Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 5, 2025

An Ultra-Short Term Forecasting Method for Massive Distributed Photovoltaics Considering Spatial-Temporal Correlation

He Yu*, Yingchun Wang, Wei Wei, Li Ye, Jun Li, Yalan Wang

Measurement Center, State Grid Hubei Marketing Service Center, Hubei, China
*Corresponding author's email: yuhe_stategridhb@126.com

Abstract. The output of geographically adjacent distributed photovoltaic (PV) units exhibits strong temporal and spatial correlations. PV operators can select representative PV units from many distributed PVs and install real-time data transmission equipment at these locations. By leveraging the correlation of neighboring PV outputs, an efficient forecasting method for large-scale PV can significantly reduce the cost of real-time communication for PV data. In this paper, a framework for an ultra-short-term forecasting method for massive distributed PVs is proposed. Firstly, the initial PV output sequence is decomposed by multivariate variational mode decomposition. Then, based on the decomposed sequence results, the K-medoids algorithm is utilized to categorize the distributed PV units into distinct clusters, with data transmission systems positioned at cluster centers. Finally, a distributed PV ultra-short-term forecasting network is constructed using a dynamic graph convolution and gated cycle unit structure, fully considering the correlation of adjacent distributed PV outputs. The experimental results demonstrate that the proposed ultra-short-term forecasting framework can efficiently plan real-time PV communication equipment and achieve high-precision forecasting of large-scale distributed PVs.

Key words. Photovoltaic power forecasting, Spatial-temporal correlativity, Multivariate variational mode decomposition, Graph convolutional network, Gated cycle unit

1. Introduction

According to the International Renewable Energy Agency's World Energy Transition Outlook, solar photovoltaic (PV) and onshore wind power are projected to generate a combined capacity of 8.5 terawatts by 2030 [1]. In recent years, the installed capacity and power generation of distributed PVs have risen rapidly. This growth trend is driven by a combination of factors, including declining costs of PV technology, supportive government policies, and increasing awareness of the environmental benefits associated with renewable energy. However, as the proportion of distributed PVs connected to the grid increases, the intermittency and volatility of PV output have posed significant challenges to the stable operation of the power grid [2].

The proliferation of distributed PVs has catalyzed the transformation of traditional distribution network (TDN) into active distribution network (ADN) incorporating distributed energy resources (DERs), including distributed PVs, electric vehicle (EV), and energy storage (ES) as shown in Figure 1. Notwithstanding these advancements, there is growing concern for revenue loss caused by the high penetration of weather-dependent distributed PVs in ADN [3]. This is mainly due to the intermittent and fluctuating nature of distributed PVs. Distributed PV output is highly uncertain, often influenced by many factors such as irradiance, temperature, wind speed, and surrounding environment, and can lead to fluctuations in power supply. These fluctuations, if not properly managed, can strain the power grid, potentially leading to issues such as voltage instability, frequency deviations, and difficulties in load balancing. Therefore, addressing the challenges posed by the integration of distributed PVs into the grid is crucial for ensuring the reliability and efficiency of the overall power system.

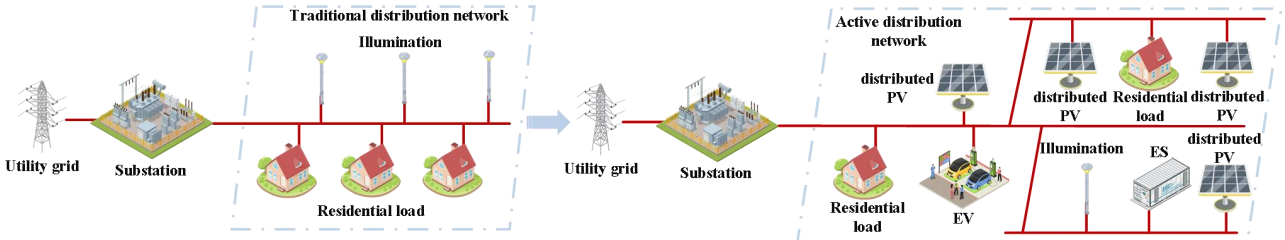


Figure 1. The transformation of TDN towards ADN.

PV operators typically install real-time data acquisition and transmission equipment at the PV units, commonly referred to as smart inverters [4]. These smart inverters are advanced equipment that not only convert the direct current (DC) generated by solar panels into alternating current (AC) suitable for the grid but also collect and transmit real-time data on PV output. Leveraging the real-time PV output data collected, PV operators can conduct high-precision ultra-short-term forecasts and assist utilities in real-time power dispatch adjustments [5]. This capability is crucial for maintaining grid stability and optimizing the integration of renewable energy sources. However, the economic cost of large-scale real-time data transmission is substantial, Installing and maintaining data transmission systems at each PV unit requires significant investment in hardware, network infrastructure, and ongoing maintenance. Additionally, there is a risk of data leakage, making it difficult for distributed PV operators to install data transmission systems at each PV unit. These challenges make it difficult for distributed PV operators, especially smaller ones, to install data transmission systems at every PV unit. Given the strong correlation between the output of geographically proximate distributed PVs, a single PV output sequence can reflect the output of surrounding PVs. This spatial correlation can be leveraged to reduce the number of data transmission points needed for accurate forecasting. Therefore, smart inverters can be installed at PV units with representative output to support efficient forecasting of a large number of distributed PVs. By selecting a few key PV units as data collection points, operators can gather sufficient information to make accurate predictions for the entire distributed PV network.

Currently, research on short-term PV forecasting mainly focuses on single PV forecasting, failing to establish an effective framework for large-scale distributed PV forecasting. This gap in research highlights the need for new approaches that can handle the complexities of distributed PV systems. There is a universal classification method of dividing the forecasting time range of PV output. It is divided into 3 categorizes: medium-term and long-term forecasting (when forecasting time range is between 1 month and 1 year) which focuses on strategic planning, resource allocation, market operations, short-term forecasting (when forecasting time range is between 4 hours and 3 days) which aids in operational planning, load balancing, maintenance scheduling, and market trading over a period of days, and ultra-short-term forecasting (less than 4 hours) which is essential for real-time grid management, renewable energy integration, emergency response, and demand response programs over a period of hours. There are significant differences in feature selection and model construction for PV forecasting across different time scales. Ultra-short-term PV forecasting, in particular, is characterized by high volatility and uncertainty due to the rapid changes in weather conditions and solar irradiance. This type of forecasting usually adopts satellite cloud image data, real-time temperature, and irradiance [6] as input features, characterized by high volatility and uncertainty. These features provide a detailed and up-to-date picture of the current and near-future conditions affecting PV output. However, the high volatility and uncertainty of these inputs require sophisticated models that can handle rapid changes and provide accurate predictions within a very short time frame. This makes ultra-short-term forecasting particularly challenging but also highly valuable for real-time grid management and power dispatch adjustments.

The ultra-short-term PV forecasting can be categorized into two main methods: physical methods [7] and statistical methods [7-11]. Physical methods rely on detailed meteorological data and the physical principles governing solar energy conversion, aiming to predict PV output based on environmental conditions such as solar irradiance, temperature, and cloud cover. These methods are highly accurate when precise meteorological data is available but can be limited by the complexity of atmospheric models and the availability of real-time data. On the other hand, statistical methods do not consider the specific physical mechanisms of PV units. Instead, they build forecasting models based on historical PV output data, leveraging patterns and trends within the data to make predictions. These methods are widely used due to their high accuracy in forecasting. [8] considers the real-time sky images and solar irradiance historical data. To figure out their correlation, it adopts the transformer model to process image, and carries out multi-step irradiance forecasting using multi-modal coupling analysis, integrating visual and temporal data to enhance prediction accuracy. Another approach is that PV output data can be components with low and high frequencies. Components with different frequencies are forecasted using convolutional neural network (CNN) separately [9]. This method leverages the strengths of CNNs to handle the varying dynamics of PV output. [10] proposes that Integrated Empirical Mode Decomposition (EEMD) can also construct PV modal components. With the decomposed sequences, Autoregressive Moving Average (ARMA) can be combined with Long Short-Term Memory (LSTM) to work as a forecasting model. The reason why the two methods are combined is that ARMA model is responsible for identifying the general trend, while the LSTM handles autocorrelation fluctuations in PV output, thereby enhancing the overall forecasting accuracy. Accurate meteorological predictions serve as a fundamental pillar for ultra-short-term PV forecasting. It has been proposed that the development of an effective feature selection module for forecasting, the implementation of multi-objective intelligent optimization techniques, and the utilization of combined models can collectively enhance the robustness and precision of PV forecasting [11]. The complex temporal correlation of PV output is considered, with a particular focus on the impact of meteorological factors on PV output [12]. This method helps establish a comprehensive feature selection framework. Along with the mechanism of rolling time series network, the method integrates a neural network called Enhanced Time Convolutional Neural Network (TCN), so as to forecast the ultra-short-term multi-step outputs of PV systems. This integrated approach not only considers the immediate meteorological conditions but also the historical context,

providing a robust framework for accurate and reliable PV forecasting.

The power output of PV units that are located next to each other tends to be highly correlated. This phenomenon is particularly significant as large-scale distributed PV systems are increasingly integrated into the power grid. The integration of these systems has led to a surge in research focusing on ultra-short-term PV power forecasting, with a particular emphasis on accounting for geographical correlation. By analyzing the spatial-temporal correlation of PV power, suitable adjacent PV power stations are identified based on satellite image data. Subsequently, the meteorological characteristics and output data from these neighboring PV stations are utilized to create input features [13]. It has been proposed that considering the interaction between temporal and spatial features can enhance forecasting accuracy. To this end, a two-flow model and an attention mechanism can be employed to extract sky image features. Subsequently, the power forecasting is completed by a progressive structure and a time series model [14]. A stacked dilated convolutional network (DCN) helps to extract spatial features of new energy. To fully obtain the features on a spatial-temporal base and further utilize them in forecasting, together with integrated attention mechanism, attention-time convolutional network is constructed [15]. The characteristics of input data are obtained by self-attention mechanism (SA), which allows the model to focus on the most relevant parts of the input data. Furthermore, to makes full use of the spatiotemporal correlation of PV output when building an ultra-short-term forecasting model, [16] proposes integrating a encoder-decoder structure (ACGRU) called bidirectional convolutional gated cycle unit (BiConvGRU). The models mentioned above consider adjacent PV units and excavate their spatial correlations, but they are primarily designed for single PV forecasting and are not easily applicable to large-scale distributed PV forecasting scenarios. This limitation highlights the need for more comprehensive models that can handle the complexities and scale of distributed PV systems.

In this paper, we propose a framework for ultra-short-term forecasting of large-scale distributed PVs. To excavate the original sequence into sub-modes with different frequency centers, Multivariate variational mode decomposition (MVMD) is an effective choice. Based on the decomposed sequence results, to determine the placement of data transmission systems by clustering, the K-medoids algorithm is employed. A ultra-short-term forecasting network for massive distributed PV is constructed using a dynamic graph convolution and gated cycle unit (GRU) structure, fully considering the correlation of adjacent distributed PV outputs. The main contributions of this paper are summarized as follows.

(1) An integrated framework for ultra-short-term forecasting of massive distributed PVs is proposed. The framework addresses the challenges of limited communication equipment installation capacity and high costs associated with real-time data acquisition and transmission by strategically selecting locations for data transmission equipment. This approach ensures efficient and accurate forecasting while optimizing resource allocation.

(2) A novel data pre-processing method using MVMD is introduced in this paper. This method effectively decomposes multi-dimensional PV sequences into distinct modes while preserving the output correlation between adjacent power stations. By enhancing the quality and structure of the input data, this pre-processing step significantly improves the accuracy and efficiency of subsequent forecasting tasks.

(3) A dynamic graph convolutional network combined with a gated recurrent unit (DGCN-GRU) model is developed in this paper. This advanced model architecture captures both temporal and spatial correlations within distributed PV output, thereby enhancing the accuracy of ultra-short-term PV forecasting.

Additionally, this paper provides insights for future work, including determining the optimal number of communication equipment to balance forecasting accuracy and system economics, and developing an integrated system for equipment location selection and PV forecasting to further enhance model performance. The structure of this paper is organized as follows: Section II introduces a data pre-processing method based on multivariate variational mode decomposition. Section III describes the method of using K-medoids to determine the planning scheme of communication of equipment by clustering. Section IV presents a short-term forecasting scheme for large-scale PVs based on DGCN-GRU. Section V verifies the forecasting effectiveness of the proposed scheme using actual PV data. Finally, Section VI summarizes the entire paper and discusses future research directions.

2. Method: Data Pre-processing Based on Multivariate Variational Mode Decomposition

In this part, a comprehensive data pre-processing framework based on MVMD is established to enhance the accuracy and efficiency of large-scale distributed PV forecasting. This section is primarily divided into two subsections. Section A introduces the fundamental concept and mathematical formulation of variational mode decomposition (VMD), highlighting its effectiveness in decomposing non-stationary time series data into manageable sub-modes. Section B extends the discussion to MVMD, detailing how it addresses the challenges of multi-dimensional data and improves the decomposition process by considering the correlations among multiple PV output sequences.

A. Variational Mode Decomposition

VMD [17] is frequently employed to address non-stationary time series that are highly complex and unstable. The factors influencing PV output are intricate and highly uncertain. Using these factors directly as

input features can negatively impact the accuracy of forecasting models. Hence, data pre-processing is essential to break down the original data into manageable sub-mode sequences and effectively tackle the mode aliasing issue present in traditional empirical mode decomposition. The core of VMD lies in solving variational problems, decomposing sub-modes with distinct center frequencies, and minimizing the total bandwidth of these sub-modes. The alternating direction method of multipliers is used to iteratively update the center frequency of each mode. The sub-modes are then demodulated to their fundamental frequencies, and the information of the sub-modes and center frequencies is extracted. VMD boasts a robust theoretical foundation and outperforms other modal decomposition algorithms in terms of noise robustness, modal decomposition performance, and sampling effectiveness. VMD assumes that each sub-mode has a fixed center frequency and limited bandwidth. The objective function aims to minimize the sum of the bandwidths of the sub-modes, subject to the constraint that the sum of the sub-modes equals the original input signal. The detailed steps are as follows:

(1) The Hilbert transform is applied to derive the unilateral spectrum of each sub-mode function: The frequency spectrum of each mode is then shifted to its corresponding fundamental frequency band, as in (1).

$$s_+(t) = s(t) + j\Gamma s(t) = A(t)\mathrm{e}^{j\phi(t)} \quad (1)$$

Where $A(t)$ is the amplitude function, $\Gamma s(t)$ is a time-varying signal phase function, $\phi(t)$ is a phase function, affected by phase modulation, $s_+(t)$ is a unilateral spectrum signal generated by the Hilbert transform. $\Gamma s(t)$ can be calculated as (2).

$$\Gamma s(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{s(\tau')}{\tau-\tau'}\mathrm{d}\tau' \quad (2)$$

Suppose $\hat{s}_+(\omega)$ is the Fourier transform of $s_+(t)$, it can be written as (3).

$$\begin{aligned}\hat{s}_+(\omega) &= \int_{-\infty}^{\infty} s_+(t)\mathrm{e}^{-j\omega t}\mathrm{d}t \\ &= \left(1+\mathrm{sgn}(\omega)\right)\hat{s}_+(\omega)\end{aligned} \quad (3)$$

Where $\mathrm{e}^{-j\omega t}$ is the exponential term of the center frequency.

(2) The gradient norm of the demodulated signal is computed, serving as an estimate for the total bandwidth of the sub-modes. Based on this, a constrained variational problem is formulated. In this paper, the L2 norm is utilized, as in (4).

$$\min_{\{u_m\},\{\omega_m\}}\left\{\sum_m\left\|\partial_t\left[(\delta(t)+\frac{j}{\pi t})*s_m(t)\right]\mathrm{e}^{-j\omega_m t}\right\|_2^2\right\} \quad (4)$$

Where $\delta(t)$ is the Dirac distribution function, symbol $*$ is a convolution multiplier, $\{s_m\} := \{s_1,\ldots,s_M\}$ means the sub-modes, $\{\omega_m\} := \{\omega_1,\ldots,\omega_M\}$ represents each center frequency, and m means the number of decomposed modes. The constraint ensures that the sum of the sub-modes reconstructs the original signal, as in (5).

$$\text{s.t.}\quad \sum_m s_m = g \quad (5)$$

Where g represents the original input sequence.

(3) The initial constrained variational problem is converted into an unconstrained one by incorporating Lagrange multipliers and quadratic penalty factors. The quadratic penalty term helps maintain reconstruction accuracy in the presence of Gaussian noise. The resulting Lagrange expression is presented as in (6).

$$L(\{s_m\},\{\omega_m\},\mu) := p\sum_m\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*s_m(t)\right]\mathrm{e}^{-j\omega_m t}\right\|_2^2 + \left\|g(t)-\sum_m s_m(t)\right\|_2^2 + \left\langle\mu(t), g(t)-\sum_m s_m(t)\right\rangle \quad (6)$$

Where p represents quadratic penalty factor, and μ is the Lagrange multiplier.

(4) VMD adopts alternate direction method of multipliers (ADMM) to calculate the variational problem and update s_m and ω_m alternately. The iteration formula of s_m is as in (7).

$$\hat{s}_m^{n+1}(\omega) = \frac{\hat{g}(\omega)-\sum_{i\neq m}\hat{s}_i(\omega)+\frac{\hat{\mu}(\omega)}{2}}{1+2p(\omega-\omega_m)^2} \quad (7)$$

The iterative formula of ω_m is as in (8).

$$\omega_m^{n+1} = \frac{\int_0^{\infty}\omega\left|\hat{s}_m(\omega)\right|^2\mathrm{d}\omega}{\int_0^{\infty}\left|\hat{s}_m(\omega)\right|^2\mathrm{d}\omega} \quad (8)$$

Where symbol $\wedge$ represents the Fourier transform.

B. Multivariate Variational Mode Decomposition

To tackle the challenge of forecasting large-scale distributed PV power requires a sophisticated approach to handle the complexity of the data. The input to the forecasting model consists of multiple PV power time series, meaning that the multivariate signal must be decomposed into distinct modes. There is a strong correlation between the PV output of geographically adjacent locations. However, current multivariate modal decomposition methods face several issues:

(1) Correlation Between Sequences: It is challenging to account for the correlation between different sequences. Traditional methods often fail to capture the intricate relationships between multiple PV power time series, leading to less accurate forecasts.

(2) Frequency Alignment: Frequency alignment problems arise when different signals share the same frequency. This can lead to mode mixing and aliasing, which further complicates the decomposition process.

If traditional multivariate mode decomposition is applied to process the massive distributed PV output sequences, issues such as sampling effects, noise sensitivity, and insufficient mode separation are likely to occur. To overcome these challenges, this paper employs MVMD [18] to generate power data for multiple PV units. This method is an extension of the VMD algorithm into the multivariate domain. Like VMD, it decomposes multivariate time series data into multiple modal components with limited bandwidth and fixed center frequencies, modulates the signals using the Hilbert transform, and aims to minimize the total bandwidth as the objective function, subject to the constraint that the sub-modes can reconstruct the original signal. While VMD performs variable optimization in a one-dimensional space, MVMD requires signal modulation in a multidimensional space to effectively enhance the continuity and mode alignment capabilities of the model, making it more suitable for handling complex multivariate data. The objective function of the variational problem corresponding to MVMD is designed to minimize the total bandwidth of the decomposed modes while ensuring that the sum of the sub-modes reconstructs the original signal. The objective function is as (9).

$$\min_{\{s_m\},\{\omega_m\}} \left\{ \sum_m \sum_b \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * s_{m,b}(t) \right] \mathrm{e}^{-j\omega_m t} \right\|_2^2 \right\} \quad (9)$$

Where $s_{m,b}(t)$ means the decomposed time signal, and b is the number of original signals. The constraint is about each original signal, as in (10).

$$\text{s.t.} \ \sum_m s_{m,b}(t) = g_b(t), \ \ b = 1, 2, \cdots, B \quad (10)$$

By employing MVMD, this paper aims to provide a more accurate and robust method for decomposing the complex multivariate PV power time series, thereby enhancing the overall forecasting performance. Also, Lagrange multipliers is introduced to transform the original constrained variational problem, as in (11).

$$L\left(\{s_{m,b}\}\{\omega_m\}, \mu_b\right) := p \sum_m \sum_b \left\| \partial_t \left[s_+^{m,b}(t) \mathrm{e}^{-j\omega_m t} \right] \right\|_2^2 + \sum_b \left\| g_b(t) - \sum_m s_{m,b}(t) \right\|_2^2 + \sum_b \left\langle \mu_b(t), g_b(t) - \sum_m s_{m,b}(t) \right\rangle \quad (11)$$

ADMM algorithm is used to calculate each sub-mode and corresponding center frequency iteratively. The iterative formula of decomposed signal in sub-mode is as in (12).

$$\hat{s}_{m,b}^{n+1}(\omega) = \frac{\hat{g}_b(\omega) - \sum_{i \neq m} \hat{s}_{i,b}(\omega) + \frac{\hat{\mu}_b(\omega)}{2}}{1 + 2p(\omega - \omega_m)^2} \quad (12)$$

The iterative formula of the center frequency is as in (13).

$$\omega_m^{n+1} = \frac{\sum_b \int_0^\infty \omega \left| \hat{s}_{m,b}(\omega) \right|^2 \mathrm{d}\omega}{\sum_b \int_0^\infty \left| \hat{s}_{m,b}(\omega) \right|^2 \mathrm{d}\omega} \quad (13)$$

For the Lagrange multiplier μ_b, its iterative formula is as in (14).

$$\hat{\mu}_b^{n+1}(\omega) = \hat{\mu}_b^n(\omega) + \beta \left[\hat{x}_b(\omega) - \sum_m \hat{s}_{m,b}^{n+1} \right] \quad (14)$$

Where β is an update coefficient, used to mitigate the influence of Lagrange transformation on the mode decomposition results.

3. Method: Addressing Communication Equipment

Real-time PV power series and meteorological data are commonly utilized as input features for ultra-short-term forecasting. PV operators typically install real-time data acquisition and transmission equipment at PV units to enhance forecasting accuracy. However, for large-scale distributed PV forecasting, installing data transmission systems at every PV unit would significantly increase equipment costs. Additionally, constructing an extensive real-time communication system would incur maintenance expenses and pose security risks. Given that

PV units in close geographical proximity exhibit highly similar power outputs, operators can select a representative PV unit from a large number of distributed PVs to be forecasted. This representative unit can provide insights into the current status and future trends of PV output in the region based on its output information.

In this paper, clustering is used as equipment location algorithm. Clustering is an unsupervised classification algorithm, which can classify similar objects into different clusters according to the specific characteristics of the data [19,20]. It can be primarily divided into hierarchical, density-based, and other methods. Equipement addressing is a typical application of clustering algorithms. In this section, large-scale distributed PVs are grouped into several clusters according to their processing characteristics, with real-time data transmission equipment installed at the cluster centers. The PV unit at the cluster center can reflect the overall output characteristics of the cluster. The primary goal of PV clustering here is to identify typical PV units rather than classify PVs, making it different from conventional clustering problems.

The distributed PV data collection point should be a representative set of PV units that can capture the output power characteristics of the entire PV system. To achieve this goal, K-medoids [21] cluster analysis is used to divide all PV units into groups based on the geographical distribution of distributed PVs. In each group, the data collection point is one of the most representative units, with the smallest sum of distances to all other units in the same category. The traditional K-means [22] clustering algorithm is easy to apply but it only classifies objects based on centroids, which may lead to the result of lacking physical meaning, and is easily influenced by outliers. In contrast, the K-medoids algorithm iteratively optimizes the cluster center point, offering greater stability. The cluster center can be directly used as the equipment installation point. Therefore, K-medoids is selected to determine the equipment installation point, with the clustering feature being the PV power sequence of the training set. The Euclidean distance [23] is chosen as the distance metric, as in (15).

$$l_{ij} = \sqrt{\sum_{n=1}^{q} \left(Y_{i,n} - Y_{j,n} \right)^2} \quad i = 1, \cdots, n; j = 1, \cdots, n \quad (15)$$

Calculate the indicator r_i according to the distance, as in (16).

$$r_i = \sum_{j=1}^{n} \frac{l_{ji}}{\sum_{l=1}^{n} l_{jl}}, \quad i = 1, \cdots, n \quad (16)$$

Assuming n distributed PV units to be forecasted, m PV units installed with real-time data transmission equipment are called representative PV units, that is, the clustering number is m . The detailed steps of the K-medoids algorithm are as follows:

(1) Initialization of clustering center: Compute the distances and r_i values among the clustering objects. Select m PV units with the smallest r values as the initial clustering centers. Perform the initial clustering by assigning each PV unit to the cluster whose center is closest to it. After the initial clustering, compute a distance index for each cluster.

(2) Update clustering center: Based on the updated clustering results, identify the point within each cluster that has the minimum total distance to other points in the cluster and set it as the new clustering center.

(3) Update the clustering: Recalculate the distance between each PV unit and the new representative PV unit of its cluster Based on these recalculated distances, reassign each PV unit to the cluster whose new center is closest to it. After updating the clustering, recompute the sum of distances between each PV unit and the representative PV unit of its clustering. Compare the new sum of distances with the previous sum. If the distance sum remains unchanged, the algorithm terminates. Otherwise, return to Step 2 to continue updating the clustering center and reassigning clusters.

4. Method: Ultra-Short-Term Forecasting

In this part, ultra-short-term forecasting model is established. This section is primarily divided into three subsections. Section A provides a training model of the neural network utilized, namely Graph convolutional neural network (GCN). Section B proposes the principle and calculation method of the GRU model to achieve an efficient and high-precision time series prediction method. Section C depicts the global forecasting framework of the massive distributed PV, and shows the flow chart.

A. Graph Convolutional Neural Network

GCN [24] is often mentioned when talking about a neural network structure commonly used to process graph data. A GCN is a type of neural network specifically designed to handle data that is naturally represented as graphs. The graph convolutional layer within a GCN is capable of extracting spatial features from the data while preserving the topological information embedded in the input features. This ability to retain the structural relationships between nodes is what sets GCNs apart from traditional fully connected neural networks, which do not account for the underlying graph structure of the data. The graph convolutional layer operates by aggregating information from a node's neighborhood, effectively capturing the spatial dependencies within the graph. This process allows the network to learn more meaningful representations of the nodes, which in turn enhances the overall fitting performance of the model. Figure 2 illustrates the basic structure of a single-layer graph convolutional network,

which retains the graph characteristics of the input data, setting it apart from traditional fully connected neural networks. In the context of large-scale distributed PV power forecasting, the geographical proximity of PV power stations plays a vital role. The outputs of adjacent PV power stations are highly correlated due to shared environmental conditions and similar operational characteristics. This correlation is a key factor in improving the accuracy of PV power forecasts. Real-time data transmission systems are installed at representative PV power stations. This representative station serves as a central point for data collection, providing critical power information that is used to forecast the output of other PV units. By leveraging the power information from this representative station and the known correlation between PV outputs, the forecasting model can make more accurate predictions. This approach underscores the importance of considering the geographical correlation of PV output, as it significantly enhances the effectiveness of the forecasting framework.

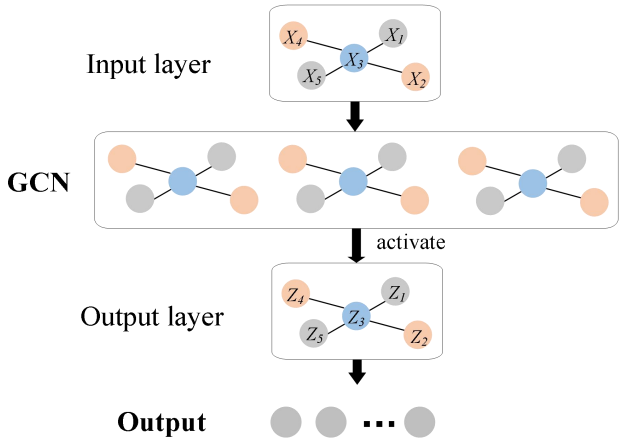


Figure 2. Structure of GCN.

Suppose network V describe massive distributed PVs, mathematically written as a graph with two components to describe the geographic correlation of PV output, U and D, thus $V=(V,D)$. In the two components, $U=\{u_1,u_2,\cdots,u_M\}$ is a set of PV units, with M being the number of them. The another component D is a set of edges. Further we need to introduce an adjacency matrix $E \in R^{M\times M}$ as the spatial correlation of PVs, representing the connection relationship.

The following formulas illustrate the calculation process of GCN according to the single-layer graph convolutional network in Figure 2. Given the adjacency matrix E and input feature matrix F, the graph convolution layer can be expressed as in (17).

$$N^{(1)}=\zeta\left(\tilde{G}^{-\frac{1}{2}}E\tilde{G}^{-\frac{1}{2}}N^{(1)}\varphi^{(1)}\right) \quad (17)$$

Where $\tilde{G}:=\sum_j \tilde{E}_{ij}$ is a degree matrix, N is the output in the neural network that is iterated every time, here $N^{(1)}$ is the output of layer 1, $\varphi^{(1)}$ is the corresponding weight coefficient, and Sigmoid function is chosen as activation function $\zeta(\cdot)$. The output layer and output sequence of the second iteration can be represented as in (18).

$$N^{(2)}=\zeta\left(N^{(1)}\varphi^{(2)}\right) \quad (18)$$

Similarly, the output layer and output sequence of the third iteration as in (19).

$$N^{(3)}=\zeta\left(N^{(2)}\varphi^{(3)}\right) \quad (19)$$

Where $N^{(2)}$ and $N^{(3)}$ represent the output of layer 2 and layer 3 of the network, that is, the output layer and output sequence; $\varphi^{(2)}$ and $\varphi^{(3)}$ represent the weight coefficients of layers 2 and 3. The data conversion layer converts the graph data into a forecasted PV power series.

In traditional GCN networks, the static adjacency matrix is typically generated using geographic location data and the topological layout. Given that PV output power is highly uncertain and volatile, based on the PV output sequence from the previous day, this paper proposes constructing a dynamic adjacency matrix to establish a DGCN network. The calculation formula of the dynamic adjacency matrix is presented as in (20).

$$C(i,j)_l=1-\frac{\mathrm{d}\left(Y_{i,l-1},Y_{j,l-1}\right)}{\max\limits_{m,k\in[1,n]}\mathrm{d}\left(Y_{m,l-1},Y_{k,l-1}\right)} \quad (20)$$

Where $C(i,j)_l$ is an adjacency matrix, l here means the matrix is adopted in the l-th sample day, i and j means the element in row i and column j, thus $C(i,j)_l$ as a whole means the correlation of the number i and j PV units in the l-th sample day geographically. Y is an eigenmatrix; $\mathrm{d}\left(Y_{i,l-1},Y_{j,l-1}\right)$ represents the Euclidean distance of the output sequence. n represents the number of PV units.

Based on the adjacency matrix C calculated above, the DGCN model constructs convolution layer, which considers the dynamics of the graph structure. The basic formula of DGCN is as in (21).

$$N_t^{(1)}=\zeta\left(\tilde{G}_t^{-\frac{1}{2}}E_t\tilde{G}_t^{-\frac{1}{2}}N_t^{(1)}\varphi^{(1)}\right) \quad (21)$$

Where $\widetilde{G_t}:=\sum_j \widetilde{E_{ij,t}}$ represents degree matrix at time step t, $N_t^{(1)}$ represents the output of layer 1 of the neural network at time step t.

In DGCN, the structure of the graph can change over time, that is, the adjacency matrix C can have different values at different time steps. DGCN handles these dynamics by applying GCN at each time step, while considering the time series information, thus the correlation of adjacent PV output is accurately described.

B. Gated Cycle Unit

Ultra-short-term PV output exhibits strong temporal correlations, output is characterized by strong temporal correlations, meaning that the power output at a given moment is closely related to the outputs in the immediate preceding moments. This temporal dependency is crucial for accurate forecasting, as it allows models to capture the short-term dynamics of PV power generation. However, GCN alone struggles to achieve high-precision time series forecasting results due to their limited ability to model temporal dependencies. To address this limitation, this paper proposes a high-precision forecasting framework that integrates both spatial and temporal correlations of large-scale distributed PV output. This framework is based on dynamic graph convolutional networks, which extend the capabilities of traditional GCNs by incorporating dynamic elements that can adapt to changing conditions over time. For the problem of long time series forecasting, both Long short-term memory (LSTM) and GRU models [25] have been widely recognized for their effectiveness in addressing the issues of gradient disappearance and gradient explosion that are common in traditional recurrent neural networks (RNNs) [26]. These advanced models are designed to handle long-term dependencies and maintain stable gradients during training, which is essential for accurate forecasting. Between these two models, GRU stands out for its simpler structure, fewer parameters, and higher training efficiency. GRU incorporates update gate and reset gate mechanisms to modulate the influence of prior states on the current state and to selectively forget information. Consequently, the GRU model is chosen for time series forecasting in this paper. The detailed calculation process is as follows:

(1) The gated units v_t and q_t, are calculated based on input feature x_t and historical information variable f_{t-1}. The calculation formula of gated unit v_t is as in (22).

$$v_t = \zeta\left(W_v \cdot \left[f_{t-1}, x_t\right] + e_v\right) \quad (22)$$

The calculation of gated unit q_t is as in (23).

$$q_t = \zeta\left(W_q \cdot \left[f_{t-1}, x_t\right] + e_q\right) \quad (23)$$

Where, e_v and e_q represent deviation vector of reset unit and update unit respectively, W_v and W_q represent the weight matrixes, and $\zeta(\cdot)$ denotes activation function.

(2) Calculate hidden state s_t, as in (24).

$$s_t = \tan h\left(W_s \cdot \left[v_t \cdot f_{t-1}, x_t\right] + e_s\right) \quad (24)$$

Where W_s is a corresponding weight matrix of units in hidden state, and e_s is also the deviation vector.

(3) Based on the above results, calculate the output sequence of the GRU network, as in (25).

$$f_t = \left(1 - q_t\right) \cdot f_{t-1} + q_t \cdot s_t \quad (25)$$

C. Global forecasting framework

The ultra-short-term forecasting framework for massive distributed PVs introduced is composed of three integral components, each playing a crucial role in the overall prediction process. These components include data preprocessing, equipment planning, and the core PV ultra-short-term forecasting module.

In the data preprocessing stage, MVMD is employed to decompose the complex output sequences of multiple distributed PV units into more manageable and interpretable sub-modes, which not only simplifies the data structure but also enhances the clarity of the underlying patterns within the PV output sequences. For the equipment planning module, the clustering feature serves as the basis for this strategic placement, and is derived from the decomposed modes obtained in the previous step. The placement of equipment is determined by K-medoids algorithm. In ultra-short-term forecasting module of PVs, input features are constructed in accordance with the clustering results. Subsequently, the dynamic adjacency matrix is computed, capturing the spatial relationships between the different PV units. Finally, the DGCN-GRU forecasting model is established. The DGCN component effectively captures the spatial correlations between the PV units, while the GRU component handles the temporal dynamics. Together, they provide a powerful tool for forecasting the ultra-short-term output of massive distributed PV systems. After the model is established, network fitting is performed to optimize the model parameters and ensure the best possible prediction performance. The flow chart of forecasting model is shown in Figure 3.

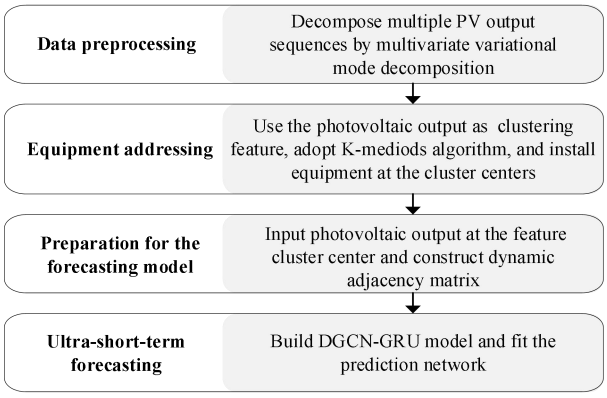


Figure 3. Forecasting model flow chart.

5. Materials and Results

In this part, real PV data is employed to validate the efficacy of the proposed algorithm. This section is primarily divided into three subsections. Section A provides a detailed description of the datasets and data preprocessing utilized. Section B illustrates the modal decomposition outcomes of the PV data and the results of equipment placement. Section C showcases the results of the PV ultra-short-term forecasting.

A. Dataset Description

Based on the historical operation information of 10 distributed PVs in Xiangyang, Hubei Province, this paper uses PV output data recorded every ten minutes from August 2, 2024 to October 31, 2024. The dataset is partitioned such that the initial 75% serves as the training set, while the remaining 25% functions as the test set. During the training phase of the fundamental PV forecasting model, given the intermittent nature of PV output, only power data from 6:00 to 19:00 local time were utilized, meaning each sample day comprised 84 data points. The original input features were characterized by a power sequence from 2 hours prior to each distributed PV forecasting moment, encompassing 12 data points. Following the decomposition of the original data into sub-modes, these sub-modes were employed as clustering features, and the K-medoids algorithm was applied to select the locations for real-time communication equipment installation. The data transmission system was situated at the cluster center. The output information from the PV units at the cluster center was then selected to form the input features for the forecasting model, enabling the ultra-short-term forecasting of distributed PV power. Both the input features and PV output were normalized to mitigate the influence of installed PV capacity on the forecasting model. This paper employs three indexes as metrics to assess forecasting accuracy, namely mean squared error (MSE) [27], root mean square error (RMSE) [28], mean absolute error (MAE) [29], and R-squared value(R^2) [30]. I_{RMSE} is the index of RSE, as in (26).

$$I_{\mathrm{RMSE}}=\sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(R_s-Y_s\right)^2} \quad (26)$$

I_{MAE} is the index of MAE, as in (27).

$$I_{\mathrm{MAE}}=\frac{1}{S}\sum_{s=1}^{S}\left|R_s-Y_s\right| \quad (27)$$

I_{MSE} is the index of MSE, as in (28).

$$I_{\mathrm{MSE}}=\frac{1}{S}\sum_{s=1}^{S}\left(R_s-Y_s\right)^2 \quad (28)$$

Also, I_{R^2} is the index of R^2 value, as in (29).

$$I_{R^2}=1-\frac{\sum\left(R_s-Y_s\right)^2}{\sum\left(R_s-\bar{R}_s\right)^2} \quad (29)$$

Where in the four formulas of calculating index, S is the length of time series, or time steps of forecasting, Y_s indicates the forecasting output, and R_s is the data recorded.

B. Decomposition and Clustering

MVMD algorithm is used to pre-process the PV output sequence with sub-modes setting to 4. The mode decomposition effect for PV0 (i.e., the distributed PV unit labeled as 0) is illustrated in Figure 4.

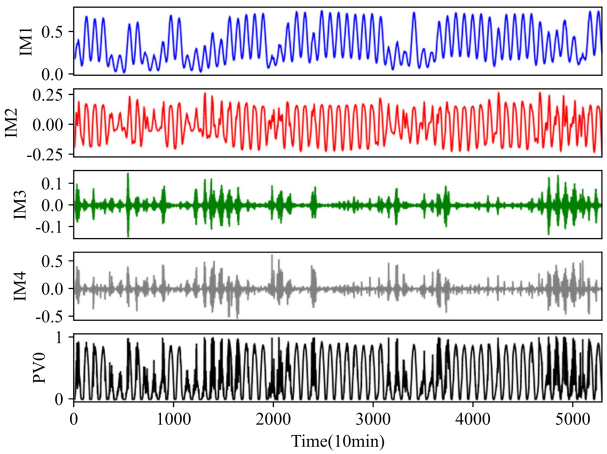


Figure 4. Mode decomposition results of PV0.

As observed from the Figure 4, the original data exhibits high complexity and significant uncertainty. MVMD can decompose it into sub-modes with varying center frequencies, resulting in more stable sequences. This decomposition enhances the efficiency of subsequent processes, such as equipment placement and PV forecasting modeling. The quantity of real-time data transmission systems is influenced by multiple factors, including equipment installation costs and forecasting accuracy requirements. In this study, 4 clustering centers with real-time communication equipment installed are chosen, so the number of equipment is determined as 4.

The cluster centers are determined by K-medoids algorithm. The ultra-short-term forecasting of distributed PVs are then carried out based on the real-time data collected. Using the silhouette index, the influence of change in number of clustering centers on the clustering effect is quantitatively compared. As shown in Figure 5, the closer the silhouette index is to 1, the better the clustering effect. It is optimal within the range of a small number of communication equipment when using K-medoids clustering method to select 3 centers. Also, using the clustering method has a better effect than randomly choosing the placement of communication equipment. Ultimately the identified clustering centers in this experiment are: PV3, PV5, and PV9.

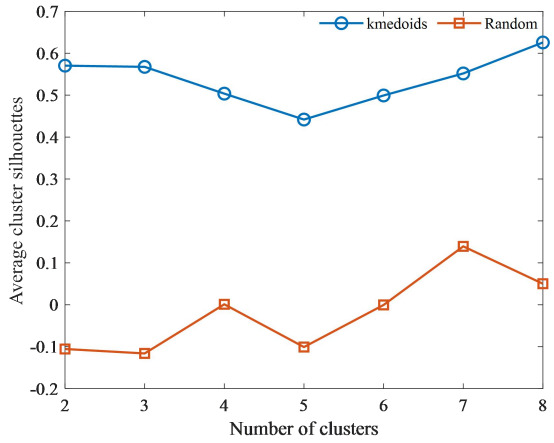


Figure 5. Quantitative analysis results of clustering effect.

C. Ultra-Short-Term Forecasting

In the DGCN-GRN model built in this paper, the DGCN and GRU components are set to 1 layer, with 256 nodes set in each layer separately. The learning rate is 0.001, the batch size is 84, the epochs is 200, the validation set ratio is 0.2, the discard rate is 0.3, and the solution uses the Adam optimizer. In this paper, the multi-layer perceptron (MLP), DGCN-MLP, plain GRU and LSTM are utilized as benchmark methods for comparison with DGCN-GRU.The MLP model features a neural network with two hidden layers, DGCN-MLP represents a cascaded network structure combining DGCN and MLP, whereas plain GRU and LSTM are commly-used baseline methods. The overall forecasting results for large-scale distributed PV output are presented in Table 1. The distribution of forecasting errors are illustrated in Figure 6. From these results, the following observations can be made:

(1) The proposed method achieves the highest forecasting accuracy among the three comparison methods, demonstrating superior performance under all three-accuracy metrics. Specifically, the RMSE value of 0.0845 indicates that our forecasting method significantly enhances the accuracy of ultra-short-term forecasting under the background of massive distributed PVs.

(2) DGCN-MLP outperforms the MLP model in forecasting accuracy, suggesting that incorporating the spatial correlation of distributed PV output effectively improves forecasting precision.

Table 1. Forecasting results of different models.

METHOD	Accuracy parameter			
	RMSE(p.u.)	MSE(p.u.)	MAE(p.u.)	R^2 (p.u.)
MLP	0.0972	0.0094	0.0513	0.6454
DGCN-MLP	0.0929	0.0086	0.0840	0.8967
plain GRU	0.1018	0.0104	0.0554	0.8852
LSTM	0.1119	0.0125	0.0629	0.8593
Our method	**0.0845**	**0.0071**	**0.0389**	**0.9063**

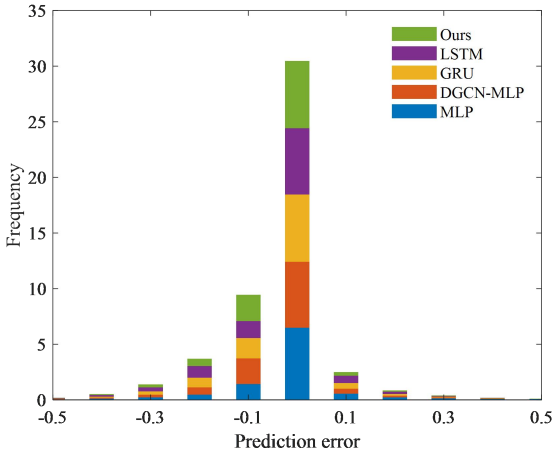
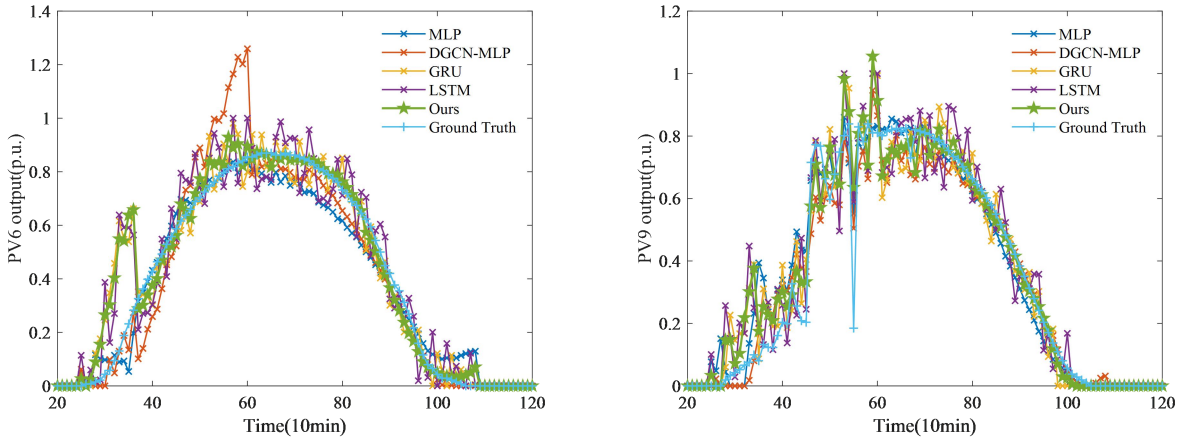


Figure 6. Distribution of forecasting errors of different methods.



(a) PV6 output

(b) PV9 output

Figure 7. PV forecasting output.

Figure 7 illustrates the ultra-short-term forecasting performance for PV6, which is equipped with a data transmission system, and PV9, which is a typical PV unit without such equipment. This comparison more vividly highlights the effectiveness of the proposed algorithm. As shown in Figure 7, the following conclusions can be drawn:

(1) By fully accounting for both temporal and spatial correlations in large-scale distributed PV output, the forecasting model presented in this paper outperforms the other two comparative models in terms of forecasting accuracy;

(2) PV6, which is fitted with real-time data transmission equipment, can precisely capture the trends in PV output changes, resulting in high forecasting accuracy;

(3) This paper does not incorporate meteorological forecasts as input features. Instead, it relies solely on feedback data from nearby PV data transmission points to forecast the output of ordinary PV units.

Consequently, the forecasting accuracy is influenced by the proximity of these data transmission points, and the model struggles to accurately forecast short-term fluctuations in PV output. Furthermore, according to the forecasting results, it can be seen that:

(1) PV3, PV5, and PV9, which are equipped with real-time data transmission systems, enable the forecasting model to capture the status and trends of PV output more effectively. As a result, their forecasting accuracy is significantly higher than that of other PV units lacking real-time monitoring;

(2) Among PV units with installed data transmission system, PV4 has the highest forecasting error (RMSE=0.0345), while PV9 shows the lowest forecasting error (RMSE=0.0593). Among PVs without real-time information transmission system, PV6 achieves the highest forecasting accuracy, whereas PV8 has relatively low forecasting accuracy.

Therefore, it can be concluded that the effectiveness of PV forecasting is influenced by several factors, including judging the necessity of installing real-time data transmission equipment, the inherent characteristics of the PV unit, and meteorological conditions of different cases.

Furthermore, by gradually removing or modifying the components in the framework and observing the impact of these changes on the overall performance, it is verified that the overall performance of the framework proposed in this paper is superior. The following experimental groups are designed: The complete framework; Group 1: without the component of MVMD data decomposition; Group 2: the placement of communication equipment is randomly determined rather than by K-medoids clustering; Group 3: using GCN-GRU network model instead of DGCN-GRU model. The forecasting results of these experiments are shown in Table 2.

Table 2. Forecasting results of ablation experiments.

EXPERIMENT	Accuracy parameter			
	MSE(p.u.)	MAE(p.u.)	R^2 (p.u.)	Time(s)
Complete framework	**0.0071**	**0.0389**	0.9063	31.7093
Group1	0.0093	0.0496	0.9055	31.7307
Group2	0.0081	0.0606	**0.9197**	31.7870
Group3	0.0100	0.0773	0.9032	**31.6103**

It can be seen from Table 2 that the complete framework scores the highest in MSE and MAE, Group 2 performs the best in the R-square value, and Group 3 has the shortest running time. However, the R-square value of the complete framework is close to that of Group 2, and the running time of the complete framework is not much different from that of Group 3. This indicates that the complete framework has good operating efficiency while maintaining a high forecasting accuracy. When comprehensively considering all performance indicators, the PV output forecasting framework proposed in this paper, combining the MVMD sequence decomposition method, the K-medoids clustering method and the DGCN-GRU forecasting model, shows the superiority of its overall performance, achieves a balance between performance and operating time, and verifies the effectiveness of each component. This indicates that it is a solution that is both applicable and efficient in practical applications.

6. Conclusion and Discussion

This paper proposes a comprehensive framework for addressing the challenge of ultra-short-term forecasting of large-scale distributed PV systems, encompassing both data transmission equipment placement and PV forecasting. Given the constraints posed by the costs associated with data acquisition and transmission equipment, PV manufacturers typically have the capacity to install communication equipment in only a limited number of PV units to aid in PV forecasting. This paper operates under the assumption that the quantity of communication equipment is fixed. Firstly, MVMD algorithm is employed to process multi-dimensional PV sequences, effectively separating distinct modes while preserving the output correlation between adjacent power stations. Then, utilizing the PV output sequence as the clustering feature, the K-medoids algorithm is applied to cluster large-scale distributed PV units. The cluster centers are designated as real-time data transmission points, thereby completing the selection of equipment locations. Finally, the DGCN-GRU model is used to analyze distributed PV output by obtaining the temporal and spatial correlations of them, so that the accuracy of ultra-short-term PV forecasting can be enhanced.

Future work can be carried out around the following contents. First, determine the optimal number of communication equipment to strike a balance between the accuracy of PV forecasting and the economic viability of system operation; Then, develope an integrated system for equipment location selection and PV forecasting. In this paper, the tasks of location selection and forecasting are treated separately. Integrating these tasks into a cohesive system may have the potential to further enhance model performance.

Acknowledgement

This work is supported by Science and Technology Project of State Grid Hubei: Research and Application of Forecasting and Aggregation Control Technology for Low-Voltage Distributed Photovoltaic Clusters (B31543238466).

References

[1] The International Renewable Energy Agency (IRENA), Renewable Energy Statistics 2024, International Renewable Energy Agency, Abu Dhabi. ISBN: 978-92-9260-614-5

[2] J.G. Kassakian, R. Schmalensee, G. Desgroseilliers, T.D. Heidel, K. Afridi, A.M. Farid, et al. The future of the electricity grid: an interdisciplinary MIT study, Cambridge, MA, Tech. ISBN: 978-0-9828008-6-7

[3] A. Baviskar, K. Das, M. Koivisto, A.D. Hansen. Multi-Voltage Level Active Distribution Network With Large Share of Weather-Dependent Generation. IEEE Transactions on Power Systems, 2022, 37(6), 4874-4884. DOI: 10.1109/TPWRS.2022.3154613

[4] Z. Fang, Y.Z. Lin, S.J. Song, C. Li, X.F. Lin, Y.B. Chen. State estimation for situational awareness of active distribution system with photovoltaic power plants. IEEE Transactions on Smart Grid, 2021, 12(1), 239–250. DOI: 10.1109/TSG.2020.3009571

[5] C. Lyu, S. Eftekharnejad. Probabilistic Solar Generation Forecasting for Rapidly Changing Weather Conditions. IEEE Access, 2024, 12, 79091-79103. DOI: 10.1109/ACCESS.2024.3407778

[6] C.Z. Huang, M.Y. Yang. Memory long and short term time series network for ultra-short-term photovoltaic power forecasting. Energy, 2023, 279, 127961. DOI: 10.1016/j.energy.2023.127961

[7] X. Zhu, R. Ju, X. Cheng, Y. Ding, H. Zhou. A very short term prediction model for photovoltaic power based on numerical weather prediction and ground-based cloud images. Automation of Electric Power Systems, 2015, 39(6), 4-10. DOI: 10.7500/AEPS20140409004

[8] J.X. Liu, H.X. Zang, L.L. Cheng, T. Ding, Z.N. Wei, G.Q. Sun. A Transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting. Applied Energy, 2023, 342, 121160. DOI: 10.1016/j.apenergy.2023.121160

[9] J.C. Yan, L. Hu, Z. Zhen, F. Wang, G. Qiu, Y. Li. Frequency-domain decomposition and deep learning based solar PV power ultra-short-term forecasting model. IEEE Transactions on Industry Applications, 2021, 57(4), 3282-3295. DOI: 10.1109/TIA.2021.3073652

[10] Y.X. Jiang, L.W. Zheng, X. Ding. Ultra-short-term prediction of photovoltaic output based on an LSTM-ARMA combined model driven by EEMD. Journal of Renewable and Sustainable Energy, 2021, 13(4), 046103. DOI: 10.1063/5.0056980

[11] J.Z. Wang, Y. Yu, B. Zeng, H.Y. Lu. Hybrid ultra-short-term PV power forecasting system for deterministic forecasting and uncertainty analysis. Energy, 2024, 288, 129898. DOI: 10.1016/j.energy.2023.129898

[12] H. Xiao, W.T. Zheng, H. Zhou, W. Pei. Ultra-short-term photovoltaic power prediction based on improved temporal convolutional network and feature modeling. CSEE Journal of Power and Energy Systems, 2024, 1-12. DOI: 10.17775/CSEEJPES.2023.06130

[13] F. Wang, X.X. Lu, S.W. Mei, Y. Su, Z. Zhen, Z.B. Zou, et al. A satellite image data based ultra-short-term solar PV power forecasting method considering cloud information from neighboring plant. Energy, 2022, 238, 121946. DOI: 10.1016/j.energy.2021.121946

[14] H.X. Zang, D.H. Chen, J.X. Liu, L.L. Cheng, G.Q. Sun, Z.N. Wei. Improving ultra-short-term photovoltaic power forecasting using a novel sky-image-based framework

considering spatial-temporal feature interaction. Energy, 2024, 293, 130538. DOI: 10.1016/j.energy.2024.130538

[15] J.K. Liang, W.Y. Tang. Ultra-short-term spatiotemporal forecasting of renewable resources: An attention temporal convolutional network-based approach. IEEE Transactions on Smart Grid, 2022, 13(5), 3798-3812. DOI: 10.1109/TSG.2022.3175451

[16] G. Wang, S.Y. Sun, S.Y. Fan, Y.N. Liu, S.X. Cao, R.Q. Guan. A spatial–temporal data-driven deep learning framework for enhancing ultra-short-term prediction of distributed photovoltaic power generation. International Journal of Electrical Power & Energy Systems, 2024, 160, 110125. DOI: 10.1016/j.ijepes.2024.110125

[17] K. Dragomiretskiy, D. Zosso. Variational mode decomposition. IEEE Transactions on Signal Processing, 2013, 62(3), 531-544. DOI: 10.1109/TSP.2013.2288675

[18] N. Rehman, H. Aftab. Multivariate variational mode decomposition. IEEE Transactions on Signal Processing, 2019, 67(23), 6039-6052. DOI: 10.1109/TSP.2019.2951223

[19] R.A. Jarvis, E.A. Patrick. Clustering Using a Similarity Measure Based on Shared Near Neighbors. IEEE Transactions on Computers, 1973, C-22 (11), 1025-1034. DOI: 10.1109/T-C.1973.223640

[20] D. Butina. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. Journal of Chemical Information and Computer Sciences, 1999, 39(4), 747-750. DOI: 10.1021/ci9803381

[21] H.S. Park, C.H. Jun. A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications, 2009, 36(2), 3336-3341. DOI: 10.1016/j.eswa.2008.01.039

[22] C. Boutsidis, A. Zouzias, M.W. Mahoney, P. Drineas. Randomized Dimensionality Reduction for k -Means Clustering. IEEE Transactions on Information Theory, 2015, 61(2), 1045-1062. DOI: 10.1109/TIT.2014.2375327

[23] G.C. Cardarilli, L. Di Nunzio, R. Fazzolari, A. Nannarelli, M. Re, S. Spano. N-Dimensional Approximation of Euclidean Distance. IEEE Transactions on Circuits and Systems. II, Express Briefs, 2020, 67 (3), 565-569. DOI: 10.1109/TCSII.2019.2919545

[24] T.N. Kipf, M. Welling. Semi-supervised classification with graph convolutional networks. arXiv Preprint arXiv:1609.02907, 2016. DOI: 10.48550/arxiv.1609.02907

[25] J.Y. Chung, C. Gulcehre, K.H. Cho, Y.S. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555, 2014. DOI: 10.48550/arxiv.1412.3555

[26] N. Mohajerin, S.L. Waslander. Multistep Prediction of Dynamic Systems With Recurrent Neural Networks. IEEE Transaction on Neural Networks and Learning Systems, 2019, 30 (11), 3370-3383. DOI: 10.1109/TNNLS.2019.2891257

[27] H.L. Tan, Z.G. Li, Y.H. Tan, S. Rahardja, C.H. Yeo. A Perceptually Relevant MSE-Based Image Quality Metric. IEEE Transactions on Image Processing, 2013, 22(11), 4447-4459. DOI: 10.1109/TIP.2013.2273671

[28] L. Mentaschi, G. Besio, F. Cassola, A. Mazzino. Problems in RMSE-based wave model validations. Ocean Modelling (Oxford), 2013, 72, 53-58. DOI: 10.1016/j.ocemod.2013.08.003

[29] J. Qi, J. Du, S.M. Siniscalchi, X.L. Ma, C.H. Lee. On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. IEEE Signal Processing Letters, 2020, 27, 1-1. DOI: 10.1109/LSP.2020.3016837

[30] T. Raykov, L. Calvocoressi. Model Selection and Average Proportion Explained Variance in Exploratory Factor Analysis. Educational and Psychological Measurement, 2021, 81(6), 1203-1220. DOI: 10.1177/0013164420963162