

Improved Probability Prediction of Distributed Photovoltaic Power Based on Optimized KMC-Vine Copula[#]

Xiaotong Yang¹, Zuan Fu¹, Jie Shi^{1*}

¹ University of Jinan

(Corresponding Author: Jie Shi email: sps_shij@ujn.edu.cn)

ABSTRACT

Distributed photovoltaic (DPV) power generation has increasingly emerged as one of the most promising renewable energy sources due to its extensive distribution and flexible installation. However, its output is highly influenced by solar irradiation conditions, exhibiting diurnal and seasonal variations, with weather conditions being the most significant influencing factor. To ensure the secure and stable operation of the power system and maintain an appropriate balance between supply and demand, predicting DPV power output presents a viable solution. This paper proposes an improved KMC-Vine Copula model for probabilistic prediction of DPV power by integrating spatio-temporal influencing factors and the prediction results of centralized PV power. Using copula correlation theory, the spatio-temporal dependencies between centralized and distributed PV systems are figured out. Conventional single copula functions are inadequate in capturing complex correlations. Thus, an optimized copula model is introduced to address the above limitation. However, due to its inability to represent high-dimensional dependencies, a flexible Vine Copula model is constructed as the final prediction framework. Through the case study, compared with the optimized copula model, the proposed model improved by 26.87% (interval width), 41.6% (reliability), and 39.23% (sharpness), respectively, under three evaluation indexes.

Keywords: terms—distributed photovoltaic, k-means clustering, optimized copula, quantile regression, vine copula

1. INTRODUCTION

In recent years, with the advancement of the “dual carbon” strategy and the continuous expansion of new energy installations, distributed photovoltaic (DPV) generation has become a key pillar of China’s clean energy system [1]. By the end of 2024, new installations

had increased the country’s cumulative photovoltaic (PV) capacity to 886.66 million kilowatts, including 118 million kilowatts of newly installed distributed PV capacity [2]. However, DPV generation exhibits significant randomness, volatility, and intermittency, as its output is strongly influenced by solar radiation and meteorological conditions [3]. As the penetration of DPV continues to grow within distribution networks, large-scale integration has led to issues such as voltage fluctuations, reverse power flows, and protection coordination challenges, thereby affecting the safety and economic operation of the system [4]. Consequently, high-precision DPV power forecasting is considered essential to ensure the secure operation of active distribution networks.

The accuracy and robustness of prediction models have been substantially enhanced with the advancement of deep learning and hybrid modeling techniques. Spatio-temporal modeling of multi-site DPV output has been achieved using a spatio-temporal graph attention network [5]. A comprehensive review of distributed solar forecasting methods that integrate remote sensing and deep learning was presented in Ref. [6]. In Ref. [7], a personalized federated generative adversarial network was proposed, which improves model generalization while preserving data privacy.

Simultaneously, a shift in DPV forecasting from deterministic approaches to interval and probabilistic predictions has been observed. In Ref. [8], Bayesian bootstrap and quantile regression techniques were applied for real-time probabilistic forecasting, enabling the quantification of uncertainty for dispatch and risk management. In [9], prediction errors were incorporated into voltage regulation through distributed Model Predictive Control (MPC), highlighting the critical impact of prediction accuracy on the economic efficiency and safety of high-penetration DPV operations. Consequently, DPV forecasting research is progressing from single-point deterministic approaches toward multi-source data fusion and probabilistic modeling,

[#] This is a paper for the 17th International Conference on Applied Energy (ICAE2025), December 8-12, 2025, Bangkok, Thailand.

thereby enhancing the flexibility and security of active distribution networks.

Compared with centralized photovoltaic systems [10], DPV generation is characterized by numerous geographically dispersed generation points. Its output exhibits strong spatial correlation and non-stationarity, influenced by variations in terrain and meteorological conditions [11]. The coordinated variations across multiple nodes are difficult to capture using traditional single-point prediction models, thereby affecting overall forecasting accuracy and system stability. To address this issue, a method for output curve correction based on the reconstruction of daily cumulative generation was proposed by Ref. [12], through which the power consistency of distributed plants was significantly improved. An improved self-feedback iterative strategy was utilized by Ref. [13], through which integrated prediction – control optimization was achieved. At the system level, a two-stage stochastic robust model combined with model predictive control (MPC) was employed by Ref. [14], with the critical influence of prediction accuracy on economic and safety performance under high-penetration DPV conditions being validated. A flexible modeling framework for describing dependencies among random variables is provided by copula theory, enabling variable correlations to be characterized without relying on specific distributional assumptions [15]. In Ref. [16], higher-order moment analysis was conducted on the joint distribution of wind and solar power outputs using multivariate copulas, followed by an assessment of system flexibility. However, traditional single-copula models face limitations in high-dimensional scenarios, struggling to accurately capture complex dependency structures among multiple nodes.

Ref. [17] proposes a KMC-optimized Copula model to overcome the limitations of the single copula model and enhance the accuracy of distributed photovoltaic power prediction. Ref. [18] introduces a photovoltaic power interval prediction method based on an optimized backpropagation (BP) neural network by Simulated Annealing Modified Particle Swarm Optimization (SAPSO-BP) and quantile regression. Ref. [19] proposes a short-term interval probability prediction of photovoltaic power based on similar daily clustering and the clustering QR-CNN-BiLSTM model.

Existing quantile regression methods are typically combined with machine learning models that support only single quantile regressions. The Vine Copula model can flexibly and accurately describe the joint probability distribution of high-dimensional random variables and

has an analytical form and strict monotonicity [20]. These advantageous properties help address the challenges of complex modeling and the issue of “quantile crossing” present in the current methods.

Considering the advantages of the Vine Copula model and weather classification, an improved probability prediction of distributed photovoltaic power based on a novel KMC-Vine Copula method is proposed in this paper.

2. METHOD

2.1 Improved Dynamic Time warping(IDTW)

In time series tasks such as power prediction, model accuracy and stability are directly impacted by data quality. The model's learning capability is compromised, systematic bias is introduced, and prediction reliability is reduced by missing or anomalous data. It was indicated by Ref. [21] that the completeness and predictive performance of multivariate time series can be significantly enhanced by deep imputation. It was demonstrated by Ref. [22] that the modeling and prediction outcomes of univariate time series are substantially influenced by different imputation methods. Therefore, prediction reliability is ensured by scientific data imputation, which is considered a critical prerequisite.

The Improved Dynamic Time Warping (IDTW) is designed through the integration of Dynamic Time Warping (DTW) and K- Nearest Neighbor (KNN) concepts, allowing the spatiotemporal characteristics of power data from multiple PV units within a PV system to be comprehensively utilized. The Euclidean distance in traditional KNN is replaced by the DTW distance to measure the similarity between the time series of the unit to be interpolated and its neighboring units. On this basis, the most similar PV stations are selected for weighted imputation, by which missing intervals are robustly estimated. Through this process, temporal alignment is maintained and spatial collaborative information is incorporated, resulting in high accuracy and strong robustness. Consequently, a more complete and reliable data foundation is provided for subsequent prediction models.

In this study, complete power sequences from five adjacent monitoring stations are selected as candidate reference sequences. Sequence similarity is evaluated by computing the sum of DTW distances between the selected sequences and the valid data segments before and after the target missing segment. The parameter k is assigned the value L . The selected PV stations are

denoted as $p=\{1,2,\dots,L\}$. Where B is the missing segment sequence. The calculation formula is given as follows:

$$B = \frac{1}{L} \sum_{p=1}^L B^p \quad (1)$$

The reconstructed data are then used as inputs for weather clustering and copula modeling.

2.2 Copula and Vine Copula

Copula theory was proposed by Sklar in 1959 that the copula function could describe the correlation between edge distribution functions [23].

Unlike traditional joint distribution modeling, the copula approach separates the marginal distributions from the dependence structure, allowing the construction of a joint distribution regardless of the specific forms of the marginal variables.

Typical types of copula include the elliptical family and several Archimedean variants [16]. Owing to its flexibility and strong compositional property, the Archimedean Copula has become a popular tool for analyzing the dependence between random variables [17]. Among them, the Frank Copula exhibits symmetric tail dependence, responding equally to extreme values at both ends of the distribution.

To better capture the tail dependence characteristics, an improved copula model based on the Archimedean family was developed [24], as shown in :

$$C_t(u_1, \lambda_1, \lambda_2, \lambda_3) = A * C_t(u_1, \lambda_1) + B * C_c(u_1, \lambda_2) + C * C_c(u_1, \lambda_3) \quad (2)$$

According to the formula, the optimized copula coefficient is shown in the Ref. [17].

The fitting accuracy of bivariate dependency structures is effectively improved by the optimized copula model, and better flexibility is provided compared with single copula functions. However, its weight coefficients are fixed during the modeling process, which limits the ability of the model to dynamically represent the relative importance and interaction among different copula components. Consequently, the global optimal solution may not be obtained by the optimized copula model when complex and high-dimensional dependence relationships are dealt with. To overcome the above limitations, the Vine Copula model is introduced in this paper, which allows for a hierarchical decomposition of multivariate dependencies into multiple pairwise copulas. Thereby, greater flexibility and accuracy are provided in capturing nonlinear and asymmetric relationships among variables.

The Vine Copula decomposes a multivariate joint probability density function into a cascade of multiple binary copula density functions [25-26]. It is based on the

combination of the joint probability distribution function and copula theory. The commonly used vine structures are C-vine and D-vine.

The C-vine model has a star structure, which has high precision when one variable is strongly correlated with other variables, while the other variables are weakly correlated with each other. The figure below shows a C-vine model by taking a temperature (U_1), PV power at site 1 (U_2), and PV power at site 2 (U_3) as an example.

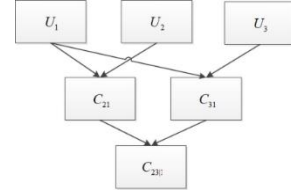


Fig. 1 Vine Copula

When the C-vine structure is used to analyze the correlation of multidimensional correlation variables, its joint probability distribution is expressed as follows:

$$f(x_1, \dots, x_n) = \prod_{k=1}^n f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,j+1, \dots, j-1} [F(x_j | x_1, \dots, x_{j-1}), F(x_{j+1} | x_1, \dots, x_{j-1})] \quad (3)$$

Where $f_i(x_i)$ and $F_i(x_i)$ respectively represent the probability density function and cumulative distribution function of X_i . In the formula, the conditional distribution has the following properties:

$$F(x | v) = \frac{\partial C[F(x | v_{-j}), F(v_j, v_{-j})]}{\partial F(v_j | v_{-j})} \quad (4)$$

Where: v_j represents the j variable in vector v ; and v_{-j} represents the remaining vector v after removing v_j .

2.3 Conditional Quantile Regression of Photovoltaic Power

The Vine Copula model provides a flexible and precise description of the dependency between photovoltaic power and its influencing factors, such as irradiance and temperature, through an analytical framework. From this relationship, the conditional quantile function of photovoltaic power is derived, forming the basis for the corresponding quantile regression model. Accordingly, the conditional probability distribution of PV power is represented analytically based on the dependence structure captured by the Vine Copula.

Taking the C-vine structure shown in Fig. 1 as an example, the 2+1 dimensional conditional probability distribution function $F(x_3 | x_2, x_1)$ of PV power U_2 of

site 1 with respect to conditional variables (PV power U_3 and temperature U_1 of site 2) can be expressed as:

$$\begin{cases} F(x_3 | x_1) = h_{12}[F(x_1), F(x_3)] \\ F(x_2 | x_1) = h_{13}[F(x_1), F(x_3)] \\ F(x_3 | x_1, x_2) = h_{231}[F(x_2 | x_1), F(x_3 | x_1)] \end{cases} \quad (5)$$

The relation between function $h(u, v)$ and copula function is $h(u, v) = \partial C(u, v) / \partial v$.

Based on the variable relationship expressed in (5), the regression equation of the conditional quantile of photovoltaic power can be derived. Given the values of each conditional variable: $x_1 = R_1, x_2 = R_2$, the regression equation is used to calculate the sub-point α of photovoltaic power [27].

$$\begin{cases} F(R_2 | R_1) = h_{12}[F(R_1), F(R_2)] \\ F(\hat{q}_\alpha | R_1) = h_{231}^{-1}[\alpha, F(R_2 | R_1)] \\ F(\hat{q}_\alpha) = h_{13}^{-1}[F(\hat{q}_\alpha | R_1), F(R_1)] \\ \hat{q}_\alpha = F^{-1}(F(\hat{q}_\alpha)) \end{cases} \quad (6)$$

Where: $h^{-1}(\cdot)$ represents the inverse function of $h(\cdot)$; \hat{q}_α represents the α quantile estimate of photovoltaic power.

2.4 Evaluation Indicators

Ensuring the reliability of prediction results is the primary objective of probabilistic prediction. According to the definition of the reliability evaluation index, this index describes the deviation between the nominal coverage and the empirical coverage of the predict interval [28].

$$\eta_{1-\beta} = \begin{cases} 1 & \hat{q}_{\beta/2, n} < y_n < \hat{q}_{1-\beta/2, n} \\ 0 & \text{other} \end{cases} \quad (7)$$

$$\varphi_{1-\beta} = \frac{1}{N} \sum_{n=1}^N \eta_{1-\beta, n} \quad (8)$$

$$R_{1-\beta} = \varphi_{1-\beta} - (1 - \beta) \quad (9)$$

Where: $\eta_{1-\beta, n}$ is the indicator variable, corresponds to the n the prediction point under confidence $1-\beta$; y_n represents the true value of the n the prediction point; $\hat{q}_{\beta/2, n}$ and $\hat{q}_{1-\beta/2, n}$ represent the $\beta/2$ quantile and $1-\beta/2$ quantile predicted value of the n predicted point, respectively; $\varphi_{1-\beta}$ represents the experience coverage rate, $1-\beta$ represents the experience coverage rate; N represents

the total number of prediction points; $R_{1-\beta}$ indicates the reliability indicator.

The sharpness index reflects the average width of the probability prediction interval, and the smaller the sharpness, the lower the uncertainty of the probability prediction result. In the case of confidence $1-\beta$, the prediction results $\hat{q}_{\beta/2, n}$ and $\hat{q}_{1-\beta/2, n}$ of paired sub-points express the probability interval, then the calculation process of the sharpness index is

$$\delta_{1-\beta, n} = \hat{q}_{1-\beta/2, n} - \hat{q}_{\beta/2, n} \quad (10)$$

$$S_{1-\beta} = \frac{1}{N} \sum_{n=1}^N \delta_{1-\beta, n} \quad (11)$$

Where: $\delta_{1-\beta, n}$ represents the confidence interval width of the n the prediction point under the confidence degree $1-\beta$; $S_{1-\beta}$ represents the sharpness index, reflecting the interval width of confidence $1-\beta$.

2.5 Overall prediction Process

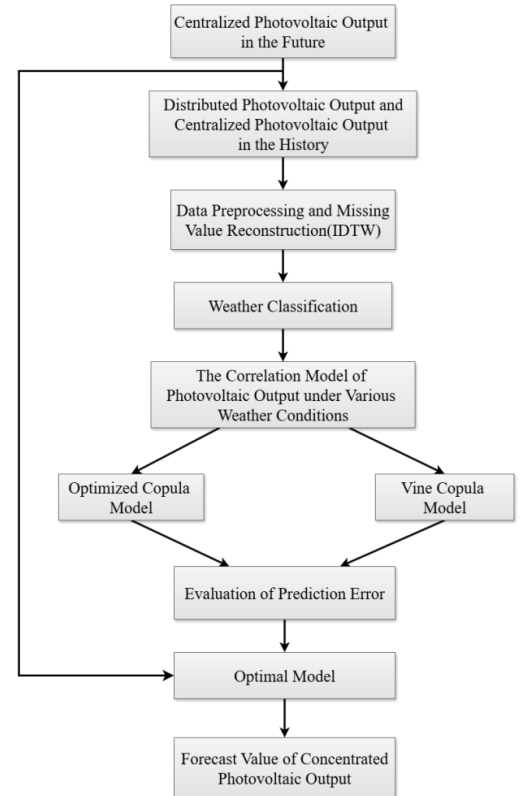


Fig. 2 prediction Flow Chart

3. CASE STUDY

This study employs photovoltaic generation and meteorological datasets obtained from an Australian

photovoltaic plant, recorded at 10-minute intervals. The distance between individual PV units is about 2.6 km. Data from 2019 and 2020 are utilized to validate the proposed conditional probability estimation method for DPV output.

By replacing the Euclidean distance in KNN with DTW distance, it is found that when $k = 2$, dynamic similarities between time series can be more effectively captured, while matching errors caused by temporal misalignment in traditional Euclidean distance are avoided. A comparison of the IDTW-based imputation method with other approaches is presented in the figure.

In Figure (a), the IDTW-based imputation closely matches the original power sequence, smoothly bridging missing segments and accurately capturing power trends. In Figure (b), the KNN-based imputation aligns with the overall trend but deviates at local extrema, leading to reduced smoothness. In Figure (c), the MICE-based results recover steady regions well but exhibit delayed responses to sudden power changes, producing slightly smoother values. In Figure (d), the MF-based imputation yields a relatively smooth sequence; however, significant deviations occur in the missing segments, making the dynamic characteristics of power variations difficult to reflect.

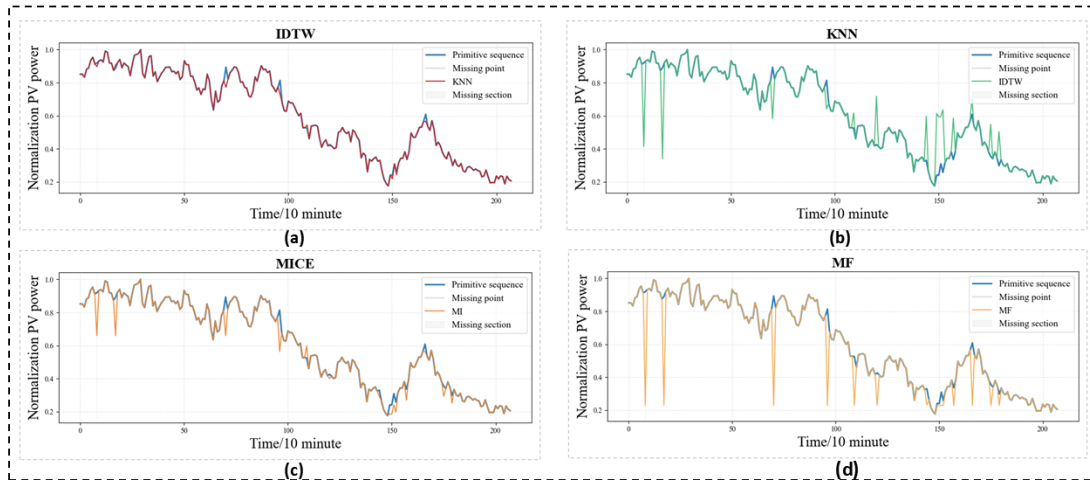


Fig. 3 Comparison of different imputation effects

Table 1 presents the correlation coefficients between meteorological variables and power under Pearson, Spearman, and Kendall tests. Where v1 is temperature, v2 is wind speed, v3 is Relative Humidity, v4 is Pressure, v5 is Irradiation. Based on these results, relative humidity, pressure, and irradiation were selected as the meteorological inputs for weather classification. To classify the weather conditions, a three-dimensional K-means clustering algorithm was applied, while the corresponding cluster centers are summarized in Table 2.

Table. 1 Three correlation coefficients for each meteorological

| Type | The correlation coefficient | | | | |
|----------|-----------------------------|------|------|------|------|
| | V1 | V2 | V3 | V4 | V5 |
| Pearson | 0.09 | 0.08 | 0.12 | 0.29 | 0.48 |
| Spearman | 0.03 | 0.05 | 0.08 | 0.29 | 0.41 |
| Kendall | 0.03 | 0.03 | 0.05 | 0.20 | 0.35 |

Table. 2 Clustering Center Data

| The weather types | The weather factor | | |
|-------------------|--------------------|----------|------------------------|
| | Relative Humidity | Pressure | Short-wave irradiation |
| R | 12.14 | 941.88 | 37.44 |

| | | | |
|---|------|--------|-------|
| G | 21 | 936.99 | 32.54 |
| B | 9.41 | 938.50 | 43.01 |

Referring to Table 2, the analysis of the clustering centers for temperature, humidity, and wind speed shows that the R-type category can be characterized as cloudy, the G-type as overcast, and the B-type as sunny [17]. Once the weather categories were determined, a spatial correlation model was developed using the copula function.

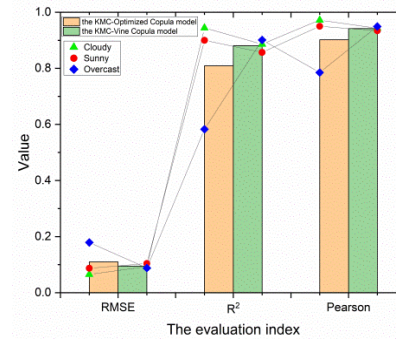


Fig. 4 Comparison of point prediction errors under different models

As depicted in Fig. 4, the KMC-Optimized Copula model outperforms the KMC-Vine Copula model under cloudy and sunny conditions, while the latter performs better in overcast weather. On average across all weather types, the KMC-Vine Copula model shows superior overall accuracy. Based on these findings, the copula models corresponding to different weather conditions were applied, using the predicted centralized photovoltaic output as input to estimate the distributed photovoltaic power.

Subsequently, a conditional probabilistic prediction model was established based on the point forecasts of distributed PV output. The probabilistic prediction results of the KMC-Optimized Copula and KMC-Vine Copula models are compared in the figure below.

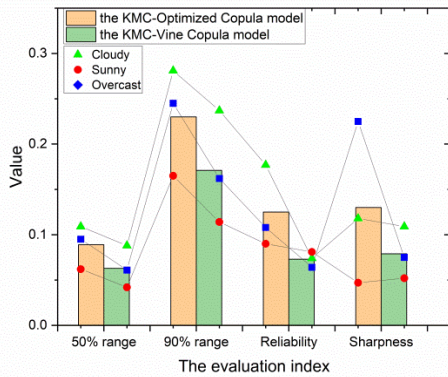


Fig. 5 Comparison of probabilistic prediction results of different models

As shown in the figure, the KMC-Vine Copula model performs better than the KMC-Optimized Copula model under all three weather conditions. For the 50% and 90% prediction intervals, its accuracy improves by about 25% on average. In terms of reliability and sharpness, the performance increases by around 40%, except for a slight decline under sunny conditions.

Accordingly, the copula models for different weather types use the predicted centralized photovoltaic output as input to estimate the distributed photovoltaic output. Representative results are shown in Fig. 6.

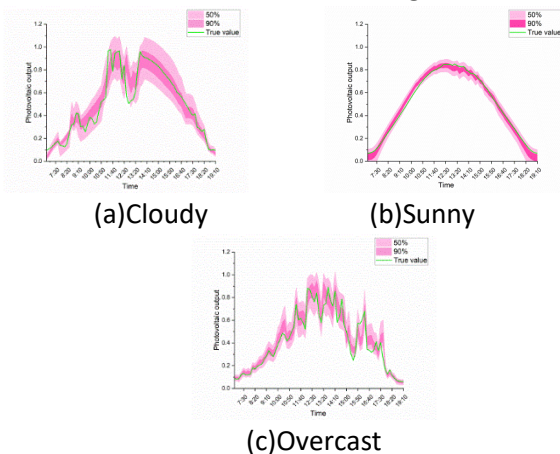


Fig. 6 Confidence interval of PV output point prediction, actual value and probability prediction for three weather types

To further explore the influence of selected meteorological factors on the KMC-Vine Copula model, meteorological variables of irradiation and temperature were considered here to analyze the probabilistic prediction results of the Vine Copula model. As shown in the figure below, the probability prediction effect diagram under a 90% confidence interval is drawn after temperature and radiation are respectively added to the model. The figure shows the average true value, the average predicted value and average predicted interval under the three kinds of weather.

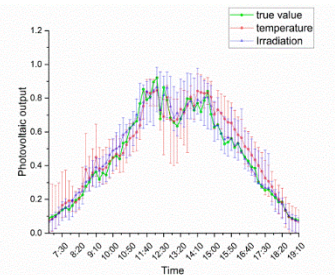


Fig. 7 Graph of prediction Results of Vine-Copula Model under Different Meteorological Variables

As can be seen from the figure above, the average prediction interval of the KMC-Vine Copula model after the addition of the irradiation variable is better than that of the model after the addition of temperature. The prediction interval is smooth without big fluctuation, and the overall prediction effect is stable. The comparison of the evaluation indexes of the prediction results of different meteorological variables is shown in the following table.

Table. 3 Comparison of Vine Copula Probability prediction Results Under Different Meteorological Variables

| Vine Copula | Average width of 50% predict range | Average width of 90% predict interval | R | S |
|-------------|------------------------------------|---------------------------------------|-------|-------|
| V1 | 0.069 | 0.201 | 0.123 | 0.089 |
| V2 | 0.063 | 0.171 | 0.073 | 0.079 |

As can be seen from the above table, the prediction model with the addition of the irradiation variable is better than the model with the addition of the temperature variable in various evaluation indexes.

In conclusion, compared with the KMC-Optimized Copula model, the KMC-Vine Copula model is optimized by 26.87%, 41.6%, and 39.23% in average predict interval

width, reliability, and sharpness, respectively. In addition, the prediction result of the KMC-Vine Copula model is better than that of the temperature variable when considering the addition of the irradiation variable.

4. CONCLUSION

A probabilistic prediction framework for DPV based on the KMC-Vine Copula model is proposed in this study. To enhance the reliability of input data, an IDTW based imputation method is introduced to reconstruct missing values in meteorological and power series. In this method, the Euclidean distance in the traditional KNN algorithm is replaced with DTW distance, enabling better temporal alignment and smoother data recovery, thereby providing cleaner inputs for model training.

Using measured data from an Australian PV station, the proposed KMC-Vine Copula model is developed to characterize the nonlinear dependence between centralized and distributed PV power outputs under various weather conditions. Clustering is first conducted based on meteorological variables such as humidity, pressure, and irradiance, after which distinct dependence structures are constructed for each weather type.

Experimental results demonstrate that the proposed model achieves superior performance in both point and probabilistic prediction. Compared with the KMC-Optimized Copula model, the KMC-Vine Copula approach achieves improvements of 26.87%, 41.6%, and 39.23% in interval width, reliability, and sharpness, respectively. By integrating IDTW and Copula-based dependence modeling, the accuracy, stability, and generalization capability of DPV power prediction are effectively enhanced.

REFERENCE

[1] Ahsan, F., Dana, N.H., Sarker, S.K. et al. Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review. *Protection, Control and Modern Power Systems* 2023;8:43.

[2] National Energy Administration. Statistical Data on China's 2024 New Photovoltaic Grid-Connected Capacity. Retrieved from <http://www.nea.gov.cn/20250221/e10f363cabe3458aa f78ba4558970054/c.html>

[3] Yin, X., Lei, M. Jointly improving energy efficiency and smoothing power oscillations of integrated offshore wind and photovoltaic power: a deep reinforcement learning approach. *Protection, Control and Modern Power Systems* 2023;8:25.

[4] Karimi M., Mokhlis H., Naidu K., et al. Photovoltaic penetration issues and impacts in distribution network – A review. *Renewable and Sustainable Energy Reviews* 2016;53:594-605.

[5] Simeunović J., Schubnel B., Alet P., et al. Interpretable temporal-spatial graph attention network for multi-site PV power prediction. *Applied Energy* 2022;327:120127.

[6] Chu Y., Wang Y., Yang D., et al. A review of distributed solar prediction with remote sensing and deep learning. *Renewable and Sustainable Energy Reviews* 2024;198:114391.

[7] Deng F., Wang J., Wu L., et al. Distributed photovoltaic power prediction based on personalized federated adversarial learning. *Sustainable Energy, Grids and Networks* 2024;40:101537.

[8] Bozorg M., Bracale A., Carpita M., et al. Bayesian bootstrapping in real-time probabilistic photovoltaic power prediction. *Solar Energy* 2021;225:577-590.

[9] Ma C., Xiong W., Tang Z., et al. Distributed MPC-Based Voltage Control for Active Distribution Networks Considering Uncertainty of Distributed Energy Resources. *Electronics* 2024;13(14):2748.

[10] Li J., Mu G., Zhang J. et al. Dynamic economic evaluation of hundred megawatt-scale electrochemical energy storage for auxiliary peak shaving. *Protection, Control and Modern Power Systems* 2023;8:50.

[11] Wang J., Xie N., Huang C. et al. Two-stage stochastic-robust model for the self-scheduling problem of an aggregator participating in energy and reserve markets. *Protection, Control and Modern Power Systems* 2023;8:45.

[12] Qiao H., HAN J., Li Z., et al. An accurate calculation method of line loss based on reconstructing a photovoltaic outputcurve by daily cumulative power generation. *Power System Protection and Control* 2024;52(24):161-168

[13] Yang J., Yang T., Luo L. et al. Tracking-dispatch of a combined wind-storage system based on model predictive control and two-layer fuzzy control strategy. *Protection, Control and Modern Power Systems* 2023;8:58.

[14] Qin M., Yang Y., Zhao X. et al. Low-carbon economic multi-objective dispatch of integrated energy system considering the price fluctuation of natural gas and carbon emission accounting. *Protection, Control and Modern Power Systems* 2023;8:61.

[15] Nelsen R. B. An introduction to copulas. 1th ed. New York: Springer; 2013.

[16] Liu Y., Chen H., Tian Y., et al. Flexibility assessment of a power system based on higher-moment analysis of

an investment portfolio. *Power System Protection and Control* 2024;52(5):116-127

[17] Fu Z., Cheng X., Shi J., et al. Probability estimation of distributed photovoltaic power prediction based on KMC-optimized copula. *2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia) 2022*; 1257-1262.

[18] Jia D., Lv Q., Lin F.. Photovoltaic power interval prediction based on SAPSO-BP and quantile regression. *Power System Protection and Control* 2021;49(10):21-26.

[19] Wang K., Du H., Jia R., et al. Short-term interval probability prediction of photovoltaic power based on similar daily clustering and QR-CNN-BiLSTM model. *High Voltage Engineering* 2022;48(11):4372-4384.

[20] Kraus D., Czado C.. D-vine copula based quantile regression. *Computational Statistics & Data Analysis* 2017;110:1-18.

[21] Kazijevs M., Samad M.D.. Deep imputation of missing values in time series health data: A review with benchmarking. *Journal of Biomedical Informatics* 2023;144:104440.

[22] Niako N., Melgarejo J.D., Maestre G.E. et al. Effects of missing data imputation methods on univariate blood pressure time series data analysis and prediction with ARIMA and LSTM. *BMC Med Res Methodol* 2024;24:320.

[23] Wang S., Ping C., Xue G., et al. Research and analysis of factors affecting the output of photovoltaic power. *Electrical Engineering* 2018;19(8):68-71.

[24] Ji F., Cai F., Wang J.. Wind power correlation analysis based on hybrid copula *Electeic Power Systems* 2014;38(2):1-5.

[25] Qiu Y., Yubo O.. Modeling of multi-dimensional wind and PV farm output correlation based on mixture vine copula structures and its application in reactive power optimization. *Power System Technology* 2017;41(3):792-798.

[26] Dissmann J., Brechmann E.C.. Selecting and estimating regular vine copula and application to financial returns. *Computational Statistics & Data Analysis* 2013;59:52-69.

[27] Xu B., Xu Q., Huang Y.,et al. Day-ahead probabilistic prediction of photovoltaic power based on vine copula quantile regression. *Power System Technology* 2021;45(11):4427-4434.

[28] Van der Meer D.W., Widén J., Munkhammar J., et al. Review on probabilistic prediction of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews* 2018;81:1484-1512.