

Article

Power Reconstruction and Quantitative Analysis of Photovoltaic Cluster Fluctuation Characteristics Considering Cloud Movement Time Lag

Gangui Yan ¹, Jianshu Li ^{1,*}, Aolan Xing ² and Weian Kong ²¹ Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, China² School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China

* Correspondence: 13583797569@163.com; Tel.: +86-13583797569

Abstract

The power fluctuation of large-scale photovoltaic (PV) clusters is significantly affected by cloud movement. Aiming at the engineering reality that meteorological observation data are generally lacking for most power stations in wide-area PV clusters, as well as the problem that existing models overfit second-order high-frequency noise such as microscopic cloud deformation, this paper proposes a disturbance reconstruction and smoothing effect quantification method for PV clusters focusing on the first-order dominant meteorological component. First, a clear-sky model is introduced as a deterministic trend filter to extract the purely random disturbance sequence that induces grid-connection risks from the measured output power. Second, the dimensionality reduction modeling concept of “macro-advection dominance and microscopic deformation filtering” is established: the PV cluster is finely partitioned by fusing Dynamic Time Warping (DTW) and geographical distance, and a cross-space inversion of the macro-cloud velocity vector is realized, driven by pure power data using the Time-Lagged Cross-Correlation (TLCC) algorithm, thus constructing a disturbance power generation model that accounts for the phase misalignment of power output. Independent verification based on measured data in Jilin Province shows that the 95% confidence interval of the power reconstructed only by the first-order advection characteristics can cover 90.2% of the measured fluctuations, and the reconstruction error of the fluctuation standard deviation—an indicator that determines the system reserve demand—is merely 5.9%. This verifies that the macro-cloud displacement is the absolute dominant factor governing the extreme fluctuations of PV clusters. Finally, a normalized Smoothing Factor (SF) characterizing the “reserve capacity release ratio” is constructed, and combined with its statistical indicators, it is used to quantitatively evaluate the smoothing benefits provided by different spatial layout schemes. Under data-constrained conditions, the method proposed in this paper verifies the engineering rationality that microscopic meteorological noise can be safely neglected at the macro-PV cluster scale, providing a reliable quantitative basis for the safe grid expansion and peak-shaving energy storage capacity sizing of high-proportion PV bases.

Keywords: PV cluster; fluctuation characteristics; smoothing effect; affinity propagation clustering



Academic Editor: Ahmed Abu-Siada

Received: 4 February 2026

Revised: 27 February 2026

Accepted: 3 March 2026

Published: 11 March 2026

Correction Statement: This article has been republished with a minor correction to the Funding statement. This change does not affect the scientific content of the article.

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

In recent years, with the accelerating deterioration of the ecological environment and the increasing depletion of fossil fuels, solar energy has been widely adopted as a clean and abundant renewable energy source [1]. As the capacity and density of PV in distribution grids continue to increase, the stochasticity and intermittency of PV output bring numerous risks to the secure operation of the grid [2]. Due to the interaction between cloud movement and geographical distribution, the power fluctuations of PV clusters exhibit complex spatiotemporal non-stationarity. Therefore, exploring the evolution law of output fluctuation and the spatial smoothing effect of PV clusters is of significant practical importance for the scientific allocation of peak-shaving resources in regional power grids.

Scholars both domestically and internationally have conducted multi-perspective studies on the regularity of PV output fluctuations. Reference [3] utilized a Self-Organizing Map clustering method to analyze the output curves of PV stations within a region. This finding highlights the weather dependency of the spatial correlation of PV output. Reference [4] constructed several quantitative fluctuation metrics, including average fluctuation intensity, reverse fluctuation frequency, and moving fluctuation density. References [5,6] analyzed PV fluctuations from the perspective of different time scales. To analyze the impact of photovoltaics with fluctuating characteristics on power system operations, References [7,8] employed historical data to conduct time-series production simulations. References [9,10] fitted solar irradiance based on observational results from extensive meteorological data and by modeling various meteorological factors, in turn indirectly obtaining the PV output sequence through the photoelectric conversion relationship. Reference [11] proposed a method to directly generate PV power station output, which is based on Markov chains and models the historical power data of PV stations. Compared to individual power stations, power generation authorities are more concerned with the aggregate output characteristics after PV clusters are integrated. Reference [12] found that as the scale of the PV cluster increases, the volatility of the cluster decreases and the output becomes smoother. This phenomenon, known as the smoothing effect of PV output power, is of great significance for exploring PV grid integration. Although quantitative research on the smoothing effect of PV cluster output is relatively scarce, it can draw upon research on wind power, which is also a new energy source. Reference [13] indicates that the smoothing effect of wind power increases with the scale of the wind farm. Reference [14] explored the power smoothing characteristics of wind power clusters from multiple angles, such as fluctuation intervals, characteristic curves, and spectrum analysis. Reference [15] quantified the relationship between the number of wind farms, regional diameter, and the smoothing effect based on correlation analysis.

To delve into the complex spatiotemporal evolution of PV output, artificial intelligence technologies have advanced rapidly in recent years. Accordingly, advanced deep learning prediction models, including graph-based clustering [16,17] and spatiotemporal graph neural networks [18,19], have been widely applied to the mining of non-stationary time series. These models exhibit excellent fitting performance in capturing micro-meteorological features at the single-station level. However, for large-scale PV clusters, high-frequency microscopic random disturbances such as localized cloud deformation are not only extremely difficult to predict accurately due to their strongly chaotic properties, but also naturally smoothed and offset during spatial aggregation. Attempting to forcibly fit such redundant second-order noise will not only significantly increase the risk of model overfitting, but also deviate from the core concerns of macro-level power grid dispatch. In addition, most PV stations are only equipped with ground meteorological observation stations, making it difficult to obtain upper-air meteorological parameters. Even if the corresponding parameters can be obtained, there exists a complex functional relationship between the ambient

wind speed and the actual moving speed of cloud clusters, leading to an excessively high cost of overall data acquisition and calculation. This makes it difficult for such methods to be widely used in regional PV power generation scenarios.

Furthermore, the existing quantitative evaluation metrics in most studies are mostly absolute physical quantities directly dominated by the total installed capacity (e.g., maximum ramp rate), which makes it difficult to mathematically decouple the “capacity scale effect” from the “spatial topology smoothing benefit”. This limitation restricts current research on the smoothing effect of PV clusters to mostly simple qualitative descriptions, and there is a lack of a quantitative framework that can break through the interference of installed capacity and objectively compare the ability of different geographical siting schemes to suppress the extreme tail risk of the system.

To address the above issues, this paper establishes an engineering dimension-reduction modeling concept of “macro-scale advection dominance and micro-scale deformation filtering”. Based on large-scale measured data from PV clusters, a power reconstruction and fluctuation characteristic quantification framework for PV clusters considering cloud movement time lag is proposed. First, the physical clear-sky model is introduced as a deterministic trend filter to accurately separate the astronomical background component with no grid integration risk. Then, a dual-dimensional metric integrating Dynamic Time Warping (DTW) distance and geographic distance is used to realize fine-grained zoning of heterogeneous PV clusters. On this basis, the Time-Lagged Cross-Correlation (TLCC) algorithm is adopted to invert the macro-scale cloud velocity vector across space, and a disturbance power reconstruction model accounting for the phase misalignment of power output is constructed. Rigorous out-of-sample independent validation results confirm that the high-precision coverage of the core fluctuation interval can be achieved only by capturing the first-order advection characteristics of cloud clusters, which fundamentally overcomes the synchronous resonance error of traditional upscaling methods. Finally, a normalized Smoothing Factor (SF) unaffected by the installed capacity base is proposed, and its physical meaning is rigorously mapped to the “reserve capacity release ratio” of the power grid. By calculating the corresponding statistical indicators, the smoothing benefits brought by the spatial distribution of PV clusters with different installed capacities are quantitatively evaluated, which provides theoretical support for the macro-quota determination of regional reserve capacity and system expansion planning in high-PV-penetration areas.

2. Analysis of PV Station Power Characteristics

2.1. Deterministic Trend Extraction and Random Disturbance Separation

The output power level of a PV station directly depends on the amount of solar radiation it receives. This radiation quantity is primarily governed by two types of factors: one is the astronomical and geographical factor determined by the relative motion of the sun and earth, possessing strict periodicity; the other is the weather factor induced by cloud activity, possessing strong uncertainty [20]. Accordingly, the PV daily output $P_j(t)$ can be theoretically decomposed into two core components:

$$P_j(t) = P_{j,\text{Reg}}(t) + P_{j,\text{Rad}}(t) \quad (1)$$

where: $P_{j,\text{Reg}}(t)$ characterizes the clear-sky output of the daily PV station, representing the theoretical maximum output determined by the station’s geographical location (latitude, longitude, altitude) under ideal cloudless conditions, with its variation following the laws of the earth’s rotation and revolution, presenting a clear deterministic trend of daily and seasonal cycles; $P_{j,\text{Rad}}(t)$ is the disturbance daily output, characterizing the minute-to-hour-

level random fluctuation components introduced by meteorological processes such as cloud movement, aggregation, and dissipation, which cause shading and attenuation of solar radiation. It should be noted that the actual PV output is also affected by factors including aerosol scattering, PV module soiling, temperature gradients, and inverter response delay. However, at the macro-dispatch timescale of 15 min and above, which is the primary focus of power grid operation, factors such as module soiling and temperature present as ultra-low-frequency smooth offsets, while inverter delay and local shading are naturally smoothed as high-frequency microscopic noise during the spatial aggregation of the PV cluster. Therefore, the dual decomposition proposed in this paper serves as an engineering dimension-reduction approach, which directly strips irrelevant frequency bands and accurately locks onto the large-scale advection shading process of cloud layers that poses the greatest threat to the grid's instantaneous ramp events.

Modeling the clear-sky output curve $P_{j,\text{Reg}}(t)$ is key to implementing power decomposition. This paper adopts the PV station power model proposed in [21], which comprehensively considers factors such as solar irradiance, PV panel parameters, and ambient temperature, and effectively characterizes the attenuation effect of weather conditions on solar radiation by introducing the clearness index. Under ideal sunny and cloud-free conditions, the clearness index reaches its upper limit of 0.864 [22]. Once information such as the longitude, latitude, installed capacity, and PV panel tilt angle of the PV station is determined, the clearness index is set to this upper limit to simulate the maximum atmospheric transmittance of solar radiation under ideal cloud-free conditions. Thus, the clear-sky output of any station in the region at any time can be calculated. On this basis, the disturbance component sequence $P_{j,\text{Rad}}(t)$ can be separated by subtracting the theoretical clear-sky output from the measured PV output, as shown in the following equation:

$$P_{j,\text{Rad}}(t) = P_j(t) - P_{j,\text{Reg}}(t) \quad (2)$$

Through the above data preprocessing process, the random disturbance component $P_{j,\text{Rad}}(t)$, which truly dominates the short-term ramp pressure of the power grid, can be extracted. In the subsequent chapters of this paper, all spatiotemporal correlation analysis, clustering division, and power reconstruction modeling are carried out for this disturbance sequence, so as to directly target the physical origin of the power fluctuation of the PV cluster.

2.2. Analysis of Fluctuation Characteristics of Disturbance Components

Figure 1 displays the disturbance output $P_{j,\text{Rad}}(t)$ curves for multiple stations in this region under different weather conditions. As shown, under uniform and stable weather conditions such as sunny and rainy days, the disturbance outputs of various stations exhibit strong synchronization and high spatial correlation. However, under cloudy weather, the station outputs show significant spatiotemporal heterogeneity, with outputs presenting significant fluctuations and the correlation between stations weakening markedly.

The primary cause of this phenomenon lies in the movement of cloud layers and changes in their morphology. However, the formation and evolution of cloud clusters possess significant physical inertia, and changes in their morphology often require a certain time span to manifest [22]. For cloud clusters with a spatial scale of tens of kilometers and an evolution time scale of 15 min, the horizontal advection motion driven by the ambient wind field is the dominant process, while the morphological deformation and splitting of cloud clusters are secondary processes with relatively weak impacts. Therefore, the dominant factor causing differences in disturbance output between adjacent stations is the time lag effect induced by cloud movement. This characteristic of “similar waveform but shifted phase” results in a significant weakening of spatial correlation under direct observation.

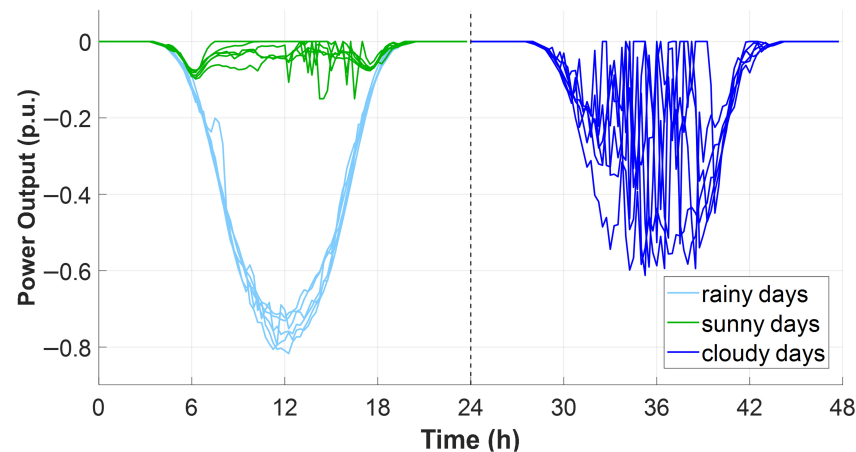


Figure 1. Disturbance output of stations under different weather conditions.

To quantitatively assess the disturbance output characterized by time delays caused by cloud movement, this paper introduces the TLCC coefficient as an evaluation metric. The TLCC refers to the correlation coefficient calculated between one time series after lagging by k periods and another time series. Its calculation formula is as follows:

$$\rho(k) = \frac{\sum_{t=t_0}^n (p_{e,i}(t-k) - \bar{p}_{e,i})(p_{e,j}(t) - \bar{p}_{e,j})}{\left[\sum_{t=t_0}^n (p_{e,i}(t-k) - \bar{p}_{e,i})^2 \cdot \sum_{t=t_0}^n (p_{e,j}(t) - \bar{p}_{e,j})^2 \right]^{1/2}} \quad (3)$$

where: $p_{e,i}(t)$ and $p_{e,j}(t)$ are two different time series, i and j , respectively; n is the length of the time series; $\bar{p}_e$ is the mean value of the error time series; t_0 is the start time of the time series; and $\rho(k)$ is the cross-correlation coefficient of the two time series at lag k , with values ranging between -1 and 1 .

To quantitatively analyze the degree of uncertainty impact caused by cloud layers and eliminate the interference of diurnal trends on correlation analysis, the ratio of disturbance output to clear-sky output is defined as the relative disturbance output $\omega_j(t)$ to effectively distinguish fluctuation differences among different stations:

$$\omega_j(t) = \frac{P_{j,\text{Rad}}(t)}{P_{j,\text{Reg}}(t)} \quad (4)$$

By calculating the maximum TLCC coefficient and its corresponding time lag for the relative disturbance output $\omega_j(t)$ between any two stations, the direction and speed of cloud movement between them can be inferred, thereby enabling a quantitative analysis of the cloud movement path.

Figure 2 illustrates the TLCC coefficients of the relative disturbance output sequences between Station A and Stations B and C, located at different distances within the region. From the perspective of spatial distribution, only the sequence between Station A and the closer Station B shows significant lagged correlation. This result further reflects that when cloud layers propagate over short distances, their movement direction and speed are relatively stable, resulting in high lagged correlation between adjacent station output sequences. However, as the distance between stations increases, the movement state and morphology of cloud clusters are more prone to change during long-distance propagation. Consequently, even when considering time lag, it remains difficult to maintain a high correlation between station output sequences.

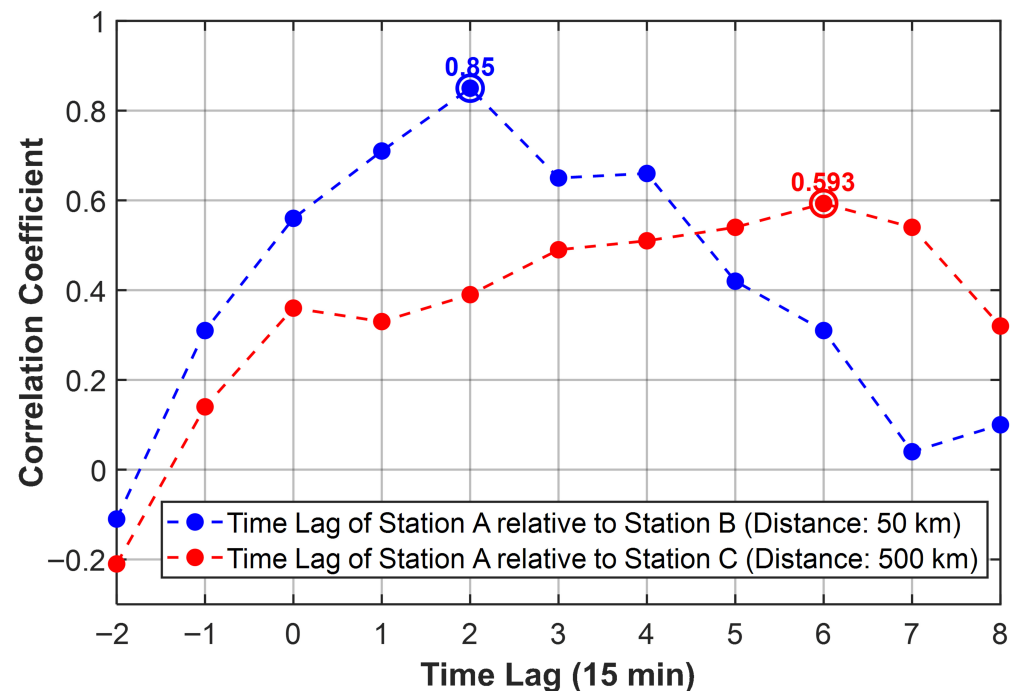


Figure 2. TLCC coefficients of disturbance power between stations at different distances.

The above analysis indicates that to generate the disturbance power of the PV cluster more accurately, a reasonable sub-region division must be carried out for the widely distributed PV cluster. The core purpose of the division is to group PV stations that are dominated by the same cloud movement process and share the common characteristic of “similar waveform with phase misalignment” in disturbance output, and to fully consider the core impact of cloud movement in the disturbance power modeling of each sub-region.

3. PV Cluster Disturbance Power Reconstruction Model Considering Cloud Movement Correction

Through the characteristic analysis in Chapter 2, it is known that when evaluating the total fluctuation power of a PV cluster, the traditional capacity-proportional upscaling reconstruction method inherently implies an extreme physical assumption: “all stations in the region fluctuate synchronously and instantaneously.” This assumption holds basically under uniform and stable weather conditions, with negligible model errors. However, under cloudy weather conditions, this approach artificially erases the internal smoothing effect of the cluster, leading to false sharp jumps in the reconstructed aggregated power.

To address this issue and accurately characterize the dominant component of the cluster’s output smoothing benefit, this paper proposes a “coordinated partitioning and stepwise reconstruction” method for generating cluster disturbance power. First, based on the similarity of power fluctuation patterns and geographical proximity, large-scale PV stations are divided into several sub-regions with similar fluctuation characteristics, which are dominated by the same cloud movement process. Subsequently, within each sub-region, a cloud movement model is introduced to quantify the time lag effect between stations. The disturbance output sequences of each station are corrected and reconstructed using selected representative stations, and the total power of the sub-region is generated by aggregation. Finally, by superimposing the power of all sub-regions, the total disturbance power of the entire cluster can be generated more accurately.

3.1. Sub-Region Division Based on Affinity Propagation (AP) Clustering with Dual-Dimensional Distance Metric

3.1.1. Derivation of Similarity Matrix S Integrating Dual-Dimensional Distance Metrics

This section treats the PV disturbance output data sequence samples to be clustered as different data points in the AP clustering algorithm, and the optimal cluster formed by each type of data point is regarded as the sub-region division result. For the measured data of n PV stations with a time series length of r in the region, the sample matrix P is established:

$$P = (p_{i,t})_{n \times r} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,r} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,r} \\ \vdots & \vdots & & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,r} \end{bmatrix} \quad (5)$$

where $p_{i,t}$ represents the measured output data of station i at the time stamp t sampling point.

Based on the measured data sample matrix P , the distance matrix H between any two PV station output data is constructed:

$$H_{u,w} = (1 - k_{\text{cdi}}) \times d_{\text{DTW}}(P_u, P_w) + k_{\text{cdi}} \times d(L_u, L_w) \quad (6)$$

where $d_{\text{DTW}}(P_u, P_w)$ represents the DTW distance between the normalized PV disturbance output data sample of the station u and the station w . By allowing non-linear warping and alignment of the time axis, DTW can effectively match fluctuation patterns with phase differences; a smaller distance value indicates more similar fluctuation patterns between the two sequences [23]; $d(L_u, L_w)$ is the Euclidean distance calculated from the normalized latitude and longitude of the two stations. For the Jilin Province region with relatively flat terrain studied in this paper, this two-dimensional distance is sufficient to characterize the spatial topology; if extended to complex mountainous areas with highly heterogeneous terrain, additional normalized correction of the altitude dimension is required. H is a symmetric matrix with main diagonal elements of 0. The Cluster Distance Index (CDI) is calculated through clustering computations performed under different k_{cdi} values, where k_{cdi} is the weight coefficient used to adjust the weight proportion of geographical location similarity. When the CDI reaches its minimum value, the corresponding k_{cdi} value is the optimal weight. The CDI is a commonly used clustering evaluation index that comprehensively considers intra-class compactness and inter-class dispersion. Its calculation formula is as follows:

$$\text{CDI} = \frac{1}{\hat{d}_c} \sqrt{\sum_{k=1}^K \hat{d}_\Omega} \quad (7)$$

where: $\hat{d}_c$ represents the average distance between cluster centers; $\hat{d}_\Omega$ represents the average distance within class K . The similarity matrix S is constructed using the square of the negative Euclidean distance:

$$S_{u,w} = -H_{u,w}^2 \quad (8)$$

By adjusting the weight balance through the adaptive k_{cdi} value, the optimal similarity matrix S integrating the DTW distance of power curves and the geographical Euclidean distance can be obtained, paving the way for the next clustering step.

3.1.2. AP Clustering

AP clustering is an adaptive clustering method based on information propagation. By iteratively updating attractiveness and availability between data points, the algorithm adaptively determines cluster centers, which are all actual data points. The algorithm only

takes the similarity matrix S and preference parameter as inputs, is insensitive to initial values, and has good robustness, making it a commonly used clustering method. Figure 3 shows the flowchart of the algorithm.

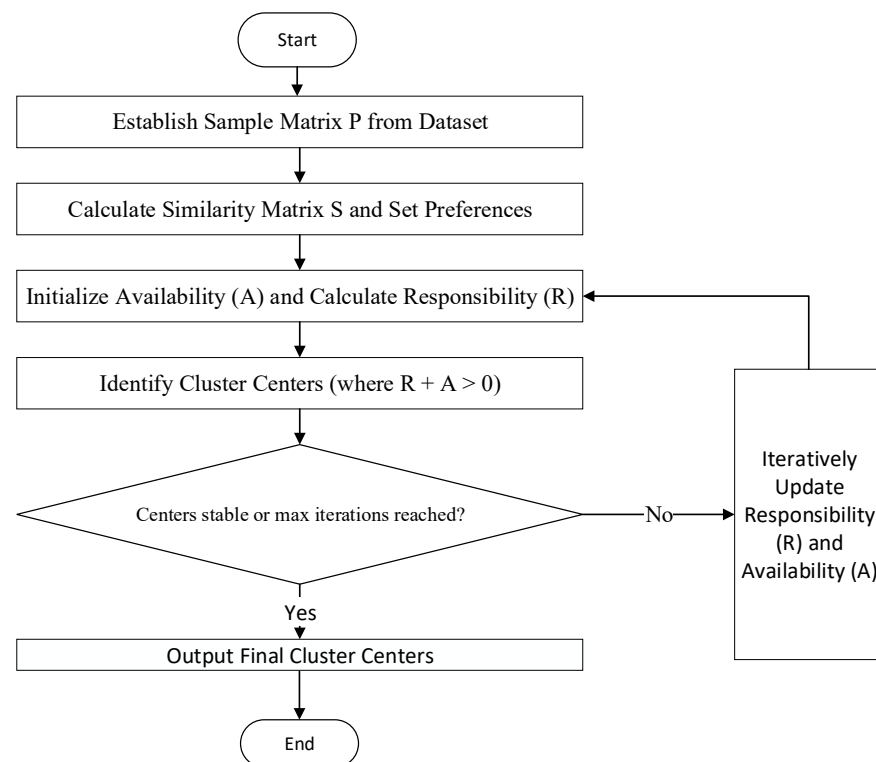


Figure 3. Flowchart of the AP clustering algorithm.

It should be noted that the time complexity of the AP clustering algorithm is quadratic with respect to the sample size. For the provincial regional power grid (with a scale of tens to hundreds of nodes) targeted in this paper, the computational cost can be completed within seconds under conventional computing power; if it is necessary to extend to a wide-area interconnected power grid containing thousands of nodes, coarse-grained pre-clustering or hierarchical strategies can be introduced to ensure the scalability of the algorithm.

Based on the disturbance relative output sequences and corresponding latitude and longitude coordinates of each station, this paper constructs the similarity matrix S , which reflects the similarity between stations. The matrix is then input into the AP clustering algorithm to achieve refined sub-region division of the station cluster. It should be noted that PV output fluctuation characteristics are affected by weather conditions, and sub-region boundaries themselves have a certain degree of fuzziness. The division result in this section can be regarded as the most representative zoning scheme for this region in the sense of historical data statistics, providing a stable and reliable framework for subsequent power generation.

3.2. Generation of Sub-Region Disturbance Power Under Cloudy Events Based on Cloud Movement Correction Algorithm

After completing the sub-region division, the core prerequisite for high-precision generation of regional PV power under cloudy scenarios is to accurately characterize the variation characteristics of cloud clusters. To construct a simplified model that retains the main physical mechanism and has engineering applicability, based on the analysis of the inertia characteristics of cloud cluster motion in Section 2.2, this study makes the following core assumption: within a sub-region with relatively consistent fluctuation characteristics,

the variation trend of macro-scale cloud clusters is approximated as equivalent uniform linear motion. This assumption aims to capture the first-order smoothing effect caused by the overall movement of cloud clusters, while temporarily ignoring the second-order randomness caused by the morphological changes in cloud layers and local noise. According to the law of large numbers in complex systems, this part of the second-order random disturbance will regress to the mean during spatial aggregation. Therefore, the uniform linear motion model is the optimal physical approximation for the “overall characteristics of the PV cluster”, which can capture the core source of the smoothing effect with the lowest model complexity. Based on this assumption, a power correction algorithm based on cloud movement tracking is proposed below. For each sample (a single cloudy event day in a single sub-region), this method obtains a system of linear equations by solving the optimal TLCC coefficients of multiple effective station pairs in the sample. By solving this system of equations, the cloud movement velocity vector is derived to correct the output time series of each power station, thus enabling a more accurate restoration of the spatiotemporal smoothing effect of disturbance power caused by cloud movement. The specific steps are as follows:

1. Calculation of TLCC Coefficients for Station Pairs under Cloudy Scenarios. Calculate the optimal TLCC of the relative disturbance output $\omega_j(t)$ between all station pairs in the sub-region for each sample.

2. TLCC Threshold Selection. The screening of effective station pairs is the core prerequisite for the inversion of cloud movement characteristics, and the threshold of the TLCC coefficient is the key parameter determining sample validity. The threshold selection must simultaneously meet two core principles: first, it can effectively filter out false correlations caused by local random noise and scattered cloud blockage, ensuring that the output fluctuation of the selected samples is dominated by the same large-scale cloud advection process; second, it will not lead to insufficient effective sample size and loss of statistical representativeness due to an excessively high threshold. The setting of this threshold can essentially automatically identify and filter out chaotic scenarios caused by the rapid evolution of cloud systems or overlapping multiple cloud clusters, ensuring that cloud velocity inversion is only performed on reliable samples.

Based on the above principles, this paper establishes a TLCC threshold gradient sensitivity analysis framework to determine the optimal threshold suitable for the target study area: the reasonable interval of strong significant correlation [0.6, 0.8] is selected, and multiple groups of gradient thresholds are set with a step size of 0.05. In the subsequent case study section, a comparative verification of the full gradient thresholds will be carried out with the number of effective samples meeting the rationality requirements and the average width of the 95% confidence interval of the velocity component as the core evaluation indicators, to finally determine the optimal threshold for the region.

3. Selection and Robustness Verification of Effective Station Pairs. Based on the above threshold determination framework, this paper screens effective station pairs in accordance with the following standardized rules to provide a reliable sample basis for the subsequent inversion of cloud movement characteristics:

First, select station pairs whose optimal TLCC coefficient of relative disturbance output is not lower than the regional optimal threshold, and all selected pairs must pass the significance t-test with $p < 0.05$, to ensure that the correlation mainly comes from the same large-scale cloud advection process.

Second, the displacement direction of the selected station pairs must cover at least two quadrants of the sub-region to avoid ill-posed inversion caused by unidirectional sampling.

Finally, if the number of effective station pairs is less than three under this threshold, the TLCC threshold gradient is appropriately relaxed, but the hard constraint of the $p < 0.05$ significance test is retained.

In this section, leave-one-out cross-validation is used to verify the robustness of the cloud velocity inversion results to the selection of station pairs: for a sample containing n effective station pairs, the k -th station pair is sequentially eliminated ($k = 1, 2, \dots, n$), and the remaining $n - 1$ station pairs are used to invert the leave-one-out optimal velocity vector, and a total of n groups of independent inversion results are obtained. By calculating the coefficient of variation in velocity magnitude $CV_{|V|}$ and the maximum direction deviation $\Delta\theta_{\max}$ of all groups, the dispersion degree of the inversion results is quantified, and the high-leverage outliers that play a decisive role in the results are verified and eliminated.

4. Establishment and Verification of Cloud Movement Model

The cloud movement within the sub-region is regarded as a uniform linear motion process, and its velocity can be expressed as

$$V = (V_x, V_y) \quad (9)$$

where: V_x represents the eastward velocity component; V_y represents the northward velocity component.

For the i -th station pair (Station A and Station B) in the sample, if the calculated time lag indicates that the cloud moves from A to B, the displacement vector between them is

$$L = (L_x, L_y) \quad (10)$$

where: L_x represents the eastward displacement component from A to B; L_y represents the northward displacement component from A to B.

Let θ be the angle between the velocity vector and the displacement vector; then the measured projection velocity of the i -th station pair is

$$V_{\text{meas,proj},i} = \frac{|L|}{\Delta t} = |V| \cos(\theta) \quad (11)$$

where $|V| \cos\theta$ is the component of velocity in the direction of displacement.

The dot product of the velocity vector V and the displacement vector L can be expressed as

$$V \cdot L = |V||L| \cos(\theta) = \frac{|L|^2}{\Delta t} \quad (12)$$

Expanded into vector form:

$$V_x \cdot L_{xi} + V_y \cdot L_{yi} = s_i \quad (13)$$

L_{xi} and L_{yi} are the distance vectors between the pair of stations i ; s_i represents the square of the modulus of the distance vector between the pair of stations i divided by the time lag at that time. When there are n groups of effective station pairs, a system of linear equations can be obtained:

$$\begin{cases} V_x \cdot L_{x1} + V_y \cdot L_{y1} = s_1 \\ V_x \cdot L_{x2} + V_y \cdot L_{y2} = s_2 \\ \vdots \\ V_x \cdot L_{xn} + V_y \cdot L_{yn} = s_n \end{cases} \quad (14)$$

Expressed in matrix form:

$$Y = XV \quad (15)$$

When $n > 2$, the system of equations is usually overdetermined, and the least squares method can be used to obtain the optimal solution of the cloud velocity vector V . Then, the theoretical projection velocity in the displacement direction of the i -th station pair is

$$V_{\text{theo,proj},i} = \frac{V \cdot L_i}{|L_i|} \quad (16)$$

To quantitatively verify the rationality of the hypothesis made by the model, this paper establishes a projection residual statistical analysis method to quantify the deviation between cloud cluster motion and the uniform linear motion hypothesis. The specific method is as follows:

To eliminate the influence of different cloud velocity magnitudes on the residuals, the relative projection residual is used to quantify the velocity consistency of a single station pair:

$$\delta = \frac{|V_{\text{meas,proj},i} - V_{\text{theo,proj},i}|}{|V|} \quad (17)$$

For a single cloudy event in a single sub-region, the average relative projection residual is defined as the mean of the residuals of all effective station pairs in the sample, which serves as the core evaluation indicator for the velocity consistency of the sample:

$$\bar{\delta} = \frac{1}{n} \sum_{i=1}^n \delta_i \quad (18)$$

The smaller $\bar{\delta}$ is, the smaller the deviation is between the measured projection velocity of all station pairs and the theoretical projection velocity under the uniform linear motion hypothesis, meaning that the hypothesis is more consistent with the actual cloud movement.

To unify the hypothesis verification scale for all samples, corresponding to the statistical significance of the 95% confidence level of linear fitting, this paper defines samples that simultaneously meet the average relative projection residual $\bar{\delta} \leq 15\%$, the velocity magnitude coefficient of variation $CV_{|V|} \leq 15\%$ from leave-one-out cross-validation, and the maximum direction deviation $\Delta\theta_{\text{max}} \leq 15^\circ$ as valid samples that meet the rationality requirements and conform to the uniform linear motion hypothesis. The selection of the threshold here is an optimal trade-off made based on the conventional tolerance standards of meteorological prediction and power system engineering statistics: in terms of physical geometry, the direction deviation of 15° strictly limits the cloud movement within the half-width of a standard meteorological wind direction sector (30°), ensuring high consistency of the advection trajectory; statistically, the 15% threshold of residual and coefficient of variation can effectively strip the severe nonlinear deformation caused by serious cloud cluster tearing or overlapping of multiple cloud systems, while avoiding the sharp reduction in effective sample size and the loss of macro-statistical representativeness caused by overly strict thresholds (e.g., 5%).

5. Derivation of Sub-region Disturbance Power. After calculating the cloud movement parameters, a representative station within the sub-region is first selected. The station with the average power at the median value is selected as the representative station [24]. Then, based on the geometric relationship between the representative station, the target station, and the cloud velocity vector, the physical time lag can be calculated. The relative disturbance output $\omega_j(t)$ of the representative station is then used as a “template” describing the optical characteristics of the cloud layer, and the “template” is shifted according to the time lag. Finally, the shifted “template” is applied to the clear-sky output of the target station itself to reconstruct its absolute disturbance power $P_{j,\text{Rad}}(t)$, and the total

disturbance power of the sub-region is obtained by accumulating the powers of all stations in the sub-region.

3.3. Uncertainty Quantification of Disturbance Power Reconstruction Model

The inversion uncertainty of the cloud velocity vector is the core source of model uncertainty. In this paper, thus quantification is carried out from two dimensions: the statistical uncertainty of parameter estimation and the inherent uncertainty of model assumptions, and the coupling calculation of the two types of uncertainty is realized.

Based on the least squares estimation results in the previous section, the covariance matrix of the cloud velocity vector V is

$$\text{Cov}(V) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (19)$$

where $\hat{\sigma}^2$ is the unbiased estimation variance of the residuals.

To fully cover the uncertainty of cloud velocity inversion, this paper takes the model uncertainty caused by the uniform linear motion assumption, namely the average relative projection residual, as the scale correction coefficient of the cloud velocity probability distribution. The covariance matrix of the least squares estimation is corrected to obtain the total covariance matrix coupled with full uncertainty:

$$\text{Cov}_{\text{total}}(V) = \left(1 + \frac{\bar{\delta}}{100\%}\right)^2 \cdot \text{Cov}(V) \quad (20)$$

When $\bar{\delta} = 0$, the cloud cluster motion fully conforms to the uniform linear motion assumption, and the total covariance matrix degenerates into a pure statistical covariance matrix; when it increases, the model uncertainty increases, the total covariance matrix is amplified synchronously, and the confidence interval is widened synchronously.

Based on the coupled total covariance matrix, the 95% confidence intervals of the eastward and northward velocity components can be obtained to clarify the statistical uncertainty boundary of cloud velocity inversion:

$$\begin{aligned} V_x &\in \left[\hat{V}_x - t_{0.975, n-2} \cdot \sqrt{\text{Cov}_{\text{total}}(V)_{11}}, \hat{V}_x + t_{0.975, n-2} \cdot \sqrt{\text{Cov}_{\text{total}}(V)_{11}} \right] \\ V_y &\in \left[\hat{V}_y - t_{0.975, n-2} \cdot \sqrt{\text{Cov}_{\text{total}}(V)_{22}}, \hat{V}_y + t_{0.975, n-2} \cdot \sqrt{\text{Cov}_{\text{total}}(V)_{22}} \right] \end{aligned} \quad (21)$$

where $t_{0.975, n-2}$ is the 97.5% quantile of the t-distribution with the degree of freedom $n - 2$, corresponding to the 95% confidence level.

The uncertainty of cloud velocity inversion will propagate to the final reconstruction results of sub-region disturbance power through the phase correction process. This paper adopts the Monte Carlo simulation method commonly used in the power system field. Based on the least squares estimation results, the cloud velocity components V_x and V_y are set to follow a two-dimensional normal distribution, with the mean value being the optimal least squares estimation value and the covariance matrix being the total covariance matrix $\text{Cov}_{\text{total}}(V)$ coupled with full uncertainty calculated by Equation (20), so as to realize the synchronous propagation of statistical uncertainty and model uncertainty. Based on the above probability distribution, $N = 1000$ independent random samplings are carried out to obtain the corresponding cloud velocity vector samples; for each group of cloud velocity samples, the sub-region disturbance output reconstruction process in the previous section is repeated to obtain 1000 independent reconstructed power sequences. The 2.5% and 97.5% quantiles of the empirical distribution of the obtained sequences are calculated

respectively to obtain the 95% confidence interval of the reconstructed power and quantify the uncertainty boundary of the reconstruction results.

4. Quantitative Evaluation of Cluster Power Fluctuation Characteristics

For the cluster disturbance reconstruction model, its fluctuation characteristics are essentially manifested as the smoothing effect caused by the difference in the time sequence of disturbance output of each station due to cloud movement. Therefore, this section adopts the standard deviation of the power time series as the core indicator to measure volatility.

Two types of total power curve standard deviations are defined and calculated:

(1) The actual relative disturbance power, ω_{raw} , is the relative disturbance output obtained by the arithmetic superposition of the disturbance outputs of all stations in the cluster (solid blue line in Figure 4). Its standard deviation, $SD(\omega_{\text{raw}})$, reflects the total system disturbance fluctuation after the smoothing effect has actually occurred.

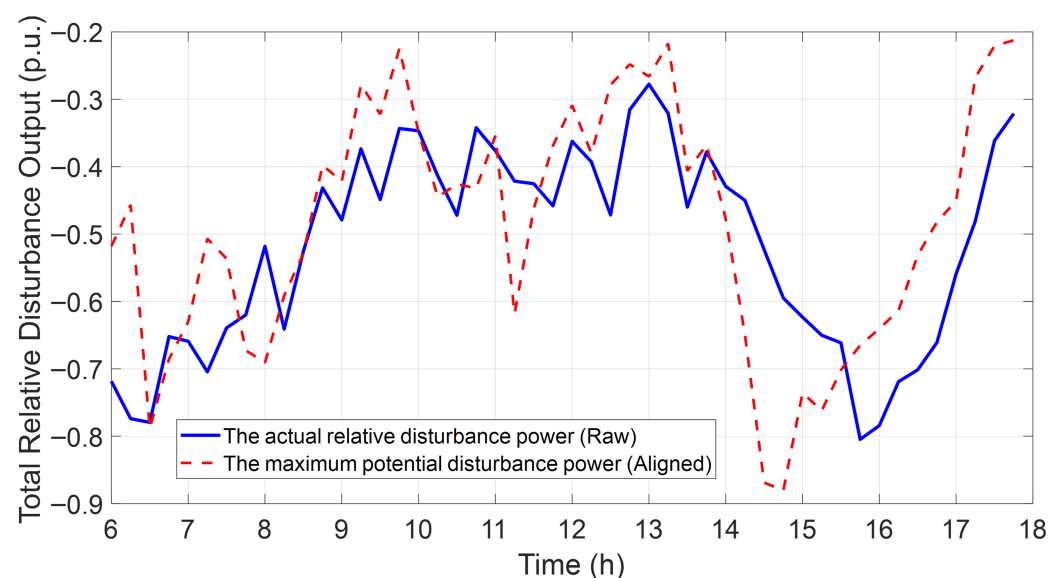


Figure 4. Comparison of actual aggregated disturbance power and maximum potential disturbance power.

(2) The maximum potential disturbance power, ω_{aligned} , is the relative disturbance output obtained after spatiotemporally aligning the disturbance outputs of various stations using the cluster representative station via the cloud velocity vector (dashed red line in Figure 4). This simulates the extreme scenario where a cloud cluster affects all stations in the cluster at the same instant. Its standard deviation, $SD(\omega_{\text{aligned}})$, represents the theoretical maximum fluctuation baseline without any smoothing effect.

Based on the above two indicators, this paper constructs SF to quantitatively evaluate the intensity of the smoothing effect. The calculation formula is as follows:

$$SF = 1 - \frac{SD(\omega_{\text{raw}})}{SD(\omega_{\text{aligned}})} \quad (22)$$

The value of this factor ranges within the interval [0, 1]. The closer the SF value is to 1, the more the actual fluctuation falls below the theoretical maximum fluctuation, proving that the temporal difference in cloud disturbances across space has been effectively canceled out, and the stronger the smoothing effect of the cluster is. Conversely, the closer the SF value is to 0, the more it implies that all stations fluctuate nearly synchronously, indicating a weaker smoothing effect.

In the long-term planning of power systems, the operational envelope of the system is often determined by extreme harsh scenarios. The theoretical maximum fluctuation ω_{aligned} generated by spatiotemporal alignment of cloud velocity vectors in this paper physically simulates the extreme scenario of a cliff-like “co-directional resonance” drop in output caused by the cloud cluster synchronously shielding the entire cluster in an instant. Although the occurrence probability of this scenario is low, it defines the physical upper limit of the theoretical maximum spinning reserve requirement of the regional power grid.

Based on this, the SF defined in this paper is endowed with a clear power grid engineering significance. According to the classic 3σ reserve configuration principle in automatic generation control of power grids, the frequency modulation reserve capacity demand of the system is proportional to the standard deviation of power fluctuation. If the smoothing effect is not considered, the limit reserve capacity R_{max} required by the system is three times $\text{SD}(\omega_{\text{aligned}})$, and the actual reserve demand R_{actual} after smoothing is reduced to three times $\text{SD}(\omega_{\text{raw}})$. Then the reserve capacity saved by the power grid due to the spatial and geographical dispersion of the PV cluster can be derived as follows:

$$\Delta R = R_{\text{max}} - R_{\text{actual}} = 3 \times \text{SD}(\omega_{\text{aligned}}) \times \left(1 - \frac{\text{SD}(\omega_{\text{raw}})}{\text{SD}(\omega_{\text{aligned}})}\right) \quad (23)$$

Combined with Equation (22), it can be obtained that

$$\Delta R = R_{\text{max}} \times \text{SF} \quad (24)$$

It can be seen from this that the physical essence of SF is the “reserve capacity release ratio” generated by the difference in the spatial and geographical distribution of the PV cluster, which means that the power grid dispatching department can safely reduce the reserve capacity or energy storage investment in the corresponding proportion by virtue of the spatial smoothing effect.

Compared with traditional evaluation indicators such as the maximum ramp rate or Variability Index, the SF constructed in this paper has significant analytical advantages. Most traditional fluctuation indicators are absolute physical quantities directly dominated by the total installed capacity base, which makes it difficult to mathematically decouple the “capacity scale effect” from the “spatial topology smoothing benefit” when comparing expansion schemes of different scales. As a normalized relative metric (value range of 0~1), SF can purely and objectively quantify the per-unit capacity smoothing dividend brought by the dispersion of the PV cluster by introducing the limit fluctuation of spatiotemporal alignment as the reference benchmark, thus providing a standardized evaluation scale for comparing the anti-risk ability of different spatial layout scenarios.

Considering the strict requirements of long-term macro-planning of power grids for high reliability, this paper deeply calculates the probability distribution characteristics of SF under historical cloudy scenarios in the region, which provides an intuitive quantitative basis for grid expansion and capacity determination under a high proportion of PV grid integration:

Mean_SF: Reflects the expected value of reserve capacity that can be saved by the cluster on average during long-term operation and represents the average economic benefit of the cluster expansion scheme.

P5_SF: Represents the minimum guaranteed reserve reduction that the cluster can stably provide under the extreme harsh cloudy scenarios with 95% confidence. In the practical macro-capacity sizing planning of reserve energy storage for power grids, constrained by the system security bottom line, planning departments typically make one-time reliable

investments based on the extreme probability boundary. Based on this, the actual reserve capacity that the system ultimately needs to configure can be expressed as

$$R_{invest} = R_{max} \times (1 - P5_SF) \quad (25)$$

This formula indicates that when comparing and selecting expansion scenarios with different geographical topologies, the siting scenario with a higher P5_SF can exempt the power grid from more energy storage construction costs by virtue of its guaranteed smoothing capacity under the worst-case scenarios.

The quantitative evaluation flowchart of the fluctuation characteristics is shown in Figure 5. For a proposed station, determining its sub-region attribution is the key to simulating its output and calculating the cluster smoothing effect. Through the method in Section 3.1, the proposed station can be assigned to an existing sub-region based on geographical coordinates. However, when the location is exactly at the boundary zone of multiple already-divided sub-regions, its zoning possesses inherent uncertainty. Therefore, it should be assumed that this station belongs to each possible sub-region respectively, and the SF of the entire cluster should be calculated.

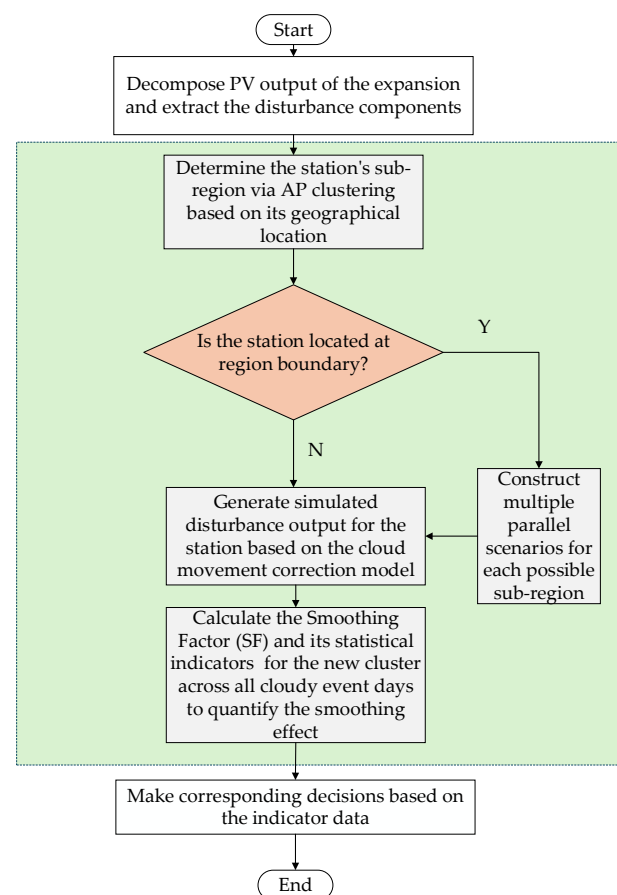


Figure 5. Technical framework for quantitative assessment of PV cluster fluctuation characteristics.

5. Case Study

In this section, the full-year measured power data (sampling interval: 15 min) of 27 PV stations in Jilin Province from 2020 to 2022 are used for verification. The output data from 2020 to 2021 are taken as the training set to obtain the sub-region division results and the reasonable TLCC threshold of the cloud velocity inversion model; the data from 2022 are used as the test set, which is completely independent of the training set, to verify the reconstruction accuracy of the model and the quantification of the smoothing

effect. The geographical distribution of the stations is shown in Figure 6. This region has significant spatial heterogeneity caused by cloud movement, which is suitable for verifying the modeling method proposed in this paper.

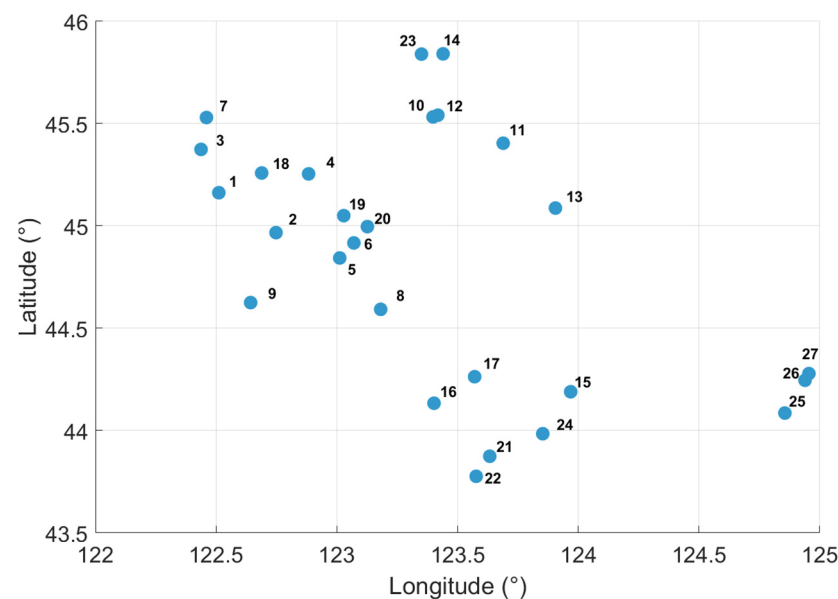


Figure 6. Geographical distribution of the 27 PV stations in the study area.

5.1. Validation of Sub-Region Division and TLCC Threshold Sensitivity Analysis

5.1.1. Evaluation of Sub-Region Division Effect

The rationality of sub-region division is the basis for subsequent time lag correction. To verify the effectiveness of the proposed method, this paper selects the training set of all stations in the region for simulation analysis and designs the following three sub-region division methods for comparison:

Method 1: Perform AP clustering using the fusion of latitude/longitude and output sequences of distributed PV stations in the region (the proposed method).

Method 2: Perform AP clustering using the latitude/longitude of distributed PV stations in the region ($k_{cdi} = 1$).

Method 3: Perform AP clustering using the output sequences of distributed PV stations in the region ($k_{cdi} = 0$).

This paper adopts the Davies–Bouldin Index (DBI) and Silhouette Coefficient (K) as evaluation metrics for clustering performance. A smaller DBI value indicates better clustering results; a Silhouette Coefficient closer to 1 represents better cohesion and separation within the clusters.

From the data in the comparison Table 1, it can be seen that the sub-region division result of Method 1 is the best. This aligns with the fact that distributed PV stations within the same PV system are relatively clustered and possess similar output characteristics. The division results are shown in Figure 7 below.

Table 1. Comparison of clustering performance metrics for different division methods.

Method	DBI	K
Proposed Method	0.783	0.777
AP Clustering using Geographical Coordinates	0.852	0.711
AP Clustering using Output Sequences	1.195	0.584

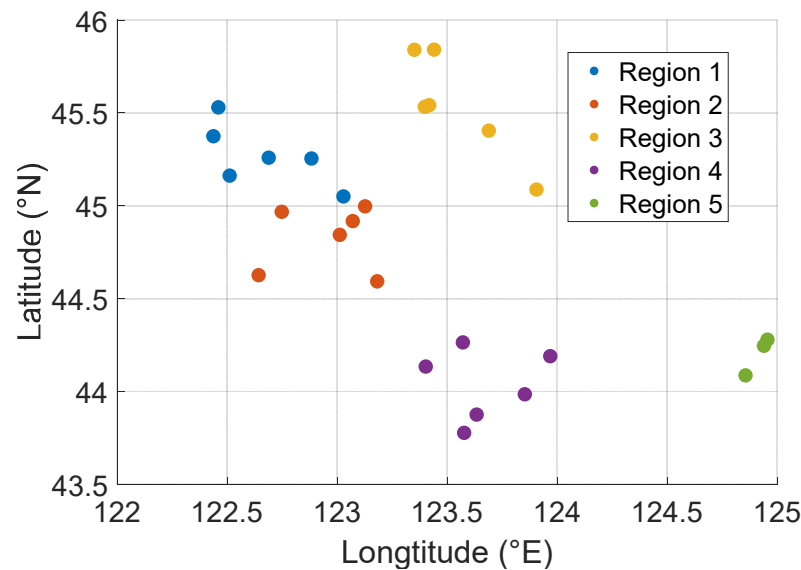


Figure 7. Result of PV cluster sub-regional division based on AP clustering.

5.1.2. TLCC Threshold Sensitivity Analysis and Validation of Uniform Linear Motion Hypothesis

In this paper, all sub-regions divided in the study area are taken as the research objects, and the statistical samples cover all effective cloudy events in the training set of each sub-region. Full-sample sensitivity analysis is carried out for five groups of TLCC gradient thresholds [0.6, 0.65, 0.7, 0.75, 0.8]. Effective station pairs are selected according to different thresholds, and the average width of the 95% confidence interval of the cloud velocity component and all valid samples meeting the rationality requirements are calculated. The statistical results are shown in Table 2.

Table 2. Full-sample statistical results under different TLCC thresholds.

TLCC Threshold	Average Width of 95% Confidence Interval of Velocity Component (km/h)	Proportion of Valid Samples Meeting Rationality Requirements (%)
0.6	4.18	70.5
0.65	3.52	74.3
0.7	3.01	84.2
0.75	3.61	71.6
0.8	4.79	68.7

Full-sample statistical results demonstrate that a TLCC threshold of 0.7 for the study area yields the narrowest 95% confidence interval of cloud velocity and the highest proportion of valid samples meeting rationality criteria, thus securing well-defined uncertainty boundaries and favorable statistical stability for the inversion results. Table 3 shows the full-sample statistical results of all sub-regions under this threshold.

It can be seen from the table that more than 80% of the samples in this region are valid under the threshold of 0.7, which proves that the uniform linear motion hypothesis of cloud clusters has sufficient statistical rationality under the spatial scale of the sub-regions divided in this paper and the 15 min time resolution. For Sub-region 6 with a small number of stations, the inversion stability is weaker than that of other sub-regions with more stations due to the limited number of effective station pairs. After adopting the gradient threshold relaxation scheme, the proportion of valid samples reaches 75%, which proves

that this adaptive scheme can effectively ensure the rationality of the inversion results for small-scale sub-regions.

Table 3. Full-sample statistical validation results of the uniform linear motion hypothesis.

TLCC Threshold	Proportion of Valid Samples Meeting Rationality Requirements (%)
Sub-region 1	88.1
Sub-region 2	84.6
Sub-region 3	91.1
Sub-region 4	82.0
Sub-region 5	85.0
Sub-region 6	75.0
Full Region Summary	84.7

5.2. Validation of the Effectiveness of Disturbance Output Reconstruction Model

Based on the training results of the training set in the previous section, the sub-region division is completed, and 0.7 is selected as the optimal TLCC threshold for the study area. The 2022 output data of all stations in Sub-region 4 are used as the test set to verify the effectiveness of the disturbance output reconstruction model. This study first identifies cloudy event days by classifying daily types into sunny-like days, rainy days, and cloudy days with the amplitude parameter and standard deviation of output fluctuation as the core features, referring to the method in Reference [3]. Subsequently, a typical cloudy event day is selected as the validation case for the effectiveness of the cloud movement correction algorithm. By calculating the optimal TLCC coefficients of multiple effective station pairs in the sub-region (Table 4), the optimal cloud velocity vector in this scenario is inverted, with a magnitude of 33.17 km/h and a direction of 166.8° east of north. Based on the uncertainty analysis framework established above, the 95% confidence interval of the cloud velocity vector is calculated synchronously: the confidence interval of velocity magnitude is [30.21 km/h, 36.13 km/h], and the confidence interval of movement direction is [153.7°, 179.9°], as shown in Figure 8.

Table 4. Parameters of station pairs used for cloud speed calculation.

Movement Direction	L _x (km)	L _y (km)	Time Lag (15 min)
Station 16 → Station 21	18.471	−28.812	6
Station 15 → Station 24	−9.231	−22.814	4
Station 15 → Station 21	−26.793	−35.045	6
Station 17 → Station 15	31.686	−8.187	3

Based on the inverted optimal cloud velocity vector, Station 15, whose average power is at the median value, is selected as the representative station. The representative station is used to reconstruct the time sequence of the disturbance component of the remaining stations in the sub-region, to obtain the deterministic reconstruction result of the disturbance power of this sub-region. Meanwhile, based on the coupled uncertainty framework established above, random sampling and batch reconstruction are carried out. The 2.5% and 97.5% quantiles of the empirical distribution are calculated for the reconstructed power results at each 15 min time step within the day, and finally the 95% confidence interval of the sub-region disturbance power reconstruction results is obtained, as shown in Figure 9.

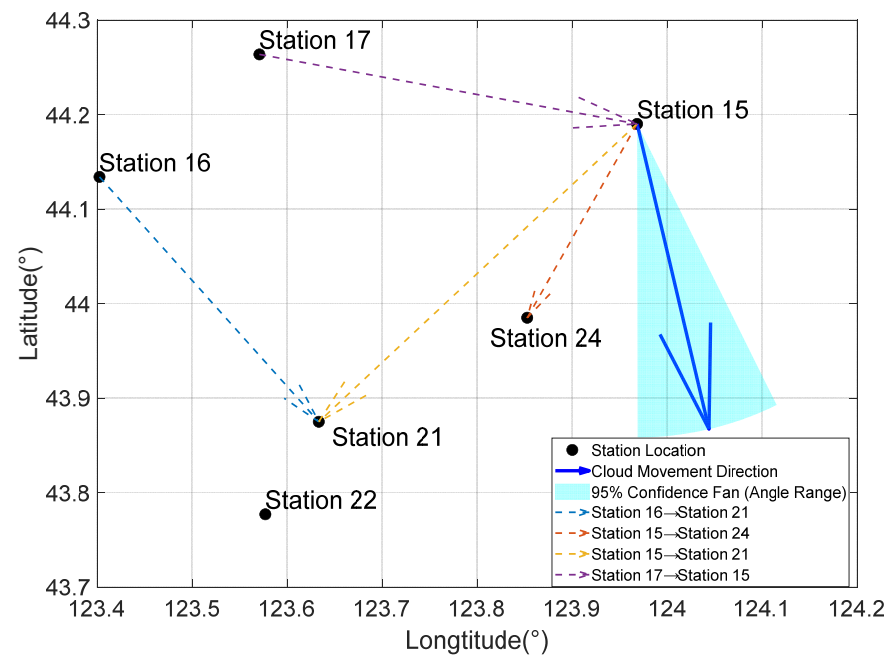


Figure 8. Inferred cloud movement direction and power output lag direction between stations.

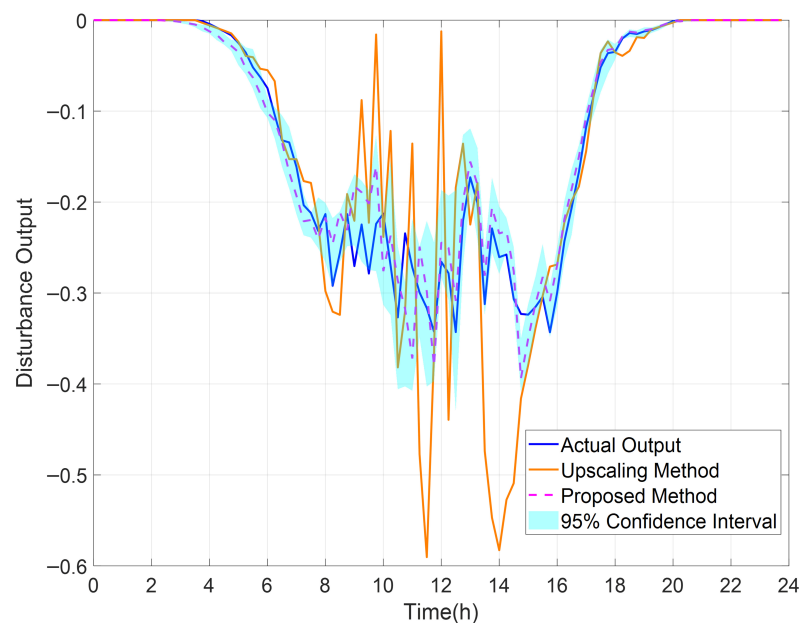


Figure 9. Comparison of reconstructed disturbance power from different models with measured data.

The experimental results show that between 11:45 and 12:00, the disturbance power estimated by the traditional method fluctuates drastically with an obviously high extreme value, while the results of the proposed method are in good agreement with the actual small fluctuations. This indicates that the cluster presents a significant smoothing characteristic due to spatial effects during this period, which the traditional method fails to capture. At around 12:30, the actual power gap is large and the smoothing effect is weakened, while the estimation result of the proposed method still has a small error with the actual value, which verifies its applicability in various scenarios.

To further verify the robustness of the model in long-term independent application, this paper calculates the out-of-sample prediction errors of the traditional upscaling method and the proposed reconstruction method for all cloudy event days of each sub-region in

the test set (2022), with the results shown in Table 5. In addition, to quantify the model's ability to capture the uncertainty boundary and real fluctuation characteristics, the average Prediction Interval Coverage Probability of the 95% confidence interval of the reconstructed power and the average relative error of the fluctuation standard deviation are further calculated, with the results shown in Table 6.

Table 5. Comparison of disturbance power reconstruction errors of different models on cloudy days.

Method	Mean Absolute Error (%)	Root Mean Square Error (%)
Upscaling method	12.51	15.69
Proposed method	6.03	7.09

Table 6. Validation of uncertainty boundary and fluctuation amplitude accuracy of the disturbance power reconstruction model.

Indicator	Value (%)
Average Coverage Probability of 95% Confidence Interval	90.2
Average Error of Standard Deviation	5.9

It can be seen from Table 5 that the reconstruction accuracy of the proposed method is significantly better than that of the traditional upscaling method. This verifies that correcting the phase misalignment of power output through cloud velocity tracking can effectively avoid the co-directional resonance error caused by the “synchronous fluctuation assumption” of the traditional method.

It can be seen from Table 6 that the average coverage rate of the 95% confidence interval generated by the model reaches 90.2%. This statistically confirms that the first-order macro-fluctuation caused by cloud advection is absolutely dominant, while the uncovered 9.8% residual objectively maps the unmodeled second-order high-frequency noise. This out-of-sample generalization performance statistically proves the universality and rationality of the core hypothesis. Although the model actively filters out this part of the microscopic noise, the average error of the standard deviation of the reconstructed sequence is only 5.9%. Given that the standard deviation directly determines the reserve capacity demand of the system, this result not only confirms the physical law that the second-order noise is naturally smoothed and offset during spatial aggregation but also proves the engineering effectiveness of the dimension-reduction reconstruction strategy of “locking the first-order dominant component” in the macro-reserve evaluation of the power grid.

5.3. Verification of Cluster Fluctuation Characteristic Analysis

To quantify the long-term impact of integrating stations at different geographical locations on the fluctuation characteristics of the existing cluster, this paper designates Station 22 and Station 23 within the region as proposed stations, and the system formed by the remaining stations as the baseline cluster (original cluster). Two expansion scenarios are designed for comparative verification:

Scenario 1: Integrate Station 22 into the baseline cluster.

Scenario 2: Integrate Station 23 into the baseline cluster.

To evaluate the smoothing effect brought by cloud movement, the coordinates of each sub-region cluster center (Exemplar) generated by AP clustering in the training set are extracted, the proposed station is divided into the existing sub-regions through its geographical coordinates, and the cloud movement correction model is applied to simulate its disturbance output in 2022. Subsequently, the SF and its statistical indicators of the two

schemes under all cloudy event days of the year are calculated respectively, with the results shown in Figure 10 and Table 7.

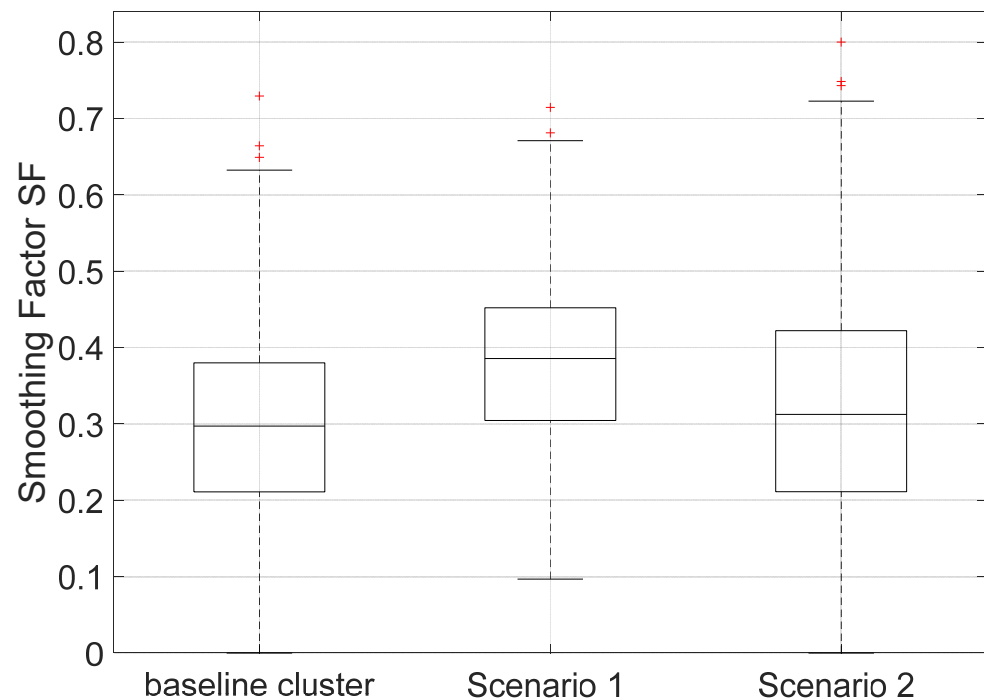


Figure 10. Box plot of the smoothing factor for the PV cluster under different expansion scenarios.

Table 7. Comparison of smoothing factor metrics under different expansion scenarios.

Indicator	Baseline Cluster	Scenario 1	Scenario 2
Mean_SF	0.293	0.378	0.321
P5_SF	0.096	0.197	0.075

The case results show that Scenario 1 presents the optimal system-level smoothing optimization benefit: its Mean_SF increases to 0.378, and P5_SF rises to 0.197. This indicates that the integration of Station 22 not only improves the long-term average reserve capacity release ratio of the cluster but also significantly enhances the reliability bottom line of the smoothing effect under extreme meteorological shocks. In contrast, although Scenario 2 slightly increases Mean_SF, P5_SF, which characterizes the safety bottom line, drops to 0.075. This phenomenon indicates that the proposed Station 23 is prone to “co-directional resonance” of output with the existing cluster due to its specific spatial location. The in-phase superposition of local fluctuations exacerbates the tail risk of the system, forcing the power grid to reserve more spinning reserve capacity to cope with extreme instantaneous ramp events.

6. Conclusions

Aiming at the engineering challenges that high-frequency micro-noise is difficult to accurately predict and lacks the mapping of macro-physical constraints of power grids in the quantification of fluctuation characteristics of large-scale PV clusters, this paper proposes a disturbance power reconstruction and smoothing effect quantitative evaluation framework for PV clusters focusing on the first-order dominant meteorological component. Through the spatiotemporal independent verification and analysis of cross-year measured data from 27 PV stations in a certain area of Jilin Province, the main conclusions are drawn as follows:

The dimensionality reduction modeling concept of “macro-advection dominance and micro-deformation filtering” is established. In this paper, the physical clear-sky model is used as a deterministic trend filter to extract the disturbance components that truly dominate the fluctuation characteristics. On this basis, through cluster division with fused dual-dimensional distance metrics and cross-space cloud velocity vector inversion, the phase misalignment of output within the PV cluster is effectively corrected, which fundamentally overcomes the co-directional resonance error of the traditional upscaling method caused by the “instantaneous synchronous fluctuation assumption”.

Strict out-of-sample independent verification confirms the model’s ability to capture the extreme uncertainty boundary. The experimental results show that the 95% confidence interval of the power reconstructed only by the first-order cloud advection characteristics can cover up to 90.2% of the measured fluctuation energy, and the reconstruction error of the key indicator determining the system reserve demand (fluctuation standard deviation) is only 5.9%. This statistically proves that local second-order high-frequency noise is easily offset naturally during spatial aggregation, and locking the macro-advection trend can restore the real fluctuation boundary of the PV cluster with high accuracy under limited-data conditions.

A quantitative evaluation system for fluctuation characteristics based on the Smoothing Factor is constructed. In this paper, the physical meaning of the SF is directly mapped to the “reserve capacity release ratio” of the power grid. To meet the multi-dimensional requirements of long-term power grid planning, this paper not only quantifies the expected long-term average economic benefits of the PV cluster through the mean value of SF but also relies on its probability distribution characteristics and uses the extremely low probability quantile to accurately define the guaranteed smoothing capacity provided by the spatial topology of the PV cluster under extreme severe cloudy scenarios. It acutely identifies the tail risk implied by local co-directional resonance, providing reliable decision support for the grid expansion and macro-capacity determination of reserve energy storage in high-proportion PV bases.

The model proposed in this paper shows good engineering applicability in pure data-driven scenarios lacking upper-air cloud image observation. However, in regions with more complex cloud system dynamic mechanisms (such as mountainous areas with severe terrain relief or equatorial regions dominated by local thermal convection), micro-deformation and vertical movement of clouds may periodically become the dominant factors, leading to the expansion of the error boundary of the first-order advection assumption. Future research will carry out cross-domain generalization verification based on multi-climate zone data, attempt to lightly integrate the second-order deformation characteristics extracted from satellite cloud images into the existing framework, and quantify the in-depth impact of spatial smoothing effect on the actual operation cost of the power system.

Author Contributions: Conceptualization, G.Y.; methodology, G.Y.; formal analysis, J.L.; data curation, W.K.; writing—original draft preparation, J.L.; visualization, J.L.; writing—review and editing, A.X.; supervision, G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Project of the Headquarters of State Grid Corporation of China, grant number 4000-202399368A-2-2-ZB.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to the fact that the PV output data in this article are actual measurement data from 27 PV stations in Jilin Province, China, in 2022, and some data (e.g., detailed operational parameters) involve commercial confidentiality and grid security sensitivity.

Acknowledgments: The authors would like to thank all individuals and organizations that provided support and assistance for this study and the preparation of the manuscript. During the preparation of this manuscript, the authors used a large language model developed by Google for the purpose of translating the manuscript from Chinese to English. The core research, including the methodology, data analysis, and conclusions, is the original work of the authors.

Conflicts of Interest: The authors declare that this study received funding from Science and Technology Project of the Headquarters of State Grid Corporation of China. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Abbreviations

The following abbreviations are used in this manuscript:

AP	Affinity Propagation
CDI	Cluster Distance Index
DBI	Davies–Bouldin Index
DTW	Dynamic Time Warping
PV	Photovoltaic
SF	Smoothing Factor
TLCC	Time-Lagged Cross-Correlation

References

1. Yu, S.W.; Zheng, Y.L.; Li, L.X. A comprehensive evaluation of the development and utilization of China's regional renewable energy. *Energy Policy* **2019**, *127*, 73–86. [\[CrossRef\]](#)
2. Li, K.; Gong, M.; Li, Z.; Huang, C. Classified identification and estimation of behind-the-meter distributed photovoltaic panels using high-resolution aerial imagery. *IEEE Trans. Ind. Appl.* **2025**, *61*, 1604–1612. [\[CrossRef\]](#)
3. Wang, J.; Huang, Y.H.; Li, C. Time Series Modeling Method for Multi-photovoltaic Power Stations Considering Spatial Correlation and Weather Type Classification. *Power Syst. Technol.* **2020**, *44*, 1376–1384. [\[CrossRef\]](#)
4. Zhang, W.; Liu, Z. Simulation and analysis of the power output fluctuation of photovoltaic modules based on NREL one-minute irradiance data. In Proceedings of the 2013 International Conference on Materials for Renewable Energy and Environment, Chengdu, China, 19–21 August 2013. [\[CrossRef\]](#)
5. Liu, S.; Wu, S.Y.; Liu, W.L.; Liu, C.L.; Wang, X.; Kang, J.Y. Capacity optimization of wind-solar pumped storage power generation system considering uncertainties. *J. Hydroelectr. Eng.* **2024**, *43*, 43–56. [\[CrossRef\]](#)
6. Zhenwei, W.; Xiaoping, J.; Huimeng, M.A.; Suliang, M. Study on fluctuations characteristics of photovoltaic power output in different time scales. *Mod. Electr. Power* **2014**, *31*, 58–61. [\[CrossRef\]](#)
7. Ummels, B.C.; Gibescu, M.; Pelgrum, E.; Kling, W.L.; Brand, A.J. Impacts of Wind Power on Thermal Generation Unit Commitment and Dispatch. *IEEE Trans. Energy Convers.* **2007**, *22*, 44–51. [\[CrossRef\]](#)
8. da Silva, A.M.L.L.; Sales, W.S.; da Fonseca Manso, L.A.; Billinton, R. Long-Term Probabilistic Evaluation of Operating Reserve Requirements with Renewable Sources. *IEEE Trans. Power Syst.* **2010**, *25*, 106–116. [\[CrossRef\]](#)
9. Zou, K.; Agalgaonkar, A.P.; Muttaqi, K.M. Distribution System Planning with Incorporating DG Reactive Capability and System Uncertainties. *IEEE Trans. Sustain. Energy* **2012**, *3*, 112–123. [\[CrossRef\]](#)
10. Zhang, Y.; Shen, Y.B.; Shi, G.Y. Short-term Forecasting Technology of Solar Energy for Concentrating Solar Power. *Autom. Electr. Power Syst.* **2016**, *40*, 158–167. [\[CrossRef\]](#)
11. Ekström, J.; Koivisto, M.; Mellin, I.; Millar, R.J.; Lehtonen, M. A Statistical Model for Hourly Large-Scale Wind and Photovoltaic Generation in New Locations. *IEEE Trans. Sustain. Energy* **2017**, *8*, 1383–1493. [\[CrossRef\]](#)
12. Chen, X.; Zhang, L.Z.; Yang, M.; Zhu, Y. A method for AGC reserve capacity analysis considering photovoltaic power fluctuation characteristics. *Autom. Electr. Power Syst.* **2015**, *39*, 16–21, 52. [\[CrossRef\]](#)
13. Cui, Y.; Mu, G.; Liu, Y.; Yan, G. Spatiotemporal distribution characteristic of wind power fluctuation. *Power Syst. Technol.* **2011**, *35*, 110–114. [\[CrossRef\]](#)
14. Yin, J.N.; Ge, Y.F.; Gao, K. Analysis on clustering effect of wind generations. *Electr. Meas. Instrum.* **2015**, *5*, 104–108. [\[CrossRef\]](#)
15. Liu, Y.; Tian, R.; Zhang, D.; Zhang, X.; Zhou, J. Analysis and application of wind farm output smoothing effect. *Power Syst. Technol.* **2013**, *37*, 987–991. [\[CrossRef\]](#)
16. Li, G.; Zhao, Z.J.; Dai, H.N.; Wang, F.; Sun, S.M.; Cheng, Y. Short-term Power Forecasting Method for Photovoltaic Clusters Considering Weather In consistency and Intraday Output Characteristic Changes. *High Volt. Eng.* **2026**, *37*, 1–18. [\[CrossRef\]](#)

17. Huang, C.Z.; Liu, Y.D. Short-Term Power Prediction Method for Photovoltaic Power Plants Based on ECA-MTGNN Integration. *Control Inf. Technol.* **2025**, *68*–75. [[CrossRef](#)]
18. Zhao, W.B.; Yang, Q.; Chen, Y.Y. Ultra-short-term Source-load Joint Forecasting Method Based on Group Cross-gating Mechanism and Graph Convolutional Neural Network. *High Volt. Eng.* **2026**, 1–14. [[CrossRef](#)]
19. Yang, F.Y.; Wan, Y.N.; Yang, G.H.; Li, F.Y.; Zhang, M.Q.; Du, W.C. Short Term Photovoltaic Power Generation Prediction based on OVMD-CNN-BiLSTM. *Acta Energiæ Solaris Sin.* **2025**, *46*, 333–342. [[CrossRef](#)]
20. Xia, L.F.; Li, J.M.; Zhao, L.; Ai, X.M.; Fang, J.K.; Wen, J.Y.; Xie, H.L. A PV Power Time Series Generating Method Considering Temporal and Spatial Correlation Characteristics. *Proc. CSEE* **2017**, *37*, 1982–1993. [[CrossRef](#)]
21. Cui, Y.; Li, H.Q.; Yan, G.G.; Zhang, J.T.; Mu, G.; Wang, M.C. An Optimization Method to Determine Integrated Power Transmission Capacity of Clustering Photovoltaic Plants Based on Clustering Effect. *Power Syst. Technol.* **2015**, *39*, 3491–3496. [[CrossRef](#)]
22. Hollands, K.G.T.; Huget, R.G. A probability density function for the clearness index, with applications. *Sol. Energy* **1983**, *30*, 195–209. [[CrossRef](#)]
23. Salvador, S. Toward Accurate Dynamic Time Warping in Linear Time and Space. *Intell. Data Anal.* **2007**, *11*, 561–580. [[CrossRef](#)]
24. Fu, Z.H.; Mu, G.; Wu, F.Z.; Yang, M.; Yan, G.G. Modeling and Analysis of Power Fluctuation Amplitude Evolution Mechanism of Wind Farm Cluster Based on the Amplitude-Duration Characteristics of Single Farm Power Waveform. *Proc. CSEE* **2025**, *45*, 124–138. [[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.