International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

Short-Term Wind Power Forecasting Based on Data Quality Awareness and Global-Local Feature Collaboration

Dian Wang

[1]North China Electric Power University, No. 2 Beinong Road, Changping District, Beijing, China
Email: *wangdian614151910[at]163.com*

Abstract: ***Short-term wind power forecasting is essential for improving power system scheduling and renewable energy integration. However, practical wind farm data are frequently affected by outliers, missing values, redundant meteorological variables, and highly nonlinear temporal characteristics. To address these challenges, this paper proposes a short-term wind power forecasting framework based on data-quality awareness and global local feature collaboration. DBSCAN is employed for outlier detection, KNN imputation is adopted to recover missing values, and correlation analysis is performed to select informative meteorological variables. An improved Transformer incorporating multi-head differential attention is then combined with a temporal convolutional network to jointly capture global temporal dependencies and local fluctuation patterns. Experiments conducted on a real-world SCADA dataset demonstrate that the proposed approach consistently outperforms CNN, TCN, Transformer, and Informer models under multiple forecasting horizons. At the 48-step forecasting horizon, the proposed method reduces MAE and RMSE by 11.7% and 14.5%, respectively. Ablation studies further verify the effectiveness of each component. The proposed framework provides an effective and robust solution for short-term wind power forecasting.***

Keywords: short-term wind power forecasting; data quality awareness; global–local feature collaboration; improved Transformer; temporal convolutional network

1. Introduction

With the global transition toward low-carbon energy systems and the continuous growth of renewable energy installations, wind power has become one of the most important renewable energy sources in modern power systems. However, wind power generation is jointly affected by meteorological conditions, terrain characteristics, seasonal variations, and turbine operating states. As a result, wind power output usually exhibits strong randomness, intermittency, and fluctuation. As the penetration of wind power in power grids continues to increase, forecasting errors may directly affect frequency regulation, reserve capacity allocation, economic dispatch, and electricity market operation. Therefore, developing accurate short-term wind power forecasting models is essential for improving renewable energy accommodation and maintaining the secure and stable operation of power systems [1–4].

Short-term wind power forecasting generally aims to predict power variations over the next several hours to several days. Compared with medium- and long-term forecasting, short-term forecasting places greater emphasis on the model's ability to respond to recent meteorological changes and power fluctuations. Compared with ultra-short-term forecasting, it also requires the model to maintain reliable trend-tracking capability over longer forecasting horizons. In real-world wind farm SCADA data, outliers, missing values, and heterogeneous variables with different scales often coexist. If raw data are directly fed into forecasting models, abnormal samples may weaken the model's ability to learn real operating patterns, missing values may disrupt temporal continuity, and redundant features may introduce irrelevant noise.

Existing wind power forecasting methods can be broadly divided into physical models, statistical models, and data-driven models. Physical models rely on numerical weather prediction, terrain parameters, and aerodynamic modeling. Although they have a certain degree of interpretability, their modeling process is complex and sensitive to initial conditions. Statistical models, such as ARIMA and SVM, are relatively simple to implement, but their representation ability is limited when dealing with nonlinear, high-dimensional, and non-stationary wind power sequences. In recent years, deep learning methods have been widely applied to short-term wind power forecasting because of their ability to automatically learn complex temporal features [5–16]. Among them, Transformer models can capture long-range dependencies through self-attention mechanisms, while temporal convolutional networks (TCNs) can model local temporal patterns through causal and dilated convolutions. Both structures show promising potential for wind power forecasting.

Nevertheless, using Transformer alone may cause the model to overlook local fluctuation details, whereas using TCN alone may be insufficient for modeling global dependencies across distant time periods. In addition, the standard encoder–decoder architecture of Transformer was originally designed for sequence-to-sequence modeling tasks. When directly transferred to wind power forecasting, it may introduce structural redundancy and error accumulation. To address these limitations, this paper investigates short-term wind power forecasting from two perspectives: data quality awareness and global–local feature collaboration. Without changing the definition of the forecasting task, data cleaning, correlation-based feature selection, improved Transformer-based global modeling, and TCN-based local modeling are integrated into a unified forecasting procedure.

The main contributions of this paper are summarized as follows.

1) A data-quality-aware wind power data processing procedure is proposed. DBSCAN is used to identify abnormal samples, KNN imputation is adopted to recover missing values, and correlation analysis is employed to select key meteorological variables, thereby reducing the influence of low-quality data and redundant features on the forecasting model.
2) A global–local feature collaborative forecasting model is constructed. An improved Transformer with multi-head differential attention is introduced to extract long-range global dependencies, while a TCN module is incorporated to capture local temporal fluctuation features.
3) Multi-step forecasting experiments are conducted on a real-world single-turbine wind farm dataset. The proposed method is compared with CNN, TCN, Transformer, and Informer models, and ablation experiments are further performed to verify the effectiveness of each module.

2. Related Work

2.1 Wind Power Forecasting and Data Quality Processing

Wind power forecasting methods can be classified according to either forecasting horizons or modeling strategies. In terms of time scale, they are generally divided into ultra-short-term, short-term, and medium- to long-term forecasting. In terms of modeling strategy, they can be broadly categorized into physical models, statistical models, and hybrid forecasting models. Physical forecasting methods are usually established based on wind farm geographic conditions, numerical weather prediction, and turbine aerodynamic parameters. They are suitable for newly built wind farms with limited historical data, but they are sensitive to terrain conditions, boundary settings, and initial parameters. Statistical forecasting methods learn the mapping relationship between historical observations and output power, and have been widely used in short-term forecasting tasks. However, their ability to handle complex nonlinear relationships is limited. Hybrid forecasting methods integrate data preprocessing, feature selection, and multiple forecasting models, thereby improving adaptability to complex wind power sequences [5–7].

Data quality is an important factor affecting the accuracy of wind power forecasting. Variables such as wind speed, wind direction, temperature, air pressure, humidity, and actual power are continuously collected by SCADA systems. These measurements may be affected by sensor failures, communication interruptions, extreme weather conditions, and turbine shutdowns, resulting in outliers or missing values. Existing studies commonly use statistical thresholds, clustering-based detection, and interpolation methods to handle data quality problems. DBSCAN can identify abnormal points according to sample density and is suitable for detecting outliers in irregularly distributed wind farm operating data [17]. KNN imputation recovers missing values by using information from neighboring samples, which helps preserve the continuity of input sequences [18]. Therefore, incorporating data quality processing as a preprocessing stage can improve the reliability of subsequent temporal modeling.

2.2 Deep Learning Models for Wind Power Forecasting

Traditional machine learning models, such as artificial neural networks (ANNs), support vector machines (SVMs), and ARIMA, have been applied to wind power forecasting. ANNs have nonlinear fitting capability, but they often involve a large number of parameters, relatively slow training, and limited interpretability [8]. SVMs show certain advantages in nonlinear regression tasks with small samples, but their performance depends heavily on hyperparameters such as kernel functions and penalty coefficients [9]. ARIMA can describe linear dependencies in stationary time series, but it has limited adaptability to non-stationary wind power data with strong fluctuations [10].

In recent years, deep learning models such as CNN, LSTM, GRU, and Informer have been widely used in time series forecasting and wind power forecasting tasks. CNNs are effective in extracting local patterns, but they are less capable of modeling long-range temporal dependencies [13]. LSTM and GRU can model temporal dependencies through gated structures, but they still suffer from low training efficiency and memory decay when processing long sequences [11–12]. Informer improves the attention mechanism for long-sequence forecasting and has advantages in reducing computational complexity [16]. Although these methods improve forecasting performance from different perspectives, it remains necessary to simultaneously consider global trends, local fluctuations, and data quality problems in real-world wind farm data.

2.3 Global–Local Feature Collaborative Temporal Modeling

Transformer establishes dependencies between arbitrary positions in a sequence through the self-attention mechanism. Compared with recurrent structures, it has higher parallel computing efficiency and stronger global modeling capability [15]. In wind power forecasting, power sequences are affected by the lag effects of meteorological variables and continuous temporal variations, which leads to long-range correlations across different time periods. Therefore, Transformer is suitable for extracting global trend features. However, the standard Transformer was originally designed for general sequence modeling. When directly applied to wind power forecasting, it may suffer from insufficient local fluctuation modeling, structural redundancy, and interference from irrelevant contexts.

TCN is a convolution-based temporal modeling network. It uses causal convolution to avoid information leakage from future time steps, dilated convolution to enlarge the receptive field, and residual connections to improve the training stability of deep networks [14]. Compared with Transformer, TCN is more effective in capturing local variations and periodic fluctuations within short time windows. Wind power sequences usually contain both macroscopic trends and short-term disturbances. Therefore, combining the global dependency modeling capability of Transformer with the local feature extraction capability of TCN can provide a more suitable global–local collaborative modeling strategy for wind power forecasting.

3. Method

3.1 Task Definition and Overall Framework

Let $X = x_1, x_2, \ldots, x_L$ denote the multivariate meteorological and operational feature sequence of a single wind turbine within a historical observation window, where $x_t \in \mathbb{R}^d$ represents the d-dimensional input feature vector at the t-th sampling time. The objective of short-term wind power forecasting is to predict the future power sequence $\hat{Y} = \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_H$ based on the historical window of length L, where H denotes the forecasting horizon. In this paper, a multi-step forecasting setting is adopted, and model performance is evaluated under forecasting horizons of H =48, 64, and 96.

The proposed method is designed around two aspects: data quality awareness and global–local feature collaboration. First, outlier detection, missing value imputation, and normalization are performed on the raw SCADA data to obtain more reliable input sequences. Second, correlation analysis is used to select wind speed and wind direction features that contribute more strongly to power forecasting, reducing the interference of redundant variables. Third, the processed sequence is fed into the improved Transformer module, where multi-head differential attention is used to capture global temporal dependencies. Finally, a TCN layer is employed to further extract local fluctuation features, and a fully connected layer is used to output the predicted wind power sequence. The overall model architecture is shown in Fig. 1.

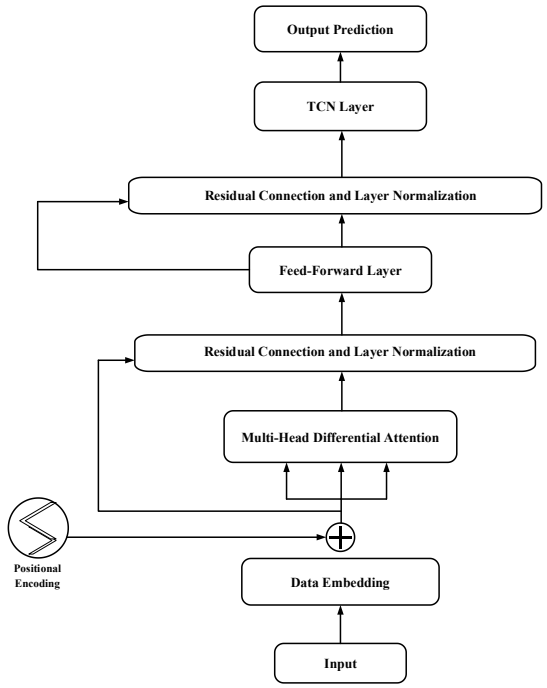

Figure 1: Architecture of the data-quality-aware and global – local feature collaborative forecasting model

3.2 Data-Quality-Aware Preprocessing

Wind farm SCADA data are often affected by sensor conditions, communication quality, and turbine operating states, which may lead to outliers and missing values. The goal of data-quality-aware preprocessing is to reduce the influence of low-quality samples on model training, enabling the forecasting model to better learn the underlying power variation patterns. In this paper, DBSCAN is first employed to detect abnormal samples [17]. This algorithm identifies core points, border points, and noise points based on the neighborhood radiu *Eps* and the minimum number of neighboring samples *MinPts*. Since it does not require the number of clusters to be predefined, DBSCAN is suitable for identifying outliers caused by abnormal operation or data acquisition errors in wind farm data.

For missing values, KNN imputation is adopted in this paper [18]. KNN imputation selects the K nearest complete samples according to sample similarity and fills the missing entries using the mean values of the corresponding features from neighboring samples. Taking Euclidean distance as an example, the distance between two samples x_i and x_j can be expressed as

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \quad (1)$$

where m is the feature dimension, and x_{ik} and x_{jk} denote the values of the $k-th$ feature in samples x_i and x_j, respectively. After outlier processing and missing value imputation, min – max normalization is used to scale variables with different units into a unified range:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Normalization prevents variables such as wind speed, wind direction, air pressure, and power from dominating model training due to differences in numerical scale, thereby improving the convergence stability of the model.

3.3 Correlation-Driven Feature Selection

Wind power is affected by multiple meteorological factors, but different variables contribute differently to actual power output. If all variables are directly used as model inputs, weakly correlated or irrelevant variables may introduce noise and increase training complexity. Therefore, this paper combines Pearson correlation coefficient and Spearman correlation coefficient for feature selection. The Pearson correlation coefficient measures the degree of linear correlation between variables, while the Spearman correlation coefficient calculates monotonic correlation based on rank information, making it more suitable for handling nonlinear relationships commonly observed in wind power data [19].

Let d_i denote the rank difference between two variables and n denote the number of samples. The Spearman correlation coefficient can be written as

$$\rho_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (3)$$

According to the correlation analysis results, wind speed variables show a strong positive correlation with actual power, wind direction variables show a certain negative correlation with actual power, while temperature, air pressure, and humidity exhibit relatively weak correlations with power. Therefore, wind speed and wind direction are mainly selected as input features in this paper to reduce the interference of redundant variables during model training.

Table 1: Main statistical characteristics of the dataset

Variable	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
Wind speed at 10 m (m/s)	6.3	3.47	0.001	21.96
Wind direction at 10 m (°)	143.1	81.97	0	355.04
Wind speed at 50 m (m/s)	7.19	4.29	0.001	22.24
Wind direction at 50 m (°)	149.6	86.15	0	355.25
Wind speed at hub height (m/s)	7.45	4.66	0.001	22.76
Wind direction at hub height (°)	157.03	87.86	0	355.61
Ambient temperature (°C)	13.49	0.49	0.001	15.33
Air pressure (hPa)	868.85	6.48	0.001	882.59
Humidity (%)	36.99	20.23	0.001	95.273
Actual power (MW)	70.31	65.99	0.03	200.05

Table 2: Correlation analysis between meteorological factors and actual power

Var.	WS10	WD10	WS30	WD30	WS50	WD50
Pearson	0.76	-0.44	0.8	-0.45	0.83	-0.47
Spearman	0.86	-0.41	0.89	-0.4	0.91	-0.42
Variable	WS70	WD70	WSH	WDH	Temp.	Press. / Hum.
Pearson	0.85	-0.48	0.85	-0.48	-0.041	-0.10/ -0.016
Spearman	0.91	-0.43	0.91	-0.43	-0.024	-0.15/ -0.012

Note: WS and WD denote wind speed and wind direction, respectively; WSH and WDH denote hub-height wind speed and hub-height wind direction.

3.4 Global–Local Feature Collaborative Forecasting Model

In the global feature modeling stage, the improved Transformer is used as the core module. The scaled dot-product attention in the standard Transformer can be expressed as

$$Attention(Q,K,V)=softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

To enhance the model's ability to select informative contexts, this paper introduces a multi-head differential attention mechanism [20]. In each attention head, two groups of query – key mappings are constructed, and the difference between the two attention maps is calculated to reduce the influence of redundant information and noisy contexts. Given the input X, its linear projections are defined as

$$\begin{aligned} &Q_1=XW_{Q_1}, Q_2=XW_{Q_2}, K_1=XW_{K_1},\\ &K_2=XW_{K_2}, V=XW_V \end{aligned} \tag{5}$$

The differential attention map is defined as

$$A_{diff}=softmax\left(\frac{Q_1K_1^T}{\sqrt{d_k}}\right)-\lambda\bullet softmax\left(\frac{Q_2K_2^T}{\sqrt{d_k}}\right) \tag{6}$$

where λ is a learnable balancing coefficient used to adjust the difference between the two attention distributions. The outputs of multiple heads are concatenated and then passed through a linear projection to obtain the final global feature representation. Compared with the standard attention mechanism, differential attention enables the model to focus more on key temporal positions and feature dimensions that are relevant to power forecasting.

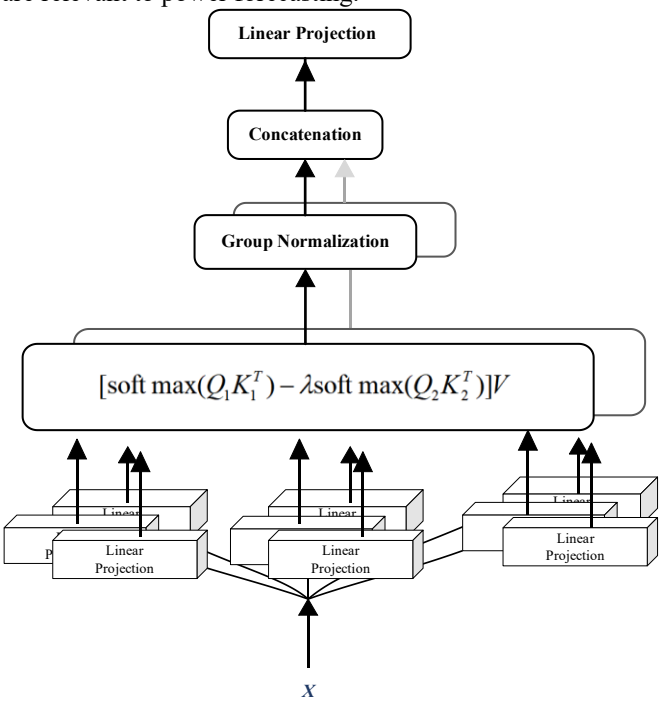

Figure 2: Structure of the multi-head differential attention mechanism

In the local feature modeling stage, TCN is introduced to further process the output of the improved Transformer. TCN uses causal convolution to ensure that future information is not used during prediction, dilated convolution to enlarge the receptive field without significantly increasing the number of parameters, and residual connections to improve the training stability of deep networks. Let k denote the convolution kernel size and d denote the dilation factor. The dilated convolution output at time t can be expressed as

$$y_t=\sum_{i=0}^{k-1}w_i\cdot x_{t-d\cdot i} \tag{7}$$

With the above design, the improved Transformer is responsible for capturing long-range dependencies and overall variation trends in wind power sequences, while TCN complements it by extracting local fluctuation information within short time intervals. These two modules form a global – local collaborative relationship, allowing the model to balance trend fitting and local error control. Finally, the model outputs the predicted power values for the next H time steps through a fully connected layer. During training, the mean squared error (MSE) is used as the loss function, and the Adam optimizer is employed to update model parameters.

4. Experiments and Results Analysis

4.1 Dataset and Experimental Settings

The experimental data were collected from the SCADA system of a real-world wind farm from January 1, 2019 to December 31, 2019. To meet the requirements of submission anonymity and data confidentiality, the specific geographical location of the wind farm is not disclosed in this paper. The dataset contains single-turbine operating data recorded at a sampling interval of 15 min, with a total of 35,040 records.

The collected variables include wind speed and wind direction at different heights, temperature, air pressure, humidity, and actual power.

The dataset is divided into training, validation, and testing sets at a ratio of 7:2:1. Specifically, data from January to August 2019 are used for training, data from September to October 2019 are used for validation, and data from November to December 2019 are used for testing.

The experiments are conducted on a Windows 11 operating system with an Intel i7 processor, 16 GB RAM, and an NVIDIA RTX4070 GPU. The model is implemented using Python 3.8. Adam is adopted as the optimizer, and mean squared error *MSE* is used as the loss function. The main hyperparameter settings are listed in Table 3. To evaluate the generalization ability of the proposed model under multi-step forecasting settings, forecasting horizons of 48, 64, and 96 are considered. CNN, TCN, Transformer, and Informer are selected as baseline models for comparison.

Table 3: Hyperparameter settings of the proposed model

Parameter	*Value*
TCN kernel size	3
TCN dilation base	2
Number of attention heads in the improved Transformer	4
Dropout rate	0.1
Batch size	128
Number of training epochs	80

Mean absolute error (MAE) and root mean square error (RMSE) are used as evaluation metrics. MAE measures the average absolute deviation between predicted and actual values, while RMSE is more sensitive to large prediction errors. Lower values of MAE and RMSE indicate higher forecasting accuracy. They are calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (9)$$

where y_i and $\hat{y}_i$ denote the actual and predicted power values, respectively, and n is the number of testing samples.

4.2 Visualization Analysis of Forecasting Results

Fig. 3 shows the predicted power curve of the proposed model and the corresponding actual power curve on a testing segment. It can be observed that the predicted results generally follow the variation trend of actual power. In particular, the proposed model achieves good fitting performance during ramp-up, ramp-down, and short-term fluctuation periods. This indicates that the data-quality-aware processing reduces the influence of abnormal inputs, while the global – local collaborative architecture enables the model to capture both overall trends and local fluctuations.

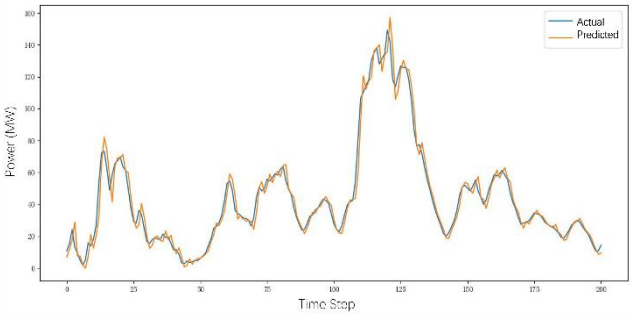

Figure 3: Comparison between the predicted values and actual values of the proposed model

Fig. 4 presents the prediction curves of different models when the forecasting horizon is 96. As the forecasting horizon increases, the uncertainty of power variation becomes more pronounced, and some baseline models exhibit obvious deviations around peaks and valleys. In comparison, the prediction curve of the proposed method is closer to the actual curve, suggesting that global – local collaborative modeling can maintain better trend forecasting capability under longer forecasting horizons.

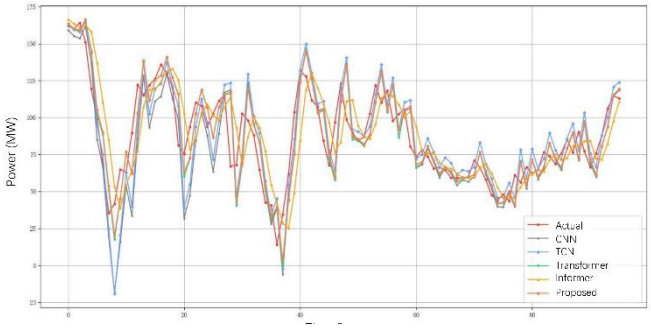

Figure 4: Comparison of forecasting results of different models when the forecasting horizon is 96

4.3 Comparison Results

The MAE and RMSE values of different models under forecasting horizons of 48, 64, and 96 are shown in Table 4. As the forecasting horizon increases, the errors of all models generally increase, indicating that information loss and error accumulation are unavoidable in multi-step forecasting. Nevertheless, the proposed method achieves the lowest MAE and RMSE under all three forecasting horizons, demonstrating stronger forecasting stability.

Table 4: Forecasting results of different models under different forecasting horizons.

Model	MAE-48	RMSE-48	MAE-64	RMSE-64	MAE-96	RMSE-96
CNN	0.836	1.014	0.923	1.192	1.324	1.626
TCN	0.812	0.958	0.862	1.02	1.073	1.248
Trans	0.761	0.915	0.814	0.961	0.983	1.137
Informer	0.724	0.852	0.786	0.938	0.884	1.072
Proposed	0.689	0.796	0.748	0.845	0.853	0.958

Compared with the standard Transformer, the proposed method reduces MAE by 9.4%, 8.1%, and 13.2% under forecasting horizons of 48, 64, and 96, respectively. The corresponding RMSE reductions are 13.0%, 12.0%, and 15.7%. These results indicate that relying solely on the global attention mechanism of Transformer is insufficient for fully

capturing local fluctuations in wind power sequences. The introduction of TCN provides complementary local temporal features, while the multi-head differential attention mechanism suppresses interference from redundant contexts and makes global feature extraction more focused. As a result, the overall forecasting accuracy is improved.

4.4 Ablation Results

To analyze the contribution of each module to forecasting performance, three ablation experiments were designed using the complete model as the reference. In the first setting, the TCN module is removed, and only the improved Transformer module is retained. In the second setting, the multi-head differential attention mechanism is removed. In the third setting, the original Transformer encoder – decoder architecture is retained without structural simplification. The experimental results are reported in Table 5.

Table 5: Ablation results

Model	MAE-48	RMSE-48	MAE-64	RMSE-64	MAE-96	RMSE-96
Ablation 1	0.741	0.885	0.791	0.933	0.924	1.092
Ablation 2	0.702	0.817	0.761	0.857	0.864	1.028
Ablation 3	0.724	0.854	0.778	0.893	0.896	1.057
Proposed	0.689	0.796	0.748	0.845	0.853	0.958

As shown in Table 5, removing the TCN module leads to a clear increase in forecasting errors, indicating that the local convolutional structure plays an important role in capturing short-term periodicity and local fluctuations. Removing the multi-head differential attention mechanism also degrades model performance, which suggests that differential attention helps the model distinguish key features from irrelevant noise more effectively. In addition, retaining the original encoder–decoder structure results in higher errors, demonstrating the necessity of appropriately simplifying the Transformer architecture for short-term wind power forecasting. Overall, the complete model achieves the best performance under all forecasting horizons, verifying the effectiveness of data-quality-aware processing and global–local feature collaborative modeling.

5. Conclusion

To address the coexistence of outliers, missing values, redundant features, and nonlinear fluctuations in real-world wind farm data, this paper proposes a short-term wind power forecasting method based on data quality awareness and global–local feature collaboration. In the proposed method, DBSCAN, KNN imputation, and normalization are first used to preprocess wind power data. Pearson and Spearman correlation coefficients are then combined to select input features that are strongly correlated with actual power. Subsequently, an improved Transformer with multi-head differential attention is adopted to extract long-range global dependencies, while a TCN module is introduced to capture local temporal fluctuations, thereby enabling multi-step forecasting of future wind power sequences.

Experiments conducted on a real-world wind farm SCADA dataset show that the proposed method outperforms CNN, TCN, Transformer, and Informer under forecasting horizons of 48, 64, and 96. The ablation experiments further demonstrate that the TCN module, the multi-head differential attention mechanism, and the structural adjustment of Transformer all contribute to improved forecasting performance. Overall, the proposed method provides an effective data-driven solution for short-term wind power forecasting by improving input data quality and jointly modeling global trends and local fluctuations.

Future research can be extended in three directions. First, spatial topological relationships among multiple turbines can be incorporated to improve the adaptability of the model to wind-farm-level power forecasting. Second, probabilistic forecasting or interval forecasting methods can be introduced to quantify the uncertainty of wind power prediction. Third, lightweight deployment strategies can be further investigated to meet the real-time and interpretability requirements of online forecasting systems in wind farms.

References

[1] International Renewable Energy Agency. Renewable Capacity Statistics 2024[R]. Abu Dhabi: IRENA, 2024.
[2] Global Wind Energy Council. Global Wind Report 2024[R]. Brussels: GWEC, 2024.
[3] Hong T, Pinson P, Wang Y, Weron R, Yang D, Zareipour H. Energy forecasting: A review and outlook[J]. IEEE Open Access Journal of Power and Energy, 2020, 7: 376-388.
[4] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation[J]. Renewable and Sustainable Energy Reviews, 2014, 32: 255-270.
[5] Wang X, Guo P, Huang X. A review of wind power forecasting models[J]. Energy Procedia, 2011, 12: 770-778.
[6] Soman S S, Zareipour H, Malik O, Mandal P. A review of wind power and wind speed forecasting methods with different time horizons[C]//North American Power Symposium. IEEE, 2010: 1-8.
[7] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power[J]. Renewable and Sustainable Energy Reviews, 2009, 13(4): 915-920.
[8] Carolin Mabel M, Fernandez E. Analysis of wind power generation and prediction using ANN: A case study[J]. Renewable Energy, 2008, 33(5): 986-992.
[9] Zeng J, Qiao W. Support vector machine-based short-term wind power forecasting[C]//IEEE/PES Power Systems Conference and Exposition. IEEE, 2011: 1-8.
[10] Chen P, Pedersen T, Bak-Jensen B, Chen Z. ARIMA-based time series model of stochastic wind power generation[J]. IEEE Transactions on Power Systems, 2010, 25(2): 667-676.
[11] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
[12] Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014: 1724-1734.

[13] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
[14] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. arXiv:1803.01271, 2018.
[15] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, Polosukhin I. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.
[16] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. Informer: Beyond efficient Transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(12): 11106-11115.
[17] Ester M, Kriegel H P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996: 226-231.
[18] Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman R B. Missing value estimation methods for DNA microarrays[J]. Bioinformatics, 2001, 17(6): 520-525.
[19] Spearman C. The proof and measurement of association between two things[J]. The American Journal of Psychology, 1904, 15(1): 72-101.
[20] Ye T, Dong L, Xia Y, Sun Y, Zhu Y, Huang G, Wei F. Differential Transformer [EB/OL]. arXiv:2410.05258, 2024.