Multi-task Driven ResNet-Transformer Framework for Wind Power Forecasting with Dynamic Mode Decomposition Correction

Yiming Liu, Wei Li, Ran Ran, Yongqi Wang, Tao Xu, and Jun Tao(✉)

College of Computer Science, Sichuan University, Chengdu 610207, China
2022141460135@stu.scu.edu.cn

Abstract. Wind power generation, as one of the most critical clean energy sources and substitutes for fossil fuels, constitutes a pivotal component in achieving carbon neutrality. In power dispatch operations, the efficiency and accuracy of wind power forecasting serve as the most crucial and influential factors for grid management and operational decision-making. Current wind power prediction methodologies predominantly rely on historical single data source, while suffering from accumulating errors over extended forecasting horizons, leading to significant degradation in prediction accuracy that fails to meet grid scheduling requirements. To address the challenges of temporal dependency and error accumulation in wind power forecasting, we propose a deep learning framework based on multi-task dynamic learning and error correction. The model employs a ResNet-Transformer hybrid architecture, where Temporal Convolutional Networks (TCN) extract localized temporal features, while Transformer encoders capture long-term temporal dependencies through self-attention mechanisms, enhanced by residual connections for hierarchical feature reuse. Experimental validation on real-world wind farm datasets demonstrates that the calibrated predictions achieve a 45.686% RMSE reduction, evidencing the method's effectiveness under complex meteorological conditions. Comparative analyses further reveal substantial improvements in both short-term (<6 h) and extended-range (>24 h) forecasting performance compared to conventional baselines, establishing technical advantages for practical grid integration applications.

Keywords: Wind power forecasting · Asymmetric history-future stacking · Multi-source meteorological data fusion · Error propagation mitigation

1 Introduction

With the sustained global growth in demand for clean energy, wind power has emerged as a critical component of renewable energy systems. However, the inherent volatility of wind energy poses significant challenges for accurate forecasting, which is essential for optimizing grid operations and reducing operational costs. National strategic documents such as the "14th Five-Year Plan for Modern Energy Systems" and the "2024 Energy Work Guidelines" emphasize the urgent need for advanced forecasting technologies to ensure sustainable wind energy development.

D.-S. Huang et al. (Eds.): ICIC 2025, LNCS 15849, pp. 458–467, 2025.
https://doi.org/10.1007/978-981-96-9881-3_38

Current forecasting approaches primarily follow two technical paradigms: (1) Mathematical-statistical models represented by ARIMA, GM, and HMM, which demonstrate proficiency in historical data processing but exhibit limited adaptability to dynamic wind conditions; (2) Machine learning models exemplified by LSTM, Transformer, and CNN, which effectively capture temporal patterns yet face challenges including the absence of future meteorological information and error accumulation in long-term predictions. Prevailing research is constrained by insufficient integration of multi-source meteorological data and lack of dynamic error correction mechanisms, resulting in compromised accuracy and reliability in wind power forecasting.

Traditional statistical models struggle with nonlinear atmospheric interactions, while modern deep learning models often neglect future meteorological constraints and suffer from error accumulation. Moreover, limited use of multi-source data and lack of dynamic correction mechanisms exacerbate these challenges. Therefore, we propose a deep learning-based forecasting approach incorporating dynamic error correction. The method introduces two key components:

(1) an asymmetric history-future stacking architecture designed to facilitate multi-scale feature fusion for improved prediction accuracy;
(2) a Dynamic Mode Decomposition (DMD)-based error correction strategy that integrates historical power data with meteorological variables to enhance forecasting stability. These components aim to address existing challenges in long-term wind power prediction and contribute to the broader goal of more accurate and reliable forecasting in support of renewable energy systems.

2 Related Work

2.1 Statistical Wind Power Forecasting

Statistical models in wind power forecasting, such as ARIMA [1], GM [2], and HMM [3], are based on historical data and linear assumptions, but they face inherent limitations when applied to dynamic and nonlinear wind conditions. For instance, Li et al. [4] improved the Hidden Markov Model (HMM) for wind speed correction, achieving a 12% reduction in MAE for hourly forecasts. However, these models still fail under turbulent conditions, as the linear wind-speed-to-power conversion induces errors. Similarly, Geng et al. [5] applied grey system theory (GM(1,1)) to reduce data randomness. Nevertheless, long-term predictions continue to experience systematic deviations due to the insufficient modeling of nonlinear meteorological interactions. These methods do not fully capture the complex dynamics between weather variables and power generation.

2.2 Machine Learning-Based Temporal Forecasting

Machine learning models, particularly deep learning methods like LSTM, Transformer, and CNN, have surpassed traditional models by capturing temporal patterns. For example, Huo et al. [6] introduced mRMR-PSO-LSTM, which integrates feature selection and optimization for LSTM, achieving an R2 of 0.89 in short-term forecasting. Recent hybrid architectures, such as the CNN-LSTM framework proposed by Malakoudi et al.

[7], further demonstrate the potential of combining spatial feature extraction with temporal modeling for wind power generation prediction, achieving robust performance under varying wind regimes., these models often rely solely on historical power data, neglecting meteorological factors that can influence wind power generation. Ye et al. [8] proposed WD-IGFCM-LSTMS, leveraging wavelet decomposition and an enhanced Grey Wolf Optimizer, which improved RMSE by 19% for 72-h forecasts. To address the limitations of fixed temporal windows, Liu et al. [9] proposed a VMD-GRU-Transformer framework that decomposes wind power signals into multi-scale components, enabling adaptive modeling of ultra-short-term dynamics. However, their approach does not integrate future meteorological constraints, which remain critical for extended-range forecasting. Despite these advancements, long-term forecasts still suffer from error propagation due to fixed sliding windows and lack of consideration for future meteorological conditions.

2.3 Multi-source Meteorological Fusion

Effective integration of multiple meteorological factors is key to improving forecasting accuracy. Liu et al. [10] developed a LightGBM-based framework to extract important factors from 15-dimensional meteorological data, achieving daily-scale NRMSE = 0.23. However, their method overlooks the temporal validity of numerical weather prediction (NWP) data. He et al. [11] proposed a feature selection method using mutual information entropy and random forest regression, reducing typhoon-season error fluctuations by 35%. Nevertheless, their approach relies on "early fusion," where feature concatenation fails to capture complex interactions between past and future features, ultimately limiting the predictive performance.

2.4 Error Correction Techniques

Error correction plays a crucial role in mitigating long-term forecast propagation. Peng et al. [12] proposed a multi-stage framework using Weibull distribution for trend extraction and Kalman filtering for transient error reduction, which improved the MAE by 22% for 48-h forecasts. Li et al. [13] applied least squares regression to model errors using historical residuals. However, their linear assumptions fail to adapt to the nonlinear dynamics present in wind power forecasting. These error correction methods rely on predefined models, limiting their ability to adapt to dynamic wind regimes. However, these solutions still lack the robustness needed for complex meteorological conditions.

To address these challenges, we introduces a deep learning framework that integrates multi-task learning and dynamic error correction for wind power forecasting. Key contributions include:

Asymmetric History-Future Stacking. The proposed heterogeneous TCN-ResNet architecture, enhanced by Transformer, enables multi-scale feature fusion to improve prediction accuracy.

DMD-Driven Dynamic Correction. Dynamic Mode Decomposition (DMD) is employed to reconstruct prediction errors, offering an effective and adaptable method to reduce long-term forecast errors.

Dynamic Weighted Loss. A dynamic weighted loss function balances multi-task objectives and reduces error propagation over time.

3 Methodology

The proposed multi-task driven ResNet-Transformer framework for wind power forecasting is designed to address the inherent challenges of long-term temporal dependency, multi-scale feature fusion, and dynamic error accumulation. It integrates several key components to capture complex spatiotemporal patterns and adapt to rapidly changing weather conditions. The core structure consists of three main modules: asymmetric history-future stacking, multi-scale feature fusion through Transformer gating, and dynamic error correction using Dynamic Mode Decomposition (DMD).

3.1 Data Preprocessing and Feature Engineering

We utilizes environmental parameters and wind turbine power generation data collected by LongYuan Power Group Corporation, encompassing 11 temporal features including wind speed, direction, and temperature. The preprocessing pipeline comprises:

Outlier Removal. Filter abnormal ROUND(A.POWER,0) and YD15(wind power indicators in the dataset) values (such as 0.0 and -754.0) for every turbine.

Temporal Alignment. Fill missing values via linear interpolation at 10-min intervals.

Feature Engineering. Extract cyclical temporal features (month, day, hour, minute). Also calculate statistical features such as the mean and standard deviation of features from nearby wind turbines. This further enriches the feature dimension and provides more comprehensive and representative input features for subsequent machine learning model training.

Normalization. Apply Z-score standardization independently to historical features ($X_1 \in R^{13}$), future meteorological features ($X_2 \in R^5$), and targets ($Y \in R^2$) to prevent data leakage.

The dataset is split into 7:2:1 ratios for training, validation, and testing. Sliding window sampling generates input-output pairs:

Input Window. 480-step (20-h) historical sequences.

Output Window. 96-step (4-h) predictions.

A stride of 76 steps enhances temporal continuity.

As shown in Fig. 1, these images depict historical data distributions of meteorological parameters (humidity, pressure, wind speed, temperature, wind direction, and rounded wind speed), highlighting trends and distribution patterns over time.

3.2 Asymmetric Multi-modal Fusion Architecture

As shown in Fig. 2, the model adopts an asymmetric history-future stacking architecture with three core modules:

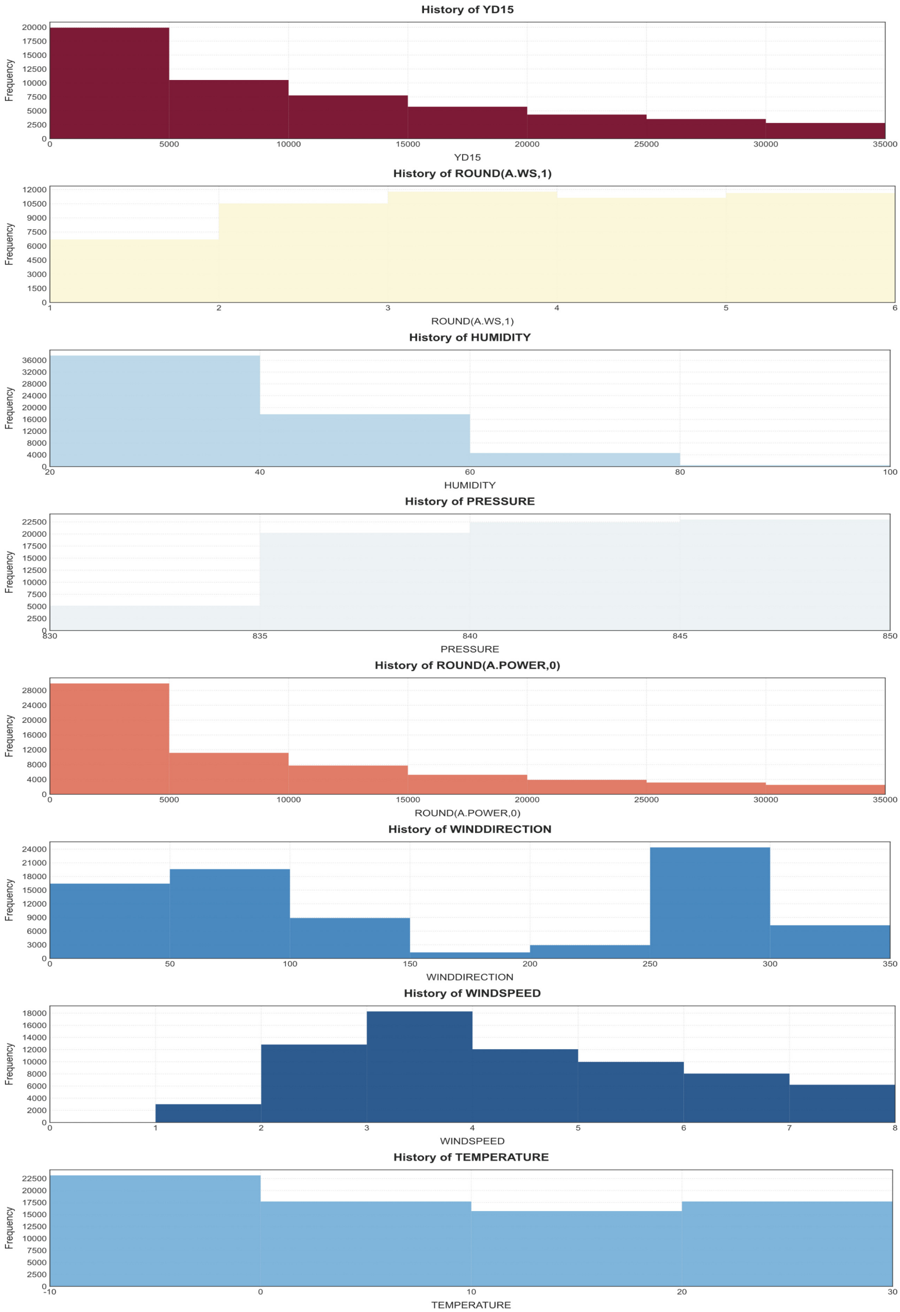

Fig. 1. Historical Data Distributions of Meteorological Parameters

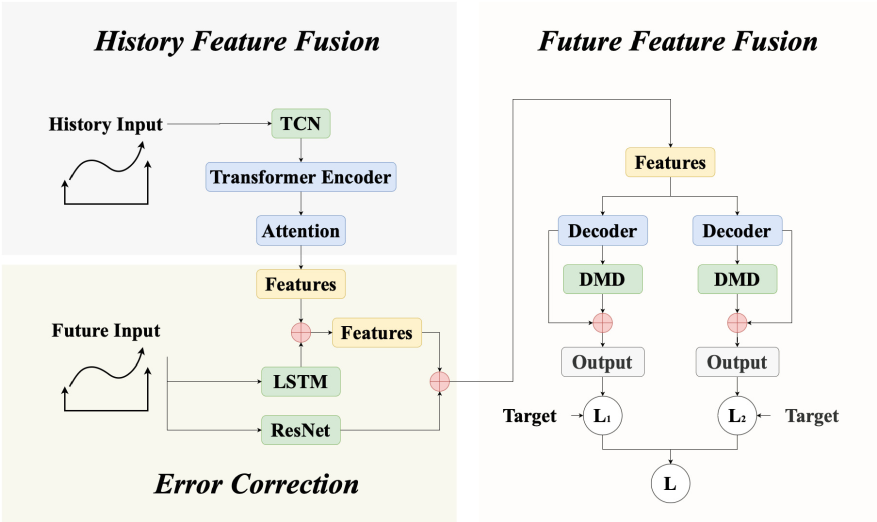

Fig. 2. Multi-modal Fusion Architecture

Historical Feature Encoder. TCN Module. Three dilated convolutional layers (kernel_size = 9) extract multi-scale local patterns with ascending channels (64 → 128 → 256), using residual connections to mitigate gradient vanishing;

Transformer Module. Using Transformer Encoder (d_model = 256, nhead = 4), where learnable positional encoding enhances temporal awareness and attention mechanisms focus on critical timestamps.

Future Feature Encoder. ResNet-1D: Processes 5D future meteorological features via 7 × 7 convolution, max-pooling, and three residual blocks (64 → 256 channels), compressed to 96D vectors through adaptive pooling.

Multi-task Fusion Layer. Dynamic Gated Fusion: Historical context vectors (256-dim) and future features (96-dim) are selectively filtered by Transformer gates before concatenation, fed into dual fully-connected branches for simultaneous prediction of ROUND(A.POWER,0) and YD15.

3.3 Dynamic Optimization and Error Correction

$$\mathcal{L}_{\text{total}} = (1 - \omega_t)\mathcal{L}_{\text{Huber}}\left(y_{\text{power}}, \hat{y}_{\text{power}}\right) + \omega_t \mathcal{L}_{\text{Huber}}\left(y_{\text{YD15}}, \hat{y}_{\text{YD15}}\right) \quad (1)$$

Dynamic Weighted Huber Loss. Where weight ω_t linearly increases from 0.5 to 0.9 during training to balance multi-task objectives.

DMD-Based Error Correction. *Error Sequence Construction:* Build Hankel matrix from test set errors.

$$H \in \mathbb{R}^{K \times L} \quad (2)$$

$$e = y_{true} - y_{pred} \tag{3}$$

Modal Decomposition. Extract dominant modes Φ and eigenvalues Λ via DMD operator A.

Error Reconstruction. Compensate predictions using reconstructed errors based on initial error e_0.

$$e_t = \Phi \Lambda^t \Phi^\dagger e_0 \tag{4}$$

4 Experiments and Results

4.1 Comparative Analysis

To evaluate the effectiveness of the proposed ResNet-Transformer framework with Dynamic Mode Decomposition (DMD) correction, we conducted a series of comparative experiments against five classical temporal modeling architectures. All models were trained and tested under identical experimental conditions to ensure a fair comparison. The experimental environment is configured as follows: It uses an NVIDIA RTX 4090 GPU for hardware support, the Adam optimizer with an initial learning rate of 0.15, and a multi-task mean squared error (MSE) loss function. Joint optimization is achieved by weighting the MSE values of the two tasks. During training, each batch consists of 512 samples. The results, presented in Table 1, reveal substantial performance gains across multiple metrics, including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), validating the superiority of the proposed approach in both short-term and long-term wind power forecasting.

Baseline Limitations. The LSTM baseline, despite its popularity for time-series tasks, exhibits significant limitations in capturing the complex temporal dynamics of wind power data, achieving an RMSE of 1.5526 Mega Watt (MW) and an MAE of 1.3607 MW. This indicates a high error rate for long-term forecasting due to limited sequence learning capability.

Hybrid Model Performance. Conventional hybrid architectures, such as CNN + LSTM and CNN + GRU, show even higher error rates, reflecting their struggle to effectively fuse spatial and temporal features. The CNN + LSTM model, for instance, suffers from a 12.14% increase in RMSE compared to LSTM alone, confirming the inherent difficulty in balancing spatial context and temporal dependencies without specialized mechanisms.

Transformer-Based Approaches. While the CNN + Transformer model demonstrates some improvement over RNN-based designs, its performance remains significantly lower than the proposed framework, with a 4.43% higher RMSE and 0.91% higher MAE, highlighting the importance of multi-scale feature fusion and dynamic error correction.

Proposed Model Superiority. The proposed framework achieves a 65.25% reduction in RMSE and a 72.83% reduction in MAE compared to the LSTM baseline, confirming its ability to capture complex temporal structures and mitigate long-term error accumulation. This significant performance boost underscores the advantages of asymmetric history-future stacking and DMD-based error correction.

Table 1. Summary of the RMSE and MAE for Contrast Experiments

Model	RMSE (MW)	MAE (MW)	RMSE (vs LSTM)	MAE (vs LSTM)
LSTM	1.5526	1.3607	–	–
RNN	1.6705	1.3872	+7.60%	+1.95%
CNN+LSTM	1.7411	1.4090	+12.14%	+3.55%
CNN+GRU	1.6248	1.3758	+4.65%	+1.11%
CNN+Transformer	1.6214	1.3731	+4.43%	+0.91%
Proposed Model	0.5394	0.3697	−65.25%	−72.83%

4.2 Ablation Study

To assess the contribution of each core component in the proposed framework, we perform a series of ablation experiments across four configurations. The Base Model serves as the baseline, employing only the original prediction architecture without incorporating future feature fusion or dynamic error correction. Building upon this, the Base + Fusion variant introduces future meteorological features to evaluate the impact of temporal information fusion, while the Base + DMD configuration integrates Dynamic Mode Decomposition (DMD) for residual error correction. Finally, the Full Model combines both enhancements—feature fusion and DMD correction—demonstrating the cumulative benefits of these components within the unified forecasting framework.

The experimental results (Table 2) demonstrate progressive performance improvements through component integration with optimal values highlighted in bold:

Table 2. Summary of the RMSE and MAE for Ablation Experiment

Configuration	RMSE (MW)	MAE (MW)	Relative Improvement (RMSE)
Base Model	1.2288	1.1248	/
Base + DMD	0.7886	0.5834	35.82%
Base + Fusion	0.8542	0.7025	30.47%
Full Model (Base + Fusion + DMD)	0.5394	0.3697	56.10%

From the experimental results, several important observations can be made:

DMD Correction Effectiveness. The standalone DMD module reduces RMSE by 35.8% and MAE by 48.1%, confirming its capability to capture residual dynamics ignored by the neural network. This aligns with wind forecasting characteristics where atmospheric persistence creates temporally correlated errors.

Feature Fusion Impact. Incorporating future meteorological features achieves 30.5% RMSE reduction, demonstrating the critical role of weather forecast integration. The

MAE improvement (37.5%) indicates better capture of extreme power fluctuations caused by weather transitions.

Synergistic Effects. The full model achieves a 56.1% reduction in RMSE compared to the base model. Notably, this improvement is not merely the additive result of individual enhancements—feature fusion and DMD correction—which yield 35.8% and 30.5% improvements respectively. Instead, the combined performance gain (exceeding either component alone but slightly less than their sum) suggests a synergistic relationship between the two mechanisms. Specifically, the feature fusion module enhances the model's capacity to learn predictive patterns from future meteorological information, while the DMD module effectively mitigates residual errors linked to system dynamics. Together, these components form a complementary structure that jointly strengthens forecasting accuracy and stability.

Error Characteristics. The MAE/RMSE ratio decreases from 0.916 (Base) to 0.685 (Full Model), indicating the combined approach particularly reduces large forecasting errors during rapid wind speed changes.

This ablation study quantitatively verifies that our dual innovation of meteorological feature fusion and Koopman operator-based correction provides multiplicative benefits for wind power forecasting. The results confirm that conventional neural architectures leave significant correctable residuals that can be addressed through physics-guided post-processing.

5 Conclusion

We introduces a multi-task learning framework with dynamic error correction for wind power forecasting, aiming to mitigate challenges related to temporal dependencies and error accumulation. The framework combines an asymmetric history-future stacking design with a ResNet-Transformer hybrid to effectively capture both local and long-range temporal features. To further enhance forecasting stability, Dynamic Mode Decomposition (DMD) is applied for error correction. Experimental results demonstrate substantial improvements, with RMSE and MAE reduced by 56.1% and 72.83%, respectively, compared to standard baselines. The proposed approach significantly outperforms existing deep learning methods, underscoring the value of multi-scale feature fusion and adaptive error correction. This framework offers a reliable and scalable solution for integrating wind power into the grid, supporting sustainable energy strategies and national goals for precision forecasting in renewable systems. Future efforts will explore architectural refinements and expanded data integration to further boost predictive performance.

References

1. Gupta, A., Sharma, K.C., Vijayvargia, A., Bhakar, R.: Very short term wind power prediction using hybrid univariate ARIMA-GARCH model. 2019 8th International Conference on Power Systems (ICPS), pp. 1–6 (2019)
2. Zhang, Y., Sun, H., Guo, Y.: Wind power prediction based on PSO-SVR and grey combination model. IEEE Access **7**, 136254–136267 (2019)

3. Ghasvarian Jahromi, K., Gharavian, D., Mahdiani, H.R.: Wind power prediction based on wind speed forecast using hidden Markov model. J. Forecast. **42**(1), 101–123 (2022)
4. Li, M., Yang, M., Yu, Y., Lee, W.J.: A wind speed correction method based on modified hidden markov model for enhancing wind power forecast. IEEE Trans. Ind. Appl. **58**(1), 656–666 (2022)
5. Geng, D., Ma, J., Zhao, Y.: Electricity generation forecasting with wind power data: a grey model approach. In: 2021 2nd International Conference on Artificial Intelligence and Information Systems (ICAIIS 2021), Article 17, pp. 1–6. Association for Computing Machinery, New York (2021)
6. Huo, X., et al.: Research of short-term wind power generation forecasting based on mRMR-PSO-LSTM algorithm. Electronics **13**(13), 2469 (2024)
7. Malakouti, S.M., Ghiasi, A.R., Ghavifekr, A.A., Emami, P.: Predicting wind power generation using machine learning and CNN-LSTM approaches. Wind Eng. **46**(6), 1853–1869 (2022)
8. Ye, L., et al.: Combined approach for shortterm wind power forecasting based on wave division and Seq2Seq model using deep learning. IEEE Trans. Ind. Appl. **58**(2), 2586–2596 (2022)
9. Liu, W., Liu, X.: Ultra-short-term wind power forecasting based on VMD-GRU-transformer. J. Res. Sci. Eng. **6**(8), 16–20 (2024)
10. Liu, S., Xiao, S., Liu, L., Liu, J.: Wind power prediction based on meteorological data visualization. GREENETS, EAI (2021)
11. He, B., et al.: A combined model for short-term wind power forecasting based on the analysis of numerical weather prediction data. Energy Rep. **8**, 929–939 (2022)
12. Peng, X., et al.: EALSTM-QR: interval wind-power prediction model based on numerical weather prediction and deep learning. Energy **220**, 119692 (2021)
13. Li, J., Liu, M., Wen, L.: Forecasting model for short-term wind speed using robust local mean decomposition, deep neural networks, intelligent algorithm, and error correction. Front. Energy Res. **11**, 1336675 (2024)