

Article

A Bi-Level Optimal Operation Model for Small-Scale Active Distribution Networks Considering the Coupling Fluctuation of Spot Electricity Prices and Renewable Energy Sources

Yu Shi ¹, Fei Lv ^{2,3,*}, Xuefeng Gao ¹, Minglei Jiang ¹, Huan Luo ^{2,3} and Ruhang Xu ^{2,3} ¹ Power Economic Research Institute of Jilin Electric Power Co., Ltd., Changchun 130021, China² School of Economics and Management, North China Electric Power University, Beijing 102206, China³ Beijing Key Laboratory of New Energy and Low-Carbon Development, North China Electric Power University, Beijing 102206, China

* Correspondence: lvfei@ncepu.edu.cn

Abstract: As the penetration rate of variable renewable energy such as wind power increases in the power system, the composition and balance of the system also change gradually. The intermittency of renewable energy poses great stability challenges to the traditional centralized generation and load-oriented transmission and distribution methods. Therefore, the Active Distribution Network Operator (ADNO) with distributed installation at the local level has a good application prospect in the new scenario. However, ADNO needs to improve its operational efficiency based on the types of local generation and storage devices and the nature of the market environment. To address this issue, this paper proposes a forecasting method that considers the coupling fluctuations of spot electricity prices and renewable energy, and a bi-level optimization operation method based on the Stackelberg game for optimizing the operation of small-scale ADNO under high wind power penetration rate. Simulation results show that the proposed methods achieve greater positive impact on the operational efficiency of ADNO than conventional methods. In addition, the proposed methods ensure the long-term profitability of ADNO, even with fluctuations in external factors.

Keywords: spot price; renewable energy; active distribution network; operation strategy



Citation: Shi, Y.; Lv, F.; Gao, X.; Jiang, M.; Luo, H.; Xu, R. A Bi-Level Optimal Operation Model for Small-Scale Active Distribution Networks Considering the Coupling Fluctuation of Spot Electricity Prices and Renewable Energy Sources.

Energies **2023**, *16*, 4507. <https://doi.org/10.3390/en16114507>

Academic Editor: Marco Merlo

Received: 1 May 2023

Revised: 25 May 2023

Accepted: 30 May 2023

Published: 3 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In order to solve the problem of carbon emissions caused by fossil fuel power generation, energy diversification has become an important trend in power generation. Countries have put forward their own energy diversification development plans. Among them, the active distribution network is a promising energy supply model. The so-called active distribution network refers to an intelligent power system that can achieve bidirectional flow of electricity, while having energy storage, small-scale power generation units, and other functions through intelligent control technology.

The potential of active distribution networks in terms of ensuring energy security and reducing carbon emissions has been widely recognized. According to data from the International Energy Agency, by 2030, over 130 million households and businesses worldwide will use active distribution networks [1]. Countries and regions such as China, the European Union, and the United States have issued relevant policies and plans to actively promote the construction of active distribution networks.

The potential of active distribution networks in ensuring energy security and reducing carbon emissions include the following aspects:

- It can achieve precise control of electricity through intelligent control technology, improve energy utilization efficiency, and reduce energy waste;
- It has energy storage capacity, which can balance the load of power grid effectively and enhance the stability of the power grid;

- Through the use of small-scale power generation units, it can better adapt to the demand for decentralized energy supply and improve the flexibility of energy supply.

Small-scale active distribution networks refer to a new type of power system that integrates distributed power sources, load control, and energy storage systems on a smaller scale based on distribution networks, forming a power system that can actively respond to the needs of the power grid. Small-scale active distribution networks have many advantages and will become an important part of future power supply systems. Small-scale active distribution networks are usually based on distribution networks with ten to dozens of nodes. They are characterized by a small scale but a very large potential base, so their market potential is also enormous. This power system can not only improve the reliability and efficiency of power supply, but also provide more services to customers. Therefore, small-scale active distribution networks have great significance for the future development of the energy field and the flexibility and low-carbon contribution of power supply systems. During peak periods of power supply, a large number of small-scale active distribution networks can quickly respond to power demand, balance the load of the power grid, and thereby improve the reliability of the power supply system. At the same time, small-scale active distribution networks promote the use of renewable energy, reduce carbon emissions, and contribute to the low-carbon development of the power system.

Nevertheless, operators of small-scale active distribution networks also face many problems. The main problems include the following three points:

1. It is difficult to influence the wholesale market price. Small-scale active distribution networks need to import electricity from the electricity wholesale market and distribute it to their customers. As price takers, operators of small-scale active distribution networks passively face price fluctuations;
2. Small-scale active distribution networks rely on controllable fossil energy or unpredictable renewable energy to provide part of their electricity internally to increase power supply flexibility. The uncertainty of variational renewable energy generation from internal and external sources can affect profitability;
3. There exists a game relationship between operators and internal customers. The strategy determined unilaterally by the operator may deviate from the equilibrium point.

This paper aims to tackle these problems through a novel conditional forecasting method and a novel bi-level optimization method. We verify the advantages of the proposed method over conventional methods in improving the operational efficiency of small-scale active distribution networks through simulation experiments, and analyze the impact of external factor fluctuations on the long-term revenue of ADNOS. This paper has important theoretical and practical value for understanding and improving the operational mechanism of small-scale active distribution networks in the context of low-carbon transformation. The rest of this paper is organized as follows: Section 2 is a literature review. The conditional forecasting method is established in Section 3. The bi-level optimization model is established in Section 4. Section 5 is a case study. The conclusions are provided in Section 6.

2. Literature Review

Active distribution networks (ADNs) are becoming increasingly important due to the integration of distributed generation, which changes power flow from unidirectional to bi-directional. In order to manage distribution networks effectively, an active network management strategy that incorporates emerging techniques of control, monitoring, protection, and communication is needed. The advantages of active distribution networks include improved operating economics and reliability of power systems and increased low-carbon performance. Several articles have explored various aspects of active distribution networks. Zhao et al. [1] provided a short review of recent advancements and identify emerging technologies and future development trends for support the active management of distribution networks. Flexible Low-Carbon Optimal Dispatch of Honeycombed Active Distribution Network (HADN) is proposed by Xiang et al. [2] as a way to increase low-

carbon benefits through appropriate HADN dispatch. Ge et al. [3] proposed a substation planning method that accounts for the widespread introduction of distributed generators in a low-carbon economy. Li et al. [4] studied a bi-level interactive optimization model for active distribution networks with microgrids, which minimizes operation cost and voltage deviation to achieve a low-carbon system. Finally, Xiao et al. [5] proposed a new Distribution Management System (DMS) framework based on security region, which aims to help operate the system close to its security boundary in order to improve the efficiency significantly within the same security standard.

To maintain the optimal operation of a network, extensive research studies have been conducted under the theme of network management. Usman et al. [6] investigated losses management strategies in active distribution networks and provided a concise yet comprehensive comparison of the most recently proposed losses management approaches and strategies. Gabash and Li [7] proposed a combined problem formulation for active-reactive optimal power flow in distribution networks with embedded wind generation and battery storage. Cortes et al. [8] proposed an iterative procedure for the optimal design of a microgrid topology in active distribution networks, which applies graph partitioning, integer programming, and a performance index for the optimal design. Kyriakou and Kanellos [9] proposed a method for the coordinated optimal operation scheduling of active distribution networks that are hosting complex microgrids comprising large building prosumers and plug-in electric vehicle aggregators. Sedghi et al. [10] presented a paper on storage scheduling for optimal energy management in active distribution networks considering uncertainties of load, wind, and plug-in electric vehicles. Jiang et al. [11] proposed a combined modeling and optimal scheduling method for active distribution networks with integrated smart buildings. Chen et al. [12] proposed a stochastic optimal operation strategy for distribution networks whose objective function considers the operation state of the distribution network. In order to enhance system resilience, Wang et al. [13] proposed a novel optimal operation strategy for an active distribution network. For the issue of optimal pricing between ADNOs, Tostado-Véliz et al. [14] proposed an equilibrium problem with an equilibrium constraints (EPECs) approach, which solves the bi-level problem associated to each participant sequentially using diagonalization. Finally, Fan et al. [15] proposed a Lyapunov optimization-based online distributed (LOOD) algorithmic framework for active distribution networks (ADNs).

Active distribution networks (ADNs) are developing rapidly due to the increasing penetration of distributed energy resources (DERs), such as renewable energy sources and electric vehicles. Therefore, forecasting is essential for network management to maintain the optimal operation of ADNs. This paragraph summarizes recent research on forecasting in ADNs. Saint-Pierre and Mancarella [16] proposed a scheduling model for active distribution system management (ADSM) to operate distribution network assets with renewable energy sources. The model considers uncertainties, market constraints, and scheduled power flows at the interface with the transmission system. The proposed ADSM approach maximizes renewable penetration, minimizes deviations from time-ahead schedules, and estimates the required level of local reserves from dispatchable generation and electrical energy storage (EES) while accounting for uncertainties. Shen et al. [17] presented a planning model for ADN that considers long-term investment costs and short-term operation conditions with EES benefits for peak load shaving and power reliability enhancement. Yu et al. [18] proposed a novel spatial-temporal graph representation method to characterize and present spatial and temporal correlations of historical load observations. The graph data are used to train a model denoted as Spatial-Temporal Synchronous Graph Convolutional Network (STSGCN) that can forecast load by learning from the graph features. Kalantar-Neyestanaki et al. [19] presented a method to determine the capability area of an ADN for the provision of both active and reactive power reserves while considering forecast errors and operational constraints of the grid and DERs. Cheng et al. [20] proposed a state estimation method for ADNs based on the forecasting of photovoltaic (PV) power generation using the Gaussian mixture model (GMM). Cong et al. [21] proposed a day-ahead active power scheduling

method that considers REG forecast errors to minimize distribution company costs and achieve optimal power flow with a hierarchical coordination optimization model based on chance constrained programming.

Overall, recent research has shown that forecasting plays a crucial role in the optimal operation of ADNs. Various methods, such as branch flow model-based relaxed optimal power flow [22], dual-horizon rolling scheduling models [23], and state estimation methods based on PV power generation forecasting, have been proposed to achieve optimal scheduling of active power and enhance the reliability and efficiency of ADNs. However, one might neglect the effects of the correlation between the fluctuation of wholesale spot price and the output of renewable energy. Moreover, one might overlook the function of ADNO time-of-use subsidy and the game relationship between ADNOs and users. This paper shows that with proper handling of these factors, ADNO may achieve significant profitability increase.

3. Conditional Wholesale Power Price Forecast for High VRE Penetration Market

3.1. Correlation between Wind Power Output and Electricity Spot Price in High Penetration Markets

The Active Distribution Network Operator (ADNO) needs to purchase electricity from the electricity spot wholesale market and sell it to retail customers. ADNO needs to decide on the dispatch strategy for the second day in advance. When the spot wholesale price is too high, ADNO should use appropriate advance dispatching to replace some of the external electricity supply with internal resources, such as renewable energy generators, fossil fuel generators, and energy storage equipment. Conversely, when the spot wholesale price is low, ADNO should optimize its dispatch strategy to maximize the utilization of cheap electricity from the external power grid. The basic operating model of ADNO is shown in Figure 1.

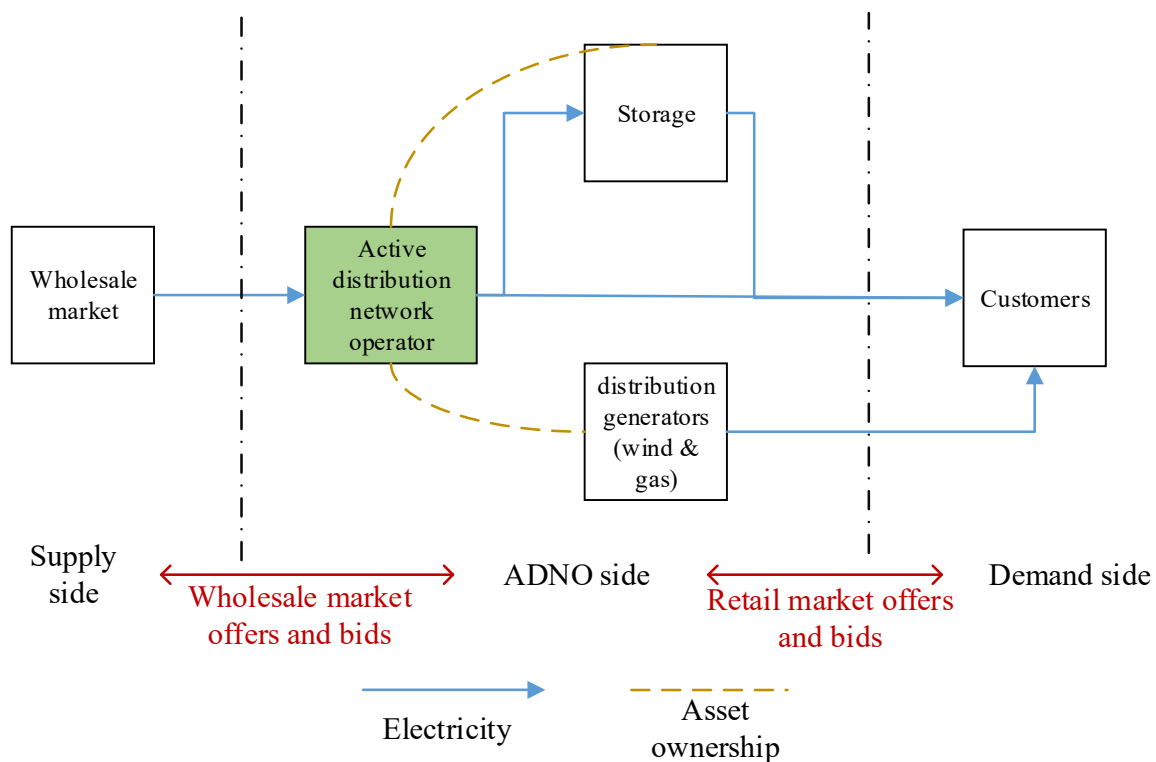


Figure 1. The basic operating model of ADNO.

For small-scale ADNOs, due to their small scale, they cannot affect market prices and can be regarded as price takers in the spot wholesale market. Therefore, they cannot

increase their profitability by influencing market prices. Therefore, accurate prediction of market prices is very important to improve their profitability.

On the other hand, appropriate renewable energy installed capacity is meaningful for reducing the operating costs of ADNOs. Variational renewable energy generation, such as photovoltaic and wind power, has very low variable costs. However, variable renewable energy (VRE) has strong intermittency and requires additional controllable distributed fuel units to supplement it, forming a power source combination. The operating sequence of ADNO is shown in Figure 2.

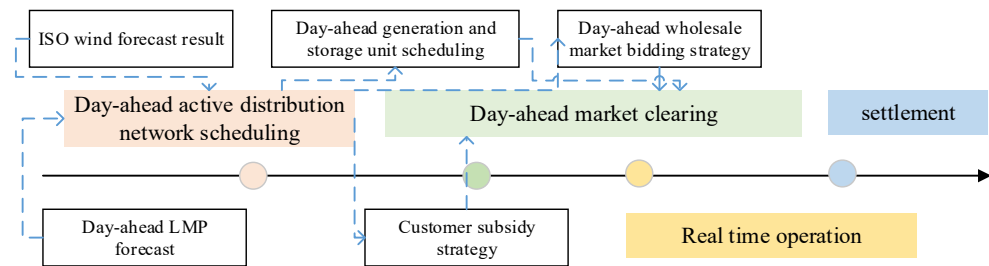


Figure 2. Basic operating sequence of ADNO.

The power generation system is undergoing a trend of low-carbonization, which means more and more large-scale VRE power generation companies are entering the wholesale market. As a result, the power spot price is increasingly affected by VRE output. In the actual market operation, the wholesale market operator Independent System Operator (ISO) will provide a day-ahead VRE output forecast based on the information of the entire system.

For ADNOs located at a certain node of the system, they face uncertainties from three parts: VRE output, spot market price, and user load. For ADNOs in a high VRE penetration power wholesale market, the first two sources may be correlated. Ignoring this correlation may lead to a decrease in accuracy of prediction.

This article suggests that ADNOs can use the network-wide VRE output published by ISO as a condition to predict the local VRE output and node electricity prices. This can significantly improve the accuracy of predictions.

Based on dataset [24], this study calculated the correlation between VRE generation and spot prices. The dataset is from ENTSO-e and consists of time-series data of wind power generation and spot market prices in the DK market for one year. This paper first standardized the data using the formula:

$$x_{t,nor} = \frac{x_t}{|X|_2} \tag{1}$$

where $|X|_2 = \sqrt{\sum_t (x_t)^2}$, x_t is the value at time t in the time series, and X refers to the entire time series. $x_{t,nor}$ is the standardized value of x_t . The normalized value of wind output and day-ahead LMP are depicted in Figure 3a. This paper examined the correlation between the two time series, by calculating the Pearson correlation coefficient at different lag steps. The calculation method is as follows:

$$\rho_{W,P,N} = \frac{cov(X_W^N, X_P)}{\sigma_W^N \sigma_P} \tag{2}$$

where $cov(X_W^N, X_P) = \sum_t (x_{t-N,W,nor} - \bar{x}_{t-N,W,nor})(x_p - \bar{x}_{t,P,nor})$. Here, X_W^N represents the time series of wind power generated N steps earlier, and X_P represents the time series of spot prices. $cov(X_W^N, X_P)$ is the covariance between X_W^N and X_P . $x_{t,nor}$ is the standardized variable of wind power generation time series, and $\bar{x}_{t,nor}$ is the mean of $x_{t,nor}$. σ is the standard deviation of X , calculated as $\sigma = \sqrt{\sum_t (x_{t,nor} - \bar{x}_{t,nor})^2}$.

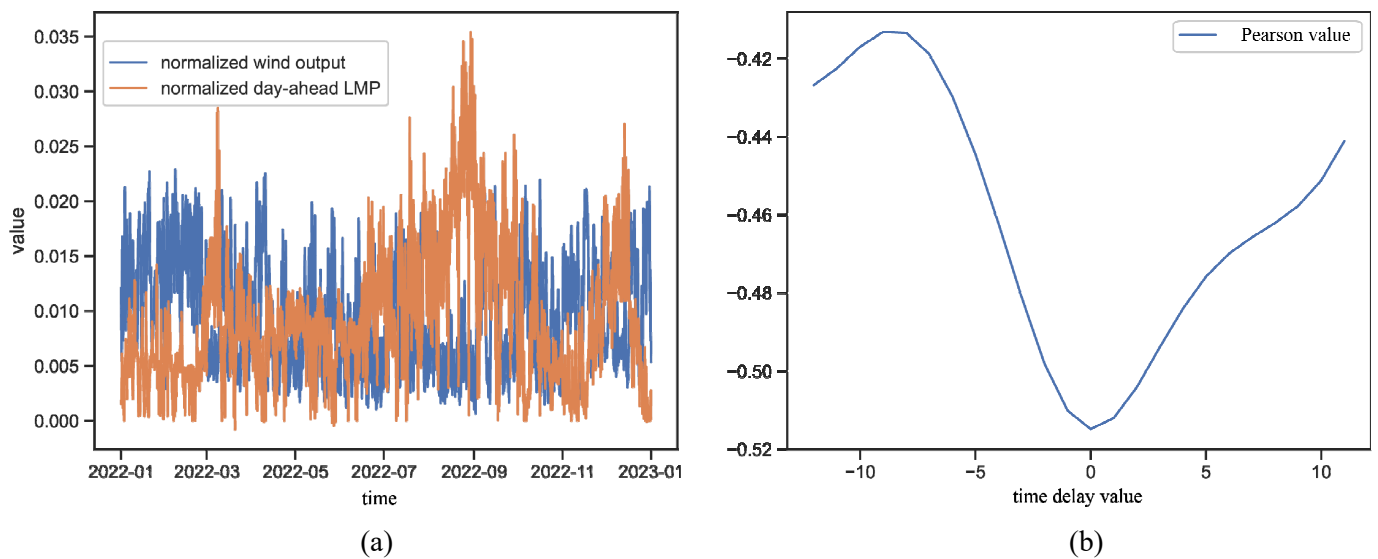


Figure 3. Pearson correlation coefficient between wind power generation and spot prices at different lag steps. (a) Shows the normalized value of wind output and day-ahead LMP; (b) shows the time-delay Pearson correlation coefficient between the two series.

The time interval of the data used in this paper is 15 min. When N takes values between -10 and 10 , the variation in $\rho_{W,P,N}$ is shown in Figure 3b. It can be seen that wind power generation and spot prices have significant negative correlation (less than -0.5) when $N = 0$.

3.2. ADNO Local Wind Power Output and Nodal Price Prediction Based on ISO Wind Power Forecast

3.2.1. Improved Transformer Time Series Conditional Prediction Model

This study constructs a prediction model based on the principle of Bayesian inference. As shown in Section 3.1, the local nodal electricity price of ADNO and the wind power output forecast of ISO are highly correlated under high penetration rates. Therefore, we built a conditional prediction model, which we apply to the prediction of ADNO local wind power output and nodal prices separately.

This paper constructs a prediction model based on the principle of maximizing the conditional likelihood probability, as shown in Formula (3).

$$p(x|y^N, \theta) = \prod_{t=t_0}^{T-1} p(x(t+1)|x(t_0), \dots, x(t), y(t_0 - N), \dots, y(t - N), \theta) \quad (3)$$

where N is the time delay step that maximizes the Pearson correlation between the two time series x and y . This paper constructs a parameter model to map this conditional probability, and the value of the parameter θ should be the value that maximizes $p(x|y, \theta)$, that is,

$$\theta^* = \underset{\theta}{\operatorname{argmax}} p(x|y^N, \theta) \quad (4)$$

Regarding the selection of the prediction parameter model, there are linear models and nonlinear models. Recent studies have shown that nonlinear models represented by deep neural networks have achieved good results in many complex conditional prediction tasks [25–27]. In deep neural networks, there are various options for such prediction tasks based on autoregressive networks (such as LSTM, GRU, etc.) and convolutional networks (such as WaveNet) [28]. Considering the specificity of this prediction task, this study adopts the transformer network for prediction. Multiple studies have shown that transformer

networks can model longer time spans and often achieve better results in price prediction tasks than autoregressive networks and convolutional networks.

This paper uses the transformer [29] deep neural network for conditional prediction. It improved the structure of the original transformer to make it more suitable for the prediction task of this study. The basic structure is shown in Figure 4.

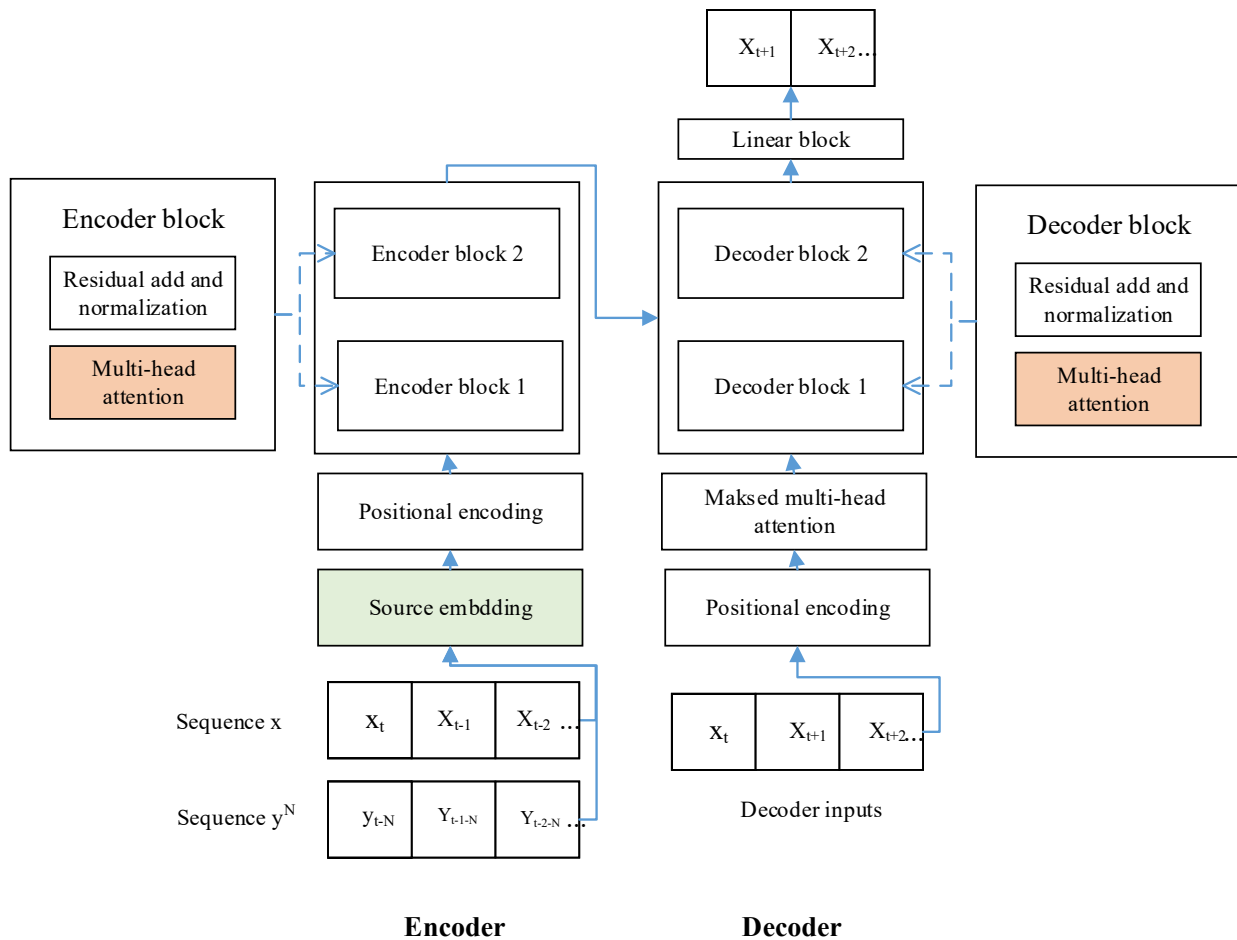


Figure 4. Transformer neural network structure used in this paper.

First, this study adjusts the input and output structure of the transformer. It first performs mean–variance standardization on all sequences to make them stationary and ensure that their values are in the range $[-1, 1]$. Then, it outputs the predicted mean and variance, rather than the probability of a specific category.

3.2.2. Multi-Head Attention Mechanism

The multi-head attention mechanism in the transformer is a mechanism that can simultaneously focus on different parts of the input sequence. The transformer uses the positional encoding module [29] to handle the temporal relationship of data. Its output matrix is used as the query matrix, key matrix, and value matrix to capture the dependencies in the time series. Let Q , K , and V be the query matrix, key matrix, and value matrix, respectively, with dimensions of $d_q \times n$, $d_k \times n$, and $d_v \times n$, where d_q , d_k , and d_v are the dimensions of the query vector, key vector, and value vector, and n is the length of the input sequence. The query vector for the i -th attention head is $q_i = QW_i^Q$, the key vector is $k_i = KW_i^K$, and the value vector is $v_i = VW_i^V$, where W_i^Q , W_i^K , and W_i^V are weight matrices. Then, the attention weight α_i for the i -th attention head is computed as follows: $\alpha_i = \text{softmax}\left(\frac{q_i^T k_j}{\sqrt{d_k}}\right)$ where j is the index of all key vectors, and $\sqrt{d_k}$ is the

normalization factor. Here, $\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$, $i = 1, \dots, K$ where z_i is the i -th element of the input, and K is the length of the input vector. Finally, the output of the multi-head attention mechanism is obtained by weighted summing all value vectors according to the attention weights: $\text{MultiHead}(Q, K, V) = \text{Concat}(\alpha_1 v_1, \alpha_2 v_2, \dots, \alpha_h v_h) W^O$ where h is the number of attention heads, and W^O is the weight matrix used to map the concatenated attention heads to the output dimension. Concat denotes the concatenation of vectors. The multi-head attention mechanism can focus on different parts of the input sequence simultaneously, which helps capture the dependencies among them. It can also handle variable-length input sequences because it does not require fixed-length inputs. Moreover, the multi-head attention mechanism can be computed in parallel, as each attention head can be computed independently, which improves computational efficiency. In time series prediction tasks, the multi-head attention mechanism in the Encoder can extract useful features from time series data. It can aggregate information from different time steps into a vector representation. The multi-head attention mechanism in the Decoder can also select the most relevant information from the previous time steps for prediction.

3.2.3. Source Embedding Mechanism

This study introduces the source embedding mechanism in the transformer. The source embedding mechanism is derived from the literature [30]. The literature [31] applied this mechanism to the power load clustering, and demonstrated that compared with traditional clustering methods, the source embedding mechanism can more effectively mine the statistical characteristics of uncertainty sources. This mechanism encodes any time series from a certain source into an embedding vector through a deep neural network. The embedding vectors of the same source are closer to their own centroid in high-dimensional space. These embedding vectors carry the statistical information of the source and can be used as input information by other neural networks for better prediction results.

In this study, the CBHG network [32] is used for encoding. For a training batch with K sources, each source has M training segments, the encoding calculation can be represented by Formula (5):

$$e_{k,m} = \frac{CBHG(x_{k,m})}{\|CBHG(x_{k,m})\| * 2} \quad (5)$$

where $e_{k,m}$ is the encoding for the m -th time series segment $x_{k,m}$ of the k -th source. $CBHG(x_{k,m})$ represents the output of the CBHG network after inputting $x_{k,m}$. $\|\cdot\|_2$ represents taking the 2-norm. In this paper, the GE2E loss is used to train this encoding network, so that for any input segment x_i of source i , an embedding vector e_i can be output, which is closer to the centroid of source i in cosine distance and away from the centroids of other sources. The principle of GE2E loss is to construct a similarity matrix S for each training batch, whose elements are $s * p, m$, where $p \in [1, \dots, K \times M]$. If $\lceil \frac{p}{M} \rceil = k$ ($\lceil \cdot \rceil$ represents rounding up), it means that these training data belong to source k and is the m -th training data. The calculation method of $s_{p,m}$ is shown in Formula (6):

$$s_{p,m} = \begin{cases} w \cdot \cos(e_{k,m}, c_k^{(-p)}) + b, \lceil \frac{p}{M} \rceil = k \\ w \cdot \cos(e_{k,m}, c_k) + b, \text{otherwise} \end{cases} \quad (6)$$

where c_k is the centroid of source k , whose calculation method is $c_k = \frac{1}{M} \sum_{m=1}^M e_{k,m}$. $c_k^{(-p)}$ is the centroid calculated when $\lceil \frac{p}{M} \rceil = k$, excluding the current training data p , $c_k^{(-p)} = \frac{1}{M-1} \sum_{m=1, m \neq (p \% M)}^M e_{k,m}$. $\cos(e_{k,m}, c_k)$ represents cosine distance, $\cos(e_{k,m}, c_k) = \frac{e_{k,m} \cdot c_k}{\|e_{k,m}\|_2 \cdot \|c_k\|_2}$, w and b are the weight and bias parameters that need to be trained. It can be seen that when $m = k$, $s_{p,m}$ calculates the distance between the time series segment and its corresponding source k centroid. When $m \neq k$, $s_{p,m}$ calculates the average distance between the time series segment and the centroids of all sources that do not belong to it. We hope that the former is

smaller and the latter is larger. Therefore, the calculation method of the GE2E loss is shown in Formula (7):

$$L(e_{p,m}) = s_{p',m} - \log\left(\sum s_{p'',k}\right) \tag{7}$$

where $\lceil \frac{p}{M} \rceil = \lceil \frac{p'}{M} \rceil = k$, $\lceil \frac{p''}{M} \rceil \neq k$. Then, the total loss value of a training batch is calculated by Formula (8):

$$LOSS = \sum L(e_{p,m}) \tag{8}$$

If a database covering enough statistical patterns is used to train this encoding network, an embedding vector e is an output for any time series outside the training set. This vector carries good statistical properties and can provide effective information for the prediction network. The principle of the source embedding mechanism is shown in Figure 5.

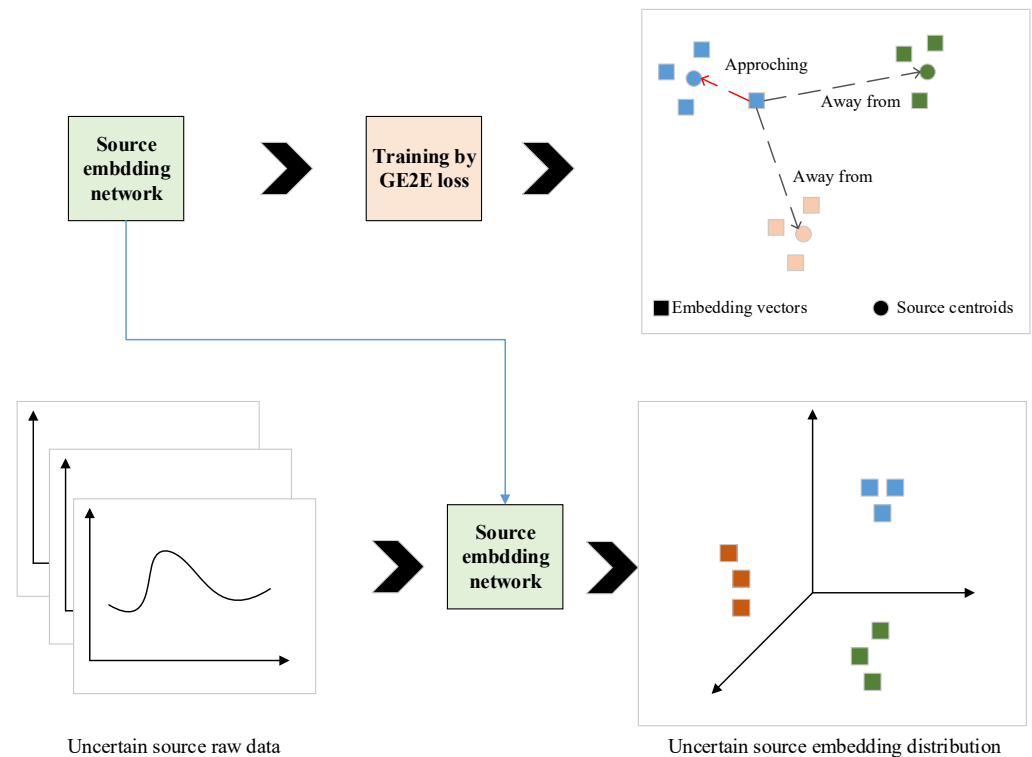


Figure 5. The principle of the source embedding mechanism.

3.3. Evaluation of Prediction Methods

This article uses the ENTSO-E database [24] to train the embedding network and transformer. The database is a data platform operated by the European Network of Transmission System Operators for Electricity (ENTSO-E), which aims to enhance the transparency and comparability of electricity market information in the European Union. The database includes spot price data for 20 markets and predicted wind power output data for ISO. This article collected local wind power data from 50 locations. The basic time interval of the data is 1 h. Based on these data, the node electricity price conditional prediction model $p_1(x_{LMP}|y_{ISO_{wind}}^N, \theta_1)$ and local wind power conditional prediction model $p_2(x_{local_{wind}}|y_{ISO_{wind}}^N, \theta_2)$ were trained. Models p_1 and p_2 use the same hyperparameters and optimizer settings, as shown in Table 1.

Table 1. Hyperparameters and optimizer settings for models p_1 and p_2 .

Parameter Name	Parameter Size
Forecasting network batch size	64
Transformer encoder block number	2
Transformer decoder block number	2
Transformer attention head number	8
Input time series length	72
Input feature	Short-time Fourier transform (STFT)
Output time series length	24
CBHG	Same as in [32]
Source embedding training batch size	64
optimizer	Adam [32]
Learning rate	0.001

This article randomly selected 10% of the data from the original dataset as the test set. Table 2 lists the evaluation results of the proposed method and traditional non-conditional probability-based LSTM [33], CNN [33], and sequence to sequence (seq2seq) [34] models on the test set. This article labeled node electricity price prediction as task1 and local wind power prediction as task2.

Table 2. Comparison of the performance of our prediction method and traditional methods on the test set.

Model	Task1		Task2	
	RMSE ¹	MAE ²	RMSE	MAE
LSTM [33]	1.003	0.836	0.909	0.840
CNN [33]	0.752	0.753	0.756	0.674
Seq2seq [34]	0.630	0.616	0.535	0.513
proposed	0.540	0.553	0.469	0.453

¹ Root mean squared error. ² Mean average error.

4. Bi-Level Optimization Model for Equilibrium Operation Strategy

4.1. Short-Term and Long-Term Operational Issues Faced by ADNO

Figure 1 shows the basic operating mode of ADNO, which is to buy electricity from the wholesale market and sell it to customers within the distribution network. We assume that the wholesale market is sufficiently competitive and that a single ADNO cannot affect the market price. Therefore, ADNO needs to make accurate predictions about the spot prices in the wholesale market. On the other hand, ADNO can influence customers' demand by providing subsidies, as customers have different levels of price elasticity for electricity demand. Customers have some rigid demand that is almost unaffected by monetary incentives (such as essential electricity consumption for work and lighting), and some flexible demand that is influenced by possible subsidy amounts (such as demand for entertainment and comfort). Therefore, ADNO needs to predict the rigid demand and influence the flexible demand through subsidies. Since ADNO's customers are limited and small-scale, they respond to ADNO's subsidies individually, and it is difficult to define the collective response of customer groups to ADNO's subsidy strategy using a specific statistical distribution method. In the short term, ADNO needs to optimize its internal power generation and energy storage scheduling based on the predicted spot market prices, local renewable energy output, and rigid loads. To find ADNO's short-term equilibrium operation strategy, this paper proposes a bi-level optimization model based on the Stackelberg game. This model is constructed for ADNO in the day-ahead scheduling using the predicted spot market prices, wind power generation output, and load. Besides predicting renewable energy output and spot market prices accurately and setting subsidies appropriately, ADNO also faces the challenge of determining the optimal capacity of energy storage, internal renewable energy generation units, and fossil fuel generation units when

investing in network assets. This is ADNO's long-term operational strategy problem. Therefore, based on the short-term optimal solution, this paper introduces an efficiency measurement indicator. When this indicator reaches the marginal optimal value, it can determine ADNO's optimal asset size in the long term.

4.2. Hedging Strategies of ADNO for Spot Wholesale Electricity Prices and Renewable Energy Output Fluctuations

4.2.1. Configuring Energy Storage Systems

Configuring energy storage systems is critical for ADNO. Energy storage systems allow ADNO to store electricity when spot wholesale prices are low or when there is excess renewable energy within the system, and release the stored electricity during opposite conditions to reduce costs. Assuming ADNO has S energy storage systems, and each system is numbered s , then $s \in [1, \dots, S]$. Similarly, assuming a day is divided into T time slots, and each time slot is numbered t , then $t \in [1, \dots, T]$. Let $P_{s,t}^{STRC}$ be the charging power of energy storage system s at time t , and $P_{s,t}^{STRD}$ be the discharging power of energy storage system s at time t . The following constraints apply, as shown in Equation (9):

$$\begin{cases} 0 \leq P_{s,t}^{STRC} \leq (1 - b_{s,t})U_s^{STRC} \\ 0 \leq P_{s,t}^{STRD} \leq b_{s,t}U_s^{STRD} \end{cases} \quad \forall t \in [1, \dots, T], s \in [1, \dots, S] \quad (9)$$

where U_s^{STRC} is the maximum charging power of s , and U_s^{STRD} is the maximum discharging power of s . $b_{s,t}$ is a binary variable that indicates whether s is charging (0) or discharging (1) at time t .

Furthermore, assuming the energy storage system s has a maximum capacity of U_s^{ESTR} , then the following constraints apply, as shown in Equation (10):

$$\begin{cases} e_{s,t}^{STR} = e_{s,t-1}^{STR} + (P_{s,t}^{STRC} - P_{s,t}^{STRD})\Delta t \\ 0 \leq e_{s,t}^{STR} \leq U_s^{ESTR} \end{cases} \quad \forall t \in [1, \dots, T], s \in [1, \dots, S] \quad (10)$$

where $e_{s,t}^{STR}$ represents the total stored electricity of s at time t , and Δt represents the time interval.

4.2.2. Configuring Distributed Fossil Energy

Suppose ADNO has I fossil energy installed internally, and the label of each fossil energy unit is i , then $i \in [1, \dots, I]$. A typical distributed fossil energy installation is a small gas turbine. The goal of installing internal power generation is to supplement the insufficient external power or to replace the expensive external power. Because these installations are relatively small, this article considers a simple linear cost model for fossil energy installation i as shown in Formula (11).

$$Cost_i = c_i \cdot P_{i,t}^{GEN} + w_{i,t} \cdot O_i, \forall t = 2, \dots, T \quad (11)$$

There are constraints on fossil energy unit i as shown in Formula (12):

$$L_i^{GEN} z_{i,t} \leq P_{i,t}^{GEN} \leq U_i^{GEN} z_{i,t} \quad (12)$$

where $Cost_i$ represents the total production cost of i , c_i is the marginal cost of i , $P_{i,t}^{GEN}$ is the active power output of i at time t , and O_i is the start-up cost of i . $z_{i,t}$ is a 0–1 variable. When it is 0, it means that i is in the shutdown state at time t ; when it is 1, it means that i is in the running state at time t . L_i^{GEN} is the minimum output of unit i , and U_i^{GEN} is the maximum output of unit i . There is:

$$w_{i,t} = \begin{cases} 1, z_{i,t} - z_{i,t-1} = 1, \forall t = 2, \dots, T \\ 0, otherwise \end{cases} \quad (13)$$

The variable $w_{i,t}$ is also a 0–1 variable. When $w_{i,t} = 1$, i is started at time t , and its start-up cost needs to be calculated. Since i is a small unit, this article ignores its shutdown cost. There are multiple operating constraints for fossil energy installation i . In the active distribution network, the advantage of distributed fossil energy is that it can adjust its output quickly to cope with the fluctuations of renewable energy output and the spot price fluctuations of main grid electricity. Therefore, this article considers the ramp constraint as shown in Formula (14).

$$L_i^{RAM} \leq P_{i,t}^{GEN} - P_{i,t-1}^{GEN} \leq U_i^{RAM}, \forall t = 2, \dots, T \quad (14)$$

where L_i^{RAM} is the minimum climbing amount of unit i between units, and U_i^{RAM} is the maximum climbing amount of unit i between units.

4.2.3. Configuring Distributed Wind Power Generation

ADNO has V distributed wind power generation units installed internally, each wind power generation unit is numbered v , and v is within $[1, \dots, V]$. The advantage of ADNO's distributed wind power generation installation is that wind power generation has almost zero marginal cost and can replace the electricity provided by the main grid at very low cost. However, distributed wind power generation also has strong randomness and uncontrollability. This article predicts the local wind power generation output through the method described in Section 3.

4.3. Customer Electricity Demand and Network Constraints

4.3.1. Customer Electricity Demand

This article assumes that ADNO's customers have two types of loads: rigid loads (such as essential electricity consumption for work and lighting) and flexible loads (such as loads that can be shifted in time). Rigid loads are random and independent of other factors. Flexible loads depend on prices. Assuming ADNO has J customers, the electricity demand of customer j at time t is shown in Formula (15):

$$Demand_{j,t} = \overline{d}_{j,t} + \tilde{d}_{j,t}, \forall j \in [1, \dots, J], t \in [1, \dots, T] \quad (15)$$

where $Demand_{j,t}$ is the total active power demand of customer j at time t , $\tilde{d}_{j,t}$ is the rigid demand of customer j at time t , and $\overline{d}_{j,t}$ is the flexible demand of customer j at time t . As the rigid demand of customers is not significantly correlated with the wind power output forecast of ISO, this article uses the method in reference [35] to predict their rigid demand.

This article assumes that the retail market is fully accurate and that ADNO cannot affect the retail electricity price $p_{j,t}$. For flexible demand, ADNO can influence it through certain subsidy strategies, as follows:

$$\overline{d}_{j,t} \in D_{j,t}(sub_{j,t}), \forall j \in [1, \dots, J], t \in [1, \dots, T] \quad (16)$$

where $D_{j,t}$ is a set of demand responses to subsidies, and $sub_{j,t}$ is ADNO's subsidy strategy at time t . It can be seen that ADNO can influence the flexible demand of customers by subsidizing them when the spot price is low or when the wind is strong.

This article assumes that for all $j \in [1, \dots, J], t \in [1, \dots, T]$, $D_{j,t}$ is a linear interval of $d_{j,t}$, i.e.:

$$\overline{L}_{j,t} \leq \overline{d}_{j,t} \leq \overline{U}_{j,t}, \forall j \in [1, \dots, J], t \in [1, \dots, T] \quad (17)$$

where $\overline{L}_{j,t} = 0$. $\overline{U}_{j,t}$ is influenced by p_t , as shown in Formula (18):

$$\overline{U}_{j,t} = \overline{U}_{j,t}^{max} + (sub_{j,t} - sub_{max})Ela_{j,t}, \forall j \in [1, \dots, J], t \in [1, \dots, T] \quad (18)$$

where $\overline{U}_{j,t}^{max}$ is the maximum flexible demand of j at time t ; $sub_{j,t}$ is the electricity subsidy that j can receive at time t ; sub_{max} is the maximum subsidy value. $Ela_{j,t}$ is the elasticity coefficient of the power demand subsidy for j at time t .

4.3.2. Network Constraints

The voltage level of the distribution network is relatively low, and the impact of transmission loss on ADNO operation needs to be considered. This paper constructs a network model constraint based on the standard distribution network branch model [36]. The power and voltage of nodes have linear constraints as shown in Formula (19):

$$\begin{cases} P_n = (P_{mn} - |I_{mn}|^2 \cdot r_{mn}) - \sum_{k \in n} P_{nk} \\ Q_n = (Q_{mn} - |I_{mn}|^2 \cdot x_{mn}) - \sum_{k \in n} Q_{nk} \\ |U_n|^2 = |U_m|^2 - 2(r_{mn} \cdot P_{mn} + x_{mn} \cdot Q_{mn}) + (r_{mn}^2 + x_{mn}^2) \cdot |I_{mn}|^2 \end{cases} \quad (19)$$

where P_n is the active power injected into node n , P_{mn} is the active power flowing from node m to node n , Q_n is the reactive power injected into node n , and Q_{mn} is the reactive power flowing from node m to node n . $|I_{mn}|$ is the amplitude of the current flowing from node m to node n . r_{mn} is the resistance of the line mn , and x_{mn} is the reactance of the line mn . Expression $k \in n$ means the set of nodes connected to node n on the network. $|U_n|$ is the voltage amplitude of node n . There are safety constraints for on-line voltage and circuit current:

$$\begin{cases} |I_{mn}| \leq I_{mn}^{max} \\ |U_n|^{min} \leq |U_n| \leq |U_n|^{max} \end{cases} \quad (20)$$

At the same time, there is also a non-convex power balance constraint as shown in Formula (21):

$$|I_{mn}|^2 \cdot u_m = P_{mn}^2 + Q_{mn}^2 \quad (21)$$

In this regard, this paper intends to perform second-order cone relaxation on this constraint to make it a convex constraint. As shown in Formula (22):

$$\left\| \left[2P_{mn} \ 2Q_{mn} \ |I_{mn}|^2 - |U_{mn}|^2 \right]^T \right\|_2 \leq |I_{mn}|^2 + |U_{mn}|^2 \quad (22)$$

The literature [37] shows that for distribution networks that meet certain conditions (such as IEEE33), after the above relaxation, the optimal solution of the new problem is also the optimal solution of the original problem. After second-order cone relaxation, the difficulty of solving the calculation is greatly reduced.

4.4. Stackelberg Game and Bi-Level Optimization Model

As mentioned in Sections 4.1 and 4.3.1, there are two major assumptions.

Assumption 1. *The competition in the wholesale market is enough to be sufficient, and a single ADNO cannot affect the market price.*

Assumption 2. *The retail market is fully accurate and ADNO cannot affect the retail electricity price.*

Under these assumptions, a Stackelberg Game based bi-level optimization model can be established.

As mentioned in Section 4.1, the game between ADNO and customers can be modeled using the Stackelberg game. Based on the contents of Sections 4.2 and 4.3, this text constructs a bi-level optimization model. In this article, we consider the following decision variables for ADNO in the current operating plan: $P_{i,t}^{GEN}$, P_t^{REA} , $P_{s,t}^{STRC}$, $P_{s,t}^{STRD}$, $w_{i,t}$, and $b_{s,t}$. Here, $q_{i,t}^{GEN}$ represents the active power dispatch of the i -th distributed fossil unit by ADNO at time t ; P_t^{REA} represents the power purchase amount from the wholesale market by ADNO

at time t ; $P_{s,t}^{STRC}$ represents the charging dispatch of the s -th distributed fossil unit by ADNO at time t ; $P_{s,t}^{STRD}$ represents the discharging dispatch of the s -th distributed fossil unit by ADNO at time t ; $w_{i,t}$ represents whether ADNO starts the i -th distributed fossil unit at time t , where 0 means not started and 1 means started; and $b_{s,t}$ represents whether the s -th energy storage device started at time t is in charging state (0) or discharging state (1).

For customers, they need to make demand decisions at a certain price level after ADNO makes operating decisions. According to Section 4.3.1, the demand variable for customers is $d_{j,t}$, which represents the elastic electricity demand.

ADNO and customers have different decision objectives. ADNO aims to maximize revenue, while customers hope to minimize electricity costs. Therefore, a bi-level game model can be constructed for this Stackelberg game as shown in Formula (23). In Formula (23), the objective function of the upper-level problem is the maximization of the revenue of the ADNO. $\sum_{j=1}^J \sum_{t=1}^T p_t \cdot (\bar{d}_{j,t} + \tilde{d}_{j,t})$ is the total electricity sale income. $\sum_{t=1}^T (\hat{p}_t \cdot P_t^{REA} + \sum_i^I (c_i \cdot P_{i,t}^{GEN} + O_i \cdot w_{i,t}) + \sum_{j=1}^J sub_{j,t})$ is the total cost, which contains the electricity import cost $\hat{p}_t \cdot P_t^{REA}$, fossil generation cost (consider start-up and shut-down cost) $\sum_i^I (c_i \cdot P_{i,t}^{GEN} + O_i \cdot w_{i,t})$, and demand subsidy cost $\sum_{j=1}^J sub_{j,t}$. The objective function of the upper-level problem is the minimization of the total cost of customers. As for the constraints, constraint (a) corresponds to the conditions stated in Section 4.2.1. Section 4.2.1 describes the constraints on operating variables $P_{s,t}^{STRC}$ and $P_{s,t}^{STRD}$, and the variable set determined by these constraints is defined as Ω_{STR} . Therefore, $P_{s,t}^{STRC}$, $P_{s,t}^{STRD}$, and $b_{s,t}$ are elements of Ω_{STR} . Constraint (b) corresponds to the conditions stated in Section 4.2.2. Section 4.2.2 describes the constraints on the operating variable $P_{i,t}^{GEN}$, and the variable set determined by these constraints is defined as Ω_{GEN} . Therefore, $P_{i,t}^{GEN}$ and $w_{i,t}$ are elements of Ω_{GEN} . Constraint (b) corresponds to the conditions stated in Section 4.3.2. Section 4.3.2 describes the network constraints, and the variable set determined by these constraints is defined as Ω_{NET} . Therefore, $P_{i,t}^{GEN}$, P_t^{REA} , $P_{s,t}^{STRC}$, $P_{s,t}^{STRD}$, $w_{i,t}$, and $b_{s,t}$ are elements of Ω_{NET} . Constraint (d) is the value positive constraint. Constraint (e) is the power balancing constraint. Constraint (e) corresponds to the subsidy constraint described by Formula (16). A bi-level game model can be constructed for this Stackelberg game as shown in Formula (23), and its detailed mathematical description is shown in Appendix A.

$$\begin{aligned}
 & \max_{\substack{P_{i,t}^{GEN}, P_t^{REA}, P_{s,t}^{STRC}, P_{s,t}^{STRD}, w_{i,t}, b_{s,t}, sub_{j,t} \\ j=1, t=1}} \sum_{j=1}^J \sum_{t=1}^T p_t \cdot (\bar{d}_{j,t} + \tilde{d}_{j,t}) - \sum_{t=1}^T \left(\hat{p}_t \cdot P_t^{REA} + \sum_i^I (c_i \cdot P_{i,t}^{GEN} + O_i \cdot w_{i,t}) + \sum_{j=1}^J sub_{j,t} \right) \\
 & \text{s.t.} \\
 & \begin{cases} P_{s,t}^{STRC}, P_{s,t}^{STRD}, b_{s,t} \in \Omega_{STR} \\ P_{i,t}^{GEN}, w_{i,t} \in \Omega_{GEN} \\ P_{i,t}^{GEN}, P_t^{REA}, P_{s,t}^{STRC}, P_{s,t}^{STRD}, w_{i,t}, b_{s,t} \in \Omega_{NET} \\ P_t^{REA} \geq 0 \end{cases} \\
 & \min_{\bar{d}_{j,t}} \sum_{j=1}^J \sum_{t=1}^T p_t (\bar{d}_{j,t} + \tilde{d}_{j,t}) \\
 & \text{s.t.} \\
 & \begin{cases} \sum_j^J d_{j,t} + \sum_j^J \tilde{d}_{j,t} + \sum_s^S P_t^{STRC} = P_t^{REA} + \sum_i^I P_{i,t}^{GEN} + \sum_s^S P_{s,t}^{STRD} + \sum_v^V P_{v,t}^{GEN} \\ \bar{d}_{j,t} \in D_{j,t}(sub_{j,t}) \end{cases}
 \end{aligned} \tag{23}$$

where \hat{p}_t represents the spot price at time t ; $P_{v,t}^{GEN}$ represents the output of the v -th wind power unit at time t ; $d_{j,t}$ represents the stochastic demand of the j -th customer at time t . They are predicted using the method described in Section 3 of this article.

4.5. Short-Term Equilibrium Solution and Long-Term Optimal Investment Analysis Method

It can be seen that the bi-level optimization model represented by Formula (23) has an upper layer model with a convex optimization model and a lower layer model with a linear optimization model. For such a model, the KKT conditions of the lower layer optimization can be added to the constraints of the upper layer optimization model, and the original bi-level optimization model can be transformed into a non-convex model with complementary relaxation equality constraints. The KKT transformed problem formulation is shown in Appendix A. This paper uses MOSEK (version 10) to find the global optimal solution to this problem.

The solution obtained based on this method is the short-term equilibrium solution for the active distribution network operation. In the long run, ADNO may achieve Pareto improvement through investment in energy storage, power generation, and other assets. In order to measure the degree of ADNO's Pareto improvement, this paper proposes an indicator as shown in Formula (24).

$$Eff^{ADNO} = \frac{E(Income)}{E(C^{REA} + C^{GEN} + C^{PUN})} \quad (24)$$

where $E(\cdot)$ represents the expectation. Income represents ADNO's daily income; C^{REA} represents ADNO's daily purchase cost from the wholesale market; C^{GEN} represents ADNO's daily generation cost; C^{PUN} represents ADNO's contract penalty cost caused by forecasting and scheduling deviations. Therefore, Eff^{ADNO} can be expressed as shown in Formula (25):

$$Eff^{ADNO} = \frac{E\left(\sum_{j=1}^J \sum_{t=1}^T p_t \cdot \left(\bar{d}_{j,t}^* + \tilde{d}_{j,t}^{Rlz}\right)\right) * K}{E\left(\sum_{t=1}^T \left(\widehat{p}_t^{Rlz} \cdot P_t^{*,REA} + \sum_i^I \left(c_i \cdot P_{i,t}^{*,GEN} + O_i \cdot w_{i,t}^*\right) + \sum_{j=1}^J sub_{j,t}^*\right) + C^{pun}\right)_K} \quad (25)$$

where the asterisk represents the equilibrium scheduling strategy in a period (day); Rlz represents the realization value of uncertain quantities in a period (day); the subscript K represents the total number of experiments conducted. C^{pun} is the penalty value for ADNO based on the actual value of the unknown quantity. ADNO may be penalized in the following cases: ① The spot price estimate is too low, and local electricity could have been used instead; or the spot price estimate is too high, and too much more expensive local power generation is used. ② The output of local wind turbines is overestimated, and the insufficient part needs to be supplemented with more expensive power generation; the output of local wind turbines is underestimated, and more expensive local fossil energy or electricity spot is used. ③ The load forecast is underestimated, and more expensive electricity spot or local fossil energy needs to be supplemented; the load forecast is overestimated, and expensive electricity spot is purchased, which could have been replaced by local wind power generation.

To simplify, this paper calculates C^{pun} as $C^{pun} = \sum_c C_c^{reason}$, and the relationship is shown in Formula (26):

$$C_c^{reason} = p_c^{reason} \cdot q_c^{reason} \quad (26)$$

where C_c^{reason} represents one of the above three penalties, p_c^{reason} is the corresponding unit penalty cost, and q_c^{reason} is the corresponding deviation.

5. Case Study

5.1. IEEE 33 Test System and Main Experimental Parameters

This article tests the effectiveness of the proposed method on a modified IEEE 33 active distribution network. It is a modified version of the IEEE 33 bus radial distribution network [38], which is widely used in distribution system analysis [39–41]. Its baseline configuration is shown in Figure 6, with energy storage systems installed at nodes 1 and 5.

Wind turbines are installed at nodes 4, 20, and 10, while small gas generators are installed at nodes 9, 23, and 26.

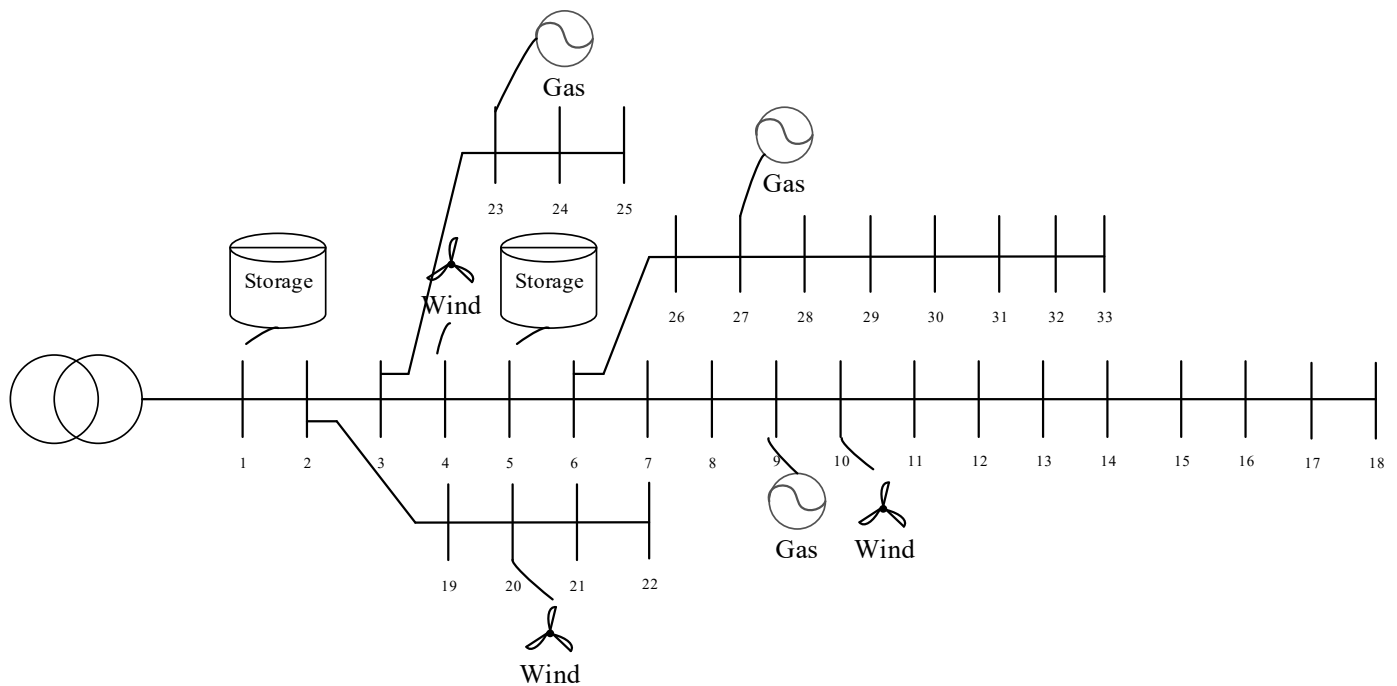


Figure 6. IEEE 33 test system used in this article.

As described in Section 3, this paper trains the conditional prediction model and source embedding model based on 90% of the data in the dataset [24], and randomly extracts 10% of the data as test data for electricity wholesale prices and wind power output. We use the node load data in the dataset as the total real-time load sequence value of the IEEE 33 active distribution network, and obtain the baseline load capacity ratio of IEEE 33 nodes from literature [42]. We distribute the real-time load data according to the capacity ratio, and obtain the real-time load sequence of each load node. The real-time output data of wind power generation is also distributed in a similar way. The main hyperparameter models of the deep neural network model used in this paper are shown in Table 3.

Table 3. The main hyperparameters of the deep neural network model used in this paper.

Parameter	Value
Encoder input dimension	72
Decoder input dimension	24
Decoder output dimension	24
Multi-head attention head number	4
Number of encoder blocks	2
Source embedding dimension	128
Encoder output dimension	128
Number of Decoder blocks	2
Number of training batch	48
Active function	Tanh
Number of training epoch	200
Training optimizer	Adam
Initial learning rate	5×10^{-5}

The baseline data used in this article is shown in Table 4.

Table 4. The baseline data used in this article.

Parameter	Value
Network type	IEEE33
Total wind turbine capacity	0.5 MW
Total energy storage capacity	0.15 MWh
Total gas turbine capacity	0.6 MW
Average variable cost of distributed gas power generation	91 USD/MWh
Average start-up cost of distributed gas power generation	USD 20
Average demand Subsidy	USD 10
Average spot price	USD 81/MWh

5.2. Ablation Experiment of The Forecast Model

According to Section 3.1., in the case of high wind power penetration, there is a high correlation between wholesale electricity price, wind power output, and ISO predicted regional wind power output level. Therefore, in Section 3, this paper proposes a prediction method based on source embedding and conditional prediction. Based on the test data, the role of source embedding and conditional prediction proposed in this paper can be tested. For a typical day, this paper compares the role of source embedding and conditional forecasting in predicting short-term spot prices and local wind power output. The results are shown in Figure 7.

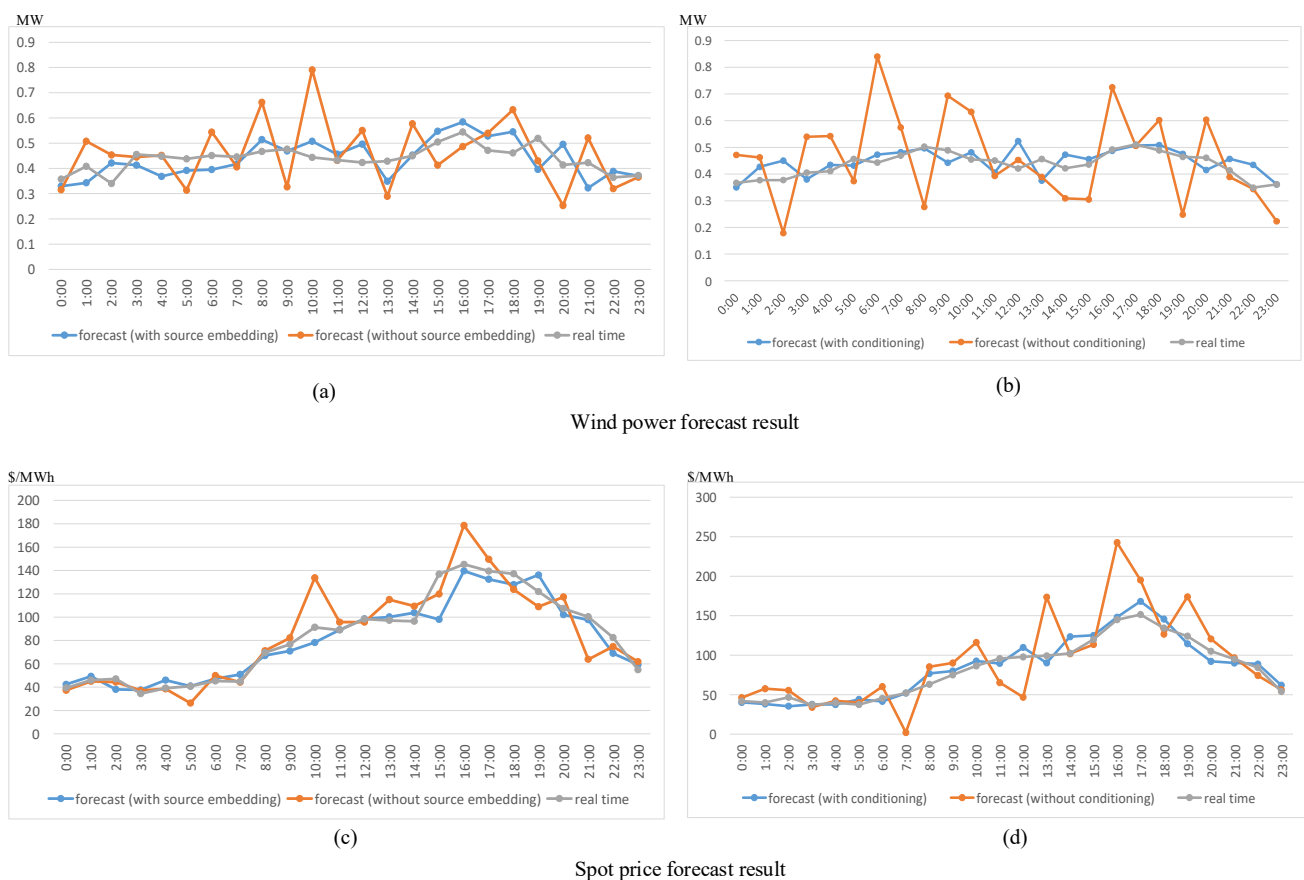


Figure 7. Forecast performance comparison curve for a typical day. (a) is the result comparison of wind power forecast with and without source embedding; (b) is the result comparison of wind power forecast with and without conditioning; (c) is the result comparison of spot price forecast with and without source embedding; (d) is the result comparison of spot price forecast with and without conditioning.

From Figure 7a,b, it can be seen that for local wind power prediction, source embedding can provide a certain improvement in prediction accuracy when conditional predictions are used. Moreover, this improvement is larger than the accuracy improvement from conditional prediction with source embedding. This may be because source embedding can effectively capture the statistical characteristics of wind power output in different geographical distributions, and the information it provides exceeds the role of conditional prediction.

From Figure 7c,d, we can see that for the spot price forecast, source embedding can provide a certain improvement in forecast accuracy when there are conditional forecasts. However, this improvement is smaller than the accuracy improvement from conditional prediction with source embedding. This may be because the spot price is greatly affected by random factors, and the statistical coding of big data can provide less effective information. However, under high wind power penetration, the system wind power has a significant impact on the spot price, which is easier to be captured by the conditional probability prediction model.

As a result, we can see that the source embedding method brings greater accuracy in wind power forecasting and the conditional method brings greater accuracy in price forecasting. When they are combined by our method, we obtain better results for both tasks.

This paper compares the control variables of these two factors in the test set and shows the results in Table 5. We denote the node electricity price prediction as task1 and the local wind power prediction as task2.

Table 5. Comparison of control variables for predictive effects.

Comparison Factor	Test Task	RMSE ¹
Without source embedding	task1	0.621
Without conditioning	task1	0.594
Without source embedding	task2	0.525
Without conditioning	task2	0.553

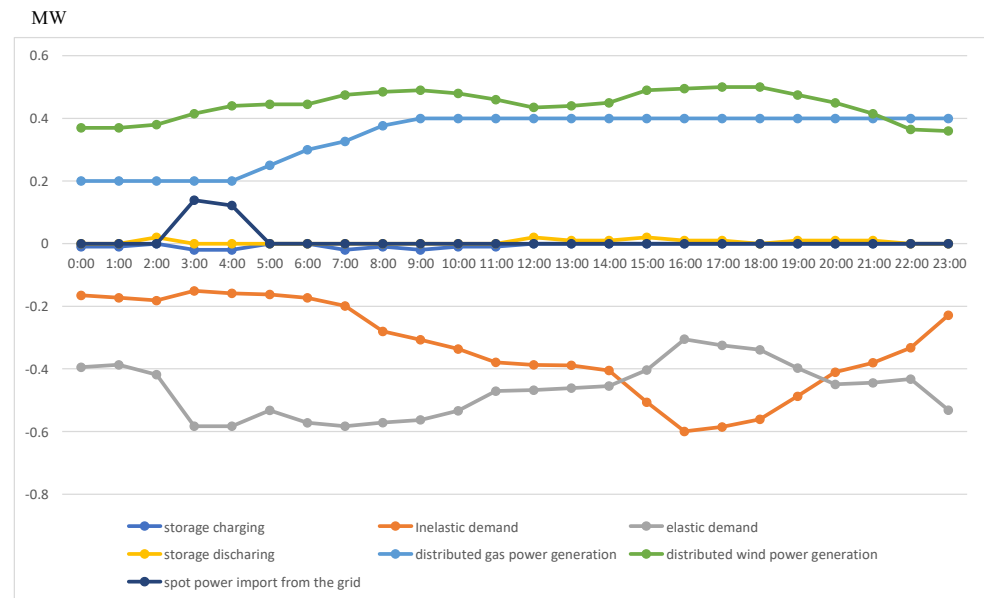
¹ Root mean squared error.

5.3. Equilibrium Scheduling and Short-Term Optimal Strategy

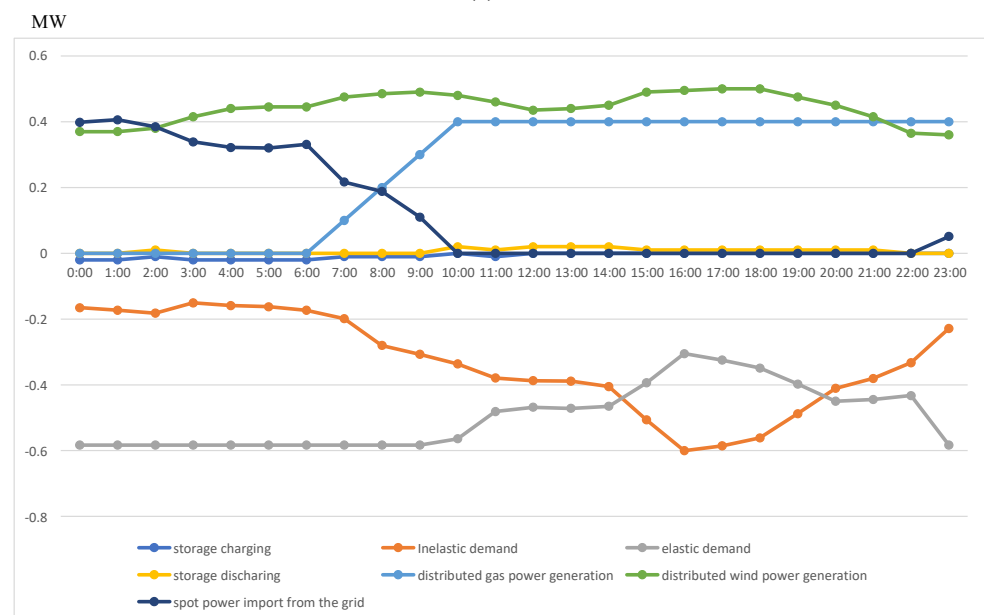
Figure 8 shows the effectiveness of the short-term equilibrium scheduling method proposed in this study. The baseline system in this paper has high wind power penetration. During the time period from 0:00 to 9:00 when the spot price is relatively low, through the method proposed in this paper, ADNO can motivate customers to increase elastic demand. At this time, a large number of elastic loads are excited to make use of cheap wind energy. Since the spot price is higher than the marginal cost of internal distributed gas-fired power generation, ADNO relies on local wind power and gas-fired power generation to meet the load demand during the peak period of electricity consumption from 9:00 to 18:00. From 0:00 to 3:00 when the spot price is the cheapest, ADNO purchases a part of the electricity spot and stores it based on energy storage equipment, and releases it during the peak period of electricity consumption from 9:00 to 18:00 to meet the load demand, as shown in Figure 8a.

If the spot price changes, ADNO's equilibrium scheduling result changes significantly. The dispatch result shown in Figure 8b is the case where the average spot price is reduced to 50% of the original. At this time, the spot price from 0:00 to 9:00 is cheap enough. During this time period, ADNO reduced local gas-fired power generation, fully imported a large amount of spot goods, and increased incentives for customers to flexibly load. It can be seen that based on the method proposed in this paper, ADNO can effectively respond to the spot price of the main network. If the main grid wind power output is too large, resulting in a drop in spot prices, ADNO can increase the consumption of main grid wind power by dispatching energy storage and stimulating loads. If the main grid power generation capacity is insufficient and the spot price rises, ADNO can dispatch local wind power and gas generators to meet the power demand, or reduce the elastic demand

of customers through incentives. In this process, clean energy has always been provided priority. The potential for this level of response is far greater than that of a single load, or of an unorganized distribution network.



(a)



(b)

Figure 8. Scheduling results on a typical day. (a) Is the dispatching result under 100% benchmark spot price; (b) is the dispatching result under 50% benchmark spot price.

From a short-term subsidy strategy point of view, ADNO also has an optimal subsidy level. It can be seen that there is an optimal value of ADNO's subsidy incentives for customers, as shown in Figure 9. The vertical axis is the ADNO efficiency index shown in Formula (24), and the horizontal axis is the level of subsidies relative to the benchmark value. Below this value, efficiency of ADNO can be improved through subsidies. However, beyond this value, increasing subsidies cannot further improve the efficiency of ADNO.

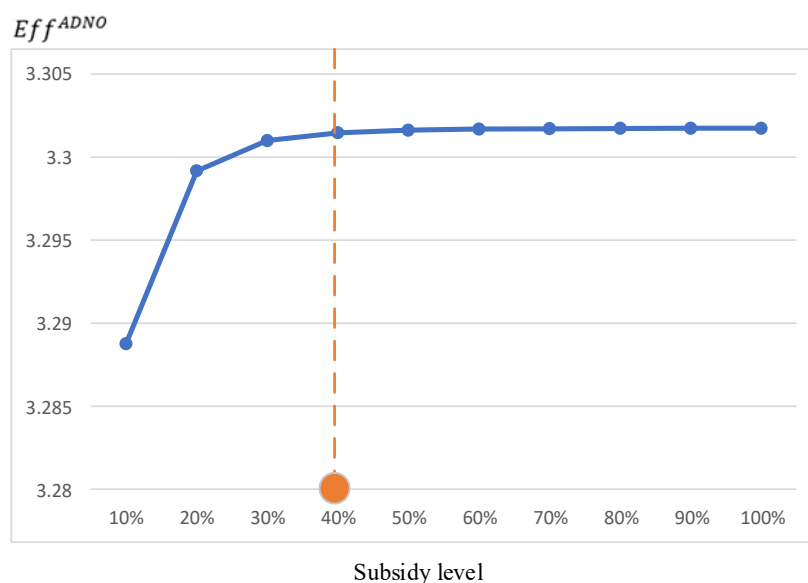


Figure 9. The influence of subsidy level to Eff^{ADNO} .

By controlling the variables, the effectiveness of the prediction method proposed in the third section and the game equilibrium scheduling method proposed in the fourth section can be demonstrated, and the results are shown in Table 6. It can be seen from the table that since the prediction method proposed in this paper can achieve lower average prediction errors, the penalty value can be reduced, thereby improving the Eff^{ADNO} index value. The two-tier optimization model proposed in this paper considers the adjustment effect of ADNO subsidies on user demand. The results show that in the equilibrium outcome, there is room for the benefit brought by the subsidy to exceed the cost of the subsidy. Compared with the single-level optimization model, the method proposed in this paper can further improve the operating efficiency of ADNO.

Table 6. Comparison study of the proposed methods.

Model Comparison *	Eff^{ADNO} Value
With LSTM [33] forecast model	2.209
With CNN [33] forecast model	2.304
With Seq2seq [34] forecast model	2.258
With proposed forecast model	2.496
With single-level (no-subsidy) optimization [43]	2.433
With proposed bi-level optimization	2.496

* All settings are consistent with the baseline other than the comparison factor.

5.4. Sensitivity to Long-Term Factors

Firstly, this article examines the impact of long-term average spot electricity price levels on the profitability indicator Eff^{ADNO} of ADNO. When the long-term average spot electricity price levels fluctuate between 10% and 100% of the benchmark values shown in Table 4, the change trend of Eff^{ADNO} is shown in Figure 10. It can be seen from the graph that the spot price has about a 50% fluctuation range on the profitability of ADNO, indicating that spot prices have a significant impact on ADNO's return on investment. It is worth noting that the impact of spot price levels on Eff^{ADNO} is marginally decreasing. In this example, when the spot price exceeds 40% of the benchmark value, it has almost no impact on Eff^{ADNO} . This is because within this range, ADNO can adjust its demand, comprehensively dispatch gas and wind power generators, and reduce power imports from the main grid to maintain the profitability at a certain level.

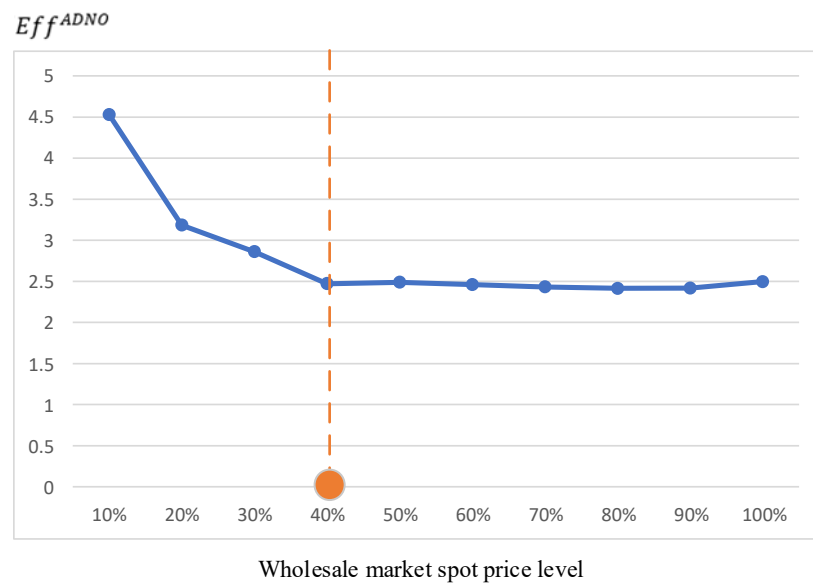


Figure 10. Sensitivity of ADNO’s profitability indicator Eff^{ADNO} to wholesale market spot price levels.

This reflects that the method proposed in this article can provide a “safeguard boundary” for ADNO against the fluctuation of spot prices in the main grid. When the fluctuation of spot prices exceeds this “safeguard boundary”, the profitability of ADNO can be guaranteed. This shows the robustness of our method to price variation.

Secondly, this article examines the impact of the elasticity coefficient $Ela_{j,t}$ of subsidies on the yield indicator Eff^{ADNO} of ADNO. When the mean elasticity coefficient of long-term average subsidies fluctuates between 10% and 100% of the benchmark value shown in Table 4, the trend of Eff^{ADNO} change is shown in Figure 11. It can be seen from the figure that the impact of the elasticity coefficient of subsidies on the yield indicator is about 20%. For ADNO, the smaller the elasticity coefficient of demand for subsidies, the higher the yield rate of ADNO. Compared with the spot price, there is also a turning point in the impact effect of the elasticity coefficient, which is about 80% of the benchmark value. After the elasticity coefficient exceeds 80% of the benchmark value, it almost has no impact on Eff^{ADNO} . This is because after the elasticity coefficient exceeds this turning point, due to the high cost of subsidies, ADNO almost no longer stimulates users’ elastic demand.

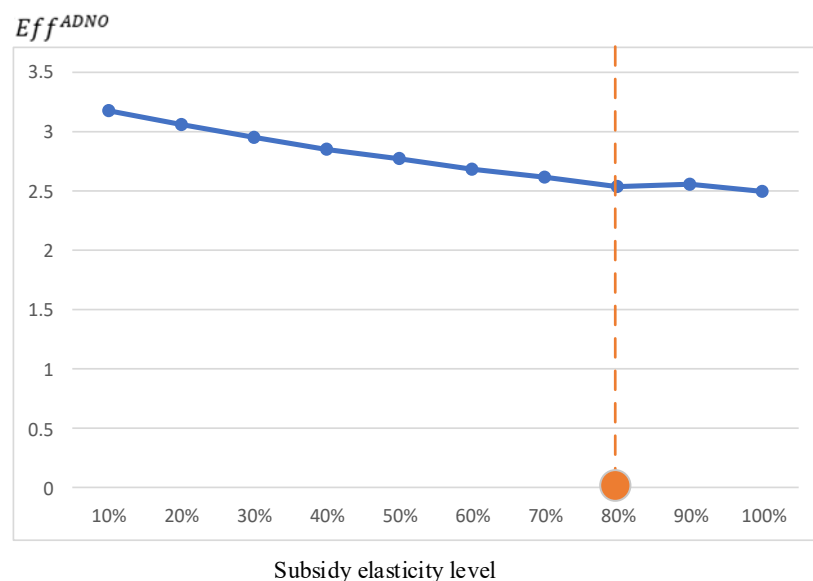


Figure 11. Sensitivity of ADNO’s profitability indicator Eff^{ADNO} to the elasticity coefficient of subsidies.

This means that as long as the elasticity coefficient of demand for subsidies is within a reasonable range, ADNO can improve its own revenue by subsidizing users. This shows the robustness of our method to the subsidy elasticity level.

This article also examines the impact of wind power installed capacity on the profitability indicator Eff^{ADNO} . When the wind power installed capacity fluctuates between 10% and 100% of the benchmark value shown in Table 4, the change trend of Eff^{ADNO} is shown in Figure 12. If the penalty cost is not taken into account, the increase in wind power installed capacity can always bring positive returns to ADNO. However, if the uncertainty caused by the increase in wind power installed capacity leads to prediction errors and economic penalties, there is also a turning point in the impact of wind power installed capacity on the return on investment of ADNO, as shown in Figure 12.

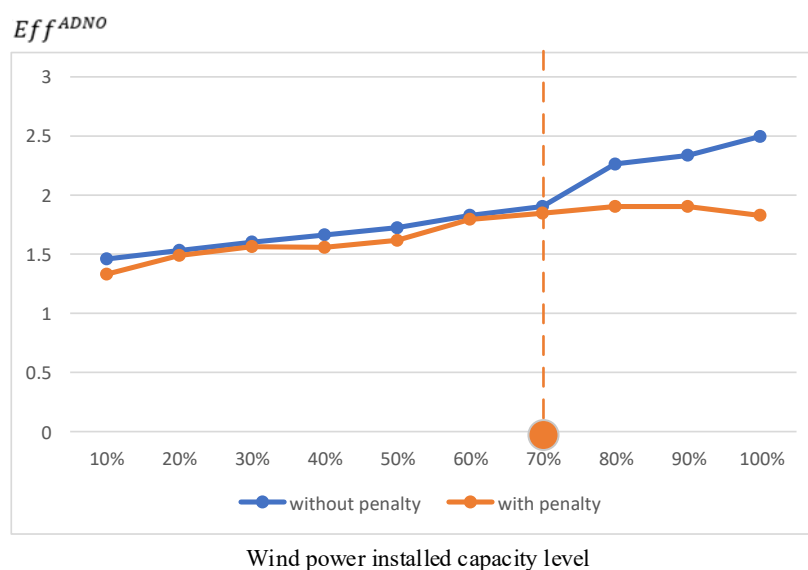


Figure 12. Sensitivity of the profitability indicator Eff^{ADNO} of ADNO to wind power installed capacity.

When the wind power installed capacity exceeds this turning point, the return on investment of ADNO will show a slow downward trend. This shows the robustness of our method to wind power installed capacity.

6. Conclusions

As the penetration rate of variable renewable energy such as wind power increases gradually in the power system, the composition and balance of the power system also changes gradually. The intermittency of renewable energy challenges the stability of the traditional centralized generation and load-based transmission and distribution methods at high proportions of renewable energy penetration. Therefore, the Active Distribution Network Operator (ADNO) with distributed installation at the local level has a good application prospect in the new scenario. However, ADNO faces the issue of how to improve operational efficiency based on the types of local power generation and energy storage installations and the nature of the market environment. In this regard, this paper proposes a forecasting method that considers the coupling fluctuations of spot electricity prices and renewable energy, as well as a bi-level optimization operation method based on Stackelberg game for operation optimization of small-scale ADNO under high wind power penetration rate. There are three major findings.

- The proposed conditional prediction method is effective. By analyzing the spot electricity price and wind power output data under high wind power penetration rate, this paper indicates the significant correlation between the two with zero time lag; therefore, this paper proposes a method based on ISO wind power prediction, which predicts the local wind power output and spot wholesale electricity price conditions.

The method combines conditional prediction, transformer deep neural network, and a source embedding method, and improves the prediction accuracy compared with traditional prediction methods. The improvement in prediction accuracy helps to reduce the penalty cost in ADNO operation and improve operational efficiency.

- A subsidy strategy may further improve the profitability of an ADNO. This paper proposes a subsidy strategy that considers the impact of ADNO on user demand. Under this subsidy strategy, a Stackelberg game is formed between the ADNO and users. This paper proposes a game model and its solution method that considers the spot electricity price of the main grid, local wind power output, local gas power generation, local energy storage, and network constraints simultaneously. Through simulation calculations, it is found that the comprehensive predictive method and bi-level optimization method proposed in this paper can indeed further improve the operational efficiency of ADNO compared with traditional methods.
- The proposed model is robust to external long-term factors. This paper compares the impact of long-term factors under the methods proposed in this paper. It is found that the methods proposed in this paper can provide a “guarantee boundary” for the fluctuation of factors such as the spot price of the main grid and the elasticity of user subsidies. When the fluctuation of factors exceeds this “guarantee boundary”, the profitability of ADNO can be guaranteed.

However, there are some aspects that need to be attended to in future studies. Firstly, the improvement is limited to the scenario where local wind power fluctuation is correlated to grid-level wind power fluctuation, otherwise the conditionality may be missing. Secondly, the computation may be time consuming as the optimization model is non-convex, especially for large-scale problems. Although this paper focuses on the off-line application, further attention may be needed for on-line applications.

Author Contributions: Conceptualization, Y.S. and R.X.; methodology, F.L.; software, F.L.; investigation, M.J.; writing—original draft, Y.S. and H.L.; writing—review and editing, M.J. and X.G.; visualization, H.L.; supervision, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by State Grid Jilin Electric Power Co., Ltd. 2022 science and technology “Jie Bang Gua Shuai” project, grant number 2022JBGS-05.

Data Availability Statement: Not applicable.

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

Nomenclatures

$P_{s,t}^{STRC}, P_{s,t}^{STRD}$	Charging and discharging power of energy storage system s at time t
U_s^{STRC}, U_s^{STRD}	Maximum charging/discharging power of s
$b_{s,t}$	Binary value indicate charging or discharging
$e_{s,t}^{STR}$	Total stored electricity of s at time t
Δt	Time interval
$P_{i,t}^{GEN}$	Active power output of i at time t
$Cost_i$	Total production cost of i
O_i	start-up cost of i
L_i^{GEN}, U_i^{GEN}	Minimum and Maximum output of unit i
L_i^{RAM}, U_i^{RAM}	Minimum and Maximum climbing amount of unit i between units
$p_{j,t}$	Retail electricity price of customer j at time t
$Demand_{j,t}$	Total active power demand of customer j at time t
$\tilde{d}_{j,t}, d_{j,t}$	Rigid and flexible demand of customer j at time t
$D_{j,t}$	Set of demand responses to subsidies
$sub_{j,t}$	ADNO’s subsidy strategy at time t
$Ela_{j,t}$	Elasticity coefficient of the power demand subsidy for j at time t .

P_n, Q_n	Active and reactive power injected into node n
P_{mn}, Q_{mn}	Active and reactive power flowing from node m to node n
$ I_{mn} $	Amplitude of the current flowing from node m to node n
$ U_n $	Voltage amplitude of node n
Ω	The variable set determined by constraints
\hat{p}_t	Spot price at time t;
$EffADNO$	ADNO's profitability indicator
K	Total number of experiments conducted
C^{GEN}	ADNO's daily generation cost
C^{PUN}	ADNO's contract penalty cost caused by forecasting and scheduling deviations
C^{PUN}	ADNO's contract penalty cost caused by forecasting and scheduling deviations
P_t^{REA}	Power purchase amount from the wholesale market by ADNO at time t

Appendix A

Formula (22) is a simplified form of the bi-level model. A detailed mathematical description of the bi-level is shown in Formula (A1).

$$\begin{aligned}
 & \max_{q_{i,t}^{GEN}, q_{i,t}^{REA}, q_{s,t}^{STEC}, q_{s,t}^{STRD}, q_{i,t}^{GEN}, q_{v,t}^{GEN}, q_{i,t}^{STRC}, q_{i,t}^{STRD}, w_{i,t}, b_{s,t}} F = \sum_{j=1}^J \sum_{t=1}^T p_t \cdot (\bar{d}_{j,t} + \tilde{d}_{j,t}) - \sum_{t=1}^T \left(\hat{p}_t \cdot q_t^{REA} + \sum_i (c_i \cdot q_{i,t}^{GEN} + O_i \cdot w_{i,t}) + \sum_{j=1}^J sub_{j,t} \right) \\
 \text{s.t.} & \left\{ \begin{aligned}
 & L_i^{RAM} \leq q_{i,t}^{GEN} - q_{i,t-1}^{GEN} \leq U_i^{RAM} \quad \forall t = 2, \dots, T \\
 & 0 \leq q_{s,t}^{STEC} \leq (1 - b_{s,t}) U_s^{STRC} \\
 & 0 \leq q_{s,t}^{STED} \leq b_{s,t} U_s^{STRD} \\
 & e_{s,t}^{STR} = e_{s,t-1}^{STR} + (q_{s,t}^{SRTC} - q_{s,t}^{STRD}) \Delta t \\
 & 0 \leq e_{s,t}^{STR} \leq U_s^{ESTR} \\
 & q_t^{REA} \geq 0 \\
 & |U_n|^2 = |U_m|^2 - 2(r_{mn} \cdot P_{mn} + x_{mn} \cdot Q_{mn}) + (r_{mn}^2 + x_{mn}^2) \cdot |I_{mn}|^2 \\
 & |I_{mn}| \leq I_{mn}^{\max} \\
 & |U_n|^{\min} \leq |U_n| \leq |U_n|^{\max} \\
 & \left\| \left[2P_{mn} \ 2Q_{mn} \ |I_{mn}|^2 - |U_{mn}|^2 \right]^T \right\|_2 \leq |I_{mn}|^2 + |U_{mn}|^2 \\
 & L_t^{GEN} z_{i,t} \leq q_{i,t}^{GEN} \leq U_t^{GEN} z_{i,t} \\
 & w_{i,t} = \begin{cases} 1, & z_{i,t} - z_{i,t-1} = 1 \\ 0, & \text{otherwise} \end{cases}
 \end{aligned} \right. \tag{A1}
 \end{aligned}$$

$$\begin{aligned}
 & \min_{\bar{d}_{j,t}, \tilde{d}_{j,t}} \sum_{j=1}^J \sum_{t=1}^T p_t (\bar{d}_{j,t} + \tilde{d}_{j,t}) \\
 \text{s.t.} & \left\{ \begin{aligned}
 & P_{n,t} \geq G_n \cdot q_t^{REA} + \sum_{j \in n} q_{i,t}^{GEN} + \sum_{s \in n} q_{s,t}^{STRD} + \sum_{v \in n} q_{v,t}^{GEN} - \sum_{s \in n} q_{i,t}^{STRC} - \sum_{j \in n} \bar{d}_{j,t} - \sum_{j \in n} \tilde{d}_{j,t} \quad (L1) \\
 & P_{n,t} \leq G_n \cdot q_t^{REA} + \sum_{j \in n} q_{i,t}^{GEN} + \sum_{s \in n} q_{s,t}^{STRD} + \sum_{v \in n} q_{v,t}^{GEN} - \sum_{s \in n} q_{i,t}^{STRC} - \sum_{j \in n} \bar{d}_{j,t} - \sum_{j \in n} \tilde{d}_{j,t} \quad (L2) \\
 & Q_{n,t} \geq G_n \cdot \phi_t^{REA} + \sum_{j \in n} \phi_{i,t}^{GEN} + \sum_{v \in n} \phi_{v,t}^{GEN} - \sum_{j \in n} \bar{\phi}_{j,t} - \sum_{j \in n} \tilde{\phi}_{j,t} \quad (L3) \\
 & Q_{n,t} \leq G_n \cdot \phi_t^{REA} + \sum_{j \in n} \phi_{i,t}^{GEN} + \sum_{v \in n} \phi_{v,t}^{GEN} - \sum_{j \in n} \bar{\phi}_{j,t} - \sum_{j \in n} \tilde{\phi}_{j,t} \quad (L4) \\
 & L_{j,t}^{ACT}(\text{sub}_{j,t}) \leq \bar{d}_{j,t} \leq U_{j,t}^{ACT}(\text{sub}_{j,t}) \quad (L5) \\
 & L_{j,t}^{REA}(\text{sub}_{j,t}) \leq \tilde{d}_{j,t} \leq U_{j,t}^{REA}(\text{sub}_{j,t}) \quad (L6)
 \end{aligned} \right.
 \end{aligned}$$

In Formula (A1), G_n is a 0–1 parameter indication whether node n is the interface node that connects the distribution network and the grid. Variable ϕ_t^{REA} is the reactive power imported by ADNO in time t. Variable $\phi_{i,t}^{GEN}$ is the reactive power generated by fossil unit i in time t. Parameter $\phi_{v,t}^{GEN}$ is the reactive power generated by VRE unit i in time t. Variable $\bar{\phi}_{j,t}$ is the lower level decision variable of reactive power consumption. Variable $\tilde{\phi}_{j,t}$ is the lower level decision variable of reactive power consumption by customer j in time t. Parameter $\tilde{\phi}_{j,t}$ is the uncontrollable reactive power consumption by customer j in time t.

The lower level problem is a relatively simple linear problem. This makes the whole problem relatively more solvable. We substitute the lower level problem with its KKT condition. The KKT transformed problem formulation is shown in Formula (A2).

$$\begin{aligned}
 & \max_{q_{i,t}^{GEN}, q_{i,t}^{REA}, q_{s,t}^{STEC}, q_{s,t}^{STRD}, q_{i,t}^{GEN}, q_{i,t}^{STRC}, q_{i,t}^{STRD}, w_{i,t}, b_{s,t}} F = \sum_{j=1}^J \sum_{t=1}^T p_t \cdot (\bar{d}_{j,t} + \tilde{d}_{j,t}) - \sum_{t=1}^T \left(\hat{p}_t \cdot q_t^{REA} + \sum_i (c_i \cdot q_{i,t}^{GEN} + O_i \cdot w_{i,t}) + \sum_{j=1}^J sub_{j,t} \right) \\
 & \text{s.t.} \\
 & \left\{ \begin{array}{l} \text{Original Upper level conditions} \\ \text{condition(L1, L2, L3, L4, L5, L6)} \\ [p_t \cdot \text{diag}] - \lambda^T \times (A \cdot x - b) - \mu^T \\ x \geq 0 \\ \lambda^T (Ax - b) = 0 \\ \mu^T x = 0 \\ \lambda \geq 0 \\ \mu \geq 0 \end{array} \right. \quad (A2)
 \end{aligned}$$

In Formula (A2), x is the original lower level decision variable vector $[\bar{d}_{1,1}, \dots, \bar{d}_{1,T}, \bar{d}_{2,1}, \dots, \bar{d}_{2,T}, \dots, \bar{d}_{J,1}, \dots, \bar{d}_{J,T}, \bar{\phi}_{1,1}, \dots, \bar{\phi}_{1,T}, \bar{\phi}_{2,1}, \dots, \bar{\phi}_{2,T}, \dots, \bar{\phi}_{J,1}, \dots, \bar{\phi}_{J,T}]^T$. diag is a $[J \cdot T] \times [J \cdot T]$ diagonal matrix with all diagonal entries equal to 1. A is a parameter matrix decided by condition (L1, L2, L3, L4, L5, and L6), which is not shown here due to its huge length. However, since the original lower problem is pure linear, matrix A is basically very simple to achieve. A is a $[4 + 8 \cdot [J \cdot T]] \times [2 \cdot [J \cdot T]]$ shaped matrix. b is a $[4 + 8 \cdot [J \cdot T]]$ shaped matrix. λ and μ are the Lagrangian coefficients. λ is a $[4 + 8 \cdot [J \cdot T]] \times 1$ shaped vector. μ is a $1 \times [2 \cdot [J \cdot T]]$ shaped vector.

References

- Zhao, J.; Wang, C.; Zhao, B.; Lin, F.; Zhou, Q.; Wang, Y. A Review of Active Management for Distribution Networks: Current Status and Future Development Trends. *Electr. Mach. Power Syst.* **2014**, *42*, 280–293. [\[CrossRef\]](#)
- Xiang, Y.; Han, W.; Zhang, J.; Liu, J.; Liu, Y. Optimal Sizing of Energy Storage System in Active Distribution Networks Using Fourier-Legendre Series based State of Energy Function. *IEEE Trans. Power Syst.* **2017**, *33*, 2313–2315. [\[CrossRef\]](#)
- Ge, S.; Wang, S.; Lu, Z.; Liu, H. Substation planning method in an active distribution network under low-carbon economy. *J. Mod. Power Syst. Clean Energy* **2015**, *3*, 468–474. [\[CrossRef\]](#)
- Li, Z.; Wang, H.; Ai, Q.; Zhang, Y. Interactive optimization between active distribution network with multi-microgrids based on distributed algorithm. *Energy Rep.* **2020**, *6*, 385–391. [\[CrossRef\]](#)
- Xiao, J.; Qibo, H.E.; Guoqiang, Z.U. Distribution management system framework based on security region for future low carbon distribution systems. *J. Mod. Power Syst. Clean Energy* **2015**, *3*, 12. [\[CrossRef\]](#)
- Usman, M.; Coppo, M.; Bignucolo, F.; Turri, R. Losses management strategies in active distribution networks: A review. *Electr. Power Syst. Res.* **2018**, *163*, 116–132. [\[CrossRef\]](#)
- Gabash, A.; Li, P. Active-Reactive Optimal Power Flow in Distribution Networks with Embedded Generation and Battery Storage. *IEEE Trans. Power Syst.* **2012**, *27*, 2026–2035. [\[CrossRef\]](#)
- Cortes, C.A.; Contreras, S.F.; Shahidehpour, M. Microgrid Topology Planning for Enhancing the Reliability of Active Distribution Networks. *IEEE Trans. Smart Grid* **2017**, *9*, 6369–6377. [\[CrossRef\]](#)
- Kyriakou, D.G.; Kanellos, F.D. Optimal Operation of Microgrids Comprising Large Building Prosumers and Plug-in Electric Vehicles Integrated into Active Distribution Networks. *Energies* **2022**, *15*, 6182. [\[CrossRef\]](#)
- Sedghi, M.; Ahmadian, A.; Pashajavid, E.; Aliakbar-Golkar, M. Storage scheduling for optimal energy management in active distribution network considering load, wind, and plug-in electric vehicles uncertainties. *J. Renew. Sustain. Energy* **2015**, *7*, 033120. [\[CrossRef\]](#)
- Jiang, T.; Li, Z.; Jin, X.; Chen, H.; Li, X.; Mu, Y. Flexible operation of active distribution network using integrated smart buildings with heating, ventilation and air-conditioning systems. *Appl. Energy* **2018**, *226*, 181–196. [\[CrossRef\]](#)
- Chen, F.; Dong, L.; Xiong, X. Research on Stochastic Optimal Operation Strategy of Active Distribution Network Considering Intermittent Energy. *Energies* **2017**, *10*, 522. [\[CrossRef\]](#)
- Wang, J.; Zheng, X.; Tai, N.; Wei, W.; Li, L. Resilience-Oriented Optimal Operation Strategy of Active Distribution Network. *Energies* **2019**, *12*, 3380. [\[CrossRef\]](#)
- Tostado-Véliz, M.; Hasanien, H.M.; Jordehi, A.R.; Turky, R.A.; Gómez-González, M.; Jurado, F. An Interval-based privacy-Aware optimization framework for electricity price setting in isolated microgrid clusters. *Appl. Energy* **2023**, *340*, 121041. [\[CrossRef\]](#)
- Fan, S.; He, G.; Zhou, X.; Cui, M. Online optimization for networked distributed energy resources with time-coupling constraints. *IEEE Trans. Smart Grid* **2020**, *12*, 251–267. [\[CrossRef\]](#)
- Saint-Pierre, T.; Mancarella, P. Scheduling model for active distribution system management with renewable energy sources. *IET Gener. Transm. Distrib.* **2017**, *11*, 1571–1579.

17. Shen, X.; Shahidehpour, M.; Han, Y.; Zhu, S.; Zheng, J. Expansion planning of active distribution networks with centralized and distributed energy storage systems. *IEEE Trans. Sustain. Energy* **2016**, *8*, 126–134. [CrossRef]
18. Yu, Q.; Li, Z. Correlated load forecasting in active distribution networks using Spatial-Temporal Synchronous Graph Convolutional Networks. *IET Energy Syst. Integr.* **2021**, *3*, 355–366. [CrossRef]
19. Kalantar-Neyestanaki, M.; Sossan, F.; Bozorg, M.; Cherkaoui, R. Characterizing the reserve provision capability area of active distribution networks: A linear robust optimization method. *IEEE Trans. Smart Grid* **2019**, *11*, 2464–2475. [CrossRef]
20. Cheng, G.; Song, S.; Lin, Y.; Huang, Q.; Lin, X.; Wang, F. Enhanced state estimation and bad data identification in active power distribution networks using photovoltaic power forecasting. *Electr. Power Syst. Res.* **2019**, *177*, 105974. [CrossRef]
21. Cong, P.; Tang, W.; Zhang, L.; Zhang, B.; Cai, Y. Day-ahead active power scheduling in active distribution network considering renewable energy generation forecast errors. *Energies* **2017**, *10*, 1291. [CrossRef]
22. Gao, H.; Liu, J.; Wang, L. Robust coordinated optimization of active and reactive power in active distribution systems. *IEEE Trans. Smart Grid* **2017**, *9*, 4436–4447. [CrossRef]
23. Saint-Pierre, A.; Mancarella, P. Active Distribution System Management: A Dual-Horizon Scheduling Framework for DSO/TSO Interface Under Uncertainty. *IEEE Trans. Smart Grid* **2017**, *8*, 2186–2197. [CrossRef]
24. ENTSO-E Transparency Platform. Available online: <https://www.entsoe.eu/data/> (accessed on 23 April 2023).
25. Hassan, M.A.; Bailek, N.; Bouchouicha, K.; Nwokolo, S.C. Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks. *Renew. Energy* **2021**, *171*, 191–209. [CrossRef]
26. Li, G.; Wang, H.; Zhang, S.; Xin, J.; Liu, H. Recurrent neural networks based photovoltaic power forecasting approach. *Energies* **2019**, *12*, 2538. [CrossRef]
27. Alipour, M.; Aghaei, J.; Norouzi, M.; Niknam, T.; Hashemi, S.; Lehtonen, M. A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration. *Energy* **2020**, *205*, 118106. [CrossRef]
28. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [CrossRef]
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, I.
30. Wan, L.; Wang, Q.; Papir, A.; Moreno, I.L. Generalized end-to-end loss for speaker verification. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Calgary, AB, Canada, 15–20 April 2018; pp. 4879–4883.
31. Ruhang, X. Efficient clustering for aggregate loads: An unsupervised pretraining based method. *Energy* **2020**, *210*, 118617. [CrossRef]
32. Wang, Y.; Skerry-Ryan, R.J.; Stanton, D.; Wu, Y.; Weiss, R.J.; Jaitly, N.; Yang, Z.; Xiao, Y.; Chen, Z.; Bengio, S.; et al. Tacotron: Towards end-to-end speech synthesis. *arXiv* **2017**, arXiv:1703.10135, 2017.
33. Ugurlu, U.; Oksuz, I.; Tas, O. Electricity price forecasting using recurrent neural networks. *Energies* **2018**, *11*, 1255. [CrossRef]
34. Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [CrossRef]
35. Zhao, Q.; Shen, Y.; Li, M. Control and bidding strategy for virtual power plants with renewable generation and inelastic demand in electricity markets. *IEEE Trans. Sustain. Energy* **2015**, *7*, 562–575. [CrossRef]
36. Antoniadou-Plytaria, K.E.; Kouveliotis-Lysikatos, I.N.; Georgilakis, P.S.; Hatziaargyriou, N.D. Distributed and Decentralized Voltage Control of Smart Distribution Networks: Models, Methods, and Future Research. *IEEE Trans. Smart Grid* **2017**, *8*, 2999–3008. [CrossRef]
37. Li, Y.; Xiao, J.; Chen, C.; Tan, Y.; Cao, Y. Service restoration model with mixed-integer second-order cone programming for distribution network with distributed generations. *IEEE Trans. Smart Grid* **2018**, *10*, 4138–4150. [CrossRef]
38. Baran, M.E.; Wu, F.F. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Trans. Power Deliv.* **1989**, *4*, 1401–1407. [CrossRef]
39. Hu, W.; Guo, Y.; Jing, J.; Chen, L. Dynamic Hosting Capacity Evaluation of DGs in Active Distribution Network. In Proceedings of the 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), IEEE, Nanjing, China, 20–23 September 2020; pp. 1–6.
40. Zhang, W.; Liu, K.; Sheng, W.; Du, S.; Jia, D. Critical node identification in active distribution network using resilience and risk theory. *IET Gener. Transm. Distrib.* **2020**, *14*, 2771–2778. [CrossRef]
41. Feroldi, D.; Rullo, P. Day-ahead optimal operation of active distribution networks with distributed generation and energy storage. *Int. J. Emerg. Electr. Power Syst.* **2022**. [CrossRef]
42. Dolatabadi, S.H.; Ghorbanian, M.; Siano, P.; Hatziaargyriou, N.D. An enhanced IEEE 33 bus benchmark test system for distribution system studies. *IEEE Trans. Power Syst.* **2020**, *36*, 2565–2572. [CrossRef]
43. Abdelouadoud, S.Y.; Girard, R.; Neirac, F.P.; Guiot, T. Optimal power flow of a distribution system based on increasingly tight cutting planes added to a second order cone relaxation. *Int. J. Electr. Power Energy Syst.* **2015**, *69*, 9–17. [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.