Optimization and Accuracy Improvement of Power Forecasting Models for Wind Farms Under Transient Weather Conditions

Yajie Li[1], Jia Shen[2], Sen Xu[2], Yuan Cao[1], Zhongyang Li[3,*], Meng Wang[3], Xuanyu Jin[3]
[1]Information and Communication Company of State Grid Xinjiang Electric Power Co., Ltd., 832000, Urumqi, Xinjiang, China
[2]State Grid Xinjiang Electric Power Co., Ltd, 832000, Urumqi, Xinjiang, China
[3]TsingRoc.ai Co., Ltds, 100083, Beijing, China

Corresponding author: Zhongyang Li (e-mail: 17757062402@163.com).

ABSTRACT Given the problems of non-stationary power time series, response lag, and amplified prediction errors due to sudden changes in wind direction under transitional meteorological conditions, this study proposes a wind power forecasting model that integrates multiscale time-series features, transition-aware attention mechanisms, physical constraints on turbine operation, and dynamic residual correction. To improve the model's ability to jointly characterize different scales of meteorological evolution and power lag characteristics, a feature extraction network with short-, medium-, and long-term branches based on TCN (Temporal Convolutional Network) is built, which incorporates bidirectional GRU and Transformer architectures; additionally, the feature weights at each scale are dynamically adjusted according to the severity of inflection points, and the prediction output is constrained by air density, power curves, operational status and ramping limits. Residual caching and time-delay gating are employed to correct for lag errors in the range of 1 to 6 sampling steps. 50,842 valid data sets from a wind farm were used for validation. The general MAE and RMSE of the model were 0.258 MW and 0.386 MW, respectively, and these were lower than those of the baseline TCN by 21.58% and 20.90%. In the transition period, the MAE and RMSE were 0.305 MW and 0.448 MW; these had been reduced by 22.19% and 23.29%. Therefore, the developed model can reduce peak deviations caused by abrupt changes in weather and improve the accuracy and stability of wind power forecasting under all operating conditions.

INDEX TERMS wind power forecasting; transitional meteorological conditions; multiscale temporal features; physical constraints; dynamic residual correction

I. INTRODUCTION

Wind power output is jointly influenced by fluctuations in wind speed, deviations in wind direction, changes in air density, and the control status of wind turbines. Particularly when meteorological parameters experience sudden increases, sudden decreases, or rapid shifts, the power time series exhibits distinct characteristics of non-stationarity, response lag, and local error amplification. Conventional forecasting methods struggle to balance overall accuracy during stable periods with dynamic adaptability during transition phases.Ammar and Xydis (2024) combined data preprocessing with deep learning to enhance the feature representation capability for wind speed forecasting; however, they did not sufficiently address the multi-timescale dependencies during meteorological abrupt changes [1] . Jamii et al. (2024) employed dimension-reduction principal component analysis and random forests for medium-term wind power forecasting, alleviating the redundancy issue of high-dimensional meteorological variables; however, static feature mapping struggles to capture the dynamic transition of feature contributions before and after transition phases [2] . Rangaraj et al. (2024) performed statistical post-processing on numerical weather prediction data using distribution scaling, thereby improving meteorological input biases; however, they did not further consider turbine inertia and operational boundary constraints when meteorological errors propagate to the power forecasting stage [3] . Parmar and Pieper (2024) used a radial basis function neural network to filter random wind speed data, enhancing noise suppression capabilities; however, they lacked a mechanism to jointly distinguish between effective changes and random perturbations in wind speed abrupt changes [4] .Kumar et al. (2024) constructed an optimized recurrent neural network to process wind power time series, improving time-series forecasting performance; however, a single recurrent structure still struggles to simultaneously capture short-term transitions, power lag, and long-period meteorological evolution [5] .

The deficiencies in the previous studies are mainly as follows: (1) Most studies use a single time window or static feature weights, which are not suitable for adapting to the rapid changes in the contribution of short-, medium- and long-scale information during transition periods. (2) Data-driven forecasting is not effectively integrated with wind turbine power curves, rated capacity, start-stop status and ramping constraints, and is prone to producing out-of-bounds values and abnormal power peaks. (3) There is a lack of specific compensation for lag errors of 1 to 6 sampling steps between meteorological changes and power response, and errors continue to accumulate during periods of sudden change. Given the above problems, a prediction model has been constructed that incorporates multi-scale, multi-branch time-series feature extraction, transition-aware dynamic attention, physical constraints on wind turbine operation, and dynamic residual correction. Based on the experimental results, the optimised model has achieved an overall MAE and RMSE of 0.258 MW and 0.386 MW, respectively, and during transition periods, these values were reduced to 0.305 MW and 0.448 MW; that is, they decreased by 21.58%, 20.90% and 22.19% compared with the baseline TCN model. Thus, the peak error during sudden changes has been reduced while maintaining operational feasibility, and more adaptive technical support for wind power forecasting and dispatch decisions in adverse weather conditions has been provided.

II. MODELING THE CHARACTERISTICS OF WIND POWER FLUCTUATIONS UNDER TRANSIENT METEOROLOGICAL CONDITIONS

Various weather conditions affect both the size and direction of wind power generated by wind turbines through all sorts of factors, such as changes in wind speed and direction, different atmospheric pressures, inertia of the turbine, etc.; therefore, at 10-minute intervals, the timing of wind speed, wind direction, temperature, atmospheric pressure, humidity and historical power data were synchronized. Sliding windows of 3, 6 and 12 steps were used to extract the mean, standard deviation, first-order difference, second-order difference, skewness and lags of 1 to 6 steps. The Transitions are as follows:

$$I_t = \alpha_1|\Delta \upsilon_t| + \alpha_2|\Delta^2 \upsilon_t| + \alpha_3|\Delta\theta_t| + \alpha_4|\Delta p_t| \tag{1}$$

where I_t is the transition intensity at time t; $\Delta \upsilon_t$ and $\Delta^2 \upsilon_t$ are the first- and second-order changes in wind speed, respectively; $\Delta\theta_t$ is the change in wind direction; Δp_t is the change in air pressure; and $\alpha_1 \sim \alpha_4$ is the normalized weight. Furthermore, I_t, the cubic term of wind speed υ_t^3, historical power P_{t-k}, turbulence intensity TI_t, and turbine status S_t are combined to form a feature vector. After missing value masking and Z-score normalization, this vector serves as the unified input for the multiscale, multi-branch prediction network [6] .

III. METHODS FOR OPTIMIZING AND IMPROVING THE ACCURACY OF WIND POWER FORECAST MODELS UNDER TRANSIENT METEOROLOGICAL CONDITIONS

A. MULTI-SCALE, MULTI-BRANCH TIME-SERIES FEATURE EXTRACTION NETWORK

In order to capture the sudden changes in wind velocity, power inertia, and long term weather evolution, a parallel time-series network with short, medium and long term branches was designed. The short-run branch uses a 2-layer dilated causal convolutional network with a kernel size of 3 and an expansion ratio of 1 and 2. The intermediate branch adopts a 12-step sequence and uses a 64 unit bidirectional GRU to model the lag response in a 2 hour time frame; the long branch takes a 48-step sequence as input and uses a 2-layer, 4-head, 128-dimensional transformer to encode 8-hour dependencies [7]. The output of the three branches is mapped into 64 dimensions via 1 × 1 convolutions, aligned along the time dimension, and concatenated:

$$H_t = \mathrm{Concat}\left(H_t^s, H_t^m, H_t^l\right) \tag{2}$$

where H_t^s, H_t^m, H_t^l represent short-, medium-, and long-scale features, respectively; C $C(\cdot)$ denotes the concatenation operation; and H_t is the 192-dimensional joint representation. At the end of each branch, a layer normalization layer and an activation rate of 0.1 are applied to suppress overfitting, and the associated features are fed into the following Transcendental Perceptual Dynamic Attention module. Figure 1 illustrates the overall process flow.

B. TRANSITION-AWARE DYNAMIC ATTENTION FUSION MECHANISM

To address the problem of nonlinear shifts in the contributions of short-, medium-, and long-scale features before and after meteorological transitions, a transition-state-driven time-scale dual-layer attention module is proposed based on the 64-dimensional output from the three branches [8]. First, the first-order and second-order changes of wind speed, angular wind speed, pressure gradient, turbulence intensity and historical power change rate are combined to form a 6-dimensional transition state vector c_t. A two-layer perceptron containing 32 neurons is then used to generate the transition gate:

$$r_t = \sigma(W_{r,2}\,\mathrm{GELU}(W_{r,1}c_t + b_{r,1}) + b_{r,2}) \tag{3}$$

where r_t is the transition sensitivity coefficient at time t, $W_{r,1}, W_{r,2}$ is the trainable weight matrix, $b_{r,1}, b_{r,2}$ is the bias term, $\mathrm{GELU}(\cdot)$ is the nonlinear activation function, and $\sigma(\cdot)$ is the sigmoid function. Then, the features from the three branches are concatenated into a 192-dimensional sequence. Four attention heads, each with 48 dimensions, are constructed, and the transition gate, the relative time

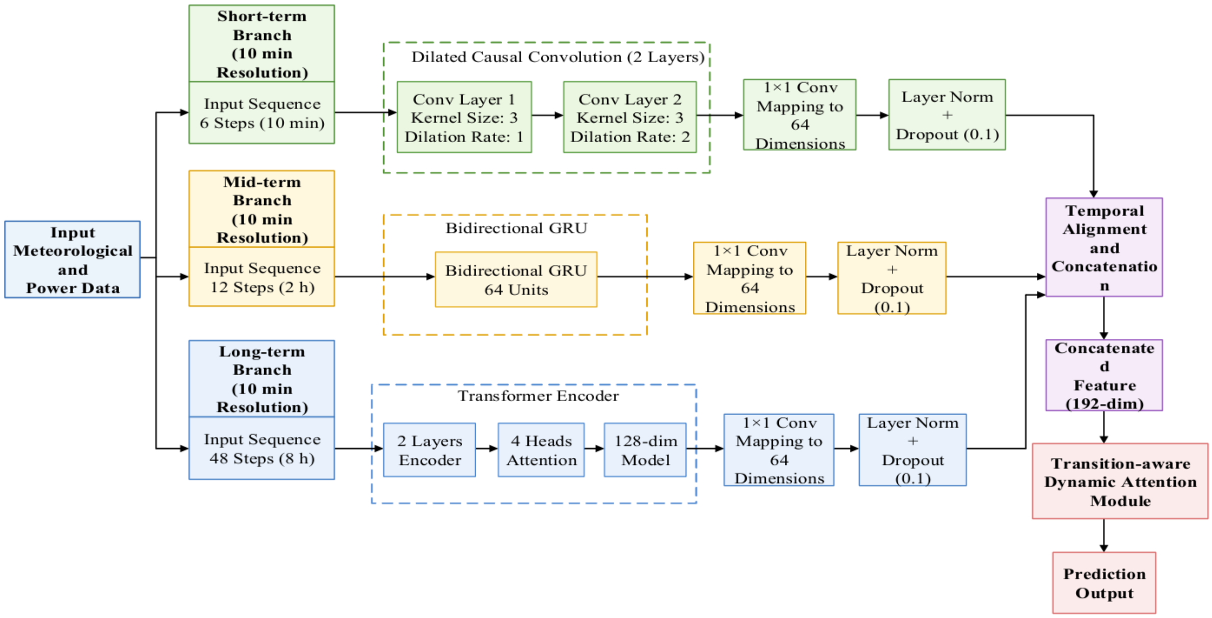

Fig. 1. Flowchart of the Multiscale, Multibranch Temporal Feature Extraction Network.

distance, and the transition neighborhood mask are embedded in the attention scores:

$$A_t^{(h)} = \mathrm{Softmax}\left(\frac{Q_t^{(h)} K_t^{(h)T}}{\sqrt{48}} + \gamma_h r_t M_t - \lambda_h D_t\right) \quad (4)$$

where $Q_t^{(h)}, K_t^{(h)}$ are the query and key matrices for the th and h th attention heads, respectively; M_t is the transition neighborhood mask; D_t is the relative time distance matrix; and γ_h and λ_h are the transition enhancement coefficient and time decay coefficient, respectively. Scale weights are further computed using the three-branch pooled features:

$$g_t = \mathrm{Softmax}\left(W_g \left[\widetilde{H}_t^s \| \widetilde{H}_t^m \| \widetilde{H}_t^l \| r_t\right] + b_g\right) \quad (5)$$

$$F_t = \mathrm{LN}\left(W_f \left[g_t^s Z_t^s \| g_t^m Z_t^m \| g_t^l Z_t^l\right] + W_o H_t\right) \quad (6)$$

In the equation, $\widetilde{H}_t^s, \widetilde{H}_t^m, \widetilde{H}_t^l$ represents the three-branch pooled features, g_t^s, g_t^m, g_t^l represents the dynamic scale weights, Z_t^l represents the temporal attention output of the th l branch, W_f represents the fusion mapping matrix, $W_o H_t$ represents the residual term, $\mathrm{LN}(\cdot)$ represents layer normalization, and F_t represents the fused features fed into the output layer that incorporates the physical constraints of wind turbine operation.

C. POWER PREDICTION OUTPUT LAYER INCORPORATING PHYSICAL CONSTRAINTS OF WIND TURBINE OPERATION

In order to solve the problems of purely data-driven outputs — for example, rated power over the limit, wrong classification of the shutdown state, and distortion of the rate of change in the transition phase — a three-tiered output layer "Neural Predictive — Mechanism Correction — Operational Domain Projection" was designed [9] . The 192-dimensional fused features F_t pass sequentially through fully connected layers with 128, 64, and 1 units, respectively. The first two layers use the GELU activation function, layer normalization, and a deactivation rate of 0.1 to generate an initial power estimate:

$$\widetilde{P}_t = P_r \sigma(W_3 \mathrm{GELU}(W_2 \mathrm{GELU}(W_1 F_t + b_1) + b_2) + b_3) \quad (7)$$

where $\widetilde{P}_t$ is the initial predicted power, P_r is the wind turbine's rated power, W_1, W_2, W_3 and b_1, b_2, b_3 are the trainable weights and biases for the three layers, respectively, and $\sigma(\cdot)$ is the Sigmoid function. Combining the 10-minute meteorological input, the air density is calculated based on atmospheric pressure ρ_t and temperature T_t:

$$\rho_t = \frac{p_t}{R_d\left(T_t + 273.15\right)} \quad (8)$$

where ρ_t is the real-time air density, R_d is the dry air gas constant, and 273.15 is used to convert from Celsius to Kelvin. The density-corrected wind speed is expressed as:

$$v_t^c = v_t \left(\frac{\rho_t}{\rho_0}\right)^{1/3} \quad (9)$$

where v_t and v_t^c represent the measured wind speed and the corrected wind speed, respectively, and ρ_0 is the standard air density. Based on the cut-in, rated, and cut-out wind speeds, a reference value for the adjustable power curve is constructed P_t^{PC}, and by incorporating start/stop, fault, power limit, and available capacity states, an operating mask is formed m_t, resulting in the final output:

$$\widehat{P}_t = m_t \Pi_{[0,P_t^{\max}]}\left(\widetilde{P}_t + \beta_t\left(P_t^{PC} - \widetilde{P}_t\right)\right) \tag{10}$$

where $\widehat{P}_t$ is the constrained predicted power, β_t is the learnable correction coefficient, $P_t^{\max}$ is the current upper limit of available power, and $\Pi(\cdot)$ is the interval projection operator. During training, a gradient penalty from the adjacent prediction step is incorporated:

$$\begin{aligned} L_{\text{phy}} = \lambda_b &\left[\max\left(0, \widehat{P}_t - P_t^{\max}\right)\right]^2 \\ &+ \lambda_r\left[\max\left(0, |\widehat{P}_t - \widehat{P}_{t-1}| - R_t^{\max}\right)\right]^2 \end{aligned} \tag{11}$$

where λ_b and λ_r are the weights for the power boundary and the gradient constraint, respectively, and $R_t^{\max}$ is the maximum allowable power change within a 10-minute step. The constrained baseline prediction sequence is further fed into the dynamic residual correction module; the complete calculation process is shown in Figure 2.

D. DYNAMIC RESIDUAL CORRECTION METHOD FOR LAG ERRORS DURING TRANSITION PHASES

To address the problem of a 1-610-second lag in the power response following a sudden increase, decrease or shift in wind speed or direction, a dynamic residual correction module based on error memory and time-delay gating was designed, and an online residual buffer of length 6 is maintained. The input sequence contains the predicted values of physical constraints, 192-dimensional fused features, transition intensity and its rate of change, wind speed rate of change, power slope, and the residuals of the previous 6 steps. First, causal convolution with a kernel size of 3 and 32 channels is used to extract local error evolution; this is then fed into a 64-unit GRU to encode the direction, magnitude and persistence of the residuals over a 60-minute period [10], [11]. The time-delay gating layer uses Softmax to generate six normalised lag weights and a Sigmoid function to adjust the amplitude; the final power is expressed as:

$$\begin{aligned} P_t^* = \Pi_{[0,P_t^{\max}]}\Bigg(&\widehat{P}_t + \sum_{k=1}^{6} \alpha_{t,k} e_{t-k} \\ &+ \beta_t \tanh(W_e h_t + b_e)\Bigg) \end{aligned} \tag{12}$$

where P_t^* is the corrected predicted power, $\widehat{P}_t$ is the physically constrained output, $\Pi_{[0,P_t^{\max}]}(\cdot)$ is the power feasibility domain projection, e_{t-k} is the predicted residual from the k th historical step, $\alpha_{t,k}$ is the corresponding time-delay weight (with the sum of the six terms equaling 1), h_t is the GRU hidden state, W_e and b_e are residual mapping parameters, and β_t is the dynamic correction coefficient. During the training phase, only historical observable residuals are used to update the parameters; in the reasoning stage, a recursive residual buffer is used to prevent the introduction of future information [12]. The relationship between the baseline error and the corrected value can be verified by a scatter graph, as illustrated in Figure 3, which can be used to determine the residual loss weight in a phased joint training.

E. PHASED JOINT TRAINING AND MODEL PARAMETER OPTIMIZATION STRATEGY

This model uses a three step progressive training method to reduce gradient conflicts due to multiple modules being updated simultaneously [13] . In the first stage, the dynamic attention and residual correction modules are frozen, and the multiscale feature network and physical constraint output layer are trained for 30 epochs, with a batch size of 64, an initial AdamW learning rate of 1×10^{-3}, and weight decay set to 1×10^{-4};In the second stage, the parameters of the attention modules are unfrozen, and the model is trained for 40 epochs with the learning rate reduced to 5×10^{-4} to ensure that the switching gates, scale weights, and power boundaries converge in a coordinated manner; in the third stage, a residual buffer of length 6 is incorporated, and the model undergoes 20 epochs of joint fine-tuning with the learning rate adjusted to 1×10^{-4} and the gradient norm clipping threshold set to 1.0. The total loss is defined as:

$$L = L_{\text{pred}} + \lambda_T L_{\text{turn}} + \lambda_p L_{\text{phy}} + \lambda_r L_{\text{res}} \tag{13}$$

In the equation, L_{pred} represents the total power prediction loss, L_{turn} represents the weighted error during the transition phase, L_{phy} represents the loss associated with power boundaries and ramping constraints, L_{res} represents the dynamic residual correction loss, and λ_T, λ_p and λ_r are the corresponding weights. The training process uses 5 cycles of linear warming and cosine annealing to update the learning rate. If the validation loss does not decrease over 10 consecutive cycles, and the optimum parameters are preserved, the training will be terminated prematurely [14]. The joint loss and learning rate changes for each stage can be illustrated using side-by-side comparison curves, as shown in Figure 4.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL DATASET, MODEL PARAMETERS, AND EVALUATION METRICS

In this experiment, a combined dataset of SCADA and meteorological monitoring data from a wind farm was constructed. Operational records spanning 12 consecutive months were selected, with a sampling interval of 10 minutes, yielding a total of 52,560 sets of raw samples, of which 50,842 sets constituted valid sample data. Each data set has 10 variables: wind speed, wind direction, temperature, atmospheric pressure, humidity, turbulence intensity, rotor speed, pitch angle, turbine status and active power. Time alignment, outlier removal and missing value imputation were carried out, and 50,842 valid samples were retained, including 2,146

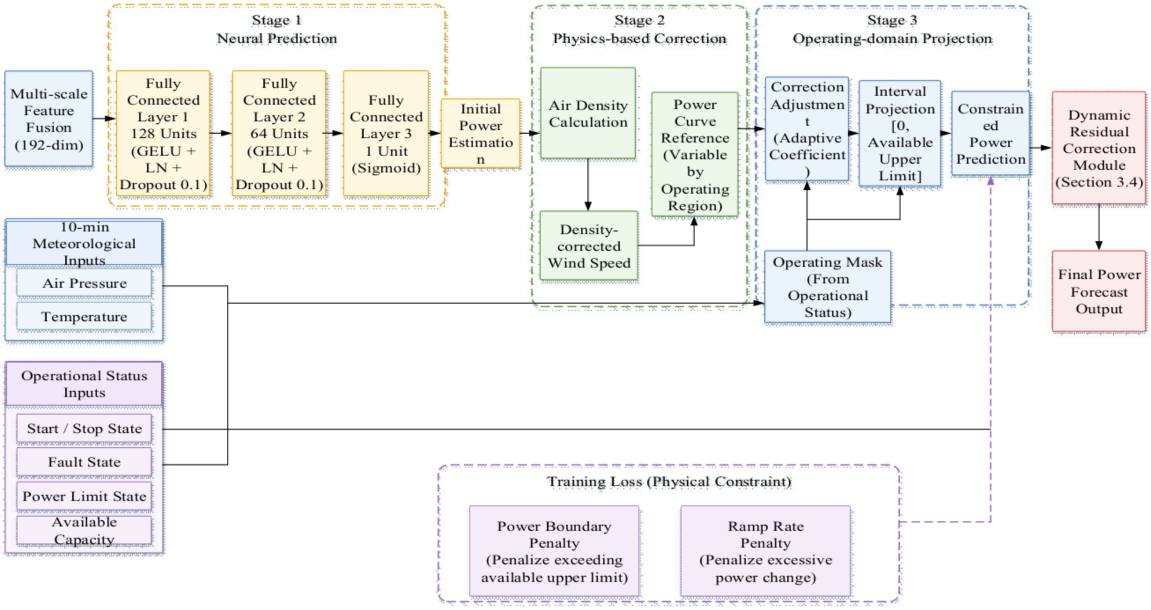

Fig. 2. Flowchart of the power forecast output layer incorporating physical constraints on wind turbine operation.

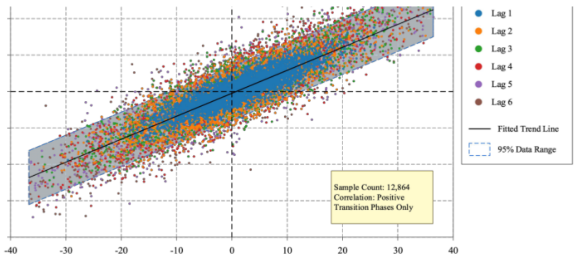

Fig. 3. Scatter plot of baseline prediction error versus dynamic residual correction during the inflection phase.

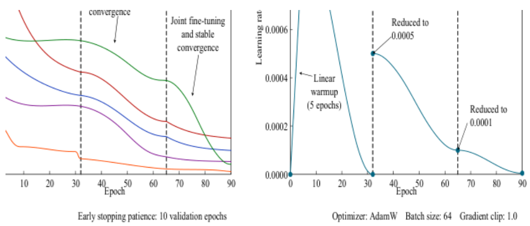

Fig. 4. Comparison of joint training loss and learning rate changes by stage.

samples of sudden wind speed increases, 1,983 samples of sudden wind speed decreases, wind direction deviations and 1,106 samples of sudden changes in atmospheric pressure. The data was divided into training, validation and test sets in the proportions of 70%, 15% and 15% respectively. The input lengths for the short, medium, and long branches were set to 6, 12, and 48, respectively; the number of attention heads was set to 4; the batch size was 64; and the initial learning rate was 1×10^{-3}. Evaluation indicators included MAE, RMSE, MAPE and R^2, and separately calculated errors for the whole period and the transition stage to assess the general forecasting performance of the model and how well it maintained accuracy during sudden changes in weather conditions [15], [16] .

B. ANALYSIS OF POWER FORECAST RESULTS UNDER DIFFERENT TRANSIENT METEOROLOGICAL CONDITIONS

In order to assess the adaptability of the model to various weather changes, 322 sets of wind speed surge samples, 297 wind velocity drop samples, 229 wind direction deviation samples, 166 air pressure variation samples were obtained. The calculation of MAE, RMSE, MAPE, and R^2 was carried out with the same input window, predicted step size, and parameter settings; the results are presented in Table 1.

As illustrated in Table 1, MAE and RMSE are 0.274 MW and 0.406 MW respectively, with a MAPE of 6.52% and 0.949. This shows that, after correcting for air density, it is relatively stable to map the variation in the atmosphere to the effective wind velocity and the power output. The MAE for the sudden wind speed rise condition is 0.286 MW, which is 0.026 MW less than the RMSE, which shows that the time pattern of the turbine’ s power ramp-up is relatively distinct; the sudden drop in wind speed is caused by the inertia of the rotor, the pitch control, and the control delay, resulting in a rise of the MAPE to 7.26%. The deviation of the wind direction showed the greatest error, with the MAE and the RMSE reaching 0.347 MW and 0.502 MW respectively, while the WHR fell to 0.927. This was primarily due to response lag in the yaw system; changes in wind direction also simultaneously alter the angle of attack, wake superposition, and turbulence distribution. Although there is a difference in error between the four states, the MAPE remains within 8.11%, suggesting that multi-scale features, transitional attention, and dynamic residual correction can maintain the continuity and effectiveness of the prediction

TABLE 1. Power Forecast Results Under Different Meteorological Transition Conditions

Transitional Weather Conditions	Number of Test Samples/Groups	MAE/MW	RMSE/MW	MAPE/%	R^2
Sudden Increase in Wind Speed	322	0.286	0.421	6.84	0.945
Sudden Drop in Wind Speed	297	0.312	0.458	7.26	0.938
Wind direction deviation	229	0.347	0.502	8.11	0.927
Sudden change in air pressure	166	0.274	0.406	6.52	0.949

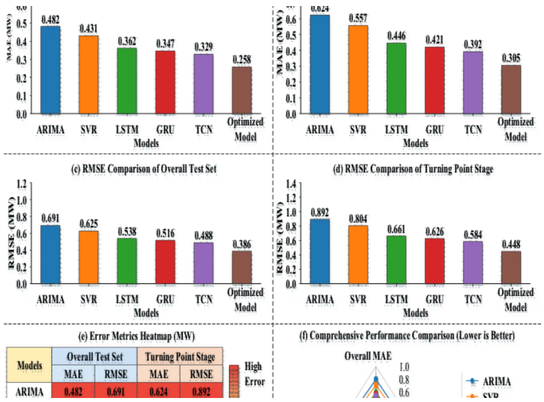

Fig. 5. Comparison of Accuracy for Different Power Forecasting Models Across the Overall and Transition Phases.

sequence [17], [18] .

C. COMPARISON OF OVERALL AND TRANSITION-PHASE ACCURACY AMONG DIFFERENT POWER FORECASTING MODELS

Using the same training set, test set, 10-minute forecast horizon and parameter optimisation conditions, ARIMA, SVR, LSTM, GRU and TCN were selected as the comparison models. The left and right figures are the MAE and RMSE of the entire test set and the transition period, respectively, as shown in Figure 5.

Figure 5 shows that all deep time series models perform better than ARIMA or SVR. The optimized model reduces the total MAE from 0.329 MW to 0.258 MW, which is a 21.58% reduction in RMSE and 20.90% in the RMSE .On the right, the error rate in the transitional phase is usually higher than in the whole period; however, the MAE and RMSE of the optimized model are reduced by 22.19% and 23.29%, respectively, compared with TCN, and by 0.141 MW and 0.213 MW, respectively, compared with LSTM. This shows that multi-scale feature extraction, dynamic fusion, physical constraint output, and lag residual correction are complementary to each other [19], [20] .

D. ABLATION OF KEY OPTIMIZATION MODULES AND ACCURACY GAIN ANALYSIS

In order to isolate the influence of each optimized module on the prediction error, the multi-scale branch, the inflection-aware attention, the physical constraint output, the dynamic residual correction were removed, and the number of training iterations, the size of the batch, the learning rate, and the data split constant [21], [22] . The results are shown in Table 2.

Table 2 shows that the MAE rose from 0.305 MW to 0.386 MW in the transitional period, and the RMSE rose to 0.558 MW, representing an increase of 26.56% and 24.55%, respectively. This shows that the dynamic scale weighting has a direct impact on the choice of the feature during the abrupt transition period. The removal of the multi-scale branch has led to a total increase in RMSE of 0.065 MW, suggesting that a single time window is difficult to take into account both the short term and the long term weather evolution. When the dynamic residual correction is removed, the RMSE in the transitional phase has increased by 0.078 MW, reflecting the fact that the delay in the power response of Steps 1-4 continues to be a significant source of error accumulation. The physical constraint module had a relatively small effect on the total error; however, its removal resulted in an increase in the RMSE of the transitional phase to 0.493 MW, suggesting that the operational border projection could suppress the abnormal power peaks [23]. In addition, the cumulative precision increase was calculated according to the sequence of integration of the modules; the results are presented in Table 3.

Table 3 indicates that after the four modules have been integrated, the error shows a steady decreasing trend. In comparison with the baseline TCN, the total MAE and RMSE were reduced by 21.58% and 20.90%, and the MAE and RMSE decreased by 22.19% and 23.29% respectively in the transitional period, respectively. Specifically, the inflection-aware attention decreased the RMSE from 0.541 MW to 0.498 MW in the inflection period, representing the greatest single phase reduction; the dynamic residual correction further reduced it to 0.448 MW, confirming the complementary relation between the feature reweighting and the lag error compensation.

E. ANALYSIS OF MODEL ROBUSTNESS UNDER DIFFERENT SEASONS AND DATA PERTURBATION CONDITIONS

The test set was divided into four groups: spring, summer, autumn and winter, and Gaussian noise of 1%, 3% and 5% was added to each group; random missing value perturbations of 5%, 10% and 20% were also applied. Missing values were replaced by the statistics of the training set and time masks. The change in model error under the condition of seasonal distribution transfer and degraded input quality was compared, and the results are presented in Table 4.

Table 4 shows that the general MAE for the fall season and the RMSE for the transition period were 0.241 MW and 0.423 MW, respectively, and thus the smallest errors. In winter, due to high-speed fluctuations in wind and density

TABLE 2. Results of Key Optimization Module Ablation Experiments

Model Configuration	Overall MAE/MW	Overall RMSE/MW	MAE/MW During Transition Phase	RMSE/MW During the Inflection Phase
Full Model	0.258	0.386	0.305	0.448
Remove multiscale branches	0.304	0.451	0.372	0.541
Remove Turn-Aware Attention	0.296	0.438	0.386	0.558
Remove physical constraint output	0.279	0.421	0.337	0.493
Remove dynamic residual correction	0.287	0.429	0.365	0.526

TABLE 3. Accuracy Changes Following the Gradual Integration of Key Optimization Modules

Model Combination	Overall MAE/MW	Overall RMSE/MW	MAE/MW During the Transition Phase	RMSE/MW at the Inflection Point
Base TCN	0.329	0.488	0.392	0.584
Incorporating multiscale branches	0.304	0.451	0.372	0.541
Add Turn-Awareness Attention	0.282	0.417	0.342	0.498
Add physical constraint output	0.271	0.401	0.326	0.474
Add dynamic residual correction	0.258	0.386	0.305	0.448

TABLE 4. Model Prediction Results Under Different Seasons and Data Distortion Conditions

Test Conditions	Overall MAE/MW	Overall RMSE/MW	MAE/MW During the Inflection Phase	RMSE/MW During Inflection Phase
Spring	0.249	0.374	0.296	0.435
Summer	0.263	0.392	0.311	0.456
Fall	0.241	0.365	0.287	0.423
Winter	0.279	0.408	0.329	0.478
1% Gaussian noise	0.265	0.395	0.314	0.461
3% Gaussian noise	0.276	0.411	0.326	0.479
5% Gaussian noise	0.292	0.432	0.344	0.503
5% random missing values	0.271	0.403	0.321	0.469
10% random missing values	0.286	0.424	0.339	0.496
20% random missing values	0.318	0.467	0.374	0.542

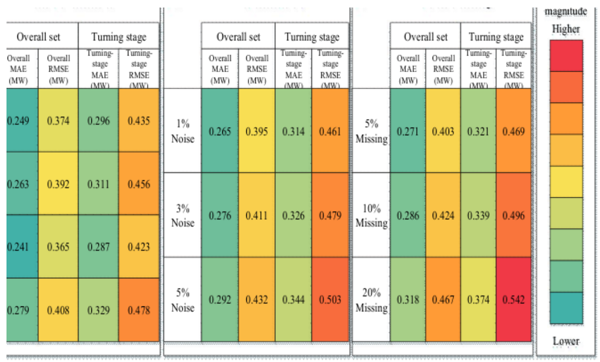

Fig. 6. Heatmap of model errors under different seasonal and data perturbation conditions.

variations of the air, as well as an increase in outages, the overall MAE rose to 0.279 MW, but this was only an 8.14% increase over the 0.258 MW obtained in the full test set. When the perturbation intensity was increased from 1% to 5%, both the overall RMSE and the RMSE during the transition phase increased from 0.395 MW and 0.461 MW to 0.432 MW and 0.503 MW, respectively; that is to say, multiscale inputs and physical constraints limited the power peak shift caused by noise. When random missing values reached 20%, the overall MAE and the MAE of the transition phase increased to 0.318 MW and 0.374 MW, respectively—an increase of 23.26% and 22.62% compared to the results without perturbation—but the prediction error did not expand suddenly. To show the gradient variations of seasons, noise intensity, missing value ratio and the four error indicators visually, please refer to Figure 6.

V. CONCLUSIONS

To address the problems of non-stationarity in the time series of wind power, response lag, and operational boundary distortion under changing weather conditions, a prediction model was constructed that integrates multi-scale, multi-branch feature extraction, transition-aware dynamic attention, wind turbine physical constraint output, and time-delay residual correction. The overall MAE and RMSE of the complete model were 0.258 MW and 0.386 MW, respectively, and these values during the transition period were reduced by 21.58%, 20.90%, 22.19%, and 23.29% compared to the baseline TCN model. Error fluctuations are continuous across all transition types, seasonal distributions, and under noise and missing data perturbations; thus, dynamic feature reweighting, operational domain projection and lag compensation jointly suppress peak deviations and cumulative errors during periods of sudden change. Future work can add numerical weather prediction uncertainty, turbine wake coupling, and cross-wind-farm transfer learning, implement online updates, and develop probabilistic interval forecasting to improve long-term adaptability and engineering applicability under conditions such as extreme weather, equipment degradation and data drift.

ACKNOWLEDGMENT

This work was supported by 5230XT250004 . (Research on Key Technologies for Short term and Ultra short term Wind Power Prediction of Wind Farms under Turning Weather Conditions Based on AI Meteorological Forecasting)

REFERENCES

[1] E. Ammar and G. Xydis, “Wind speed forecasting using deep learning and preprocessing techniques,” International journal of Green energy, vol. 21, no. 5, pp. 988–1016, 2024.
[2] J. Jamii, M. Trabelsi, M. Mansouri, et al., “Medium-term wind power forecasting using reduced principal component analysis based random forest model,” Wind Engineering, vol. 48, no. 4, pp. 597–616, 2024.
[3] A. G. Rangaraj, Y. Srinath, K. Boopathi, et al., “Statistical post-processing of numerical weather prediction data using distribution-based scaling for wind energy,” Wind engineering, vol. 48, no. 5, pp. 824–834, 2024.
[4] J. Parmar and J. Pieper, “A radial basis function neural network approach to filtering stochastic wind speed data,” Wind Engineering, vol. 48, no. 1, pp. 15–31, 2024.
[5] K. Kumar, P. Prabhakar, and A. Verma, “Forecasting wind power using Optimized Recurrent Neural Network strategy with time-series data,” Optimal Control Applications and Methods, vol. 45, no. 4, pp. 1798–1814, 2024.
[6] S. Mulewa, A. Parmar, and A. de, “A novel Bagged-CNN architecture for short-term wind power forecasting,” International Journal of Green Energy, vol. 21, no. 12, pp. 2712–2723, 2024.
[7] Y. Akbal and K. D. Ünlü, “A hybrid deep learning methodology for wind power forecasting based on attention,” International Journal of Green Energy, vol. 21, no. 16, pp. 3713–3722, 2024.
[8] M. Lydia, G. Edwin Prem Kumar, and R. Akash, “Wind speed and wind power forecasting models,” Energy & Environment, vol. 36, no. 8, pp. 3677–3713, 2025.
[9] D. Bouabdallaoui, T. Haidi, M. Derri, et al., “Multi-temporal forecasting of wind energy production using artificial intelligence models,” International Journal of Renewable Energy Development, vol. 14, no. 3, pp. 505–517, 2025.
[10] M. A. Yassen, E. S. M. El-Kenawy, M. G. Abdel-Fattah, et al., “Explainable artificial intelligence for wind power forecasting model based on long short-term memory,” Neural Computing and Applications, vol. 37, no. 19, pp. 14589–14611, 2025.
[11] T. A. Rajaperumal and C. Christopher Columbus, “Enhanced wind power forecasting using machine learning, deep learning models and ensemble integration,” Scientific Reports, vol. 15, no. 1, p. 20572, 2025.
[12] A. Pierrot and P. Pinson, “Data Are Missing Again—Reconstruction of Power Generation Data Using k k-Nearest Neighbors and Spectral Graph Theory,” Wind Energy, vol. 28, no. 1, p. e2962, 2025.
[13] A. Durap, “Explainable deep learning techniques for wind speed forecasting in coastal areas: Integrating model configuration, regularization, early stopping, and SHAP analysis,” Neural Computing and Applications, vol. 37, no. 25, pp. 21219–21257, 2025.
[14] N. Dmitrijevs, V. Komasilovs, S. Orlova, et al., “Short-term wind energy yield forecasting: A comparative analysis using multiple data sources,” Energies, vol. 18, no. 16, p. 4393, 2025.
[15] E. A. Manzano, R. E. Nogales, and A. Rios, “A Systematic Review of Wind Energy Forecasting Models Based on Deep Neural Networks,” Wind, vol. 5, no. 4, p. 29, 2025.
[16] I. U. Haq, A. Kumar, and P. S. Rathore, “Machine learning approaches for wind power forecasting: a comprehensive review,” Discover Applied Sciences, vol. 7, no. 10, p. 1139, 2025.
[17] B. Amarzaya and K. Ko, “Influence of missing wind measurements on wind turbine power production using various measure-correlate-predict methods and reanalysis datasets,” International Journal of Renewable Energy Development, vol. 14, no. 6, pp. 1213–1220, 2025.
[18] G. Dantas and J. Browell, “Seamless Short-to Mid-Term Probabilistic Wind Power Forecasting,” Wind Energy, vol. 29, no. 2, p. e70079, 2026.
[19] M. Bernabe, E. Cadenas, E. Lopez-Espinoza, et al., “Hybrid WRF–SARIMA model to improve day-ahead wind speed forecast accuracy,” International Journal of Renewable Energy Development, vol. 15, no. 1, pp. 76–88, 2026.
[20] G. Şahin, F. Kürker, A. Nur, et al., “Short-Term Wind Speed Forecasting Using Leakage-Free Time-Series Modeling and Statistical Residual Evaluation,” Sustainability, vol. 18, no. 11, p. 5623, 2026.
[21] S. M. R. Hussain, M. M. A. Baig, and M. U. Yousuf, “DVTransformer: A novel approach for multi-step wind power forecasting using spatio-temporal dynamics and predictive strategies,” AIMS Energy, vol. 14, no. 2, pp. 387–417, 2026.
[22] T. L. Nguyen, N. K. Vy, M. Y. Nguyen, et al., “WindFusion-EW: An Ensemble-Weighted Multi-Model Framework for Ultra-Short-Term Wind Power Forecasting,” Engineering, Technology & Applied Science Research, vol. 16, no. 2, pp. 33925–33930, 2026.
[23] B. Karaca, K. D. Ünlü, and S. Türkan, “Wavelet-enhanced sequence-to-sequence modeling with attention mechanism for short-term wind power forecasting,” Cybernetics and Systems, vol. 57, no. 3, pp. 485–533, 2026.