

Article ID: 1006-8775(2024)04-0390-15

Short-Term Wind Speed Forecasts over the Pearl River Estuary: Numerical Model Evaluation and Deterministic Post-Processing

SUN Xian (孙 弦)¹, SUN Lei (孙 磊)¹, LIANG Xiu-ji (梁秀姬)¹, SU Ye-kang (苏烨康)¹, HUANG Wen-min (黄文敏)¹,
KANG Hong-ping (亢红萍)¹, XIA Dong (夏 冬)^{1,2}

(1. Zhuhai Public Meteorological Service Center, Zhuhai, Guangdong 519000 China; 2. Zhuhai-Macao Collaborative Research Center for Meteorological Innovation and Application, Zhuhai, Guangdong 519000 China)

Abstract: The Pearl River Estuary (PRE) is one of China's busiest shipping hubs and fishery production centers, as well as a region with abundant island tourism and wind energy resources, which calls for accurate short-term wind forecasts. First, this study evaluated three operational numerical models, i.e., ECMWF-EC, NCEP-GFS, and CMA-GD, for their ability to predict short-term wind speed over the PRE against in-situ observations during 2018–2021. Overall, ECMWF-EC outperforms other models with an average RMSE of 2.24 m s^{-1} and R of 0.57, but the NCEP-GFS performs better in the case of strong winds. Then, various bias correction and multi-model ensemble (MME) methods are used to perform the deterministic post-processing using a local and lead-specific scheme. Two-factor model output statistics (MOS2) is the optimal bias correction method for reducing (increasing) the overall RMSE (R) to 1.62 (0.70) m s^{-1} , demonstrating the benefits of considering both initial and lead-specific information. Intercomparison of MME results reveals that Multiple linear regression (MLR) presents superior skills, followed by random forest (RF), but it is slightly inferior to MOS2, particularly for the first few forecasting hours. Furthermore, the incorporation of additional features in MLR reduces the overall RMSE to 1.53 m s^{-1} and increases R to 0.74. Similarly, RF presents comparable results, and both outperform MOS2 in terms of correcting their deficiencies at the first few lead hours and limiting the error growth rate. Despite the satisfactory skill of deterministic post-processing techniques, they are unable to achieve a balanced performance between mean and extreme statistics. This highlights the necessity for further development of probabilistic forecasts.

Key words: Pearl River Estuary; wind speed forecast; numerical model evaluation; bias correction; multi-model ensemble

CLC number: P457.5

Document code: A

Citation: SUN Xian, SUN Lei, LIANG Xiu-ji, et al. Short-Term Wind Speed Forecasts over the Pearl River Estuary: Numerical Model Evaluation and Deterministic Post-Processing [J]. *Journal of Tropical Meteorology*, 2024, 30(4): 390–404, <https://doi.org/10.3724/j.1006-8775.2024.035>

1 INTRODUCTION

Accurate wind speed forecasts are essential for a wide range of societally relevant applications, such as severe weather warnings for the general public, risk assessment and decision making in aviation, ship routing, recreational watercraft, agriculture, and so on (Tu et al. ^[1]; Wang et al. ^[2]; Jeon et al. ^[3]; Zhang et al. ^[4]). Particularly, offshore winds significantly affect marine traffic safety and wind power generation and have received increased attention (Liu et al. ^[5]; Neshat et al. ^[6]; Zhang et al. ^[7]). According to statistics (Fang et al. ^[8]), strong wind is a prominent factor in causing marine disasters and is primarily responsible for the majority of fishing vessel losses (Rezaee et al. ^[9]). Moreover, as a result of rising energy demand and environmental

concerns, renewable energy sources (such as solar, biomass, geothermal, and wind) are widely utilized worldwide for their high energy content, lack of toxicity, and sustainability (Olabi and Abdelkareem ^[10]; Gielen et al. ^[11]). Particularly, wind energy is often favored for its cleanliness and abundance (Wang et al. ^[12]; Nelson and Starcher ^[13]). Offshore wind power is more abundant than onshore, and has advantages of not occupying land resources, and not generating noise pollution for nearby residents (Bilgili et al. ^[14]; Zhang et al. ^[15]).

According to timescales, wind speed forecasts can be divided into four categories (Sharma and Singh ^[16]): ultra-short-term (minutes-to-hours), short-term (hours-to-days), medium-term (weeks-to-months), and long-term (months-to-years). This study specifically examines short-term forecasts, while the initial two timescales are primarily concerned with maritime traffic and wind power. Ultra-short-term forecasts in practice mostly utilize statistical methods, including Kalman filtering, auto-regressive models, and neural networks, to establish the internal relationship between historical observations (Lydia et al. ^[17]; Tian and Chen ^[18]), which is not the focus of this study. The numerical weather prediction (NWP) model plays a more important role than statistical models for short-term wind speed forecasts (Xu et al. ^[19];

Received 2023-05-30; **Revised** 2024-08-15; **Accepted** 2024-11-15

Funding: Science and Technology Research Project of Guangdong Meteorological Service (GRMC2021M19, GRMC2022Q16, GRMC2023M29)

Biography: SUN Xian, primarily undertaking research on meteorological application and atmospheric environment.

Corresponding author: SUN Lei, e-mail: ggqx_sl@zhuhai.gov.cn

Tascikaraoglu and Uzunoglu [20]. By formulating the governing dynamic equations and physical parameterization schemes to completely resolve the complex sea-land-air interactions in sub-scale, NWP models have demonstrated their ability to realistically reproduce the spatial and temporal patterns as well as the future trends of meteorological elements (Skamarock et al. [21]). The operational NWP model largely utilizes these methods because of their capacity to successfully address the limitations of statistical forecasting, including poor interpretability, short forecast duration, and inefficiency in capturing extremes. However, it is important to note that these models require intensive computational resources (Chen et al. [22]).

However, NWP models suffer from multi-source uncertainties from the dynamical framework, physical processes, initial and boundary conditions, and others due to the current lack of theoretical knowledge (Knutti et al. [23]). Unlike temperature and pressure, which often follow a normal distribution, wind is difficult to predict accurately due to its intrinsic randomness, intermittency, and volatility caused by complex interactions and underlying topography (Alavi et al. [24]). Thus, it is challenging to satisfy the growing demand for accurate wind speed forecasts. Bias correction is a fundamental and straightforward technique that establishes the links between model outputs and observations (Wood et al. [25]; Werner and Cannon [26]). Studies have demonstrated that it is an effective tool for reducing forecasting errors and has been extensively utilized in many scenarios (Bao et al. [27]; Li et al. [28]). For instance, Zeng et al. [29] comprehensively evaluated the capability of multiple bias correction methods for correcting the gridded wind speed forecast. They found that a two-factor model, with dynamic coefficients using a local sliding scheme, presented the optimal performance. However, the bias correction was exclusively performed on the Chinese mainland.

Multi-model ensemble (MME) has been shown to effectively reduce model uncertainty by incorporating diverse information from several numerical models or modeling approaches (Hagedorn et al. [30]; Jiang et al. [31]). Generally, MME could be categorized into linear and nonlinear approaches. Multiple linear regression (MLR), a typical representative of linear methods, has been widely applied in various applications due to its simplicity and efficiency (Pagowski et al. [32]; Wu et al. [33]). However, there are complex nonlinear relationships between the ensemble outputs from different models, physical schemes, and initial perturbations (Leduc et al. [34]). Nonlinear estimators, which are primarily based on machine learning techniques, are being employed more frequently in MME, and their effectiveness varies significantly across different applications (Crawford et al. [35]; Ahmed et al. [36]). For example, Crawford et al. [35] performed multi-GCM ensemble forecasts of monthly temperature and precipitation over part of North America, utilizing random forest (RF), artificial neural networks (ANN), multiple linear regression (MLR), and weighted

K -nearest neighbors (KNN). According to the results, RF performs best, but MLR provides skills comparable to and even superior to KNN. As far as we know, there are few applications of MME for short-term wind speed forecasts, particularly in offshore areas.

The Pearl River Estuary (PRE) is a prominent maritime transportation hub and a crucial fishery production center in China. Additionally, it has abundant island tourism and wind energy resources (Zhou et al. [37]). Consequently, there is a great demand for accurate wind speed forecasts over the PRE. Operational numeric models are a critical reference for short-term wind speed forecasts. Consequently, improving the capability of wind speed forecasts over the PRE is scientifically and operationally valuable, which is the primary motivation for this study.

This study first systematically evaluates the capability of three operational numerical models in short-term wind speed forecasts over the PRE against the in-situ measurements. Then, a range of distinct techniques (including bias correction and MME) are chosen to conduct deterministic post-processing on NWP outputs, and their effects are clarified by a comprehensive intercomparison. Furthermore, the incorporation of additional predictor variables (beyond ensemble outputs of the target variable) is proposed as a further MME-based enhancement. The remaining sections are organized as follows: Section 2 describes the study area and used data. Section 3 introduces post-processing methods, experimental settings, and evaluation metrics. Section 4 compares the performances of three models in forecasting the short-term wind speed over PRE. Section 5 compares different post-processing techniques. Sections 6 and 7 discuss and conclude the research.

2 DATA AND STUDY AREA

The Pearl River is one of China's largest rivers with respect to water discharge and sediment load (Fig. 1). The mouth of the Pearl River forms a large bay in the southeast of the delta, the Pearl River Estuary (PRE), characterized by large runoff, small tidal range, and small concentration of sediment. The rivers in the vicinity of the PRE form an intricate river network. The estuary is trumpet-shaped, with a width of approximately 5 km at the northern (upper) end and approximately 35 km at the southern (lower) end. Eight inlets, namely Modaomen, Humen, Jiaomen, Hongqimen, Hengmen, Chuandao, Shiziyang, and Wailumen, discharge the freshwater of the Pearl River into the northern part of the South China Sea. In particular, the Pearl River Delta (PRD) is a highly developed region in China, including several major cities such as Guangzhou, Shenzhen, Hong Kong, and Macao.

This study collects the wind speed measurements from 20 automatic meteorological observation stations (AMOS) distributed across the PRE region during 2018–2021. In particular, the quality control of AMOSs is constructed using a unified framework from Guangdong

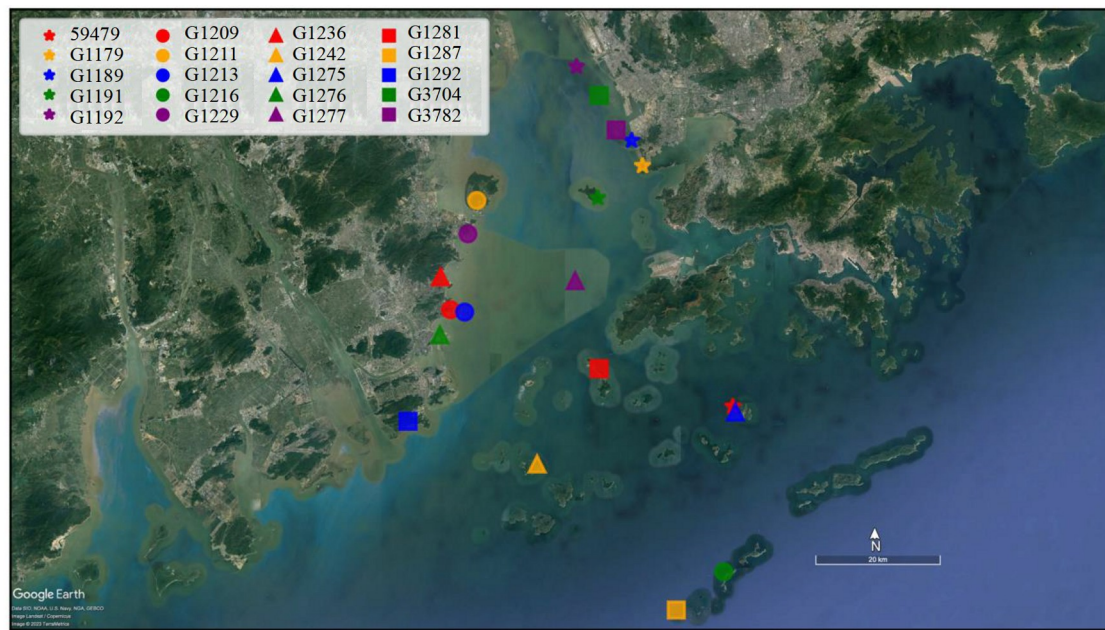


Figure 1. A Google Earth image of the Pearl River Estuary and its surrounding area, with colored markers denoting locations of wind measurement stations.

Meteorological Service. Note that the anemometers are mounted at different heights above the ground. To facilitate the comparison, the wind speed is converted uniformly to a height of 10 m using the exponential law formula (Zhi et al. [38]). The wind measurements are recorded every 5 min, and the instantaneous values at each hour are used for direct comparison with the model. Note that there is a certain percentage of data gap. To address this, we discard the months with a missing percentage exceeding 10% during 2018–2021 and then select the longest consecutive months as the analysis period for each station. Moreover, we utilize the ERA5 reanalysis data (Hersbach et al. [39]) with a resolution of 0.25° to fill the gaps. Specifically, the hourly wind speed from ERA5 is bilinearly interpolated into different stations. For the missing records, a dynamic linear regression is developed between observations (the predictand) and interpolated ERA5 (the predictor) to generate a one-step estimate, with a 5-day sliding window. Regarding NWP models, two global models, i.e., EC and GFS, from the European Centre for Medium-Range Weather Forecasts (ECMWF, hereinafter referred to as EC) and National Centers for Environmental Prediction (NCEP), and a regional model, CMD-GD, from China Meteorological Administration (CMA), are chosen for evaluation and post-processing. The global EC and GFS are characterized by a resolution of $0.1^\circ \times 0.1^\circ$ and $0.25^\circ \times 0.25^\circ$, respectively. At the same time, the regional CMA-GD has a high resolution of $3 \text{ km} \times 3 \text{ km}$, with the initial and boundary conditions derived from its global counterpart. All three models are run twice a day (0800 and 2000 UTC) to produce the short-term forecasts for the next 7–10 days. Notably, we select the 3-hourly outputs within 48 hours issued at 2000 during 2018–2021 and then interpolate

them bilinearly into 20 stations (Fig. 1).

3 METHODOLOGY

In this study, post-processing methods, including bias correction and MME, are employed to improve the accuracy of short-term wind speed forecasts over the PRE. For bias correction, we select three methods (Zeng et al. [29]), including simple bias correction (SBC), one-factor and two-factor model output statistics (MOS and MOS2). Tests reveal that a sliding window of 60 samples is sufficient for bias correction methods to produce consistent results, while ML-based ML methods require a larger number of samples. Accordingly, MME methods are evaluated using a two-fold cross-validation method. For both approaches, the dependent variable is the hourly in-situ measurements, and the independent variable is the station-interpolated outputs from a single model (bias correction) or multiple models (MME). Furthermore, the methods mentioned above are developed individually for different sites (20) and lead times (16).

3.1 Error correction

SBC assumes that the errors at different forecasting lead times are only related to the deviation at the initial time and could be eliminated by linear regression using the following expression:

$$Y'_{t+n} = Y_{t+n} + (b_0 + b_1 \times E_t) \quad (1)$$

where $E_t = O_t - Y_t$ is the deviation at the initial forecast time t , and O_t and Y_t are the observed and forecasted wind speeds, respectively. Y_{t+n} and Y'_{t+n} are the forecasted and corrected values at time $t + n$ ($n=3, 6, 9, \dots, 48$). b_0 and b_1 represent the regression coefficient and intercept constant, respectively.

In contrast, the MOS method assumes that the error at

$t + n$ is only related to the forecast at the same time and it is expressed as follows:

$$Y'_{t+n} = b_2 Y_{t+n} + b_3 \quad (2)$$

where b_2 and b_3 represent the regression coefficients and intercept constants, respectively.

The two preceding methods are single-factor corrections, whereas the MOS2 method integrates these two factors by a dual linear regression:

$$Y'_{t+n} = b_4 Y_{t+n} + b_5 E_t + b_6 \quad (3)$$

Correspondingly, b_4 and b_5 are regression coefficients, and b_6 is the constant. The above coefficients and constants can be calculated using the least-squares estimation.

3.2 Multi-model ensemble

Multiple linear regression (MLR) is a statistical technique used to determine the relationship between two or more independent variables and a dependent variable. It can be represented by the following equation:

$$Y = \sum_{i=1}^N a_i X_i + b \quad (4)$$

where a_i and b are the regression coefficients and y -intercept, which can also be solved using least-squares estimation.

Random forest (RF), gradient boosting (GB), support vector machine (SVM), and K -nearest neighbors (KNN) are all supervised machine learning algorithms that have been widely used in solving classification and regression problems. In particular, RF is an ensemble learning method that improves prediction accuracy by constructing multiple decision trees. Each decision tree is built using randomly selected features and samples. GB is a technique that constructs an ensemble model by iteratively adding weak learners (decision trees in this study) to the residual errors. The process starts with an initial prediction model and subsequently incorporates additional models iteratively to minimize the loss function. The results for both RF and GB are determined by averaging the predictions from all decision trees.

Note that RF could rank the contribution of each input feature relative to the predicted result. The relative importance (RI) is calculated based on the out-of-bag (OOB) score. In particular, the decline in accuracy caused by the change is calculated by randomly permuting the input feature (the more significant the change, the more influential the feature is), which could be represented as follows:

$$RI_{i,k} = \frac{1}{N} \sum_{k=1}^N \left[\text{MSE}(\text{OOB}_{i,k}) - \text{MSE}(\text{OOB}_{p,k}) \right]^2 \quad (5)$$

where i represents an input feature, N is the number of constructed decision trees, p represents the permuted feature, and MSE is the mean square error. Moreover, the RI is normalized to ensure that its sum is 100%.

SVM functions by locating a hyperplane that maximally separates the data points of different classes or predicts the target variable with the smallest error. In particular, the hyperplane is defined as the decision boundary that separates the data points of different

classes by the largest margin, using kernel functions, such as polynomial, sigmoid, and radial basis functions (RBF). Moreover, KNN is a non-parametric algorithm that averages the K -nearest neighbors from the training set to predict the class or value of a new data point.

3.3 Skill metrics

This study uses root mean square error (RMSE), correlation coefficient (R), forecast accuracy (FC), and threat score (TS) to quantitatively evaluate the performances of three numerical models and various post-processing methods, as expressed below:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (6)$$

$$R = \frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (7)$$

$$\text{FC} = \frac{N_{\text{Rk}}}{N} \quad (8)$$

$$\text{TS} = \frac{\text{NA}}{\text{NA} + \text{NB} + \text{NC}} \quad (9)$$

where O and P represent the observations and simulations, respectively, N is the total number of samples, $\bar{}$ denotes the operation of arithmetic average. N_{Rk} represents the number of simulations with the same wind grade (Beaufort wind scale; Beer ^[40]) as measurements. Note that the TS score is case-dependent, in which NA, NB, and NC are the numbers of true positive, false negative, and false positive events, respectively. Moreover, this study employs the Taylor diagram (Taylor ^[41]) to compare the similarity between the observations and simulations.

4 EVALUATION OF NUMERICAL MODELS

4.1 Lead-time specific

Figure 2 compares the performances of wind speed forecasts from three numerical models at various lead times. Overall, EC has the lowest RMSE values, with a mean of 2.24 m s^{-1} , for all lead times, followed by GFS (2.72 m s^{-1}) and CMA-GD (3.30 m s^{-1}). Commonly, the forecast error for a meteorological variable increases obviously as the integration proceeds. However, it does not appear to apply well to wind forecasts due to the following two factors. First, the synoptic situation governs the offshore wind, resulting in a small decrease in performance within 48 hours, which highlights the superiority of NWP over statistical methods for short-term forecasts. Moreover, the wind speed over the PRE tends to be higher during the night to next-day early morning compared to relatively smaller winds around noon. It results in the corresponding diurnal fluctuations of simulation errors, particularly for CMA-GD, which is consistent with previous studies (Schulz and Lerch ^[42]). Regarding correlation, GFS presents enhanced performance with an average R of 0.56, which is comparable to the EC ($R = 0.57$). This suggests that

systematic bias is responsible for a substantial portion of the overall biases in GFS. CMA-GD performs worse than the other two models, with an R less than 0.5 for most forecasting hours. Additional comparisons of FC reveal that EC is the superior model. However, its FC values are below 0.4 for all lead times (with a mean of 0.35), indicating the need for post-processing methods to improve performance. Furthermore, Fig. 2d employs the Taylor diagram to thoroughly evaluate the performances of three models at different lead hours. The results for each model are shown to be relatively concentrated for different lead times, revealing the consistent performances of NWP. EC and GFS present comparable skills. The major difference is that the former (latter) results are more concentrated (scattered), resulting in superior or inferior performances at different lead hours. In contrast, CMA-GD presents a less skilled performance, consistent with previous conclusions.

4.2 Station specific

Under the influence of the complex underlying topography, the surface winds over the PRE exhibit significant spatial variations (Zhou et al. [37]). Figs. 3 and 4

depict the spatial distribution of RMSE and R for three models across various lead times (3, 9, 27, and 48 hours). Each NWP model exhibits consistent spatial variations over different forecasting hours. However, their performances vary dramatically among different sites, with a larger (smaller) RMSE corresponding to a smaller (larger) R . It demonstrates the temporal consistency and spatial independence of simulation errors from NWPs, emphasizing the importance of treating different sites separately when performing post-processing techniques. Specifically, EC exhibits superior performances for smaller biases at nearly all sites for different lead times, as is also evident from the respective boxplots. GFS has greater biases than EC, especially along the coasts of PRE. Despite the highest resolution, CMA-GD's performance is unsatisfactory, indicating that resolution refinement should not guarantee a performance improvement. On the other hand, EC achieves higher R values for the majority of sites. However, for stations G1281 and G1242, with relatively stronger winds, GFS is capable of achieving a high R value (approximately 0.7), which is significantly greater than that of EC (approximately 0.5). It

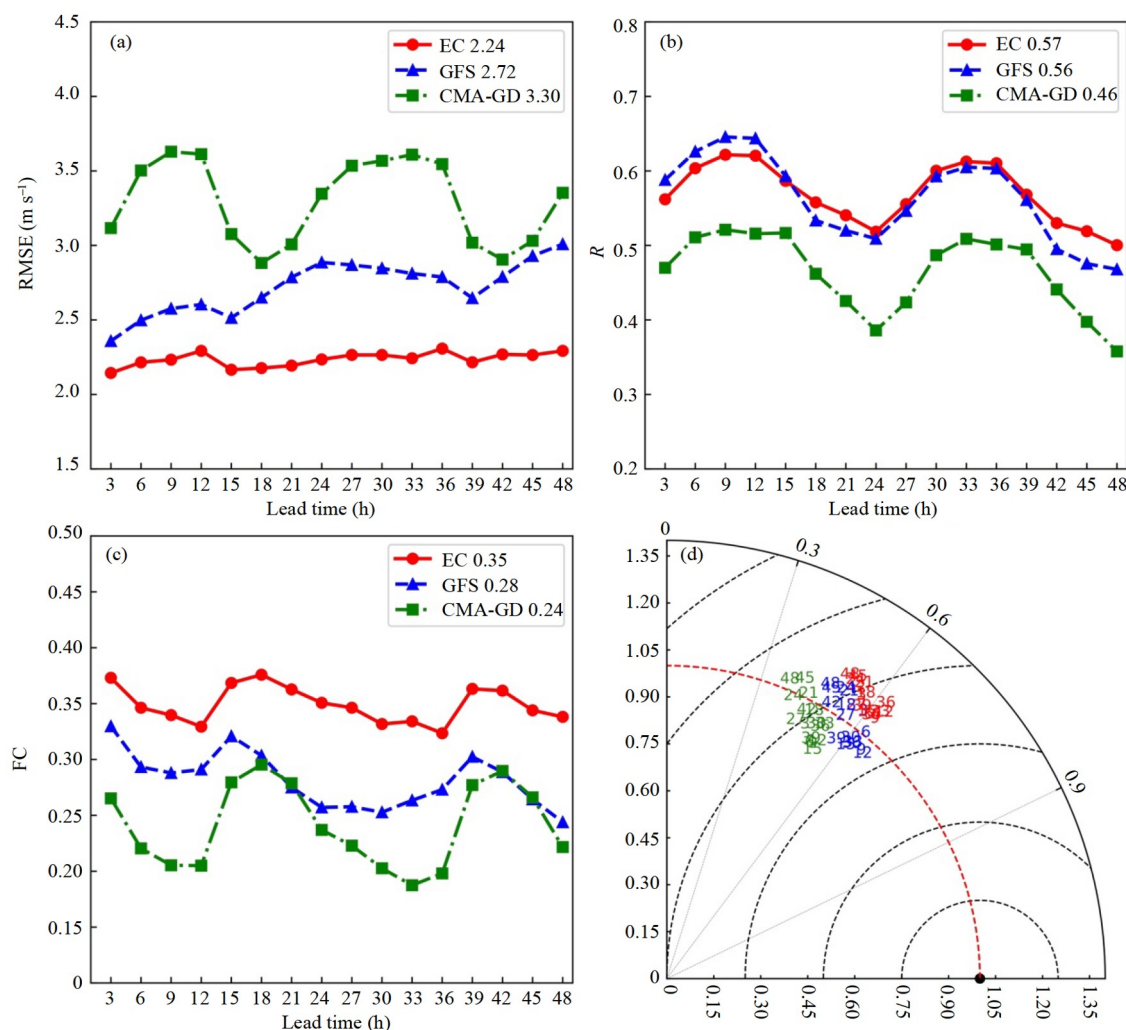


Figure 2. The comparison of (a) RMSE, (b) correlation coefficient (R), (c) forecast accuracy, and (d) Taylor diagram from three numerical models (EC, GFS, and CMA-GD) in predicting wind speeds at different lead times, with the numbers in the legend denoting lead-averaged values.

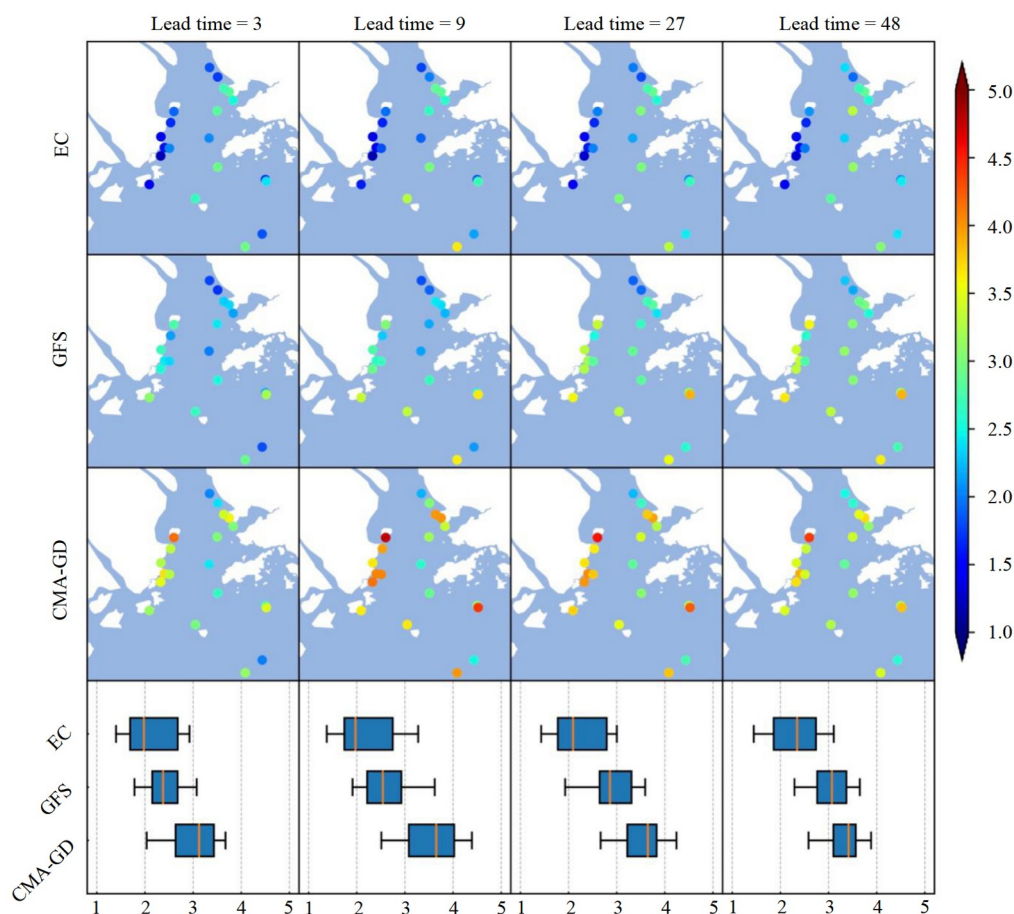


Figure 3. Geographic distribution and boxplots of RMSE of wind predictions at different lead times from three numerical models (EC, GFS, and CMA-GD). The box and whiskers cover the 25th-75th and 5th-95th percentile ranges of RMSE values, respectively, with the orange line in the box denoting the median value (same hereinafter).

demonstrates the limitations of EC in capturing strong winds. Given the distinct characteristics of NWP, it is necessary to apply MME methods to integrate their strengths effectively.

4.3 Wind-speed specific

Accurate forecasts of specific wind speeds or wind levels are essential in real-world scenarios. For instance, maritime transport focuses mainly on the occurrence of strong winds (above level 6), and the cut-in speed of offshore wind turbines is roughly 3 m s^{-1} . Fig. 5 compares the RMSE and TS from three models based on different observed wind speeds, with an interval of 1 m s^{-1} . The measurements reveal that the wind speed does not conform to the normal distribution. The observed frequency has a long tail (Alavi et al. [24]), which peaks at approximately 3 m s^{-1} wind speed and decreases rapidly as wind speed increases. For each model, the simulation error decreases as the observed wind speed increases until the turning point, after which the error grows exponentially. EC, GFS, and CMA-GD have respective turning points of 5 m s^{-1} , 7 m s^{-1} , and 8 m s^{-1} , respectively. EC presents distinct advantages in terms of the lower wind speed, which accounts for its overall satisfactory skill. Note that three NWP models present comparable performances for the

observed speed of around 8 m s^{-1} . However, when this speed is exceeded, EC's forecasting capability rapidly degrades, and a gap exists between it and the other two models, corresponding to its limitation in strong winds, as stated earlier. An additional comparison of TS scores reveals the poor performances of all models, with values less than 0.2 for each bin of the observations. The speed range of maximum values corresponds to the maximum frequency (excluding the artificially large value of the rightmost bin as a result of the infinite range). Then, it decreases rapidly with the increase in wind speed in accordance with the observed distribution curve.

5 DETERMINISTIC POST-PROCESSING

5.1 Inter-comparison

Preceding results have revealed the inadequacies of numeric models in capturing wind speed over the PRE, emphasizing the need to apply deterministic post-processing methods for improvements. This study employs two types of post-processing methods: bias correction and MME. The former considers single-factor (SBC and MOS) or two-factor (MOS2) corrections to correct the single model output directly. In contrast, the latter includes linear (MLR) and ML-based nonlinear

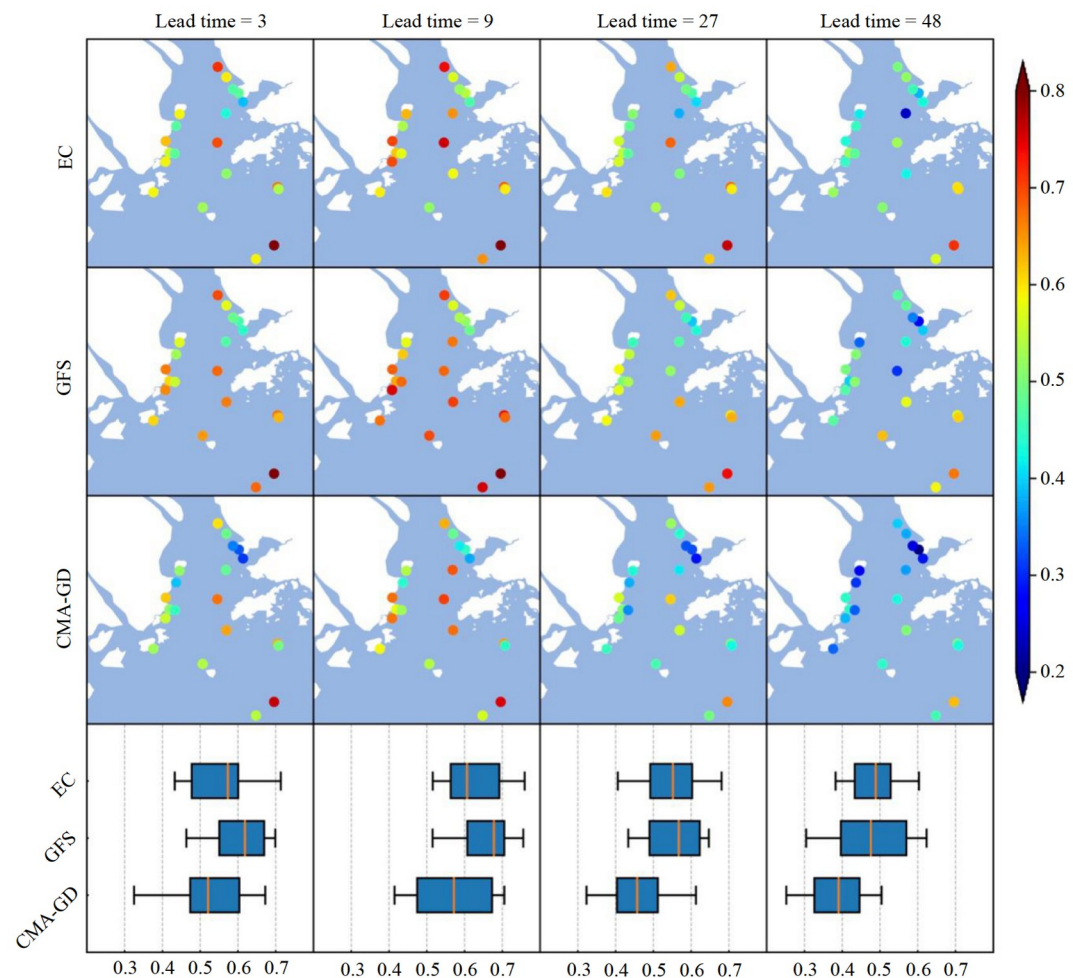


Figure 4. Same as Fig. 3 except for correlation coefficient.

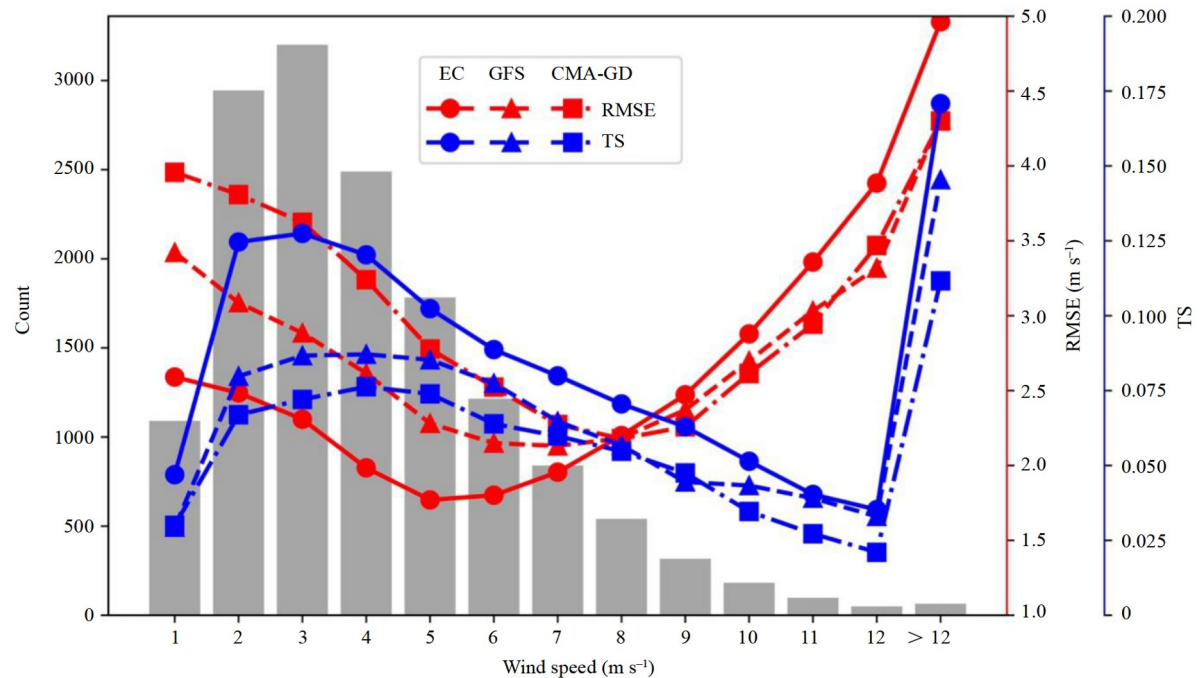


Figure 5. The comparison of RMSE (red lines) and TS score (blue lines) from three numerical models (EC, GFS, and CMA-GD) in predicting wind speeds averaged over all lead times for various bins at an interval of 1 m s^{-1} (except for the last, numbers denoting the bin centers), overlaid with the histogram of observed wind speeds.

approaches (RF, GB, SVM, and KNN), which comprehensively consider all possible models. EC is selected as the reference and bias correction target given its overall superior performance.

Figure 6 comprehensively compares the performance of various post-processing techniques at different lead times relative to EC. Regarding bias correction, each method is capable of efficiently eliminating systematic biases, thereby substantially improving the forecasting capacity of wind speed. Specifically, SBC, MOS, and MOS2 decrease (increase) the average RMSE (R) to 1.79 m s^{-1} (0.67), 1.65 m s^{-1} (0.68), and 1.62 m s^{-1} (0.70), respectively. The inter-comparison of different methods reveals the benefits of accounting errors at corresponding lead times (Zeng et al. [29]). MOS and MOS2, therefore, outperform SBC, with MOS2 performing the best across all metrics. Note that the medians of the biases (excluding EC) are almost near zero, which is attributed to the fact that the results of all statistical methods could be regarded

as bias-corrected (Sun et al. [43]). Furthermore, additional experiments excluding the sliding operation (i.e., non-dynamic) produce demonstrably inferior performances (not shown). In summary, it is of great necessity and effectiveness for bias correction techniques to consider two factors and execute sliding operations simultaneously.

All MME approaches have demonstrated substantial improvements across various lead times for all metrics. However, particular attention must be paid to the following aspects: First, different nonlinear methods are based on varying rationales but exhibit comparable performances. More importantly, the basic MLR outperforms most nonlinear methods, yielding the smallest errors and strongest correlations among MME methods, which could be attributed to greater mutual independence across different models. Second, the sliding two-factor bias correction (operated on EC) outperforms the optimal MME method (i.e., MLR), especially in the initial lead hours. It demonstrates the effectiveness of straightforward

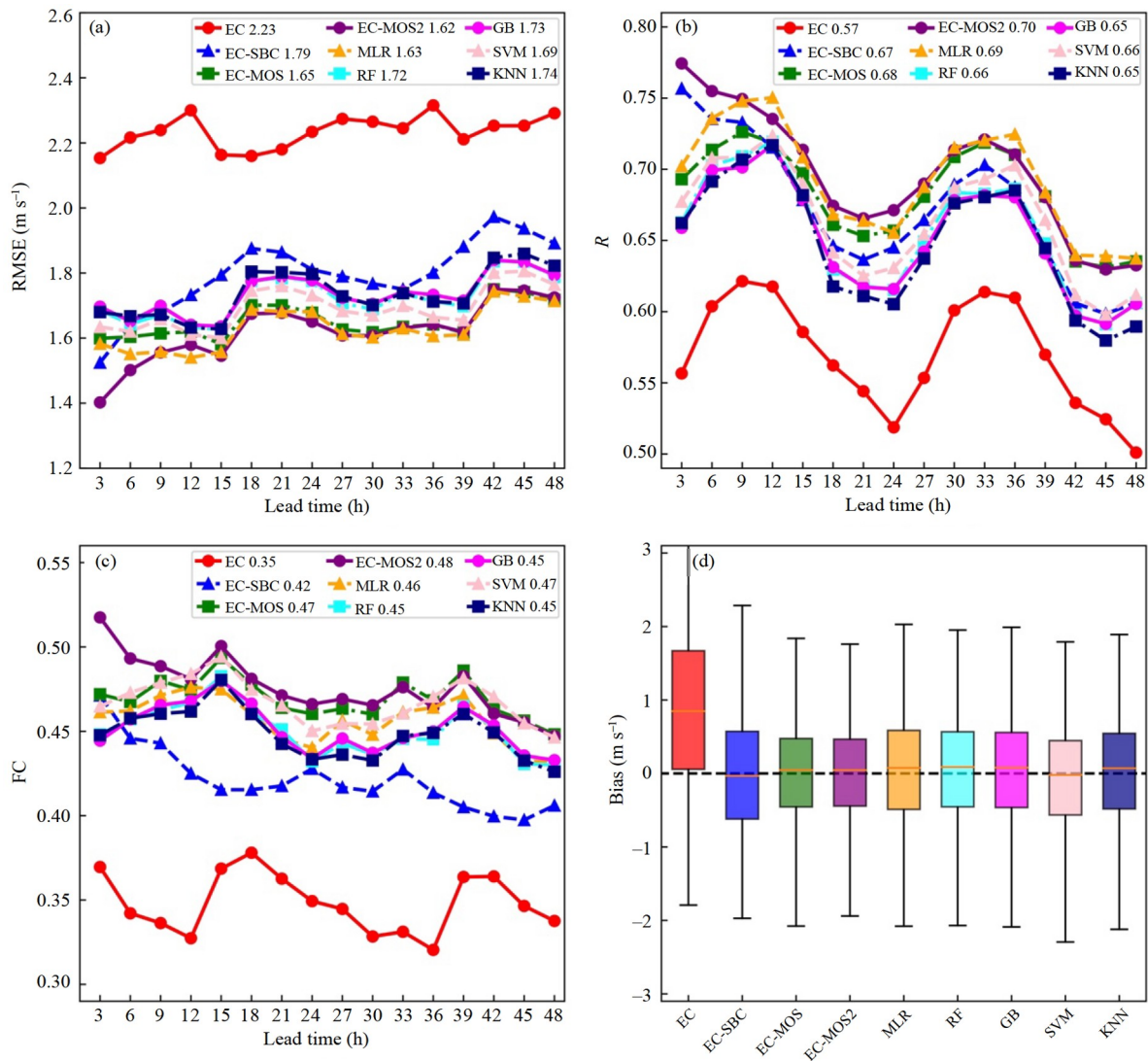


Figure 6. The comparison of (a) RMSE, (b) correlation coefficient (R), and (c) forecast accuracy (FC), from the optimal model (EC) and its bias corrections (EC-SBC, EC-MOS, and EC-MOS2), and multi-model ensembles (MLR, RF, GB, SVM, and KNN) in predicting wind speeds at different lead times, with the numbers in the legend denoting lead-averaged values, and (d) shows boxplots of the predicted errors.

bias correction and highlights the necessity to enhance MME performances further.

5.2 Added value of including additional features

Evidently, wind is characterized by scale-dependent variations and is strongly correlated with other meteorological elements. Some studies have demonstrated that the incorporation of relevant variables (beyond the target variable) as additional input features could further improve the performance of MME (Taillardat et al. [44], Taillardat and Mestre [45]). This study also examines the inclusion of temporal and meteorological features in the MME method. The time information is formulated as $YDAY = \cos\{2\pi(t-1)/365\}$ (Schulz and Lerch [42]), where t is the day of the year. GFS provides U and V components of surface winds, 2-m temperature (T2M), 2-m relative humidity (RH2M), and surface pressure (SP), which are interpolated into different stations for use.

Figure 7 compares some MME methods that exclude or include additional features to the optimal bias correction (i.e., EC-MOS2). By introducing new features, many MME methods could substantially improve their

forecasting performance for wind speed at each lead time. MLR, with the introduction of new features (MLR+), for instance, reduces the average RMSE to 1.53 m s^{-1} and increases the average R and FC to 0.74 and 0.49, respectively. Compared with the outstanding MOS2, it presents superior performance for different metrics at all times except the initial (MLR+ is slightly worse), demonstrating the advantages of MME methods in integrating different characteristics. However, the responses of ML-based methods vary significantly when additional input features are included. The performances of RF and GB also improve and are comparable to MLR, with RF performing slightly better. In contrast, the inclusion of additional features substantially degrades the performances of KNN and SVM (not shown). Consequently, this study favors ensemble-based methods (i.e., RF and GB) over others (i.e., KNN and SVM) for performing MME, particularly when considering features beyond the target variable.

Figures 8 and 9 illustrate the spatial distribution of RMSE and R from EC, as well as some post-processing

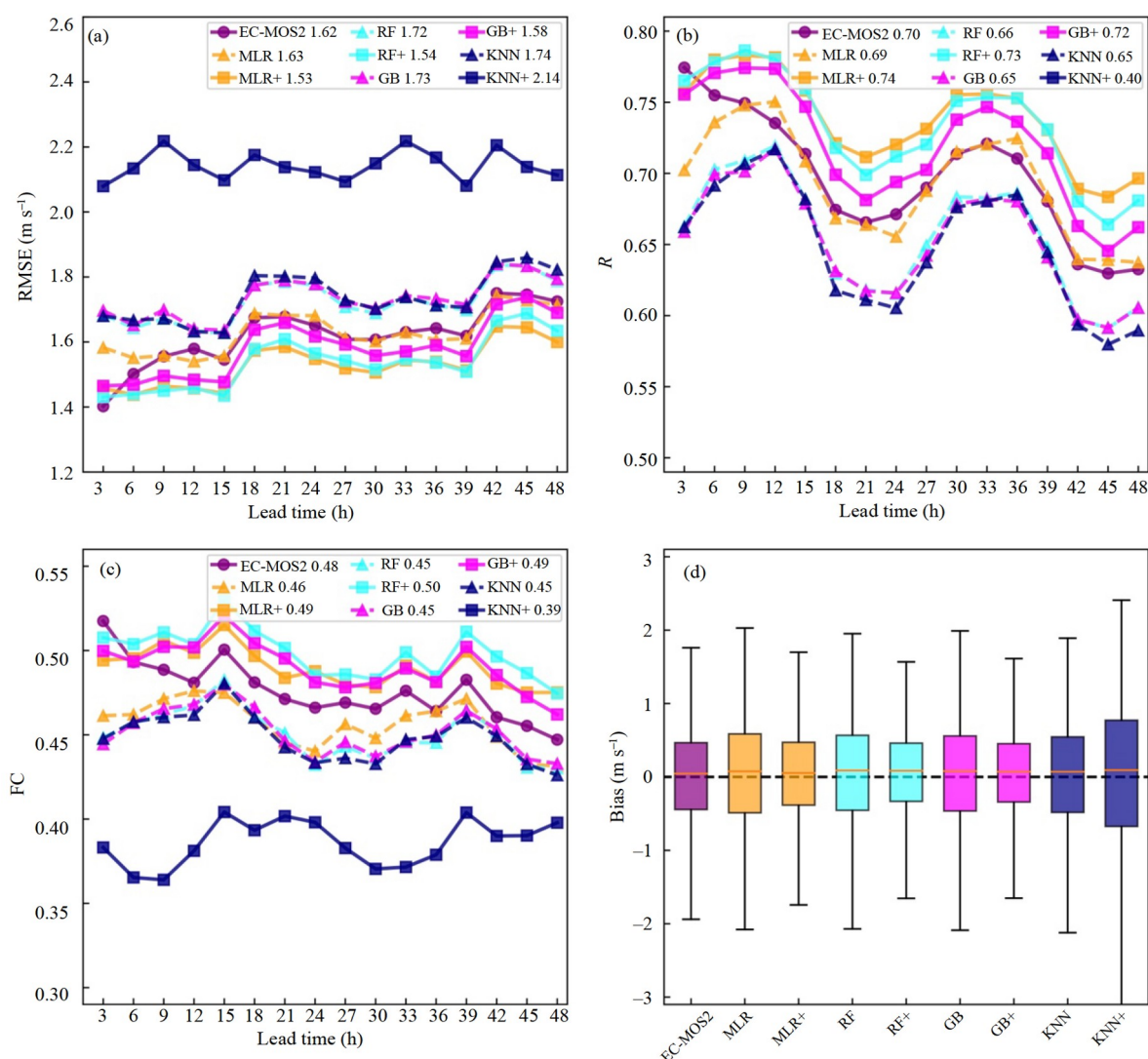


Figure 7. Same as Fig. 6 except for different methods, with “+” denoting the inclusion of additional features based on corresponding methods (same hereinafter).

techniques for various lead times. Compared with EC, the application of MOS2 results in a significant reduction (increase) of RMSE (R) for all stations at different lead times, but it still has limitations. In particular, the overall correlation decreases rapidly as integration progresses. In addition, the simulation error at certain stations (such as G1281) is growing rapidly. MLR performs slightly better than RF, but both are inferior to MOS2, especially for the initial few lead hours. When additional features are introduced, there is a significant improvement in MLR and RF at different lead times for each station, and the gaps between them are narrowed to a very small extent. It effectively improves the performance at initial lead times (comparable to MOS2) and restricts the error growth rate. Notably, the skills for stations with higher winds (such as G1281) are also enhanced.

6 DISCUSSION

Random forest can determine feature importances by measuring the decrease in node impurity weighted by the probability of reaching that node or by calculating the out-

of-bag error via permuting a feature (Calle and Urrea ^[46]). Figs. 10a and 10b depict the varying relative importance of different sites and lead times, respectively. The following essential information can be concluded: First, in the majority of instances, the importance of EC and GFS is greater than those of CMA-GD, indicating their skilled performances. However, the contribution of CMA-GD cannot be neglected, highlighting the strengths of MME methods for considering multi-source information. Second, the relative importance of each feature varies considerably between sites and lead times. It accounts for the varied performances of different models and emphasizes the need to construct models for each site and lead time separately. Third, the inclusion of additional input features is critical for enhancing MME capabilities, at least in this application. Particularly, the U or V component could exhibit the greatest importance for several sites, which is consistent with the poorly performing sites identified earlier. It also provides an excellent explanation for the substantial performance enhancements in the first few lead hours.

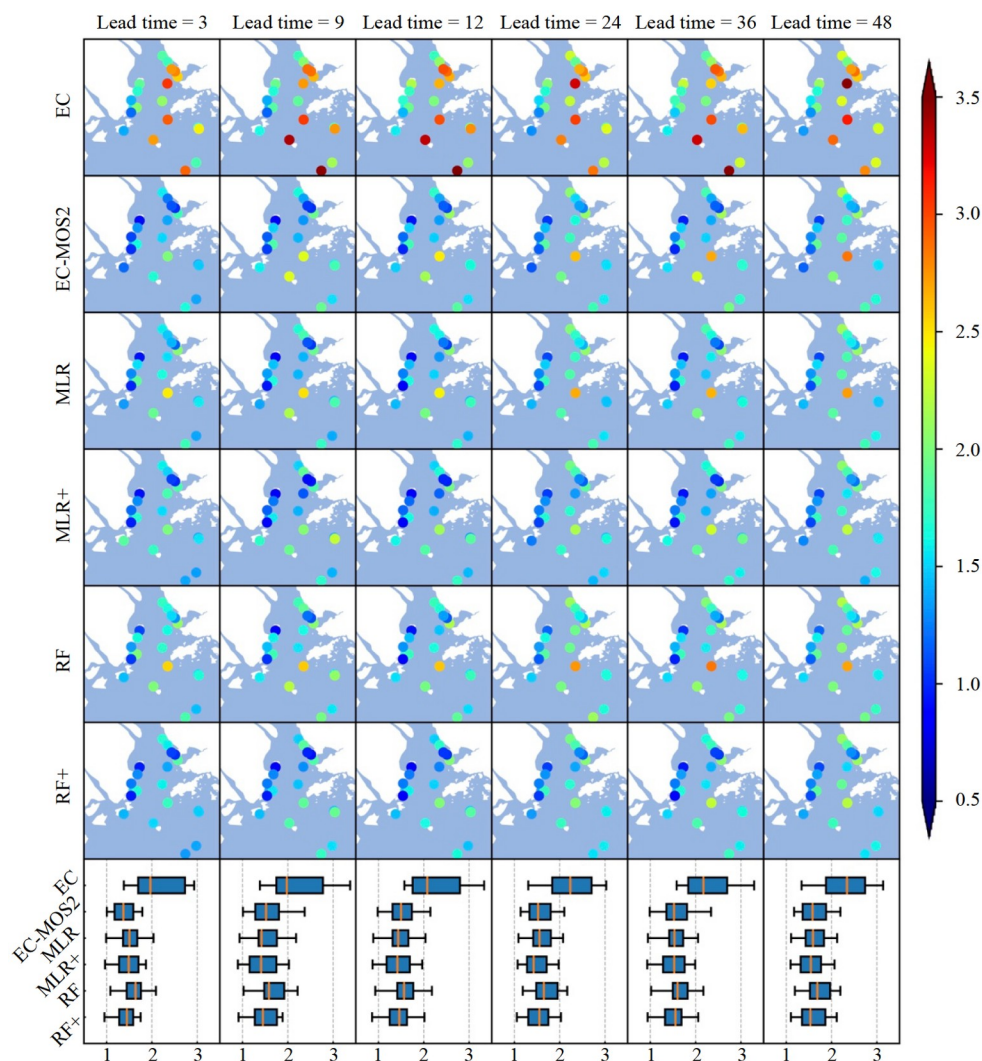


Figure 8. Geographic distribution and boxplots of RMSE of wind predictions at different lead times from the optimal numerical model (EC), its bias correction (EC-MOS2), and multi-model ensembles (MLR, MLR+, RF, and RF+).

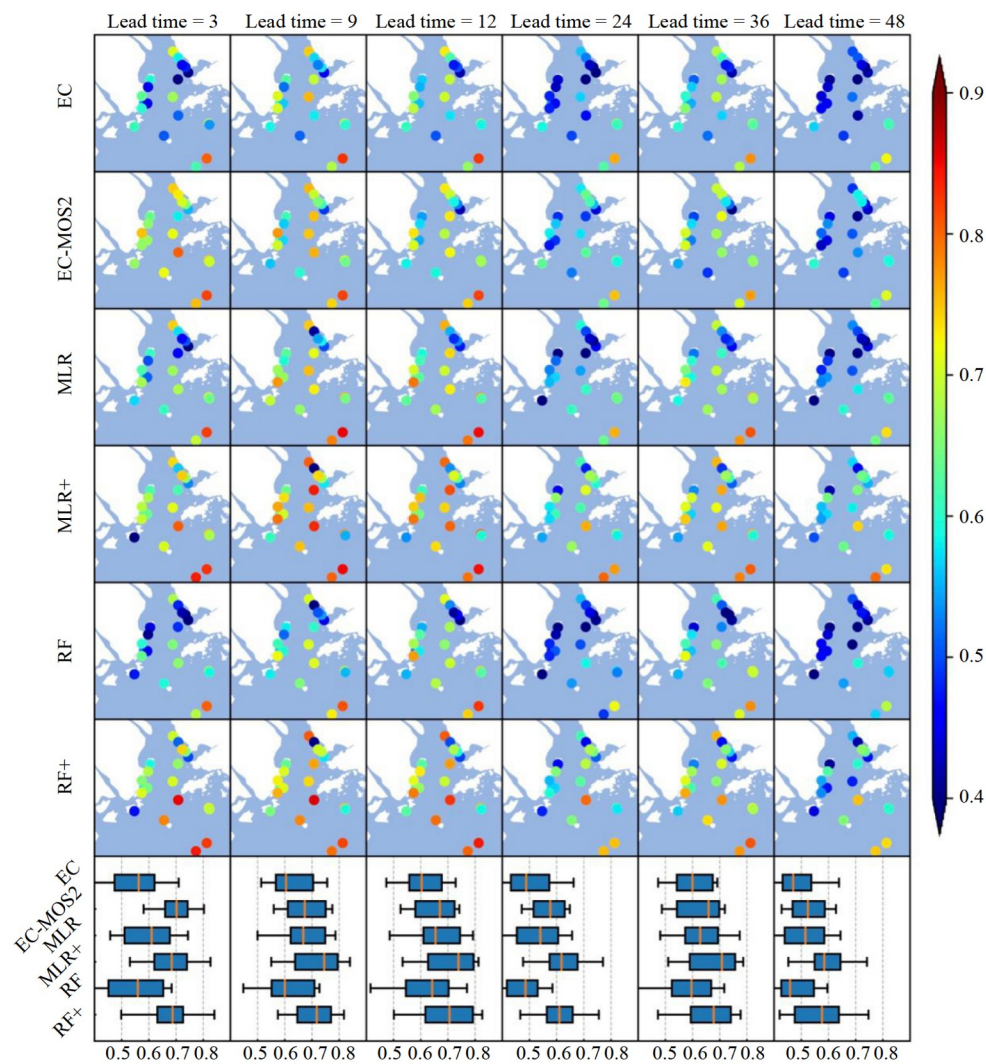


Figure 9. Same as Fig. 8 except for correlation coefficient.

It should be emphasized that bias correction and MME methods are capable of reducing simulation errors and thereby enhancing forecasting skills of wind speed over PRE, but both of them are subject to obvious limitations in accurately capturing high wind speeds. Fig. 11 depicts the variations in RMSE for two models and several post-processing techniques according to different observed ranges of wind speed. It reveals that the distribution curves from various post-processing techniques behave similarly and are almost opposite to the observed frequency distribution (Fig. 5). The higher (lower) the observed frequency, the better (worse) the simulation performance. In particular, for strong winds (above 12 m s^{-1}), the corrected results are even inferior to the raw model outputs. This is mainly due to the fact that the majority of post-processing methods are optimized to minimize the conditional mean (rather than different quantiles) of the target variable (Petscher et al. [47]). Consequently, the post-processing methods are able to achieve an overall satisfactory performance for the mean-based evaluation metrics but demonstrate no added value or even worse skills for extreme values. This study further tests the

quantile random forest (QRF) for MME. The primary difference between RF and QRF is that QRF is designed to estimate conditional quantiles (set to 0.95 in this study) of a response variable. It is demonstrated that QRF performs significantly better than RF and all other methods for winds above 8 m s^{-1} but at the expense of poorer performance in light-to-moderate winds. Thus, it is crucial to fully comprehend the characteristics of various post-processing methods and to select the appropriate one based on the specific application scenario for providing a reliable and customized meteorological service.

7 CONCLUSION

The Pearl River Estuary (PRE) is characterized by frequent shipping and fishery activities, as well as abundant island tourism and wind energy resources. Thus, accurate short-term wind speed forecasts are crucial. They are primarily dependent on NWP, which, however, are subject to multi-source uncertainties due to a lack of theoretical knowledge.

This study first comprehensively evaluates the forecasting capabilities of short-term wind speed over

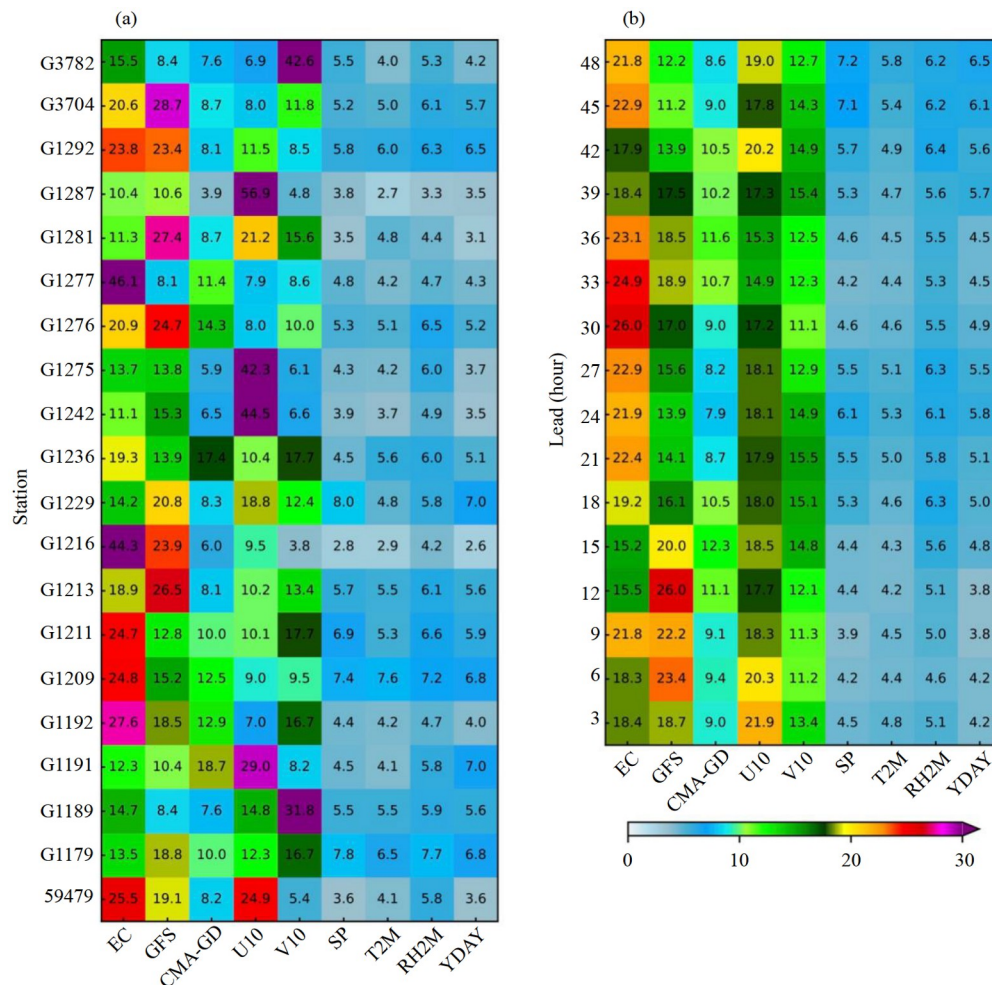


Figure 10. The portrait diagrams of relative importances for different features according to (a) different stations and (b) different lead times output from RF+. Note that the x-axis tick labels denote the wind speeds from the corresponding model, U and V components, surface pressure, 2-m temperature and relative humidity provided by GFS, and time information.

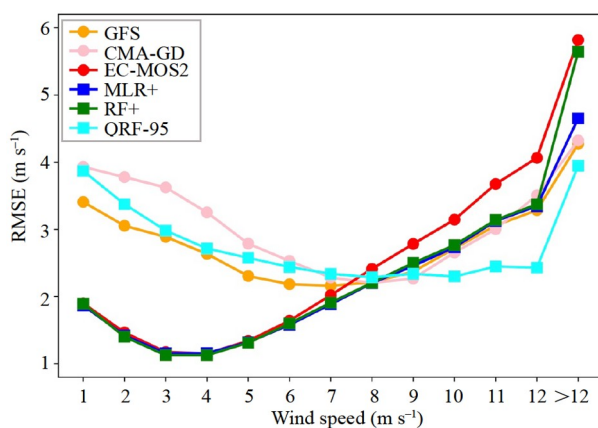


Figure 11. The comparison of RMSE from numerical models (GFS and CMA-GD), bias correction (EC-MOS2), and multi-model ensembles (MLR+, RF+, and QRF-95) for various bins at an interval of 1 m s^{-1} . Note that QRF-95 denotes the quantile regression forest at a quantile of 0.95.

the PRE from three operational NWP models, including two global models (EC and GFS) and one regional model (CMA-GD), against in-situ observations during 2018–

2021. Overall, EC outperforms other models for the majority of sites, with an average RMSE of 2.24 m s^{-1} and R of 0.57, followed by GFS (RMSE = 2.72 m s^{-1} ; R = 0.56). However, GFS performs better at a few stations with relatively stronger wind speeds, followed by CMA-GD, highlighting the inadequacies of EC in capturing high winds.

Then, a variety of two-category deterministic techniques, namely bias correction and MME, are employed to post-process the NWP outputs using a local and lead-specific scheme. In particular, bias correction considers single-factor (SBC and MOS) or two-factor (MOS2) corrections to directly correct the single model output. At the same time, MME includes linear (MLR) and ML-based nonlinear approaches (RF, GB, SVM, and KNN) that comprehensively consider all possible models. It is shown that SBC, MOS, and MOS2 decrease (increase) the lead-averaged RMSE (R) to 1.79 m s^{-1} (0.67), 1.65 m s^{-1} (0.68), and 1.62 m s^{-1} (0.70), respectively, emphasizing the need for bias correction techniques to consider both initial and lead-specific information. In comparison to EC, all MME methods have also

demonstrated substantial improvements. Simple linear MLR outperforms other MME methods, followed by RF, but is slightly inferior to MOS2, especially for the initial few lead hours.

Moreover, the incorporation of additional input features substantially improves the forecasting performance of short-term wind speed over the PRE from MLR, RF, and GB but performs in the opposite manner for SVM and KNN. When introducing new features, MLR reduces the overall RMSE to 1.53 m s^{-1} and increases R to 0.74. It also improves its performance at the initial few lead hours and restricts the error growth rate, thus outperforming MOS2 comprehensively. RF is capable of demonstrating comparable skills, whereas GB is marginally inferior. In conclusion, MLR and RF with additional features are recommended for operational short-term wind speed forecasts due to their excellent skills and without concern for performance degradation due to the inclusion of new features.

However, it should be emphasized that most deterministic post-processing methods (both bias correction and MME) are optimized with an object to minimize the conditional mean of the target and thus show unsatisfactory skills for capturing the strong winds, which are the primary concern for maritime safety. Additional use of QRF improves the performance in extreme conditions, but largely degrades the performance in light-to-moderate winds. Due to the difficulty of deterministic post-processing methods in achieving the balance between mean and extreme, future research should fully exploit the strength of multi-models to produce probabilistic forecasts with an accurate representation of forecast uncertainty using different techniques (Vannitsem et al. ^[48], Schulz and Lerch ^[42], Sun et al. ^[49]).

Acknowledgements: We thank Guangdong Meteorological Service for providing the in-situ measurements and numerical predictions.

REFERENCES

- [1] TU H, WANG X, MU L, et al. Predicting drift characteristics of persons-in-the-water in the South China Sea [J]. *Ocean Engineering*, 2021, 242: 110134, <https://doi.org/10.1016/j.oceaneng.2021.110134>
- [2] WANG Y, ZOU R, LIU F, et al. A review of wind speed and wind power forecasting with deep neural networks [J]. *Applied Energy*, 2021, 304: 117766, <https://doi.org/10.1016/j.apenergy.2021.117766>
- [3] JEON H, HARTMAN B, CUTLER H, et al. Estimating the economic impacts of improved wind speed forecasts in the United States electricity sector [J]. *Journal of Renewable and Sustainable Energy*, 2022, 14(3): 036101, <https://doi.org/10.1063/5.0081905>
- [4] ZHANG P, ZHANG J, CHEN M. Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation [J]. *Journal of Environmental Economics and Management*, 2017, 83: 8–31, <https://doi.org/10.1016/j.jeem.2016.12.001>
- [5] LIU X, LIN Z, FENG Z. Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM [J]. *Energy*, 2021, 227: 120492, <https://doi.org/10.1016/j.energy.2021.120492>
- [6] NESHAT M, NEZHAD M M, ABBASNEJAD E, et al. A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the Lillgrund offshore wind farm [J]. *Energy Conversion and Management*, 2021, 236: 114002, <https://doi.org/10.1016/j.enconman.2021.114002>
- [7] ZHANG W, LIN Z, LIU X. Short-term offshore wind power forecasting-A hybrid model based on Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and deep-learning-based Long Short-Term Memory (LSTM) [J]. *Renewable Energy*, 2022, 185: 611–628, <https://doi.org/10.1016/j.renene.2021.12.100>
- [8] FANG J, LIU W, YANG S, et al. Spatial-temporal changes of coastal and marine disasters risks and impacts in Mainland China [J]. *Ocean & Coastal Management*, 2017, 139(185): 125–140, <https://doi.org/10.1016/j.renene.2021.12.100>
- [9] REZAEI S, PELOT R, GHASEMI A. The effect of extreme weather conditions on commercial fishing activities and vessel incidents in Atlantic Canada [J]. *Ocean & Coastal Management*, 2016, 130: 115–127, <https://doi.org/10.1016/j.ocecoaman.2016.05.011>
- [10] OLABI A G, ABDELKAREEM M A. Renewable energy and climate change [J]. *Renewable and Sustainable Energy Reviews*, 2022, 158: 112111, <https://doi.org/10.1016/j.rser.2022.112111>
- [11] GIELEN D, BOSHELL F, SAYGIN D, et al. The role of renewable energy in the global energy transformation [J]. *Energy Strategy Reviews*, 2019, 24: 38–50, <https://doi.org/10.1016/j.esr.2019.01.006>
- [12] WANG Q. Effective policies for renewable energy-the example of China's wind power-lessons for China's photovoltaic power [J]. *Renewable and Sustainable Energy Reviews*, 2010, 14(2): 702–712, <https://doi.org/10.1016/j.rser.2009.08.013>
- [13] NELSON V, STARCHER K. *Wind Energy: Renewable Energy and the Environment* [M]. Boca Raton: CRC Press, 2009, <https://doi.org/10.1201/9781420075694>
- [14] BILGILI M, YASAR A, SIMSEK E. Offshore wind power development in Europe and its comparison with onshore counterpart [J]. *Renewable and Sustainable Energy Reviews*, 2011, 15(2): 905–915, <https://doi.org/10.1016/j.rser.2010.11.006>
- [15] ZHANG J, ZHANG J, CAI L, et al. Energy performance of wind power in China: A comparison among inland, coastal and offshore wind farms [J]. *Journal of Cleaner Production*, 2017, 143: 836–842, <https://doi.org/10.1016/j.jclepro.2016.12.040>
- [16] SHARMA R, SINGH D. A review of wind power and wind speed forecasting [J]. *Journal of Engineering Research and Application*, 2018, 8(7): 1–9, <https://doi.org/10.9790/9622-0807030109>
- [17] LYDIA M, KUMAR S S, SELVAKUMAR A I, et al. Linear and non-linear autoregressive models for short-term wind speed forecasting [J]. *Energy Conversion and Management*, 2016, 112: 115–124, <https://doi.org/10.1016/j.enconman.2016.01.007>
- [18] TIAN Z, CHEN H. A novel decomposition-ensemble

- prediction model for ultra-short-term wind speed [J]. *Energy Conversion and Management*, 2021, 248: 114775, <https://doi.org/10.1016/j.enconman.2021.114775>
- [19] XU W, LIU P, CHENG L, et al. Multi-step wind speed prediction by combining a WRF simulation and an error correction strategy [J]. *Renewable Energy*, 2021, 163: 772–782, <https://doi.org/10.1016/j.renene.2020.09.032>
- [20] TASCIKARAOGLU A, UZUNOGLU M. A review of combined approaches for prediction of short-term wind speed and power [J]. *Renewable and Sustainable Energy Reviews*, 2014, 34: 243–254, <https://doi.org/10.1016/j.rser.2014.03.033>
- [21] SKAMAROCK W C, KLEMP J B, DUDHIA J, et al. A description of the advanced research WRF model version 4 [R]. National Center for Atmospheric Research: Boulder, USA, 2019, <https://doi.org/10.5065/1DFH-6P97>
- [22] CHEN D, XUE J, YANG X, et al. New generation of multi-scale NWP system (GRAPES): General scientific design [J]. *Chinese Science Bulletin*, 2008, 53(22): 3433–3445, <https://doi.org/10.1007/s11434-008-0494-z>
- [23] KNUTTI R, ALLEN M R, FRIEDLINGSTEIN P, et al. A review of uncertainties in global temperature projections over the twenty-first century [J]. *Journal of Climate*, 2008, 21(11): 2651–2663, <https://doi.org/10.1175/2007JCLI2119.1>
- [24] ALAVI O, SEDAGHAT A, MOSTAFAEIPOUR A. Sensitivity analysis of different wind speed distribution models with actual and truncated wind data: A case study for Kerman, Iran [J]. *Energy Conversion and Management*, 2016, 120: 51–61, <https://doi.org/10.1016/j.enconman.2016.04.078>
- [25] WOOD A W, LEUNG L R, SRIDHAR V, et al. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs [J]. *Climatic Change*, 2004, 62(1–3): 189–216, <https://doi.org/10.1023/B:CLIM.0000013685.99609.9e>
- [26] WERNER A T, CANNON A J. Hydrologic extremes—an intercomparison of multiple gridded statistical downscaling methods [J]. *Hydrology and Earth System Sciences*, 2016, 20(4): 1483–1508, <https://doi.org/10.5194/hess-20-1483-2016>
- [27] BAO L, GNEITING T, GRIMIT E P, et al. Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction [J]. *Monthly Weather Review*, 2010, 138(5): 1811–1821, <https://doi.org/10.1175/2009MWR3138.1>
- [28] LI D, FENG J, XU Z, et al. Statistical bias correction for simulated wind speeds over CORDEX-East Asia [J]. *Earth and Space Science*, 2019, 6(2): 200–211, <https://doi.org/10.1029/2018EA000493>
- [29] ZENG X, XUE F, YAO L, et al. Comparative study of different error correction methods on model output wind field [J]. *Journal of Applied Meteorological Science*, 2019, 30(1): 49–60, <https://doi.org/201902290448971398>
- [30] HAGEDORN R, DOBLAS-REYES F J, PALMER T N. The rationale behind the success of multi-model ensembles in seasonal forecasting—I: Basic concept [J]. *Tellus A: Dynamic Meteorology and Oceanography*, 2005, 57(3): 219–233, <https://doi.org/10.1111/j.1600-0870.2005.00103.x>
- [31] JIANG Z, LI W, XU J, et al. Extreme precipitation indices over China in CMIP5 models, Part I: Model evaluation [J]. *Journal of Climate*, 2015, 28(21): 8603–8619, <https://doi.org/10.1175/JCLI-D-15-0099.1>
- [32] PAGOWSKI M, GRELL G A, DEVENYI D, et al. Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts [J]. *Atmospheric Environment*, 2006, 40(18): 3240–3250, <https://doi.org/10.1016/j.atmosenv.2006.02.006>
- [33] WU J, LU G, WU Z. Flood forecasts based on multi-model ensemble precipitation forecasting using a coupled atmospheric-hydrological modeling system [J]. *Natural Hazards*, 2014, 74: 325–340, <https://doi.org/10.1007/s11069-014-1204-6>
- [34] LEDUC M, LAPRISE R, DE ELIA R, et al. Is institutional democracy a good proxy for model independence? [J]. *Journal of Climate*, 2016, 29(23): 8301–8316, <https://doi.org/10.1175/JCLI-D-15-0761.1>
- [35] CRAWFORD J, VENKATARAMAN K, BOOTH J. Developing climate model ensembles: A comparative case study [J]. *Journal of Hydrology*, 2019, 568: 160–173, <https://doi.org/10.1016/j.jhydrol.2018.10.054>
- [36] AHMED K, SACHINDRA D A, SHAHID S, et al. Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms [J]. *Atmospheric Research*, 2020, 236: 104806, <https://doi.org/10.1016/j.atmosres.2019.104806>
- [37] ZHOU W, YANG H, FANG Z. Wind power potential and characteristic analysis of the Pearl River Delta region, China [J]. *Renewable Energy*, 2006, 31(6): 739–753, <https://doi.org/10.1016/j.renene.2005.05.006>
- [38] ZHI S Q, QIAN G M, LUO J L. A study of wind velocity varying with altitude on the coastal areas of Guangdong Province [J]. *Tropical Geography*, 2001, 21(2): 131–134, in Chinese with English abstract.
- [39] HERBACH H, BELL B, BERRISFORD P, et al. The ERA5 global reanalysis [J]. *Quarterly Journal of the Royal Meteorological Society*, 2020, 146(730): 1999–2049, <https://doi.org/10.1002/qj.3803>
- [40] BEER T. Beaufort Wind Scale [M]. BOBROWSKY P T (ed), *Encyclopedia of Natural Hazards*, Springer: Dordrecht, 2013: 42–45.
- [41] TAYLOR K E. Summarizing multiple aspects of model performance in a single diagram [J]. *Journal of Geophysical Research: Atmospheres*, 2001, 106(D7): 7183–7192, <https://doi.org/10.1029/2000JD900719>
- [42] SCHULZ B, LERCH S. Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison [J]. *Monthly Weather Review*, 2022, 150(1): 235–257, <https://doi.org/10.1175/MWR-D-21-0150.1>
- [43] SUN L, LAN Y, JIANG R. Using CNN framework to improve multi-GCM ensemble predictions of monthly precipitation at local areas: An application over China and comparison with other methods [J]. *Journal of Hydrology*, 2023, 623: 129866, <https://doi.org/10.1016/j.jhydrol.2023.129866>
- [44] TAILLARDAT M, MESTRE O, ZAMO M, et al. Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics [J]. *Monthly Weather Review*, 2016, 144(6): 2375–2393, <https://doi.org/10.1175/MWR-D-15-0260.1>

- [45] TAILLARDAT M, MESTRE O. From research to applications—examples of operational ensemble post-processing in France using machine learning [J]. *Nonlinear Processes in Geophysics*, 2020, 27(2): 329–347, <https://doi.org/10.5194/npg-27-329-2020>
- [46] CALLE M L, URREA V. Stability of Random Forest importance measures [J]. *Briefings in Bioinformatics*, 2011, 12(1): 86–89, <https://doi.org/10.1093/bib/bbq011>
- [47] PETSCHER Y, LOGAN J A, ZHOU C. Extending conditional means modeling: An introduction to quantile regression [M]. PETSCHER Y, SCHATSCHNEIDER C, COMPTON D L (eds), *Applied Quantitative Analysis in Education and the Social Sciences*, London: Routledge, 2013: 15–45.
- [48] VANNITSEM S, BREMNES J B, DEMAHEYER J, et al. Statistical postprocessing for weather forecasts—review, challenges and avenues in a big data world [J]. *Bulletin of the American Meteorological Society*, 2020, 102(3): E681–E699, <https://doi.org/10.1175/BAMS-D-19-0308.1>
- [49] SUN L, LAN Y, SUN X, et al. Deterministic forecasting and probabilistic post - processing of short - term wind speed using statistical methods [J]. *Journal of Geophysical Research: Atmospheres*, 2024, 129(7): e2023JD040134, <https://doi.org/10.1029/2023JD040134>

Citation: SUN Xian, SUN Lei, LIANG Xiu-ji, et al. Short-Term Wind Speed Forecasts over the Pearl River Estuary: Numerical Model Evaluation and Deterministic Post-Processing [J]. *Journal of Tropical Meteorology*, 2024, 30(4): 390–404, <https://doi.org/10.3724/j.1006-8775.2024.035>