter ρ. Then, the diagonal Gaussian distribution is constructed for action sampling. The sampled action $\hat{a}_t$ is constrained within $[-1, 1]$ by the action bounds to obtain the normalized action a_t executed by the agent. The calculation process is described in Equations (30)–(33):

$$\mu_t^a = \boldsymbol{W}_\mu \boldsymbol{z}^a + \boldsymbol{b}_\mu, \tag{30}$$

$$\sigma^a = \exp(\rho), \tag{31}$$

$$\hat{a}_t \sim \pi_{\theta\text{old}}(\hat{a}_t \mid \boldsymbol{o}_t) = \mathcal{N}\left(\mu_t^a, diag\left[(\sigma^a)^2\right]\right), \tag{32}$$

$$a_t = \text{clip}(\hat{a}_t, -1, 1), \tag{33}$$

where $\boldsymbol{W}_\mu$ and $\boldsymbol{b}_\mu$ denote the weight matrix and bias vector of the action output layer; θ^{old} denotes the old actor network parameter; $\pi_{\theta^{\text{old}}}(\cdot|\boldsymbol{o}_t)$ denotes the action probability distribution of the old actor network; and $\mathcal{N}(\cdot)$ denotes the Gaussian distribution.

Meanwhile, the critic network maps its feature encoding $\boldsymbol{z}^c$ to the state value $V_\phi(\boldsymbol{s}_t)$ through the value output layer, as described in Equation (34):

$$V_\phi(\boldsymbol{s}_t) = \boldsymbol{W}_v \boldsymbol{z}^c + \boldsymbol{b}_v, \tag{34}$$

where $\boldsymbol{W}_v$ and $\boldsymbol{b}_v$ denote the weight matrix and bias vector of the value output layer, and ϕ denotes the critic network parameter.

After a_t is executed in the environment, the reward r_t is calculated according to Equations (22)–(24), and the next global state $\boldsymbol{s}_{t+1}$ is obtained. The sample sequence $\boldsymbol{D} = \{\boldsymbol{o}_t, \boldsymbol{s}_t, \hat{\boldsymbol{a}}_t, r_t, \boldsymbol{s}_{t+1}\}$ is collected for network training. Furthermore, generalized advantage estimation (GAE) is employed to calculate the action advantage $\hat{A}_t$, as shown in Equations (35) and (36):

$$\hat{A}_t = \sum_{k=0}^{\infty} (\gamma\lambda)^k \delta_{t+k}, \tag{35}$$

$$\delta_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t), \tag{36}$$

where γ denotes the discount factor; λ denotes the advantage estimation factor; and δ_t denotes the temporal-difference error. After $\hat{A}_t$ is obtained, the probability ratio $r_t(\theta)$ between the new and old policies is calculated. The clipped surrogate objective function $L^{CLIP}(\theta)$ is then constructed, as shown in Equations (37) and (38):

$$r_t(\theta) = \frac{\pi_\theta(\hat{a}_t|\boldsymbol{o}_t)}{\pi_{\theta^{\text{old}}}(\hat{a}_t|\boldsymbol{o}_t)}, \tag{37}$$

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\right], \tag{38}$$

where θ denotes the new actor network parameter; $\pi_\theta(\cdot \mid \boldsymbol{o}_t)$ denotes the action probability distribution of the new actor network; $\hat{\mathbb{E}}_t$ denotes the expectation over time steps; and clip$(\cdot)$ and ϵ denote the clipping function and clipping ratio. The clipped probability ratio is constrained within $[1-\epsilon, 1+\epsilon]$ to prevent excessively large policy updates. The Adam optimizer is employed to update θ using the negative gradient $-\nabla_{\theta_k} L^{CLIP}(\theta_k)$ obtained through backpropagation to maximize $L^{CLIP}(\theta)$, as shown in Equation (39):

$$\theta_{k+1} = Adam(\theta_k, -\nabla_{\theta_k} L^{CLIP}(\theta_k)), \tag{39}$$

where k denotes the update number.

For the critic network, the mean squared error loss function $L_V(\phi)$ is constructed as shown in Equation (40):

$$L_V(\phi) = \hat{\mathbb{E}}_t\left[\left(V_\phi(s_t) - R_t\right)^2\right], \tag{40}$$

where R_t denotes the cumulative return. The Adam optimizer is employed to update ϕ using the gradient $\nabla_{\phi_k} L_V(\phi_k)$ obtained through backpropagation to minimize $L_V(\phi)$, as shown in Equation (41):

$$\phi_{k+1} = Adam(\phi_k, \nabla_{\phi_k} L_V(\phi_k)). \tag{41}$$

After a number of K PPO updating iterations, θ is synchronized to θ^{old}. The above training process is repeated until the preset number of training episodes E is reached. During online application, the local observation vector $\boldsymbol{o}_t$ is processed by the trained actor network with parameter θ^* to obtain the feature encoding $\boldsymbol{z}^a$. The action mean μ_t^a is taken as deterministic action output and constrained by the action bounds to obtain a_t, as shown in Equations (42) and (43):

$$\mu_t^a = \boldsymbol{W}_\mu^* \boldsymbol{z}^a + \boldsymbol{b}_\mu^*, \tag{42}$$

$$a_t = \text{clip}(\mu_t^a, -1, 1), \tag{43}$$

where $\boldsymbol{W}_\mu^*$ and $\boldsymbol{b}_\mu^*$ denote the weight matrix and bias vector of the action mean output layer after offline training. Ultimately, the incentive prices and multi-energy load demands in the IBIDR are obtained according to Equations (10) and (21).

4. Two-Layer DRO Dispatch Model Based on the IDT Ambiguity Set

In this section, a two-layer DRO model based on the IDT ambiguity set is presented to achieve economic dispatch of the IES considering source-load uncertainties. The outer layer model utilizes the predicted WPG output and the multi-energy load demands after the IBIDR to determine the day-ahead dispatch schedule, with the objective of minimizing dispatch costs. However, uncertainties in WPG and multi-energy load demands may cause deviations between actual operation and the day-ahead schedule. Therefore, the inner layer model performs intra-day adjustments based on the day-ahead dispatch schedule to mitigate the impacts caused by uncertainties. The interaction between the outer and inner

layer models ensures the robustness and economy of the dispatch schedule. The presented model is described in Equation (44):

$$\begin{cases} \min\limits_{\boldsymbol{x}}\{C_{\mathrm{DA}}(\boldsymbol{x}) + \max\limits_{\boldsymbol{w}\in\omega}\min\limits_{\boldsymbol{y}} C_{\mathrm{ID}}(\boldsymbol{w},\boldsymbol{y})\} \\ \text{s.t.} E_{\mathrm{DA}}(\boldsymbol{x}) = 0, I_{\mathrm{DA}}(\boldsymbol{x}) \le 0 \\ E_{\mathrm{ID}}(\boldsymbol{x},\boldsymbol{w},\boldsymbol{y}) = 0, I_{\mathrm{ID}}(\boldsymbol{x},\boldsymbol{w},\boldsymbol{y}) \le 0 \end{cases}, \tag{44}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the day-ahead dispatch schedule and intra-day regulation schedule; $\boldsymbol{w}$ denotes the uncertain variables; and ω denotes the joint uncertainty set consisting of WPG and multi-energy load uncertainties.

4.1. Construction and Transformation of the Ambiguity Set Using IDT

To characterize WPG uncertainty, this section constructs the ambiguity set using IDT and WPG samples. Let $\boldsymbol{w}_{\mathrm{WPG}} = \{w_{\mathrm{WPG},1}, w_{\mathrm{WPG},2}, \ldots, w_{\mathrm{WPG},m}\}$ denote the WPG sample set containing a number of m samples, which correspond to m possible states $w_{\mathrm{WPG},m}$ of WPG output. The probability associated with the jth state is denoted by π_j, satisfying $\sum\limits_{j=1}^{m} \pi_j = 1$. Therefore, the probability density function (PDF) of the prior distribution can be calculated using Bayesian theory, as described in Equation (45):

$$F(\boldsymbol{\pi}) = \Gamma(l)\frac{\prod\limits_{j=1}^{m} \pi_j^{l\cdot\psi_j - 1}}{\prod\limits_{j=1}^{m}\Gamma(l\cdot\psi_j)}, \forall\psi_j \in [0,1], \sum_{j=1}^{m}\psi_j = 1, \tag{45}$$

where $\Gamma(\cdot)$ denotes the gamma function; ψ_j denotes the jth prior weight coefficient; and l denotes the sample size, which is usually set to 1. The corresponding PDF of the posterior distribution can be calculated using Equation (46):

$$F(\boldsymbol{\pi}|\mathrm{N}) = \Gamma(l+\mathrm{N})\frac{\prod\limits_{j=1}^{m} \pi_j^{n_j + l\cdot\psi_j - 1}}{\prod\limits_{j=1}^{m}\Gamma(l\cdot\psi_j + n_j)}, \forall\psi_j \in [0,1], \sum_{j=1}^{m}\psi_j = 1, \sum_{j=1}^{m} n_j = \mathrm{N}, \tag{46}$$

where n_j denotes the number of observations for the jth state, and N denotes the total number of observations. Given ψ_j, the expected value of the posterior probability can be calculated using Equation (47):

$$\mathbb{E}(\pi_j) = (l\cdot\psi_j + n_j)(l+N)^{-1}. \tag{47}$$

The probability set of the jth state can be calculated using Equation (48):

$$\pi_j = \left[\underline{\mathbb{E}}(\pi_j), \overline{\mathbb{E}}(\pi_j)\right] = \left[\frac{n_j}{l+\mathrm{N}}, \frac{l+n_j}{l+\mathrm{N}}\right], \tag{48}$$

where $\underline{\mathbb{E}}(\pi_j)$ and $\overline{\mathbb{E}}(\pi_j)$ denote the minimum and maximum expected probability when $\psi_j = 0$ and $\psi_j = 1$. Therefore, the confidence interval $[\underline{\pi_j}, \overline{\pi_j}]$ can be constructed using Equation (49):

$$[\underline{\pi_j}, \overline{\pi_j}] = \begin{cases} \underline{\pi_j} = 0, \overline{\pi_j} = B_2^{-1}(\frac{1+\gamma}{2}), n_j = 0, \\ \underline{\pi_j} = B_1^{-1}(\frac{1-\gamma}{2}), \overline{\pi_j} = B_2^{-1}(\frac{1+\gamma}{2}), 0 \leqslant n_j \leqslant m \\ \underline{\pi_j} = B_1^{-1}(\frac{1-\gamma}{2}), \overline{\pi_j} = 1, n_j = m \end{cases}, \tag{49}$$

where γ denotes the confidence coefficient, and B_1 and B_2 denote the cumulative distribution function (CDF) of the Beta distribution Beta(n_j,$l + m - n_j$) and Beta($l + n_j$,$m - n_j$). However, $[\underline{\pi_j}, \overline{\pi_j}]$ may not fully cover the actual estimated interval $[\underline{w}_{\mathrm{WPG}}, \overline{w}_{\mathrm{WPG}}]$ of the uncertain variable $\boldsymbol{w}_{\mathrm{WPG}}$. Therefore, Equation (50) is further adopted to calculate its actual estimated interval:

$$[\underline{w}_{\mathrm{WPG}}, \overline{w}_{\mathrm{WPG}}] = [w_{\mathrm{WPG},1} - (1-\gamma)/2, w_{\mathrm{WPG},m} + (1-\gamma)/2], \tag{50}$$

where $\underline{w}_{\mathrm{WPG}}$ and $\overline{w}_{\mathrm{WPG}}$ denote the lower and upper bounds of the actual estimated interval. Based on $[\underline{\pi_j}, \overline{\pi_j}]$ and $[\underline{w}_{\mathrm{WPG}}, \overline{w}_{\mathrm{WPG}}]$, the ambiguity set U_{WPG} is expressed as Equation (51):

$$U_{\mathrm{WPG}} = \left\{ \pi \in U_0([\underline{w}_{\mathrm{WPG}}, \overline{w}_{\mathrm{WPG}}]) \middle| \pi(w_{\mathrm{WPG}} \le w_{\mathrm{WPG},j}) \in [\underline{\pi_j}, \overline{\pi_j}], j = 1, 2, \ldots, m \right\}, \tag{51}$$

where $U_0([\underline{w}_{\mathrm{WPG}}, \overline{w}_{\mathrm{WPG}}]$ denotes the family of probability distributions supported on $[\underline{w}_{\mathrm{WPG}}, \overline{w}_{\mathrm{WPG}}]$. The cumulative probability at each WPG state lies within the corresponding interval $[\underline{\pi_j}, \overline{\pi_j}]$. Based on the N historical observations classified into m WPG states, Equations (45)–(49) determine the lower and upper bounds of the cumulative probability at each state. These bounds form the CDF confidence band. The probability quantiles $(1-\gamma)/2$ and $(1+\gamma)/2$ are then mapped through the CDF confidence band to the WPG output domain. The resulting WPG values are taken as the lower bound $\underline{\omega}_{\mathrm{WPG}}$ and upper bound $\overline{\omega}_{\mathrm{WPG}}$, respectively. These bounds define the realization-based WPG uncertainty set ω_{WPG} used in the subsequent dispatch optimization.

The fluctuation characteristics of electricity, heat, and gas load demands also need to be considered in the dispatch model. Unlike WPG, their prediction deviations can be represented by predefined ranges without estimating the corresponding probability distributions. Therefore, this paper constructs the box-type uncertainty sets ω_{EL}, ω_{HL}, and ω_{GL} to characterize their uncertainty ranges. This treatment also reduces the computational complexity caused by modeling multiple probability distributions. Moreover, adjustable robustness parameters Γ_{WPG}, Γ_{EL}, Γ_{HL}, and Γ_{GL} are introduced to control the conservativeness of the uncertainty sets. The WPG uncertainty set and the box-type uncertainty sets of electricity, heat, and gas load demands jointly form the joint uncertainty set ω, as described in Equation (52):

$$\begin{cases} \sum_{t=1}^{T} \left[\dfrac{P_{\mathrm{WPG}}(t) - P_{\mathrm{WPG}}^{\mathrm{pre}}(t)}{\overline{\omega}_{\mathrm{WPG}}} \rho_{\mathrm{WPG}}^{+} + \dfrac{P_{\mathrm{WPG}}^{\mathrm{pre}}(t) - P_{\mathrm{WPG}}(t)}{\underline{\omega}_{\mathrm{WPG}}} \rho_{\mathrm{WPG}}^{-} \right] \le \Gamma_{\mathrm{WPG}} \\ \sum_{t=1}^{T} \left[\dfrac{P_{\mathrm{EL}}(t) - P_{\mathrm{EL}}^{\mathrm{pre}}(t)}{\overline{\omega}_{\mathrm{EL}}} \rho_{\mathrm{EL}}^{+} + \dfrac{P_{\mathrm{EL}}^{\mathrm{pre}}(t) - P_{\mathrm{EL}}(t)}{\underline{\omega}_{\mathrm{EL}}} \rho_{\mathrm{EL}}^{-} \right] \le \Gamma_{\mathrm{EL}} \\ \sum_{t=1}^{T} \left[\dfrac{Q_{\mathrm{HL}}(t) - Q_{\mathrm{HL}}^{\mathrm{pre}}(t)}{\overline{\omega}_{\mathrm{HL}}} \rho_{\mathrm{HL}}^{+} + \dfrac{Q_{\mathrm{HL}}^{\mathrm{pre}}(t) - Q_{\mathrm{HL}}(t)}{\underline{\omega}_{\mathrm{HL}}} \rho_{\mathrm{HL}}^{-} \right] \le \Gamma_{\mathrm{HL}} \\ \sum_{t=1}^{T} \left[\dfrac{G_{\mathrm{GL}}(t) - G_{\mathrm{GL}}^{\mathrm{pre}}(t)}{\overline{\omega}_{\mathrm{GL}}} \rho_{\mathrm{GL}}^{+} + \dfrac{G_{\mathrm{GL}}^{\mathrm{pre}}(t) - G_{\mathrm{GL}}(t)}{\underline{\omega}_{\mathrm{GL}}} \rho_{\mathrm{GL}}^{-} \right] \le \Gamma_{\mathrm{GL}} \end{cases}, \tag{52}$$

where $P_{\mathrm{WPG}}^{\mathrm{pre}}(t)$, $P_{\mathrm{EL}}^{\mathrm{pre}}(t)$, $Q_{\mathrm{HL}}^{\mathrm{pre}}(t)$, and $G_{\mathrm{GL}}^{\mathrm{pre}}(t)$ denote the predicted WPG output and load demands of electricity, heat, and gas; $P_{\mathrm{WPG}}(t)$ denotes the WPG output; $P_{\mathrm{EL}}(t)$, $Q_{\mathrm{HL}}(t)$, and $G_{\mathrm{GL}}(t)$ denote the load demands of electricity, heat, and gas; ρ_{WPG}^{+}, ρ_{WPG}^{-}, ρ_{EL}^{+}, ρ_{EL}^{-}, ρ_{HL}^{+}, ρ_{HL}^{-}, ρ_{GL}^{+}, and ρ_{GL}^{-} are binary auxiliary variables; $\overline{\omega}_{\mathrm{EL}}$ and $\underline{\omega}_{\mathrm{EL}}$ denote the upper and lower bands of electricity load demand; $\overline{\omega}_{\mathrm{HL}}$ and $\underline{\omega}_{\mathrm{HL}}$ denote the upper and lower bands of heat load demand; and $\overline{\omega}_{\mathrm{GL}}$ and $\underline{\omega}_{\mathrm{GL}}$ denote the upper and lower bands of gas load demand.

4.2. Modeling of the Economic Dispatch Model

4.2.1. Objective Function

According to Equation (44), the overall dispatch cost can be decomposed into the day-ahead dispatch cost C_{DA} and the intra-day regulation cost C_{ID}. C_{DA} is expressed as Equation (53):

$$\begin{cases} C_{DA} = C_{Grid,DA} + C_{FC,DA} + C_{Heat,DA} + C_{Gas,DA} \\ C_{Grid,DA} = \sum_{t=1}^{T} \left[\lambda_{Grid,DA}^{buy}(t) P_{Grid,DA}^{buy}(t) - \lambda_{Grid,DA}^{sell}(t) P_{Grid,DA}^{sell}(t) \right] \Delta t \\ C_{FC,DA} = \sum_{t=1}^{T} [\alpha_{FC,DA} P_{FC,DA}(t) + \beta_{FC,DA}] \Delta t \\ C_{Heat,DA} + C_{Gas,DA} = \sum_{t=1}^{T} \left[\lambda_{Heat} Q_{Heat,DA}^{sup}(t) + \lambda_{Gas} G_{Gas,DA}^{sup}(t) \right] \Delta t \end{cases}, \quad (53)$$

where $C_{Grid,DA}$ denotes the day-ahead electricity interaction cost; $C_{FC,DA}$ denotes the FC operation cost; and $C_{Heat,DA}$ and $C_{Gas,DA}$ denote the day-ahead heat and gas supply cost. The MT operation cost is included in $C_{Gas,DA}$ because the MT consumes gas. $\lambda_{Grid,DA}^{buy}(t)$ and $\lambda_{Grid,DA}^{sell}(t)$ denote the day-ahead electricity purchasing and selling prices; $P_{Grid,DA}^{buy}(t)$ and $P_{Grid,DA}^{sell}(t)$ denote the day-ahead electricity purchasing and selling power; Δt denotes the time step; $\alpha_{FC,DA}$ and $\beta_{FC,DA}$ denote the cost coefficients of FC; $P_{FC,DA}(t)$ denotes the day-ahead output power of FC; λ_{Gas} and λ_{Heat} denote the gas purchasing price and heat purchasing price; and $Q_{Heat,DA}^{sup}(t)$ and $G_{Gas,DA}^{sup}(t)$ denote the day-ahead purchased heat power and gas. C_{ID} can be calculated using Equation (54):

$$\begin{cases} C_{ID} = C_{LD,ID} + C_{Grid,ID} + C_{Heat,ID} + C_{Gas,ID} + C_{CGs}^{up} + C_{CGs}^{dn} + C_{WPG,ID} \\ C_{LD,ID} = \sum_{t=1}^{T} \lambda_{LD} [P_{EL}(t) - P_{EL,ID}(t) + Q_{HL}(t) - Q_{HL,ID}(t) + G_{GL}(t) - G_{GL,ID}(t)] \Delta t \\ C_{Grid,ID} = \sum_{t=1}^{T} \left[\lambda_{Grid,ID}^{buy}(t) P_{Grid,ID}^{buy}(t) - \lambda_{Grid,ID}^{sell}(t) P_{Grid,ID}^{sell}(t) \right] \Delta t \\ C_{Heat,ID} + C_{Gas,ID} = \sum_{t=1}^{T} \left[\lambda_{Heat} Q_{Heat,ID}^{sup}(t) + \lambda_{Gas} G_{Gas,ID}^{sup}(t) \right] \Delta t \\ C_{CGs}^{up} + C_{CGs}^{dn} = \sum_{t=1}^{T} \left[\lambda_{CGs}^{up} P_{CGs}^{up}(t) + \lambda_{CGs}^{dn} P_{CGs}^{dn}(t) \right] \Delta t \\ C_{WPG,ID} = \sum_{t=1}^{T} \lambda_{WPG} [P_{WPG}(t) - P_{WPG,ID}(t)] \Delta t \end{cases}, \quad (54)$$

where $C_{LD,ID}$ denotes the load-shedding penalty cost; $C_{Grid,ID}$ denotes the intra-day electricity interaction cost; $C_{Heat,ID}$ and $C_{Gas,ID}$ denote the intra-day heat and gas supply cost; $C_{CGs}^{up} + C_{CGs}^{down}$ denotes the intra-day regulation cost of controllable generations (CGs); and $C_{WPG,ID}$ denotes the wind power curtailment penalty cost. The CGs include the MT, FC, electric boiler (EB), and power to gas (P2G). λ_{LD} and λ_{WPG} denote the load-shedding penalty and wind power curtailment prices; $P_{EL,ID}(t)$, $Q_{HL,ID}(t)$, and $G_{GL,ID}(t)$ denote the intra-day electricity, heat, and gas load demands; $\lambda_{Grid,ID}^{buy}(t)$ and $\lambda_{Grid,ID}^{sell}$ denote the intra-day electricity purchasing and selling prices; $P_{Grid,ID}^{buy}(t)$ and $P_{Grid,ID}^{sell}(t)$ denote the intra-day electricity purchasing and selling power; $Q_{Heat,ID}^{sup}(t)$ and $G_{Gas,ID}^{sup}(t)$ denote the intra-day purchased heat power and gas; λ_{CGs}^{up} and λ_{CGs}^{dn} denote the upward and downward regulation penalty prices of CGs; $P_{CGs}^{up}(t)$ and $P_{CGs}^{dn}(t)$ denote the upward and downward regulation power of CGs; and $P_{WPG,ID}(t)$ denotes the injected wind power.

4.2.2. Constraints

The operation constraints considered in the optimization model are detailed in Appendix A. For convenient presentation, the above DRO model is expressed in matrix form, as shown in Equation (55):

$$\begin{cases} \min\limits_{\boldsymbol{x}} \left\{ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} + \max\limits_{\boldsymbol{w}\in\omega} \min\limits_{\boldsymbol{y}} (\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{w}) \right\} \\ \text{s.t.} \boldsymbol{C}\boldsymbol{x} = \boldsymbol{o}, \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{p} \\ \boldsymbol{E}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{y} = \boldsymbol{q}, \boldsymbol{G}\boldsymbol{x} + \boldsymbol{H}\boldsymbol{y} \le \boldsymbol{r}, \boldsymbol{I}\boldsymbol{y} \le \boldsymbol{w} \end{cases}, \tag{55}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{C}$, $\boldsymbol{E}$, $\boldsymbol{F}$, $\boldsymbol{D}$, $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{I}$ denote the coefficient matrices; $\boldsymbol{o}$, $\boldsymbol{p}$, $\boldsymbol{q}$, and $\boldsymbol{r}$ denote the constant vectors; and T denotes the matrix transpose. The detailed components of variables $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{w}$ are given in Equations (56)–(58). $\boldsymbol{x}$ is first determined based on the predicted values and remains fixed after uncertainty realization. For every admissible realization $\boldsymbol{w}$ in $\boldsymbol{\omega}$, $\boldsymbol{y}$ adjusts $\boldsymbol{x}$ and satisfies the operation constraints. The worst-case realization is identified only to determine the maximum regulation cost.

$$\begin{aligned} \boldsymbol{x} = [&P_{\mathrm{FC,DA}}(t), P_{\mathrm{MT,DA}}(t), P^{\mathrm{buy}}_{\mathrm{Grid,DA}}(t), P^{\mathrm{dis}}_{\mathrm{ESS,DA}}(t), P_{\mathrm{EB,DA}}(t), P_{\mathrm{P2G,DA}}(t), P^{\mathrm{ch}}_{\mathrm{ESS,DA}}(t), P^{\mathrm{sell}}_{\mathrm{Grid,DA}}(t), Q_{\mathrm{EB,DA}}(t), \\ &Q_{\mathrm{MT,DA}}(t), Q^{\mathrm{sup}}_{\mathrm{Heat,DA}}(t), Q^{\mathrm{dis}}_{\mathrm{HSS,DA}}(t), Q^{\mathrm{ch}}_{\mathrm{HSS,DA}}(t), G_{\mathrm{P2G,DA}}(t), G^{\mathrm{sup}}_{\mathrm{Gas,DA}}(t), G_{\mathrm{MT,DA}}(t), S_{\mathrm{CGs}}, S^{\mathrm{buy}}_{\mathrm{Grid}}(t), S^{\mathrm{sell}}_{\mathrm{Grid}}(t), \\ &S^{\mathrm{ch}}_{\mathrm{HSS}}(t), S^{\mathrm{dis}}_{\mathrm{HSS}}(t), S^{\mathrm{ch}}_{\mathrm{ESS}}(t), S^{\mathrm{dis}}_{\mathrm{ESS}}(t)] \end{aligned} \tag{56}$$

$$\begin{aligned} \boldsymbol{y} = [&P^{\mathrm{up}}_{\mathrm{FC,ID}}(t), P^{\mathrm{dn}}_{\mathrm{FC,ID}}(t), P^{\mathrm{up}}_{\mathrm{MT,ID}}(t), P^{\mathrm{dn}}_{\mathrm{MT,ID}}(t), P^{\mathrm{buy}}_{\mathrm{Grid,ID}}(t), P^{\mathrm{dis}}_{\mathrm{ESS,ID}}(t), P_{\mathrm{WPG,ID}}(t), P^{\mathrm{up}}_{\mathrm{EB,ID}}(t), P^{\mathrm{dn}}_{\mathrm{EB,ID}}(t), P^{\mathrm{up}}_{\mathrm{P2G,ID}}(t), \\ &P^{\mathrm{dn}}_{\mathrm{P2G,ID}}(t), P_{\mathrm{EL,ID}}(t), P^{\mathrm{ch}}_{\mathrm{ESS,ID}}(t), P^{\mathrm{sell}}_{\mathrm{Grid,ID}}(t), Q^{\mathrm{up}}_{\mathrm{EB,ID}}(t), Q^{\mathrm{dn}}_{\mathrm{EB,ID}}(t), Q^{\mathrm{up}}_{\mathrm{MT,ID}}(t), Q^{\mathrm{dn}}_{\mathrm{MT,ID}}(t), Q^{\mathrm{sup}}_{\mathrm{Heat,ID}}(t), Q^{\mathrm{dis}}_{\mathrm{HSS,ID}}(t), \\ &Q_{\mathrm{HL,ID}}(t), Q^{\mathrm{ch}}_{\mathrm{HSS,ID}}(t), G^{\mathrm{up}}_{\mathrm{P2G,ID}}(t), G^{\mathrm{dn}}_{\mathrm{P2G,ID}}(t), G^{\mathrm{sup}}_{\mathrm{Gas,ID}}(t), G^{\mathrm{up}}_{\mathrm{MT,ID}}(t), G^{\mathrm{dn}}_{\mathrm{MT,ID}}(t), G_{\mathrm{GL,ID}}(t), S^{\mathrm{up}}_{\mathrm{CGs}}(t), S^{\mathrm{dn}}_{\mathrm{CGs}}(t)] \end{aligned} \tag{57}$$

$$\boldsymbol{w} = [w_{\mathrm{WPG}}, w_{\mathrm{EL}}, w_{\mathrm{HL}}, w_{\mathrm{GL}}]. \tag{58}$$

4.3. Model Solution

The nested inner and outer decision variables prevent the DRO model from being solved directly. Therefore, the C&CG algorithm is adopted to decompose the original problem into a master problem (MP) and a subproblem (SP). The MP is expressed as Equation (59):

$$\begin{cases} \min\limits_{\boldsymbol{x}} (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} + \theta) \\ \text{s.t.} \theta \ge \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{w} \\ \boldsymbol{C}\boldsymbol{x} = \boldsymbol{o}, \boldsymbol{D}\boldsymbol{x} \le \boldsymbol{p} \\ \boldsymbol{E}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{y} = \boldsymbol{q}, \boldsymbol{G}\boldsymbol{x} + \boldsymbol{H}\boldsymbol{y} \le \boldsymbol{r}, \boldsymbol{I}\boldsymbol{y} \le \boldsymbol{w} \end{cases}, \tag{59}$$

where θ denotes the upper-bound auxiliary variable of the intra-day regulation cost under the worst-case scenario. The SP is expressed as Equation (60):

$$\begin{cases} \max\limits_{\boldsymbol{w}\in\omega} \min\limits_{\boldsymbol{y}} (\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{w}) \\ \text{s.t.} \boldsymbol{E}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{y} = \boldsymbol{q}, \boldsymbol{G}\boldsymbol{x} + \boldsymbol{H}\boldsymbol{y} \le \boldsymbol{r}, \boldsymbol{I}\boldsymbol{y} \le \boldsymbol{w} \end{cases}. \tag{60}$$

Subsequently, the SP maxmin(·) is transformed into an equivalent form max(·) using the strong duality theory, as shown in Equation (61):

$$\begin{cases} \max\limits_{\boldsymbol{w}\in\omega, \boldsymbol{\chi}, \boldsymbol{\delta}, \boldsymbol{\kappa}} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\zeta} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{\chi} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{\chi} + \boldsymbol{r}^{\mathrm{T}}\boldsymbol{\delta} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\delta} \\ \text{s.t.} \boldsymbol{F}^{\mathrm{T}}\boldsymbol{\chi} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{\delta} + \mathbf{I}^{\mathrm{T}}\boldsymbol{\kappa} = \boldsymbol{b} \\ \boldsymbol{\delta} \leqslant 0, \boldsymbol{\kappa} \leqslant 0, \boldsymbol{\zeta} = \boldsymbol{c} + \boldsymbol{\kappa} \end{cases}, \tag{61}$$

where $\boldsymbol{\zeta}$ denotes the auxiliary variable, and $\boldsymbol{\chi}$, $\boldsymbol{\delta}$, and $\boldsymbol{\kappa}$ denote the dual variables. However, the strong duality transformation introduces a bilinear term $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\zeta}$ into the SP. Therefore, the big-M method is adopted to linearize $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\zeta}$. The final form of the SP is expressed as Equation (62):

$$\begin{cases} \max\limits_{\boldsymbol{w}\in\omega,\boldsymbol{\chi},\boldsymbol{\delta},\boldsymbol{\kappa}} \overline{\boldsymbol{w}}\boldsymbol{\zeta}^{+} + \underline{\boldsymbol{w}}\boldsymbol{\zeta}^{-} + \boldsymbol{w}^{\mathrm{pre}}(1-\boldsymbol{\zeta}^{+}-\boldsymbol{\zeta}^{-}) + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{\chi} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{\chi} + \boldsymbol{r}^{\mathrm{T}}\boldsymbol{\delta} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\delta} \\ \text{s.t.}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{\chi} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{\delta} + \boldsymbol{I}^{\mathrm{T}}\boldsymbol{\kappa} = \boldsymbol{b} \\ \boldsymbol{\delta} \leqslant 0, \boldsymbol{\kappa} \leqslant 0, \boldsymbol{\zeta} = \boldsymbol{c} + \boldsymbol{\kappa} \\ -M(1-\omega_i^{+}) + \zeta_i \leqslant \zeta_i^{+} \leqslant M(1-\omega_i^{+}) + \zeta_i \\ -M(1-\omega_i^{-}) + \zeta_i \leqslant \zeta_i^{-} \leqslant M(1-\omega_i^{-}) + \zeta_i \\ -M\omega_i^{+} \leqslant \zeta_i^{+} \leqslant M\omega_i^{+}, -M\omega_i^{-} \leqslant \zeta_i^{-} \leqslant M\omega_i^{-} \\ \omega_i^{+} + \omega_i^{-} \leqslant 1, \sum\limits_{i}^{T}(\omega_i^{+} + \omega_i^{-}) \leqslant \Gamma \end{cases}, \tag{62}$$

where $\boldsymbol{w}^{\mathrm{pre}}$ denotes the predicted value of the uncertain variable $\boldsymbol{w}$; $\overline{\boldsymbol{w}}$ and $\underline{\boldsymbol{w}}$ denote the upper and lower bounds; $\boldsymbol{\zeta}^{+}$ and $\boldsymbol{\zeta}^{-}$ denote the positive and negative values of $\boldsymbol{\zeta}$; ω_i^{+} and ω_i^{-} denote the auxiliary variables; and M denotes a large constant. After the above transformation, the decomposed MP and SP can be iteratively solved using C&CG. The solution procedure is shown in Table 2.

$$\begin{cases} \theta \geq \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}_{s+1} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{w}_{s+1} \\ \boldsymbol{E}\boldsymbol{x} + \boldsymbol{F}\boldsymbol{y}_{s+1} = \boldsymbol{q}, \boldsymbol{G}\boldsymbol{x} + \boldsymbol{H}\boldsymbol{y}_{s+1} \leq \boldsymbol{r}, \boldsymbol{I}\boldsymbol{y}_{s+1} \leq \boldsymbol{w}_{s+1} \end{cases}. \tag{63}$$

Table 2. Solution procedure.

Step	Solution Procedure
Step 1	Set the upper bound $U_0 = +\infty$ and lower bound $L_0 = -\infty$ of the optimization problem; set the iteration number $s = 1$ and convergence gap ε. The initial scenario is $\boldsymbol{w}^{\mathrm{pre}}$.
Step 2	Solve the MP to obtain the optimal solution $(\boldsymbol{x}_s, \theta_s)$. Set the objective value $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_s + \theta_s$ obtained from the MP as the new lower bound L_s.
Step 3	Substitute $\boldsymbol{x}_s$ into the SP to obtain the optimal solution $(\boldsymbol{w}_s^{*}, \boldsymbol{y}_s^{*})$. Set the sum of the objective values $\boldsymbol{a}^{\mathrm{T}}\mathbf{x}_s + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}_s^{*} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{w}_s^{*}$ obtained from the SP and MP as the new upper bound U_s. Set $\boldsymbol{w}_{s+1} = \boldsymbol{w}_s^{*}$.
Step 4	If $U_s - L_s \leq \varepsilon$, output the optimal strategy $\boldsymbol{x}_s$. Otherwise, add the new variable $\boldsymbol{y}_{s+1}$ and Equation (63). Set $s = s + 1$. Return to Step 2.

5. Experimental Results

5.1. Experimental Setting

The experimental environments are listed in Table A1. The parameters of the CGs are listed in Table A2. The parameters of the storage systems are listed in Table A3. The trading prices of electricity are listed in Table A4. The purchasing prices of heat and gas are listed in Table A5. The penalty parameters for wind power curtailment and load shedding are listed in Table A6. The energy conversion efficiencies $\eta_{\mathrm{MT}}^{\mathrm{H}}$, $\eta_{\mathrm{MT}}^{\mathrm{E}}$, η_{EB}, and η_{P2G} are set to 0.8, 0.8, 0.9, and 0.7. The predicted WPG output and multi-energy load demands are shown in Figure A1 [10]. The IEC parameters are listed in Table A7. The training parameters of the MAPPO algorithm are listed in Table A8. The dataset adopts a modified IDEAL household energy dataset [39]. These settings are based on the cited references and practical operation conditions. The effects of key parameters are further examined through sensitivity analysis in Section 5.5.

5.2. Optimization Performance of Incentive Strategy in the IBIDR

5.2.1. Performance Comparison of RL Algorithms

To evaluate the performance of the MAPPO algorithm, MADDPG and MADQN are selected as the benchmark algorithms [25]. Specifically, MADDPG adopts an actor-critic framework based on deterministic policy gradients for continuous action spaces. Since MADQN is designed for discrete action spaces, each continuous action dimension is discretized into five levels. All algorithms are trained for 15,000 episodes using the same experimental settings. Each episode contains 24 time steps, and each time step corresponds to 1 h.

Figure 4 illustrates the convergence curves of average rewards for three algorithms during training. MAPPO achieves the fastest reward improvement, the most stable convergence, and the highest final reward due to its on-policy update and policy clipping mechanism. In contrast, MADDPG converges more slowly and exhibits larger fluctuations because its off-policy learning relies on experience replay. Historical experiences may mismatch the current policy in multi-agent environments. MADQN converges faster in the early stage but shows significant fluctuations later, and obtains the lowest final reward. Its discrete action space reduces the difficulty of early exploration but limits the precise adjustment of continuous incentive strategies.

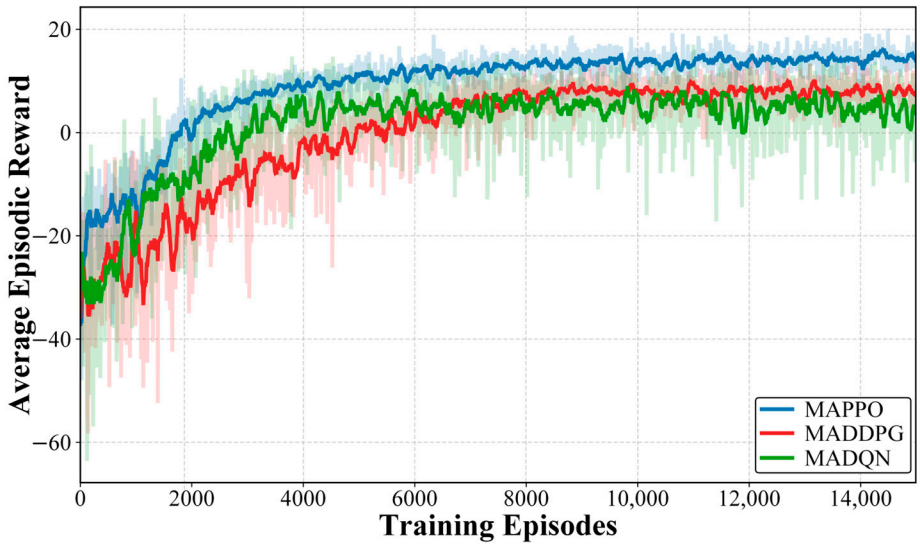


Figure 4. Convergence curves of average rewards for three algorithms.

Table 3 presents the mean and standard deviation of rewards obtained from multiple independent runs. MAPPO achieves the highest average reward and the smallest standard deviation, indicating better reward performance and training stability. In contrast, MADQN obtains the lowest mean reward and the largest standard deviation. These results further support the above analysis and demonstrate the superior performance of MAPPO.

Table 3. Comparison of average rewards among three algorithms.

Algorithms	Average Reward
MAPPO	14.63 ± 2.15
MADDPG	7.93 ± 2.40
MADQN	5.21 ± 3.62

Figure 5 illustrates the convergence of the critic network during MAPPO training. Although value loss exhibits fluctuations in the early stage, it decreases overall and stabilizes

below 0.01 after approximately 3700 episodes. This indicates the gradual convergence of the critic network and its ability to provide stable estimates of the state-value function.

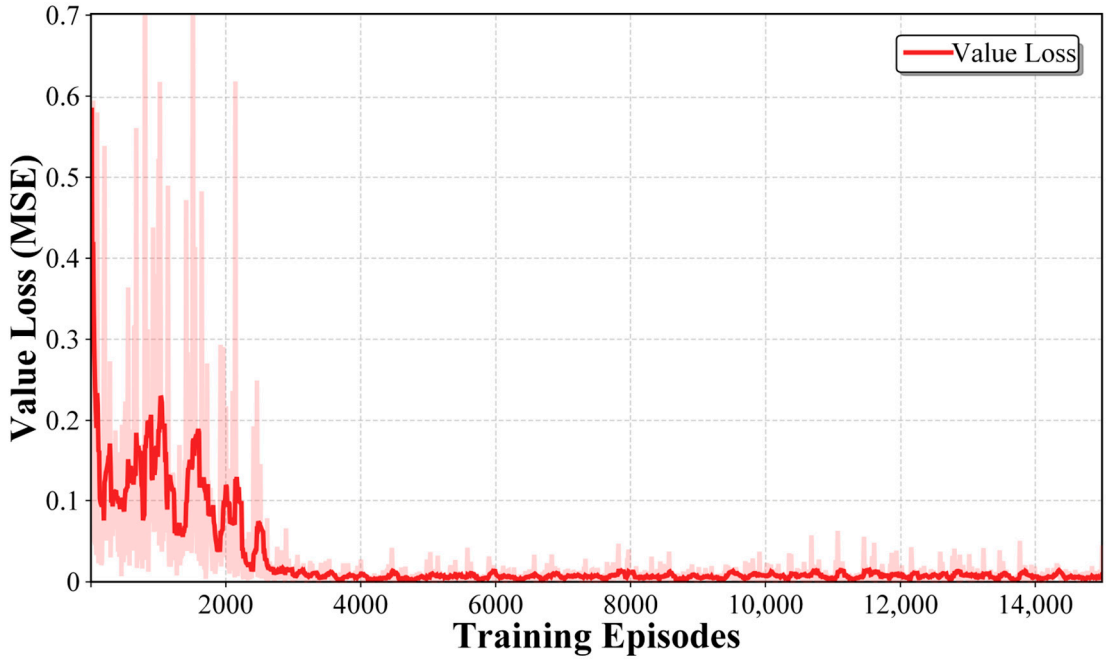


Figure 5. Value loss curve of the MAPPO algorithm.

In addition, the converged policies are tested using the same test dataset. The performance is evaluated in terms of IESO benefit, incentive cost, and testing time, with the results presented in Table 4.

Table 4. Test performance.

Algorithms	IESO Benefit/$	Incentive Cost/$	Testing Time/s
MAPPO	1360.24	2300.21	6.3
MADDPG	1115.79	3420.22	6.4
MADQN	1077.78	3037.80	6.4

Compared to those of MADDPG and MADQN, MAPPO achieves the highest IESO benefit, with improvements of 21.91% and 26.21%, respectively, while incurring the lowest incentive cost. This indicates that MAPPO can more effectively coordinate incentive decisions among electricity, heat, and gas entities. It can also allocate incentive resources according to the response requirements of different energy loads, reducing redundant incentive costs and exploiting the demand response potential. In contrast, MADDPG and MADQN exhibit weaker policy optimization performance, which limits the effectiveness of incentive resource allocation. Therefore, they require higher incentive costs to achieve similar demand response effects.

The differences in testing time among the three algorithms are relatively small. All the algorithms can complete policy decisions within a short period. This is because most computations are completed during offline training, whereas the testing stage only requires forward propagation through the trained actor networks. Overall, MAPPO achieves superior policy performance with high computational efficiency.

5.2.2. Comparison of Different Response Mechanisms

To analyze the impacts of demand-side multi-energy coupling and the bidirectional response mechanism on incentive strategy optimization in the IBIDR, three DR cases are designed for comparison:

(1) The coupling relationship between electricity, heat, and gas demand responses is not considered. IECs are only allowed to perform load reduction responses.
(2) The coupling relationship between electricity, heat, and gas demand responses is considered. IECs are only allowed to perform load reduction responses.
(3) The coupling relationship between electricity, heat, and gas demand responses is considered. The electricity demand response is extended to bidirectional response.

The benefits to the IESO and IEC under different response mechanisms are presented in Table 5.

Table 5. Benefits to the IESO and IEC under different response mechanisms.

	IESO Benefit/$	IEC Benefit/$	Incentive Cost/$	Discomfort Cost/$
DR Case 1	1180.74	207.29	1673.67	1466.38
DR Case 2	1236.97	531.32	1705.51	1174.19
DR Case 3	1360.24	402.48	2300.21	1897.73

Compared to DR Case 1, DR Case 2 significantly increases the benefit to the IEC, whereas the incentive cost to the IESO changes only slightly. This is because DR Case 1 ignores the demand-side multi-energy coupling relationship and treats the electricity, heat, and gas load responses independently, which limits the accurate evaluation of the actual response costs to IECs. By considering the coupling effects among different energy demand adjustments, DR Case 2 enables the IESO to allocate incentive resources more reasonably and reduces the discomfort cost caused by energy consumption adjustments. The benefit to the IESO is also increased by 4.76%, indicating that demand-side coupling improves the benefits to both the IESO and IEC.

DR Case 3 further introduces a bidirectional electricity response mechanism based on DR Case 2, increasing the benefit to the IESO by 9.97%. The bidirectional response mechanism allows IECs to reduce electricity loads or increase electricity consumption when WPG output is high, improving wind power accommodation and demand-side regulation capability. The expanded adjustment range also increases the discomfort cost to the IECs. Therefore, the IESO needs to provide higher incentive costs to encourage their participation in the IBIDR. Although the benefit to the IEC decreases compared to that of DR Case 2, it remains at a relatively high level. Overall, the results indicate that demand-side multi-energy coupling and bidirectional response mechanisms improve the demand-side regulation capability, which contributes to more effective incentive strategy optimization in the IBIDR.

5.2.3. Load Response Analysis Under Incentive Effects

To further analyze the incentive strategies and load response performance under the IBIDR mechanism, the dynamic incentive and response processes of electricity, heat, and gas loads are investigated. Figure 6 presents the corresponding results for electricity load, including incentive prices, electricity selling prices, electricity load demand, electricity load response, and predicted WPG output.

As shown in Figure 6, the incentive price exhibits a variation trend similar to that of WPG output over time. Meanwhile, the incentive price is also affected by the electricity selling price and load demand. For example, during 1:00–6:00, WPG output is high,

whereas load demand is low. The selling price is at the valley level, which is insufficient to encourage IECs to increase their electricity consumption. Therefore, additional incentive compensation is also required. Although the incentive price is higher than the electricity selling price during this period, it remains economically beneficial due to the reduction in wind curtailment losses. After 7:00, the electricity selling price increases, allowing IECs to obtain electricity cost savings through electricity consumption adjustment. Therefore, the incentive price generally remains below the electricity selling price. When the electricity load demand increases at 17:00 and the electricity selling price reaches the peak level at 23:00, the incentive price is still appropriately increased to alleviate the system energy supply pressure.

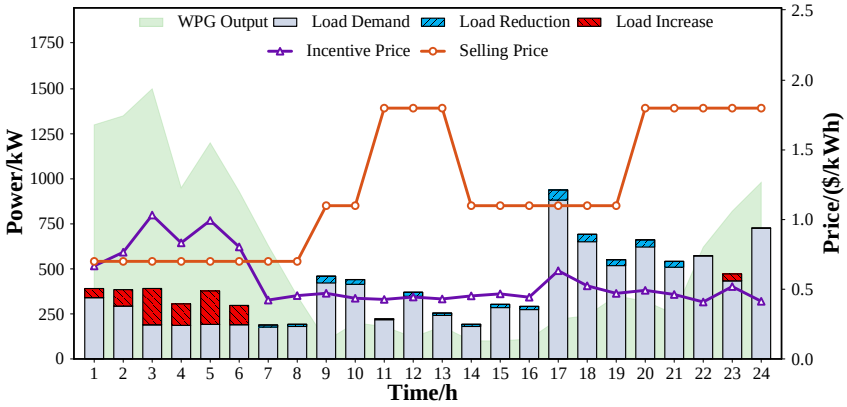


Figure 6. Dynamic incentive and response process of electricity load.

Figures 7 and 8 illustrate the dynamic incentive and response processes of heat and gas loads, respectively. Since the heat and gas loads only participate in load reduction responses, with their energy selling prices remaining constant, the incentive strategies are mainly affected by load demand variations. As the heat and gas load demands increase, the corresponding incentive prices also increase, which encourages IECs to reduce part of their loads and alleviate system energy supply pressure.

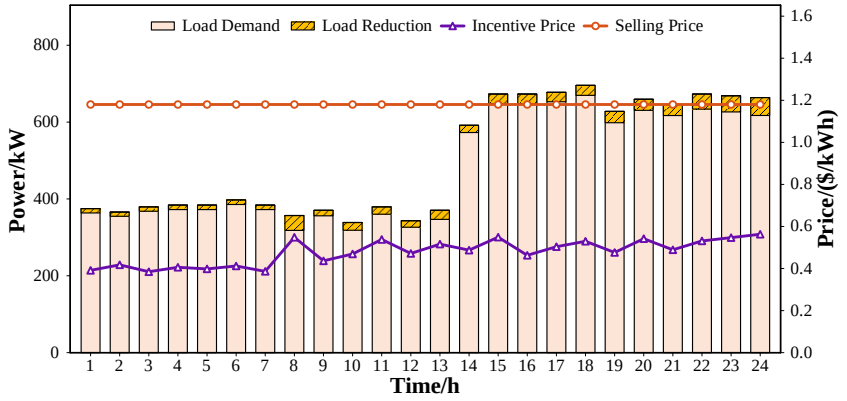


Figure 7. Dynamic incentive and response process of heat load.

The above analysis demonstrates that the MAPPO algorithm can dynamically generate incentive strategies according to changes in state information and effectively guide IECs to participate in the IBIDR.

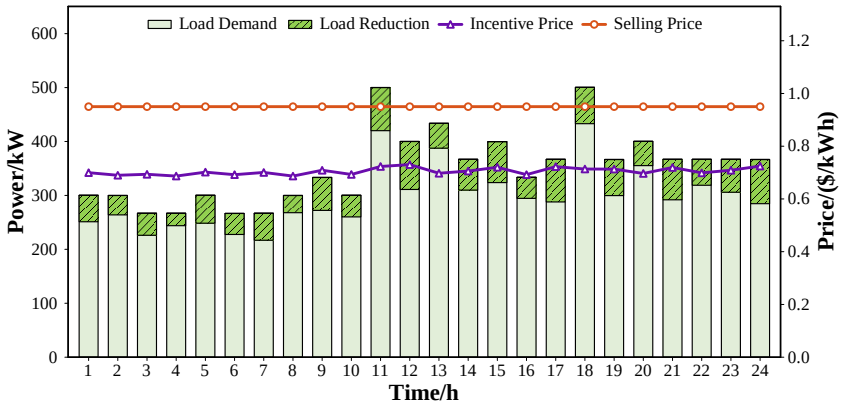


Figure 8. Dynamic incentive and response process of gas load.

5.3. DRO Dispatch Performance Based on the IDT Ambiguity Set

5.3.1. IDT Ambiguity Set Construction and Dispatch Results Under Different Sample Sizes

Based on the demand response results obtained in Section 5.2, this section analyzes the dispatch performance of the IDTDRO method for the IES under source-load uncertainties. The sample size affects the characterization of WPG uncertainty by the IDT ambiguity set. Therefore, WPG sample sets with sizes of 100, 400, 700, and 1000 samples are selected to construct the corresponding ambiguity sets at a 95% confidence level, as shown in Figure 9.

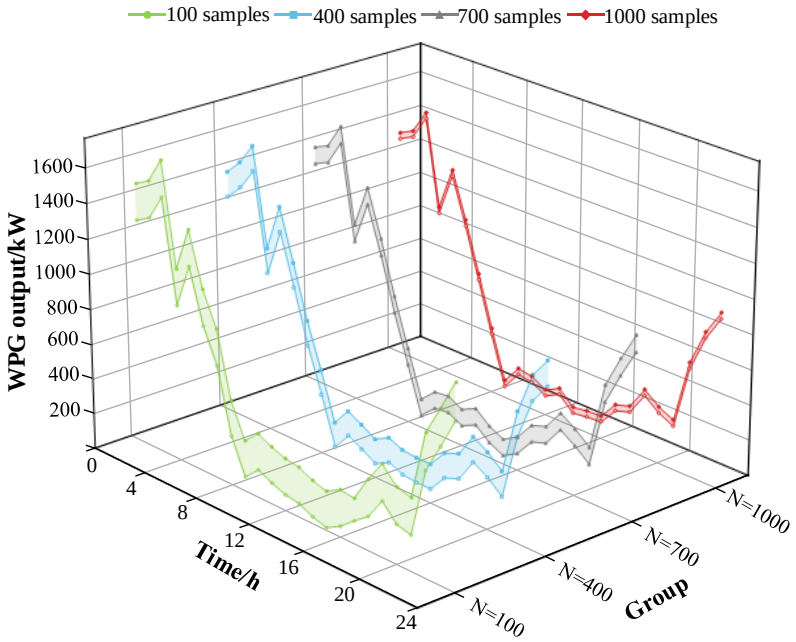


Figure 9. WPG ambiguity set.

Figure 9 shows that the IDT ambiguity sets constructed with different WPG sample sizes exhibit similar overall variation trends, reflecting the uncertainty characteristics of WPG. As the sample size increases, the ambiguity interval gradually narrows, indicating reduced conservativeness in the representation of WPG uncertainty. This is because more sample information enables more refined estimation of the uncertainty boundary. Therefore, the constructed ambiguity set gradually approaches the true probability distribution of WPG samples. The corresponding economic dispatch results are presented in Table 6.

As shown in Table 6, the total dispatch cost is the lowest when the IDT ambiguity set is constructed using 1000 WPG samples. Moreover, the total dispatch cost gradually decreases as the sample size increases. The intra-day regulation cost is also significantly

lower than the day-ahead dispatch cost. These results indicate that the IDT ambiguity set can adjust the WPG uncertainty bounds according to the available samples and reduce the intra-day regulation cost, improving the economic dispatch performance of the IES.

Table 6. Economic dispatch results with different WPG sample sizes.

Sample Size	Day-Ahead Cost/$	Intra-Day Cost/$	Total Cost/$
100	17,290.59	1564.54	18,855.13
400	16,954.21	1470.76	18,424.97
700	16,942.25	1449.39	18,391.64
1000	16,912.03	1357.76	18,269.79

5.3.2. Performance Comparison of Uncertainty Optimization Methods

To evaluate the performance of different uncertainty optimization methods, SO, RO, ARO, and IDTDRO are selected for comparison. SO generates 100 random scenarios to simulate uncertainty. RO and ARO employ box-type uncertainty sets to characterize source–load uncertainties, in which the WPG prediction deviation is set to 30% and the prediction deviations of electricity, heat, and gas load demands are set to 15%. IDTDRO constructs the WPG ambiguity set using 1000 samples and employs box-type uncertainty sets for electricity, heat, and gas load uncertainties. The adjustable robustness parameters of ARO and IDTDRO are set to 8. The dispatch costs and computation times are presented in Table 7.

Table 7. Performance comparison of uncertainty methods.

Optimization Method	Day-Ahead Cost/$	Intra-Day Cost/$	Total Cost/$	Computation Time/s
SO	17,024.96	2708.71	19,733.67	165.51
RO	17,466.45	8165.98	25,632.43	8.81
ARO	17,253.34	3827.23	21,080.57	8.98
IDTDRO	16,912.03	1357.76	18,269.79	9.47

As shown in Table 7, IDTDRO achieves the lowest total dispatch cost under source–load uncertainties. Compared to RO, the total cost is reduced by 28.72%. RO only considers the worst-case scenario within the predefined uncertainty set, which leads to conservative dispatch results and higher operation costs. ARO reduces this conservatism by introducing adjustable robustness parameters, which lower the intra-day cost. IDTDRO further incorporates statistical information from sample data to improve uncertainty characterization and reduce the intra-day cost. In terms of time consumed in optimizations, although SO achieves relatively good economic performance, it requires a large number of random scenarios, resulting in an extremely long computation time. The computation times of RO, ARO, and IDTDRO are similar; however, it can still be observed that IDTDRO achieves a lower total cost with only a slight increase in computation time.

5.4. Performance of the Two-Stage Optimization

To further verify the effectiveness of the MAPPO-IDTDRO-based two-stage optimization framework for improving the economic performance of IES operation, the following three dispatch cases are designed for comparison:

(1) Deterministic optimal dispatch of the IES without considering source–load uncertainties;
(2) Optimal dispatch based on the IDTDRO method under source–load uncertainties;
(3) Optimal dispatch by combining the MAPPO-based incentive strategy with the IDTDRO method.

The comparison results of the two-stage optimization are shown in Figure 10.

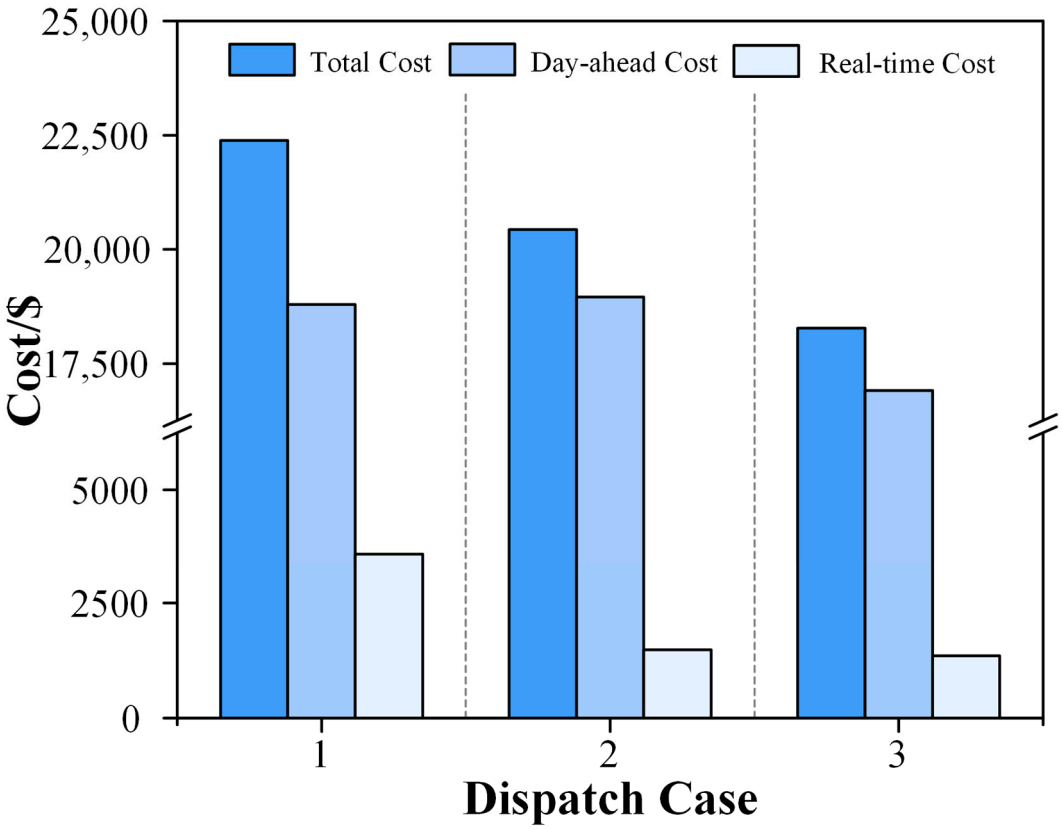


Figure 10. Performance of the two-stage optimization framework.

Figure 10 shows that the total dispatch cost decreases in three cases. Compared to that of Dispatch Case 2, Dispatch Case 1 achieves a slightly lower day-ahead cost, but its total cost is higher. This is because the deterministic optimal dispatch schedule ignores the uncertainties in the day-ahead stage, leading to a higher intra-day cost for compensating prediction deviations. Furthermore, the MAPPO-based incentive strategy is able to further reduce the total cost by 11.85% compared to that of Dispatch Case 2. This indicates that, compared to optimization dispatch only from the system side, the presented MAPPO-IDTDRO-based two-stage optimization framework can exploit demand-side regulation potential and significantly improve the economic performance of IES operation.

5.5. Sensitivity Analysis

5.5.1. Sensitivity Analysis of the PPO Clipping Parameter

The PPO clipping parameter controls the allowable variation in the probability ratio between the old and updated policies. It may affect the convergence speed and average reward of MAPPO. The performance results under different PPO clipping parameters are compared in Table 8.

Table 8. Performance comparison under different PPO clipping parameters.

PPO Clipping Parameter	Convergence Episode	Average Reward
0.10	6500	14.25
0.20	5900	14.63
0.30	6700	13.46
0.40	6300	11.91
0.50	7400	14.08

The PPO clipping parameter of 0.20 achieves the highest average reward and the fastest convergence among the tested settings. At 0.10, the restrictive clipping range limits beneficial policy updates and slows learning. Increasing the parameter beyond 0.20 reduces the clipping restriction on changes between successive policies. The resulting aggressive updates disrupt previously learned incentive strategies, contributing to the lower average rewards observed under these settings. Therefore, 0.20 is adopted to allow effective policy improvement without excessive policy changes, balancing learning efficiency and final reward performance.

5.5.2. Sensitivity Analysis of Adjustable Robustness Parameter

The adjustable robustness parameters control the conservativeness of uncertainty optimization, affecting the balance between economic performance and robustness. The performance comparison results under different adjustable robustness parameters are shown in Table 9.

Table 9. Performance comparison under different adjustable robustness parameters.

Adjustable Robustness Parameter	Day-Ahead Cost/$	Intra-Day Cost/$	Total Cost/$
4	16,794.32	486.96	17,281.28
6	16,759.65	993.11	17,752.76
8	16,912.03	1357.76	18,269.79
10	17,182.23	1620.95	18,803.18
12	17,236.54	2003.32	19,239.86

The total dispatch cost increases by 11.33% as the adjustable robustness parameter rises from 4 to 12, mainly due to higher intra-day regulation costs. A larger parameter allows more adverse source–load conditions to be considered, increasing regulation requirements and the conservativeness of dispatch. Smaller values reduce dispatch costs but provide protection against a more limited range of uncertainty. Accordingly, a moderate value of 8 is adopted as a compromise between economic performance and robustness.

5.5.3. Sensitivity Analysis of Multi-Energy Load Prediction Deviation

Multi-energy load prediction deviations are considered to examine the applicability of the dispatch method under different prediction accuracy levels in practical operation. The performance comparison results under different multi-energy load prediction deviations are shown in Table 10.

Table 10. Performance comparison under different prediction deviations.

Multi-Energy Load Prediction Deviation	Day-Ahead Cost/$	Intra-Day Cost/$	Total Cost/$
0.10	16,759.65	816.37	17,576.02
0.15	16,912.03	1357.76	18,269.79
0.20	17,203.32	1866.53	19,069.85
0.25	17,298.63	2600.99	19,899.62
0.30	17,256.98	3489.70	20,746.68

The total dispatch cost increases by 18.04% as the multi-energy load prediction deviation rises from 10% to 30%. Higher intra-day regulation costs account for approximately 84.31% of this increase. This is because larger prediction deviations widen the load uncertainty ranges, requiring greater regulation to maintain the electricity, heat, and gas

balances under adverse conditions. The presented framework accommodates these deviations through intra-day adjustments, supporting its applicability across the investigated prediction accuracy levels at the expense of higher regulation costs.

6. Discussion

The experimental results verify the contributions of this study from different perspectives. Firstly, the two-stage framework reduces the total operation cost by 11.85% compared with the framework without an IBIDR. This result demonstrates the effectiveness of coordinating day-ahead demand-side regulation with intra-day system-side dispatch. Secondly, comparison of IBIDR mechanisms shows that multi-energy coupling and bidirectional response improve operation benefits and reduce WPG curtailment. MAPPO also achieves better reward performance and training stability than MADDPG and MADQN. Thirdly, IDTDRO achieves better economic performance than the other uncertainty optimization methods. The corresponding ambiguity set is progressively refined as the sample size increases.

These results address the research issues identified in Section 1.2 and support the expected effects of the presented framework. Coordination between the day-ahead and intra-day stages improves the regulation capability and economic performance of IES operation. The IBIDR model provides greater response flexibility by considering multi-energy coupling and bidirectional response characteristics. MAPPO learns coordinated decisions through multi-agent interactions, which improves the utilization of demand-side regulation potential. The IDT ambiguity set makes full use of the statistical information contained in historical samples. Its incorporation into DRO reduces conservativeness and improves the economic performance of IES dispatch.

The experimental results also demonstrate how the presented framework addresses the research gaps identified in previous studies. The two-stage framework enables coordinated regulation between the demand side and system side. The combination of bidirectional response modeling and MAPPO-based multi-agent coordination fully exploits the regulation potential of demand-side resources. The IDT ambiguity set provides an effective approach for modeling WPG uncertainty under limited sample conditions.

In future work, the presented two-stage optimization framework will be extended to distribution networks or multiple IESs to fully exploit the coordination potential of multi-level and multi-regional regulation resources, further enhance system regulation capability, and achieve greater economic benefits.

7. Conclusions

To balance the benefits of IECs and the economic performance of IES dispatch, this paper presents a two-stage optimization framework based on MAPPO-IDTDRO. Firstly, a two-stage framework consisting of day-ahead and intra-day stages is presented to achieve coordinated regulation between the system side and the demand side. Secondly, an IBIDR model considering the multi-energy coupling relationship and bidirectional response characteristics is established. The MAPPO algorithm is employed to learn incentive strategies matched to consumer response characteristics. Finally, an IDT-based DRO method is adopted for IES economic dispatch under source-load uncertainties. The experimental results demonstrate the following:

(1) The presented MAPPO-IDTDRO-based two-stage optimization framework enables effective coordination between day-ahead incentive strategy optimization and intra-day economic dispatch. Compared to system-side-only dispatch optimization, the two-stage optimization framework reduces the economic dispatch cost of the IES by 11.85%.

(2) The IBIDR model can accurately characterize the multi-energy response characteristics of IECs. By employing the MAPPO algorithm, reasonable incentive strategies can be learned to encourage IECs to participate in IBIDRs. With the consideration of multi-energy coupling, the IESO benefit increases by 4.76%. After introducing the bidirectional response mechanism, demand-side regulation becomes more flexible, and the wind power accommodation capability of the system is enhanced. The IESO benefit is further increased by 9.97%. Compared to benchmark algorithms, MAPPO achieves superior performance in terms of training convergence speed, stability, and final policy performance. It also achieves the lowest incentive cost and increases the IESO benefit by 21.91% and 26.21% compared with MADDPG and MADQN, respectively.

(3) The presented IDT ambiguity set can effectively characterize WPG uncertainty. As the number of WPG samples increases, more statistical information becomes available. The resulting ambiguity set becomes narrower, reducing dispatch conservativeness.

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

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

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

Appendix A

Appendix A.1. Energy Balance Constraints

The day-ahead energy balance constraints are expressed as Equation (A1):

$$\begin{cases} P_{\mathrm{FC,DA}}(t)+P_{\mathrm{MT,DA}}(t)+P^{\mathrm{buy}}_{\mathrm{Grid,DA}}(t)+P^{\mathrm{dis}}_{\mathrm{ESS,DA}}(t)+P^{\mathrm{pre}}_{\mathrm{WPG}}(t)= \\ P_{\mathrm{EB,DA}}(t)+P_{\mathrm{P2G,DA}}(t)+P^{\mathrm{pre}}_{\mathrm{EL}}(t)+P^{\mathrm{ch}}_{\mathrm{ESS,DA}}(t)+P^{\mathrm{sell}}_{\mathrm{Grid,DA}}(t) \\ Q_{\mathrm{EB,DA}}(t)+Q_{\mathrm{MT,DA}}(t)+Q^{\mathrm{sup}}_{\mathrm{Heat,DA}}(t)+Q^{\mathrm{dis}}_{\mathrm{HSS,DA}}(t)=Q^{\mathrm{pre}}_{\mathrm{HL}}(t)+Q^{\mathrm{ch}}_{\mathrm{HSS,DA}}(t) \\ G_{\mathrm{P2G,DA}}(t)+G^{\mathrm{sup}}_{\mathrm{Gas,DA}}(t)=G_{\mathrm{MT,DA}}(t)+G^{\mathrm{pre}}_{\mathrm{GL}}(t) \end{cases}, \quad \text{(A1)}$$

where $P_{\mathrm{MT,DA}}(t)$ denotes the day-ahead output power of MT; $P^{\mathrm{dis}}_{\mathrm{ESS,DA}}(t)$ and $P^{\mathrm{ch}}_{\mathrm{ESS,DA}}(t)$ denote the day-ahead discharging and charging power of ESS; $P_{\mathrm{EB,DA}}(t)$ and $P_{\mathrm{P2G,DA}}(t)$ denote the day-ahead electricity consumption power of the EB and P2G; $Q_{\mathrm{EB,DA}}(t)$ and $Q_{\mathrm{MT,DA}}(t)$ denote the day-ahead heat output power of the EB and MT; $Q^{\mathrm{dis}}_{\mathrm{HSS,DA}}(t)$ and $Q^{\mathrm{ch}}_{\mathrm{HSS,DA}}(t)$ denote the day-ahead discharging and charging heat power of the HSS; $G_{\mathrm{P2G,DA}}(t)$ denotes the day-ahead gas output of P2G; and $G_{\mathrm{MT,DA}}(t)$ denotes the day-ahead gas consumption of the MT.

The intra-day energy balance constraints are expressed as Equation (A2):

$$\begin{cases} P_{\mathrm{FC,DA}}(t)+P^{\mathrm{up}}_{\mathrm{FC,ID}}(t)-P^{\mathrm{dn}}_{\mathrm{FC,ID}}(t)+P_{\mathrm{MT,DA}}(t)+P^{\mathrm{up}}_{\mathrm{MT,ID}}(t)-P^{\mathrm{dn}}_{\mathrm{MT,ID}}(t) \\ +P^{\mathrm{buy}}_{\mathrm{Grid,DA}}(t)+P^{\mathrm{buy}}_{\mathrm{Grid,ID}}(t)+P^{\mathrm{dis}}_{\mathrm{ESS,ID}}(t)+P_{\mathrm{WPG,ID}}(t)=P_{\mathrm{EB,DA}}(t)+P^{\mathrm{up}}_{\mathrm{EB,ID}}(t)-P^{\mathrm{dn}}_{\mathrm{EB,ID}}(t) \\ +P_{\mathrm{P2G,DA}}(t)+P^{\mathrm{up}}_{\mathrm{P2G,ID}}(t)-P^{\mathrm{dn}}_{\mathrm{P2G,ID}}(t)+P_{\mathrm{EL,ID}}(t)+P^{\mathrm{ch}}_{\mathrm{ESS,ID}}(t)+P^{\mathrm{sell}}_{\mathrm{Grid,DA}}(t)+P^{\mathrm{sell}}_{\mathrm{Grid,ID}}(t) \\ Q_{\mathrm{EB,DA}}(t)+Q^{\mathrm{up}}_{\mathrm{EB,ID}}(t)-Q^{\mathrm{dn}}_{\mathrm{EB,ID}}(t)+Q_{\mathrm{MT,DA}}(t)+Q^{\mathrm{up}}_{\mathrm{MT,ID}}(t)-Q^{\mathrm{dn}}_{\mathrm{MT,ID}}(t) \\ +Q^{\mathrm{sup}}_{\mathrm{Heat,DA}}(t)+Q^{\mathrm{sup}}_{\mathrm{Heat,ID}}(t)+Q^{\mathrm{dis}}_{\mathrm{HSS,ID}}(t)=Q_{\mathrm{HL,ID}}(t)+Q^{\mathrm{ch}}_{\mathrm{HSS,ID}}(t) \\ G_{\mathrm{P2G,DA}}(t)+G^{\mathrm{up}}_{\mathrm{P2G,ID}}(t)-G^{\mathrm{dn}}_{\mathrm{P2G,ID}}(t)+G^{\mathrm{sup}}_{\mathrm{Gas,DA}}(t)+G^{\mathrm{sup}}_{\mathrm{Gas,ID}}(t)= \\ G_{\mathrm{MT,DA}}(t)+G^{\mathrm{up}}_{\mathrm{MT,ID}}(t)-G^{\mathrm{dn}}_{\mathrm{MT,ID}}(t)+G_{\mathrm{GL,ID}}(t) \end{cases}, \quad \text{(A2)}$$

where $P_{FC,ID}^{up}(t)$, $P_{MT,ID}^{up}(t)$, $P_{EB,ID}^{up}(t)$, and $P_{P2G,ID}^{up}(t)$ denote the upward regulation power of the FC, MT, EB, and P2G; $P_{FC,ID}^{dn}(t)$, $P_{MT,ID}^{dn}(t)$, $P_{EB,ID}^{dn}(t)$, and $P_{P2G,ID}^{dn}(t)$ denote the corresponding downward regulation power; $P_{ESS,ID}^{dis}(t)$ and $P_{ESS,ID}^{ch}(t)$ denote the intra-day discharging and charging power of the ESS; $Q_{EB,ID}^{up}(t)$ and $Q_{MT,ID}^{up}(t)$ denote the upward regulation heat power of the EB and MT; $Q_{EB,ID}^{dn}(t)$ and $Q_{MT,ID}^{dn}(t)$ denote the corresponding downward regulation heat power; $Q_{HSS,ID}^{dis}(t)$ and $Q_{HSS,ID}^{ch}(t)$ denote the intra-day discharging and charging heat power of the HSS; $G_{P2G,ID}^{up}(t)$ and $G_{P2G,ID}^{dn}(t)$ denote the upward and downward regulation gas output of P2G; and $G_{MT,ID}^{up}(t)$ and $G_{MT,ID}^{dn}(t)$ denote the upward and downward gas consumption of the MT. The MT, EB, and P2G should satisfy the energy conversion relationships shown in Equation (A3):

$$\begin{cases} Q_{MT}(t) = P_{MT}(t)\eta_{MT}^{H}, P_{MT}(t) = G_{MT}(t)\eta_{MT}^{E} \\ G_{P2G}(t) = P_{P2G}(t)\eta_{P2G} \\ Q_{EB}(t) = P_{EB}(t)\eta_{EB} \end{cases}, \tag{A3}$$

where η_{MT}^{H} and η_{MT}^{E} denote the energy conversion efficiencies of the MT, and η_{EB} and η_{P2G} denote the energy conversion efficiencies of the EB and P2G.

Appendix A.2. Uncertain Variable Constraints

The source–load uncertain variable constraints are expressed as Equation (A4):

$$\begin{cases} 0 \le P_{WPG,ID}(t) \le P_{WPG}(t) \\ 0 \le Q_{HL,ID}(t) \le Q_{HL}(t) \\ 0 \le G_{GL,ID}(t) \le G_{GL}(t) \\ 0 \le P_{EL,ID}(t) \le P_{EL}(t) \end{cases}. \tag{A4}$$

Appendix A.3. CGs Operation Constraints

The day-ahead operation constraints of CGs are expressed as Equation (A5):

$$\begin{cases} -R_{CGs}^{dn} \le P_{CGs}(t) - P_{CGs}(t-1) \le R_{CGs}^{up} \\ S_{CGs}(t)P_{CGs}^{min} \le P_{CGs}(t) \le S_{CGs}(t)P_{CGs}^{max} \end{cases}, \tag{A5}$$

where $S_{CGs}(t)$ denotes the operation states of CGs; P_{CGs}^{max} and P_{CGs}^{min} denote the upper and lower output limits of CGs; and R_{CGs}^{up} and R_{CGs}^{dn} denote the upper and lower ramping rate limits of CGs. The intra-day regulation constraints of CGs are expressed as Equation (A6):

$$\begin{cases} S_{CGs}^{up}(t) + S_{CGs}^{dn}(t) \le 1 \\ 0 \le P_{CGs}^{up}(t) \le S_{CGs}^{up}(t)P_{CGs}^{up,max}, 0 \le P_{CGs}^{dn}(t) \le S_{CGs}^{dn}(t)P_{CGs}^{dn,max} \\ -R_{CGs}^{dn} \le P_{CGs}(t) + P_{CGs}^{up}(t) - P_{CGs}^{dn}(t) - \\ \left[P_{CGs}(t-1) + P_{CGs}^{up}(t-1) - P_{CGs}^{dn}(t-1)\right] \le R_{CGs}^{up} \\ S_{CGs}(t)P_{CGs}^{min} \le P_{CGs}(t) + P_{CGs}^{up}(t) - P_{CGs}^{dn}(t) \le S_{CGs}(t)P_{CGs}^{max} \end{cases}, \tag{A6}$$

where $S_{CGs}^{up}(t)$ and $S_{CGs}^{dn}(t)$ denote the upward and downward regulation states of CGs, and $P_{CGs}^{up,max}$ and $P_{CGs}^{dn,max}$ denote the upward and downward regulation limits of CGs.

Appendix A.4. Electricity Interaction Constraints

The day-ahead electricity interaction constraints are expressed as Equation (A7):

$$\begin{cases} S_{Grid}^{buy}(t) + S_{Grid}^{sell}(t) \le 1 \\ 0 \le P_{Grid,DA}^{sell}(t) \le S_{Grid}^{sell}(t)P_{Grid}^{max} \\ 0 \le P_{Grid,DA}^{buy}(t) \le S_{Grid}^{buy}(t)P_{Grid}^{max} \end{cases}, \tag{A7}$$

where $S_{\text{Grid}}^{\text{buy}}(t)$ and $S_{\text{Grid}}^{\text{sell}}(t)$ denote the electricity purchasing and selling states, and $P_{\text{Grid}}^{\max}$ denotes the maximum electricity interaction power. The intra-day regulation should satisfy the constraints in Equation (A8):

$$\begin{cases} 0 \le P_{\text{Grid,DA}}^{\text{sell}}(t) + P_{\text{Grid,ID}}^{\text{sell}}(t) \le P_{\text{Grid}}^{\max} \\ 0 \le P_{\text{Grid,DA}}^{\text{buy}}(t) + P_{\text{Grid,ID}}^{\text{buy}}(t) \le P_{\text{Grid}}^{\max} \end{cases}. \tag{A8}$$

Appendix A.5. Storage System Operation Constraints

The storage system operation constraints are expressed as Equations (A9)–(A12):

$$\begin{cases} S_{\text{HSS}}^{\text{ch}}(t) + S_{\text{HSS}}^{\text{dis}}(t) \le 1 \\ S_{\text{ESS}}^{\text{ch}}(t) + S_{\text{ESS}}^{\text{dis}}(t) \le 1 \end{cases}, \tag{A9}$$

$$\begin{cases} H_{\text{HSS}}(t) = (1-\delta_{\text{HSS}})H_{\text{HSS}}(t-1) + \eta_{\text{HSS}}^{\text{ch}} Q_{\text{HSS}}^{\text{ch}}(t-1) - \frac{Q_{\text{HSS}}^{\text{dis}}(t-1)}{\eta_{\text{HSS}}^{\text{dis}}} \\ E_{\text{ESS}}(t) = (1-\delta_{\text{ESS}})E_{\text{ESS}}(t-1) + \eta_{\text{ESS}}^{\text{ch}} P_{\text{ESS}}^{\text{ch}}(t-1) - \frac{P_{\text{ESS}}^{\text{dis}}(t-1)}{\eta_{\text{ESS}}^{\text{dis}}} \end{cases}, \tag{A10}$$

$$\begin{cases} S_{\text{HSS}}^{\text{dis}}(t) Q_{\text{HSS}}^{\text{dis,min}} \le Q_{\text{HSS}}^{\text{dis}}(t) \le S_{\text{HSS}}^{\text{dis}}(t) Q_{\text{HSS}}^{\text{dis,max}} \\ S_{\text{HSS}}^{\text{ch}}(t) Q_{\text{HSS}}^{\text{ch,min}} \le Q_{\text{HSS}}^{\text{ch}}(t) \le S_{\text{HSS}}^{\text{ch}}(t) Q_{\text{HSS}}^{\text{ch,max}} \\ S_{\text{ESS}}^{\text{dis}}(t) P_{\text{ESS}}^{\text{dis,min}} \le P_{\text{ESS}}^{\text{dis}}(t) \le S_{\text{ESS}}^{\text{dis}}(t) P_{\text{ESS}}^{\text{dis,max}} \\ S_{\text{ESS}}^{\text{ch}}(t) P_{\text{ESS}}^{\text{ch,min}} \le P_{\text{ESS}}^{\text{ch}}(t) \le S_{\text{ESS}}^{\text{ch}}(t) P_{\text{ESS}}^{\text{ch,max}} \end{cases}, \tag{A11}$$

$$\begin{cases} H_{\text{HSS}}^{\min} \le H_{\text{HSS}}(t) \le H_{\text{HSS}}^{\max}, E_{\text{ESS}}^{\min} \le E_{\text{ESS}}(t) \le E_{\text{ESS}}^{\max} \\ H_{\text{HSS}}(1) = H_{\text{HSS}}(24), E_{\text{ESS}}(1) = E_{\text{ESS}}(24) \end{cases}, \tag{A12}$$

where $S_{\text{ESS}}^{\text{ch}}(t)$, $S_{\text{HSS}}^{\text{ch}}(t)$, $S_{\text{ESS}}^{\text{dis}}(t)$ and $S_{\text{HSS}}^{\text{dis}}(t)$ denote the charging and discharging states of the ESS and HSS; $P_{\text{ESS}}^{\text{ch,max}}$ and $P_{\text{ESS}}^{\text{dis,max}}$ denote the maximum charging and discharging power of the ESS; $P_{\text{ESS}}^{\text{ch,min}}$ and $P_{\text{ESS}}^{\text{dis,min}}$ denote the corresponding minimum charging and discharging power; $Q_{\text{HSS}}^{\text{ch,max}}$ and $Q_{\text{HSS}}^{\text{dis,max}}$ denote the maximum charging and discharging heat power of the HSS; $Q_{\text{HSS}}^{\text{ch,min}}$ and $Q_{\text{HSS}}^{\text{dis,min}}$ denote the corresponding minimum charging and discharging heat power; $E_{\text{ESS}}(t)$ and $H_{\text{HSS}}(t)$ denote the energy storage capacities of the ESS and HSS; δ_{ESS} and δ_{HSS} denote the energy loss coefficients of the ESS and HSS; $\eta_{\text{ESS}}^{\text{ch}}$ and $\eta_{\text{HSS}}^{\text{ch}}$ denote the charging efficiencies of the ESS and HSS; $\eta_{\text{ESS}}^{\text{dis}}$ and $\eta_{\text{HSS}}^{\text{dis}}$ denote the corresponding discharging efficiencies; $E_{\text{ESS}}^{\max}$ and $H_{\text{HSS}}^{\max}$ denote the maximum energy storage capacities of the ESS and HSS; and $E_{\text{ESS}}^{\min}$ and $H_{\text{HSS}}^{\min}$ denote the corresponding minimum energy storage capacities.

Appendix B

Table A1. Experimental environment.

Term	Configuration
CPU	Intel(R) Core i7-13700F@2.10GHz
RAM	32GB
OS	Windows 11
Platform	MATLAB R2023b/Python 3.12
Solver	CPLEX

Table A2. Parameters of CGs.

CGs	P^{max}/kW	P^{min}/kW	R/kW	α	β	λ^{up}	λ^{down}
MT	600	0	240	-	-	1.35	2.375
FC	400	0	130	0.85	0	1.9	1.9
EB	250	0	160	-	-	1.00	1.00
P2G	200	0	100	-	-	1.00	1.00

Table A3. Parameters of storage systems.

Storage System	Maximum Power/kW	SOC^{max}/kWh	SOC^{min}/kWh	SOC^{0}/kWh	η^{ch}/η^{dis}	δ
ESS	300	900	40	180	1.1/0.9	0.001
HSS	300	1000	80	200	1.1/0.9	0.01

Table A4. Trading prices of electricity.

Time	Day-Ahead Stage		Intra-Day Stage	
	$\lambda^{buy}_{Grid,DA}$/(\$/kW)	$\lambda^{sell}_{Grid,DA}$/(\$/kW)	$\lambda^{buy}_{Grid,ID}$/(\$/kW)	$\lambda^{sell}_{Grid,ID}$/(\$/kW)
09:00–12:00 and 19:00–00:00	1.8	1.5	2.7	0.75
08:00–09:00 and 12:00–19:00	1.1	0.8	1.65	0.4
00:00–08:00	0.7	0.4	1.05	0.2

Table A5. Purchasing prices of heat and gas.

λ_{Heat}/(\$/kW)	λ_{Gas}/(\$/kW)
1.18	0.95

Table A6. Penalty parameters.

λ_{WPG}/(\$/kW)	λ_{LD}/(\$/kW)
1.5	0.5

Table A7. IEC parameters.

Parameters	Range
ν_e, ν_h, ν_g	[0.003, 0.006], [0.007, 0.010], [0.005, 0.008]
ϑ_e, ϑ_h, ϑ_g	[0.4, 0.5], [0.2, 0.3], [0.3, 0.4]
e_e, e_h, e_g	[0.7, 0.8]
e_o	[0.1, 0.2]

Table A8. Training parameters of MAPPO.

Parameters	Value
Discount factor	0.99
Advantage estimation factor	0.95
Clipping rate	0.2
Learning rate of actor network	0.0001
Learning rate of critic network	0.0003
Maximum training episodes	15,000

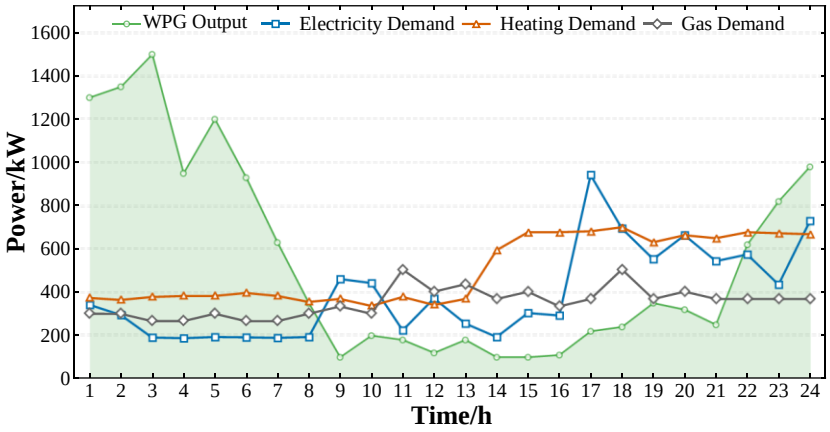


Figure A1. Predicted WPG output and multi-energy load demands.

References

1. Li, Z.; Ye, L.; Pei, M.; Song, X.; Luo, Y.; Tang, Y. Interpretable Augmented Ambiguity Set for Quantifying Regional Wind Power Uncertainty in Distributionally Robust Optimal Dispatch. *IEEE Trans. Sustain Energy* **2026**, *17*, 259–276. [CrossRef]
2. Liu, R.P.; Hou, Y.; Li, Y.; Lei, S.; Wei, W.; Wang, X. Sample Robust Scheduling of Electricity-Gas Systems Under Wind Power Uncertainty. *IEEE Trans. Power Syst.* **2021**, *36*, 5889–5900. [CrossRef]
3. Lin, Z.; Chen, H.; Chen, J.; Huang, J.; Li, M. Risk-averse Robust Interval Economic Dispatch for Power Systems with Large-scale Wind Power Integration. *CSEE J. Power Energy Syst.* **2024**, *10*, 105–116. [CrossRef]
4. Ding, J.; Xie, K.; Hu, B.; Shao, C.; Niu, T.; Li, C. Mixed Aleatory-epistemic Uncertainty Modeling of Wind Power Forecast Errors in Operation Reliability Evaluation of Power Systems. *J. Mod. Power Syst. Clean Energy* **2021**, *10*, 1174–1183. [CrossRef]
5. Chen, X.; Liu, Y.; Wu, L. Towards Improving Unit Commitment Economics: An Add-On Tailor for Renewable Energy and Reserve Predictions. *IEEE Trans. Sustain Energy* **2024**, *15*, 2547–2566. [CrossRef]
6. Wang, Y.; Hu, J.; Liu, N. Energy Management in Integrated Energy System Using Energy-Carbon Integrated Pricing Method. *IEEE Trans. Sustain Energy* **2023**, *14*, 1992–2005. [CrossRef]
7. Lyu, G.; Cao, B.; Jia, D.; Wang, N.; Li, J.; Chen, G. Optimal Scheduling of Regional Integrated Energy Systems Considering Hybrid Demand Response. *CSEE J. Power Energy Syst.* **2024**, *10*, 1208–1219. [CrossRef]
8. Fan, W.; Tan, Q.; Zhang, A.; Ju, L.; Wang, Y.; Yin, Z.; Li, X. A Bi-level optimization model of integrated energy system considering wind power uncertainty. *Renew. Energy* **2023**, *202*, 973–991. [CrossRef]
9. Zhao, A.P.; Li, S.; Xie, D.; Wang, Y.; Li, Z.; Hu, P.J.-H.; Zhang, Q. Hydrogen as the nexus of future sustainable transport and energy systems. *Nat. Rev. Electr. Eng.* **2025**, *2*, 447–466. [CrossRef]
10. Liu, Y.; Chen, X.; Wu, L.; Ye, Y. Distributionally Robust Economic Dispatch Using IDM for Integrated Electricity-heat-gas Microgrid Considering Wind Power. *CSEE J. Power Energy Syst.* **2023**, *9*, 1182–1192. [CrossRef]
11. Wu, Y.; Liu, Y. Economic Dispatching for Cyber-Physical Integrated Energy System Based on GP-WGAN and Improved Distributionally Robust Optimization. *J. Circ. Syst. Comp.* **2025**, *34*, 2550336. [CrossRef]
12. Zhang, J.; Liu, Y.; Li, H. Coordinated Optimal Configuration for Hybrid Energy Storage System Involving Differentiated Requirements from Supply-Side and Demand-Side in Microgrid. *Energies* **2026**, *19*, 2410. [CrossRef]
13. Caliano, M.; Delfino, F.; Di Somma, M.; Ferro, G.; Graditi, G.; Parodi, L.; Robba, M.; Rossi, M. An Energy Management System for microgrids including costs, exergy, and stress indexes. *Sustain. Energy Grids Netw.* **2022**, *32*, 100915. [CrossRef]
14. Salim, L.; Farhan, M.A. A supply-demand-ratio based exponential pricing scheme for coordinated peer-to-peer energy sharing in prosumer-consumer microgrids. *Energy Sustain Dev.* **2026**, *94*, 102071. [CrossRef]
15. Ren, X.-Y.; Cai, X.; Wei, B.; Ma, W. Collaborative optimization operation strategy of integrated energy system considering integrated demand response: A Stackelberg game and improved Runge Kutta optimizer. *Energy* **2026**, *344*, 139977. [CrossRef]
16. Huang, W.; Zhang, N.; Kang, C.; Li, M.; Huo, M. From demand response to integrated demand response: Review and prospect of research and application. *Prot. Control Mod. Power Syst.* **2019**, *4*, 12. [CrossRef]
17. Ma, S.; Liu, H.; Wang, N.; Huang, L.; Goh, H.H. Incentive-based demand response under incomplete information based on the deep deterministic policy gradient. *Appl. Energy* **2023**, *351*, 121838. [CrossRef]
18. Zeng, H.; Shao, B.; Dai, H.; Tian, N.; Zhao, W. Incentive-based demand response strategies for natural gas considering carbon emissions and load volatility. *Appl. Energy* **2023**, *348*, 121541. [CrossRef]
19. Alasseri, R.; Rao, T.J.; Sreekanth, K.J. Conceptual framework for introducing incentive-based demand response programs for retail electricity markets. *Energy Strategy Rev.* **2018**, *19*, 44–62. [CrossRef]

20. Hu, R.; Zhou, K.; Yin, H. Reinforcement learning model for incentive-based integrated demand response considering demand-side coupling. *Energy* **2024**, *308*, 132997. [CrossRef]
21. Liu, X.; Zhang, W.; Chi, Z.; Jiang, T.; Ji, Y. Optimal energy scheduling for multi-energy network based on incentive integrated demand response. *Sustain. Energy Grids Netw.* **2025**, *44*, 101986. [CrossRef]
22. Zheng, S.; Sun, Y.; Li, B.; Qi, B.; Zhang, X.; Li, F. Incentive-based integrated demand response for multiple energy carriers under complex uncertainties and double coupling effects. *Appl. Energy* **2021**, *283*, 116254. [CrossRef]
23. Gao, B.; Chen, C.; Qin, Y.; Liu, X.; Zhu, Z. Evolutionary Game-theoretic Analysis for Residential Users Considering Integrated Demand Response. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 1500–1509. [CrossRef]
24. Li, Z.; Yao, Y.; Nan, Z.; Shan, J.; Fu, Y. Study on integrated energy microgrid energy purchase strategy with demand-side response in market environment. *Energy* **2024**, *302*, 131724. [CrossRef]
25. Li, W.; Li, D.; Hou, S.; Xi, H.; Xiao, Y.; Yang, Q. Dual-scale optimization of integrated energy systems: A novel MAPPO-based approach for hybrid game equilibrium. *Appl. Energy* **2026**, *409*, 127468. [CrossRef]
26. Zhang, Z.; Su, Y.; Tan, M.; Cao, R. Fusing domain knowledge and reinforcement learning for home integrated demand response online optimization. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105995. [CrossRef]
27. Ahmed, F.; Arshad, A.; Rehman, A.U.; Alqahtani, M.H.; Mahmoud, K. Effective incentive based demand response with voltage support capability via reinforcement learning based multi-agent framework. *Energy Rep.* **2024**, *12*, 568–578. [CrossRef]
28. Li, T.; Xiao, Y.; Song, L. Integrating Future Smart Home Operation Platform With Demand Side Management via Deep Reinforcement Learning. *IEEE Trans. Green Commun. Netw.* **2021**, *5*, 921–933. [CrossRef]
29. Yazar, O.; Coskun, S.; Zhang, F.; Li, L.; Huang, C.; Mei, P.; Karimi, H. A novel energy management strategy for hybrid electric vehicles using deep reinforcement incentive learning. *Energy* **2025**, *334*, 137594. [CrossRef]
30. Wang, Z.; Yu, X.; Li, H.; Liu, Y.; Zang, T.; Ruan, J.; Chen, S. Federated reinforcement learning-based dynamic hosting domain for electric vehicles charging management in distribution networks with renewable energy resources. *J. Clean Prod.* **2026**, *556*, 148218. [CrossRef]
31. Liu, Z.; Zhang, Q.; Liu, Z.; Lu, Y. Distributionally robust optimization of power grid along mountain railway considering three-phase unbalance of traction load. *Int. J. Electr. Power Energy Syst.* **2026**, *178*, 111969. [CrossRef]
32. Lin, Z.; Wu, Q.; Chen, H.; Ji, T.; Xu, Y.; Sun, H. Scenarios-Oriented Distributionally Robust Optimization for Energy and Reserve Scheduling. *IEEE Trans. Power Syst.* **2023**, *38*, 2943–2946. [CrossRef]
33. Zhang, Y.; Ma, H.; Wang, Y.-Y.; He, K.-L.; Meng, N.; Chen, Q. Distributionally robust optimization of power-to-hydrogen-to-ammonia systems considering wind-solar uncertainty and electrolyzer cold-start characteristics. *Energy* **2026**, *360*, 141790. [CrossRef]
34. Cobos, N.G.; Arroyo, J.M.; Alguacil, N.; Street, A. Robust Energy and Reserve Scheduling Under Wind Uncertainty Considering Fast-Acting Generators. *IEEE Trans. Sustain Energy* **2019**, *10*, 2142–2151. [CrossRef]
35. Wu, J.; Liu, Y.; Chen, X.; Wang, C.; Li, W. Data-driven adjustable robust Day-ahead economic dispatch strategy considering uncertainties of wind power generation and electric vehicles. *Int. J. Electr. Power Energy Syst.* **2022**, *138*, 107898. [CrossRef]
36. Alismail, F.; Xiong, P.; Singh, C. Optimal Wind Farm Allocation in Multi-Area Power Systems Using Distributionally Robust Optimization Approach. *IEEE Trans. Power Syst.* **2018**, *33*, 536–544. [CrossRef]
37. Cao, Y.; Wei, W.; Chen, L.; Wu, Q.; Mei, S. Supply Inadequacy Risk Evaluation of Stand-Alone Renewable Powered Heat-Electricity Energy Systems: A Data-Driven Robust Approach. *IEEE Trans. Ind. Inf.* **2021**, *17*, 1937–1947. [CrossRef]
38. Zheng, S.; Sun, Y.; Li, B.; Qi, B.; Shi, K.; Li, Y.; Tu, X. Incentive-Based Integrated Demand Response for Multiple Energy Carriers Considering Behavioral Coupling Effect of Consumers. *IEEE Trans. Smart Grid* **2020**, *11*, 3231–3245. [CrossRef]
39. Pullinger, M.; Kilgour, J.; Goddard, N.; Berliner, N.; Webb, L.; Dzikovska, M.; Lovell, H.; Mann, J.; Sutton, C.; Webb, J.; et al. The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes. *Sci. Data* **2021**, *8*, 146. [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.