ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 38, MAY 2021, 845–861

• **Original Paper** •

An Experiment on the Prediction of the Surface Wind Speed in Chongli Based on the WRF Model: Evaluation and Calibration

Na LI[1], Lingkun RAN[*1], Dongdong SHEN[1], and Baofeng JIAO[1,2]

[1]*Key Laboratory of Cloud-Precipitation Physics and Severe Storms (LACS), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029, *China*

[2]*Key Laboratory of Meteorological Disaster, Ministry of Education/Joint International Research Laboratory of Climate and Environment Change/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing* 210044, *China*

(Received 23 June 2020; revised 11 December 2020; accepted 11 January 2021)

ABSTRACT

In this study, the ability of the Weather Research and Forecasting (WRF) model to generate accurate near-surface wind speed forecasts at kilometer- to subkilometer-scale resolution along race tracks (RTs) in Chongli during the wintertime is evaluated. The performance of two postprocessing methods, including the decaying-averaging (DA) and analogy-based (AN) methods, is tested to calibrate the near-surface wind speed forecasts. It is found that great uncertainties exist in the model's raw forecasts of the near-surface wind speed in Chongli. Improvement of the forecast accuracy due to refinement of the horizontal resolution from kilometer to subkilometer scale is limited and not systematic. The RT sites tend to have large bias and centered root mean square error (CRMSE) values and also exhibit notable underestimation of high-wind speeds, notable overestimation or underestimation of the near-surface wind speed at high altitudes, and notable underestimation during daytime. These problems are not resolved by increasing the horizontal resolution and are even exacerbated, which leads to great challenges in the accurate forecasting of the near-surface wind speed in the competition areas in Chongli. The application of postprocessing methods can greatly improve the forecast accuracy of near-surface wind speed. Both methods used in this study have comparable abilities in reducing the (positive or negative) bias, while the AN method is also capable of decreasing the random error reflected by CRMSE. In particular, the large biases for high-wind speeds, wind speeds at high-altitude stations, and wind speeds during the daytime at RT stations can be evidently reduced.

Key words: near-surface wind speed forecast, bias corrections, complex terrain

Citation: Li, N., L. K. Ran, D. D. Shen, and B. F. Jiao, 2021: An experiment on the prediction of the surface wind speed in Chongli based on the WRF model: Evaluation and calibration. *Adv. Atmos. Sci.*, **38**(5), 845–861, https://doi.org/10.1007/s00376-021-0201-4.

Article Highlights:

- Improvement of near-surface wind forecasts due to refinement of horizontal resolution is limited and not systematic in Chongli.
- Near-surface wind speed forecasts in the competition areas in Chongli suffer from great uncertainties.
- The application of postprocessing methods is an effective way to improve the forecast accuracy of near-surface wind speed.

1. Introduction

The 24th Winter Olympic and Paralympic Games will be held in China in the year 2022. The Chongli District of Zhangjiakou will host all outdoor snow sports events, including cross-country skiing, ski jumping, and six other snow sports. The competition results of these sports and the performance and safety of athletes can be considerably affected by near-surface winds. For this reason, accurate forecasts of the near-surface wind speed in the competition areas are of prime importance for the success of these games.

Enhanced observation networks and numerical weather prediction (NWP) models are two main approaches to generate the weather forecasts for the Olympic Games. With the support of the World Meteorological Organization (WMO), the Vancouver 2010, Sochi 2014, and Pyeongchang 2018 Winter Olympic and Paralympic Games were committed to

* Corresponding author: Lingkun RAN
Email: rlk@mail.iap.ac.cn

the improvement of high-resolution numerical forecasts and construction of enhanced observation networks (e.g., Joe et al., 2010; Isaac et al., 2014; Vasil'ev and Dmitrieva, 2015; Han et al., 2016, 2018; Kiktev et al., 2017; Lee et al., 2018). An enhanced observation network promotes the improvement of weather nowcasting. Those short-term forecasts are based on NWP models. However, the performance of NWP models for the prediction of near-surface variables in complex terrains remains unsatisfactory.

Due to the influence of initial conditions, atmospheric physical process parameterizations, and numerical approximations, significant errors exist in forecasts of near-surface conditions (Hanna and Yang, 2001; Zhang and Zheng, 2004). This issue becomes more complicated in complex terrains due to the inability of models to finely differentiate the structure of the topography and to successfully reproduce turbulence effects, diurnal cycles, cold air pools in basins and valleys, nocturnal low-level jets, and boundary layer stability phenomena (Shafran et al., 2000; Whiteman et al., 2001; Zhong and Fast, 2003; Cheng and Steenburgh, 2005; Liu et al., 2019). These factors impose a major influence on the accuracy of surface wind speed forecasts. The model forecast errors of the surface wind speed have originally been attributed to the approximations of atmospheric physical processes in planetary boundary layer (PBL) and land surface model (LSM) parameterization schemes. By testing the performance of different PBL schemes in simulating near-surface variables, Zhang and Zheng (2004) found that all boundary layer schemes underestimated the wind speed during the daytime and overestimated it during the nighttime. Jiménez and Dudhia (2012) argued that the Weather Research and Forecasting (WRF) model overestimation of the surface wind speed was caused by the unresolved topography, which produced an additional drag to that generated by vegetation. They introduced a parameterization scheme that considered these unresolved topographic effects in the momentum flux. However, Frediani et al. (2016) reported that this parameterization scheme penalizes moderate and high winds. Duan et al. (2018) further investigated the WRF model performance in predicting surface variables in Northwest China, where the underlying surface is very complex. It was found that the 10-m wind speed forecast errors in June are strongly correlated with the terrain, but this correlation is not as apparent in December (winter).

To better discriminate complex terrains and associated vegetation, land use types, and subgrid physical processes, a practical solution to improve numerical forecasts is to increase the horizontal model resolution. Currently, many NWP systems are operated at the kilometer scale and are being developed to the subkilometer scale (e.g., Baldauf et al., 2011; Seity et al., 2011; Clark et al., 2012; Min, 2014; Huang et al., 2017). However, whether increasing the horizontal resolution can improve the surface variable prediction accuracy remains uncertain. Verification of the real-time forecasts from the fifth-generation Pennsylvania State University–National Center for Atmospheric Research mesoscale model (MM5) by Mass et al. (2002) suggested that these forecasts were notably improved when decreasing the grid spacing from 36 to 12 km, but little improvement was achieved with further reduction from 12 to 4 km. The verification of WRF model forecasts of surface variables by Zhang et al. (2013) also revealed that surface forecasts at a fine resolution (1.33 km) do not always outperform those at a coarse resolution. Leroyer et al. (2014) extended the NWP system in Meteorological Service of Canada (MSC) to a subkilometer grid spacing (0.25 km) but did not observe much improvement, although better sea-breeze flows were reproduced. However, the wind speed forecasts of Vionnet et al. (2015) exhibited notable improvements with decreasing grid spacing for high-altitude stations exposed to or sheltered from wind.

These studies further indicate the large uncertainty in surface wind speed forecasts. Although we expect the model surface wind speed forecast to improve with increasing model resolution and enhanced topography representation, the accuracy depends on many factors. Compared to increasing the horizontal resolution, a more effective way to reduce the model forecast error and to acquire more accurate forecasts may be the application of bias-correction methods to postprocess raw numerical forecasts.

The model errors that influence the forecast accuracy include random and systematic errors. Many studies have demonstrated that these errors, particularly the systematic errors, can be predicted with statistical methods, which can then be used to calibrate the numerical forecasts. Regarding the near-surface wind speed, the available methods that can be applied for calibration include the running mean method (e.g., Hacker and Rife, 2007), model output statistics (MOS, e.g., Gneiting et al., 2005), Kalman-filtering approach (e.g., Delle Monache et al., 2008), Bayesian model averaging (e.g., Raftery et al., 2005), gene-expression algorithms (e.g., Bakhshaii and Stull, 2009), decay-averaging adjustment method (e.g., Cui et al., 2012), analogy-based (AN) method (e.g., Delle Monache et al., 2011), etc. Several of these methods have been tested in wind speed forecasting during previous Winter Olympic Games. Frogner et al. (2016) compared and evaluated three ensemble prediction systems with different grid spacings to accurately predict wintertime weather conditions in complex terrains. The relative importance of the resolution and calibration was assessed. Han et al. (2018) evaluated six postprocessing methods and found that the Bayesian model averaging method was superior in calibrating near-surface wind speed ensemble forecasts. Zhang et al. (2020) applied the decay-averaging (DA) method to correct near-surface condition forecasts from the GRAPES_3 km model, which was employed during the Pyeongchang 2018 Winter Olympic Games.

In different regions, the model performance is different, and the effectiveness of different bias-correction methods is also different. Although certain model assessment results have been obtained under complex terrain conditions in the past and comparisons of different bias-correction methods have been conducted to calibrate model forecasts, there is no assessment or bias-correction research for the Chongli area. As the Winter Olympic Games approach, it is imperative to determine whether numerical modeling and associ-

ated bias-correction methods can provide suitable near-surface wind speed forecasts in the competition venues in Chongli. For this purpose, this paper evaluates the near-surface (10 m) wind speed forecast from the WRF model in the competition venues in Chongli during the wintertime through a comparison to enhanced surface observations along the race tracks (RTs). The ability of two postprocessing methods to calibrate near-surface wind speed forecasts is simultaneously examined. Here, the two evaluated calibration methods include a first-order adaptive Kalman-filtering method, namely, the DA method and the AN method. The AN method is based on the principle that similar forecasts tend to have similar errors. With the use of several statistical parameters, all the wind speed forecasts from the model or from the postprocessing methods are evaluated in the following aspects: 1) the effect of horizontal model resolution refinement from the kilometer to subkilometer scale on improving the near-surface wind speed forecasts in the competition venues in Chongli; 2) the performance of the postprocessing methods in generating better near-surface wind forecasts in Chongli; 3) the influence of the complex terrain, diurnal cycle, and wind intensity on the performance of the model and postprocessing methods.

The paper is arranged as follows: section 2 presents the details of the evaluation, including the near-surface wind speed forecast experiments with the WRF model, the observation dataset for verification, the postprocessing methods, and the evaluation parameters. Section 3 analyzes the verification results, focusing on the above three aspects, and a summary is provided in section 4.

2. Data and methods

2.1. *Numerical model and experimental configuration*

The WRF v4.1.3 model was applied to forecast the near-surface wind speed in Chongli. Three one-way nested domains were employed with horizontal resolutions of 4.05, 1.35 and 0.45 km. These forecasts were conducted for the period from 1 January to 16 February 2019, which is the wintertime in Chongli (Fig. 1). The horizontal grid of each domain has dimensions of 361×361. The orography for the 4.05- and 1.35-km grids was obtained from the U.S. Geological Survey (USGS) 30-arc-s (approximately 0.9 km) orography dataset. The 0.45-grid topography is from Shuttle Radar Topography Mission (SRTM) data which has 3-arc-s resolution. There are 51 vertical levels in each domain.

The physical parameterization schemes employed in this study include the Thompson aerosol-aware microphysics scheme (a bulk microphysical parameterization for winter precipitation forecasts considering aerosol impacts, Thompson et al., 2008; Thompson and Eidhammer, 2014), the new version of the Rapid Radiative Transfer Model (RRTMG) radiation scheme (a radiation model by Iacono et al., 2008 considering long-lived greenhouse gases), the Noah land surface model (LSM; the most widely used land surface model with new modifications to better represent processes over ice sheets and snow covered area, Alapaty et al., 2008), and the YSU planetary boundary layer (PBL) parameterization scheme (non-local-K scheme with explicit entrainment layer, parabolic K profile in unstable mixed layer, terrain variance-related correction, and top-down mixing driven by radiative cooling, Hong et al., 2006).

The initial and lateral boundary conditions of the 4.05-km experiment are derived from the analysis and forecast data provided by the National Centers for Environmental Prediction (NCEP) Global Forecasting System (NCEP/GFS). As an experiment, the current work used the GFS data with 0.5° horizontal resolution and 3-h temporal resolution, which permits us to obtain long-term history data. The use of the GFS data with 0.25° horizontal resolution for the wind forecast is in test.

The 4.05-km experiment was initialized at 0000 UTC and run for 48 h with 3-hourly lateral boundary conditions.

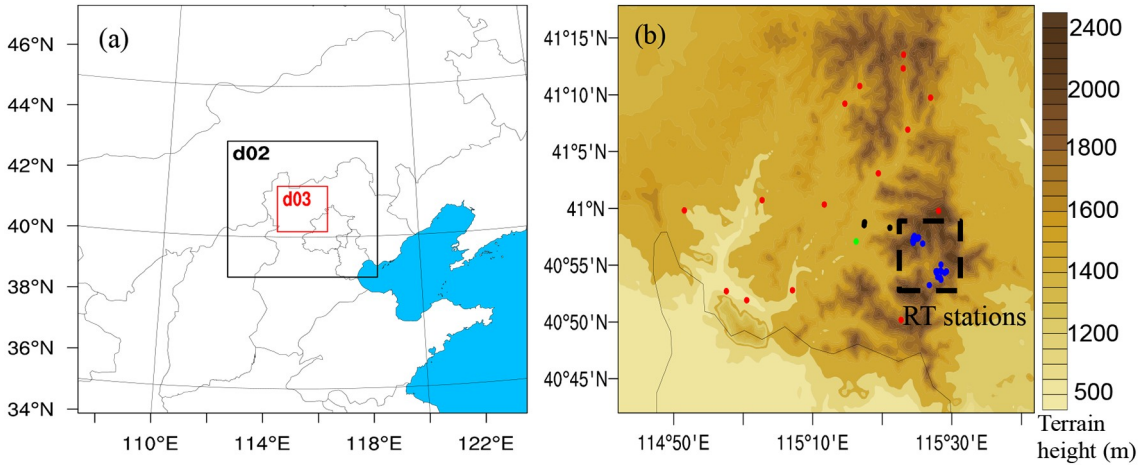


Fig. 1. Model domains for the numerical experiments: (a) the 4.05-, 1.35- and 0.45-km domains, (b) distribution of the available observation stations (solid points) superimposed on the 0.45-km-grid orography of Chongli. The solid points in (a) include 1 national station (green), 17 regional stations (red), 3 test stations (black) and 23 race track stations (blue). Apart from the single national station, the other stations are all automatic stations.

Six-hour delays were applied to the initialization of the 1.35-km (0600 UTC) and 0.45-km (1200 UTC) domains. The initial and boundary conditions of the 1.35-km run are from the prediction fields of the 4.05-km run, which are downscaled to 1.35-km grid by ndown.exe in WRF. Then, when the 1.35-km run is completed, its prediction fields further provide the initial and boundary conditions for the 0.45-km run. Lateral boundary conditions of 1.35-km and 0.45-km runs are hourly updated. All the domains have the same ending time. Therefore, for each domain, the output period progressively decreases with increasing resolution, namely, 48 h for the 4.05-km domain, 42 h for the 1.35-km domain, and 36 h for the 0.45-km domain. The 4.05-, 1.35- and 0.45-km forecasts were started once each day over the entire experimental period from 1 January to 16 February 2019. The 10-m wind speed forecasts are output each hour and remapped to the observation stations from the nearest grid point for the following evaluation and calibration procedures. The distribution of the observation stations is presented in the next subsection.

2.2. *Observation dataset in Chongli*

As one district of Zhangjiakou to the northwest of Hebei Province, Chongli is located along the border areas of the Yanshan Mountain and the Yinshan Range, characterized by several northeast-southwest-oriented mountain ranges and many valleys (Fig. 1b). The complex terrain provides favorable natural conditions for winter snow sports but also causes great difficulties in weather forecasting.

The distribution of the observation stations for the 10-m wind speed evaluation and calibration steps are shown in Fig. 1b. They include 1 national station (green), 17 regional stations (red), 3 test stations (black), and 23 stations along RTs (blue). These observation sites are separated into two groups, namely, the RT stations near the competition venues and other stations excluding the race-track (NRT) stations. The RT stations include all the stations along the RTs (the blue dots in Fig. 1b) and two regional stations (the red dots coinciding with the blue dots) near the race venues. The NRT stations include 1 national station, 15 regional stations, and 3 test stations (the other dots except for the RT stations in Fig. 1b). The performance of the WRF model and associated postprocessing methods is assessed between the RT and NRT stations. As shown in Fig. 1b, the RT stations are mainly distributed across the high-altitude area in Chongli, while the NRT stations mostly occur in valleys. The different distributions of the RT and NRT stations are defined to enable a comparison of them, considering the performance of the numerical model in the valleys and high-altitude sites (Zhang et al., 2013; Vionnet et al., 2015; Duan et al., 2018).

2.3. *Postprocessing methods for calibration*

2.3.1. *The DA method*

As a relatively easy and flexible postprocessing method, the DA method has been applied to reduce the bias of the global ensemble forecast by NCEP operation. The algorithm is mostly the same as that described in Cui et al. (2012), except that it was applied to the station sites, and the bias was estimated with the aid of available observations at each station site instead of with the analysis grid data. The same method was adopted by Zhang et al. (2020) to correct the 2-m temperature, 2-m relative humidity and 10-m wind speed forecasts from the GRAPES model during the Pyeongchang Winter Olympic Games. The two steps of this adaptive algorithm for a given station can be described as follows:

$$B_t = (1-\omega)\, B_{t_b-1} + \omega\left(F_{t_b} - O_{t_b}\right), \tag{1}$$

$$\mathrm{DA}_t = F_t - B_t, \tag{2}$$

where t is the current forecast time and t_b is the prior forecast time of t considering the latest available observation (for the forecast less than 24 h, $t_b = t-1$; however, for the forecast greater than 24 h, t_b may be $t-2$ or even earlier depending on the availability of the latest observation). F_t and O_t are the numerical forecast and corresponding observation, respectively, at time t. B_t is the decaying average bias, which is updated by combining the latest available bias $F_{t_b} - O_{t_b}$ at t_b and B_{t_b-1} at the prior forecast time of t_b with weight coefficient ω. The initial value of B_t is 0. By applying the decaying average bias B_t to the current forecast F_t, the new bias-corrected forecast DA_t at each lead time and each station can be obtained, as expressed in Eq. (2).

2.3.2. *The AN method*

The DA method utilizes all the available past errors to estimate the current bias with the latest errors assigned higher weights. In comparison, the AN method applies a sorting process to the past predictions and selects certain members with similar features to those of the current forecast [called analogs by Delle Monache et al. (2011)]. It is assumed that for a given location, analogous forecasts have similar forecast errors. Therefore, the current forecast error can be inferred from the errors of past analogous forecasts. Delle Monache et al. (2011) defined a metric to measure the similarity of past forecasts to current forecast as follows:

$$\left\|F_t, A_{t_0}\right\| = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_i} \sqrt{\sum_{j=-\Delta t}^{\Delta t} \left(F_{i,t+j} - A_{i,t_0+j}\right)^2}, \tag{3}$$

where F_t is the forecast to be corrected at given time t, A_{t_0} is the available past forecast at time t_0 before given time t, N_v is the number of analogous physical variables to measure the analogy (e.g., 10-m wind speed, 2-m temperature, etc. for 10-m wind speed corrections), $F_{i,t+j}$ is the current forecast of a given variable in a time window $2\Delta t$, A_{i,t_0+j} is the corresponding past forecast in the same time window, σ_i is the standard deviation of all the available past forecasts of a given variable, and w_i is the weight of each variable which

is determined by the correlation of the variable to the near-surface wind speed. The smaller of the above matric value, the better analogy of the history forecast to the current forecast.

With the above metric, all the available past predictions are ranked from small matric value to higher, and the best N_a analogous members are selected for bias correction of the current numerical forecast. Delle Monache et al. (2011) directly applied the corresponding observations of the best N_a analogous members to generate a new AN forecast as follows:

$$\mathrm{AN}_t = \sum_{i=1}^{N_a} \gamma_i \mathrm{OA}_i \,, \tag{4}$$

where AN_t is the AN forecast at time t for a given location, $\{\mathrm{OA}_i\}_{i=1,2,\ldots,N_a}$ are the corresponding observations to the N_a members ($N_a = 5$ in this study), and γ_i is the weight of each selected analogous member, which is defined as follows:

$$\gamma_i = \frac{\left\| F_t, A_{i,t_i} \right\|}{\sum_{j=1}^{N_a} \left\| F_t, A_{j,t_j} \right\|} \,. \tag{5}$$

The above equation implies that the observations of the more analogous members (lower analogy metric) have higher weights in AN_t.

In this study, the DA and AN algorithms are applied to calibrate the model 10-m wind speed forecasts in Chongli at each of the observation sites. Calibrations were conducted for the 4.05-km/1.35-km/0.45-km domain of the 1–48-h/1–42-h/1–36-h (1-h intervals) forecasts from 1 January to 16 February 2019 at all station sites. During the application, many sensitivity tests are carried out to optimize the correction. Regarding the DA method, a test of the weight ω is conducted, and $\omega = 8.2\%$ is adopted, which yields the best root mean square error (RMSE, introduced in the following subsection). For the AN method, an important task is to determine the physical parameters in the analogy matric [Eq. (3)], which can largely influence the efficiency of the method. In Delle Monache et al. (2011), five near-surface variables, including 10-m wind speed, 10-m wind direction, 2-m temperature, and surface pressure and humidity, are selected. However, many studies have shown that the near-surface wind speed is not just related to surface variables, but also elevated atmospheric conditions. For example, strong horizontal momentum at 700 hPa or 500 hPa may be transported downward into the near-surface layer inducing an increase of near-surface wind speed. Therefore, apart from the surface variables, eight other variables outside of the near-surface layer, such as wind speed at 700 hPa and 500 hPa, temperature at 700 hPa and 500 hPa, etc., are added as analogous physical variables. Also, their contributions to the matric are determined by their correlation coefficient c_i to the observed 10-m wind speed, that is

$$w_i = \frac{c_i}{\sum_{i=1}^{N_v} c_i} \,. \tag{6}$$

Note that c_i has been normalized using its maximum and minimum values.

2.4. *Evaluation methods*

BIAS, CRMSE, RMSE, Pearson correction coefficient (COR), and normalized standard deviation (NSD) are adopted to examine the performance of the WRF model and the three postprocessing methods in forecasting the 10-m wind speed. Their physical meanings have been introduced in detail in Taylor (2001), Delle Monache et al. (2011), and other papers on model evaluation (e.g., Vionnet et al., 2015). Their definitions in this study are as follows:

$$\mathrm{BIAS} = \overline{F} - \overline{O} \,, \tag{7}$$

$$\mathrm{CRMSE} = \sqrt{\frac{1}{N_p} \sum_{1}^{N_p} \left[\left(F_i - \overline{F} \right) - \left(O_i - \overline{O} \right) \right]^2} \,, \tag{8}$$

$$\mathrm{RMSE}^2 = \frac{1}{N_p} \sum_{1}^{N_p} (F_i - O_i)^2 = \mathrm{CRMSE}^2 + \mathrm{BIAS}^2 \,, \tag{9}$$

$$\mathrm{COR} = \frac{1}{s_f s_o} \frac{1}{N_p} \sum_{1}^{N_p} \left[\left(F_i - \overline{F} \right) \left(O_i - \overline{O} \right) \right] , \tag{10}$$

$$\mathrm{NSD} = \frac{s_f}{s_o} \,, \tag{11}$$

where F_i is the 10-m wind speed from the raw model forecast or from the three postprocessing methods for each of the 4.05-, 1.35- and 0.45-km domains under any given condition (station, forecast time, and time), O_i is the corresponding 10-m wind observation, N_p are all available pairs of prediction F_i and observation O_i, $\overline{F}$ and $\overline{O}$ are the means of the predictions and observations, respectively, over the N_p values, and s_f and s_o are the standard deviations of the predictions and observations, respectively.

BIAS measures the systematic errors of the forecast. CRMSE represents the random error other than the bias in RMSE. COR and NAD are used to plot Taylor diagrams, which reflect the correlation pattern of the predictions and observations.

3. Results

3.1. *Global evaluation*

Here, the term global implies that the evaluation is conducted across all available stations and times (including the forecast times; N_p = stations × forecast times × times) for

each of the domains and each of the forecast methods. Through this global evaluation, the overall performance of the raw, DA, and AN 10-m wind speed forecasts in the different model domains can be evaluated.

3.1.1. *The refinement of the horizontal resolution from the kilometer to subkilometer scale*

Figure 2 shows three Taylor diagrams plotted with the COR and NSD values for all stations, RT stations (only RT stations are included in the evaluation), and NRT stations (only NRT stations are included in the evaluation). To analyze the performance of the model with increasing horizontal resolution from the kilometer to subkilometer scale, the dots marked with number 1 in Fig. 2 are first explained.

In Fig. 2a, the change in COR with increasing model resolution refinement is slight. The COR values of all three domains are approximately 0.4, and the 0.45-km grid COR value is even slightly smaller than that of the 4.05-km grid, with that of the 1.35-km grid being an intermediate value. In contrast, the change in NSD is evident, with the 1.35-km grid exhibiting the best NSD approaching 1. The slightly smaller COR value of the 0.45-km grid probably occurs because this grid contains more details of the flow across the mountains and thus exhibits more disturbances in the near-surface wind speed (Raderschall et al., 2008; Vionnet et al., 2015). The circumstances at the RT and NRT stations are similar except that the COR value at the RT stations is smaller than 0.3 while the COR value at the NRT stations is larger than 0.5 (Figs. 2b–c). The relatively small COR value indicates an unsatisfactory performance of the WRF model in forecasting the 10-m wind speed at the RT stations.

BIAS and CRMSE for each of the domains and each of the forecasting methods are further calculated across all stations, RT stations, and NRT stations. Here, we still focus on the raw WRF forecasts first. As shown in Figs. 3a–b, at all stations, both BIAS and CRMSE decrease with increasing horizontal resolution down to the subkilometer scale. BIAS decreases from 0.696 m s^{-1} in the 4.05-km grid to 0.53 m s^{-1} in the 1.35-km grid and further to 0.068 m s^{-1} in the 0.45-km grid. The BIAS improvement reaches up to 90.2% from the 4.05-km grid to the 0.45-km grid.

An improvement of the 10-m wind speed forecast with increasing horizontal resolution is also observed at the RT and NRT stations, with a slightly higher improvement at the RT stations. Figure 3a reveals that in contrast to what is observed for the COR values, BIAS at the RT stations is lower than that at the NRT stations at the same resolution. At the RT stations, BIAS is reduced by 92% when the resolution is refined down to the subkilometer scale, and this reduction is 82% at the NRT stations. However, confidence in this result is somewhat uncertain due to the possible canceling of the positive and negative BIAS, which can result in a lower BIAS value. In contrast to BIAS, CRMSE at the RT stations is larger than that at the NRT stations (Fig. 3b). However, the CRMSE improvement with increasing horizontal resolution refinement at the RT stations (7.8%) is

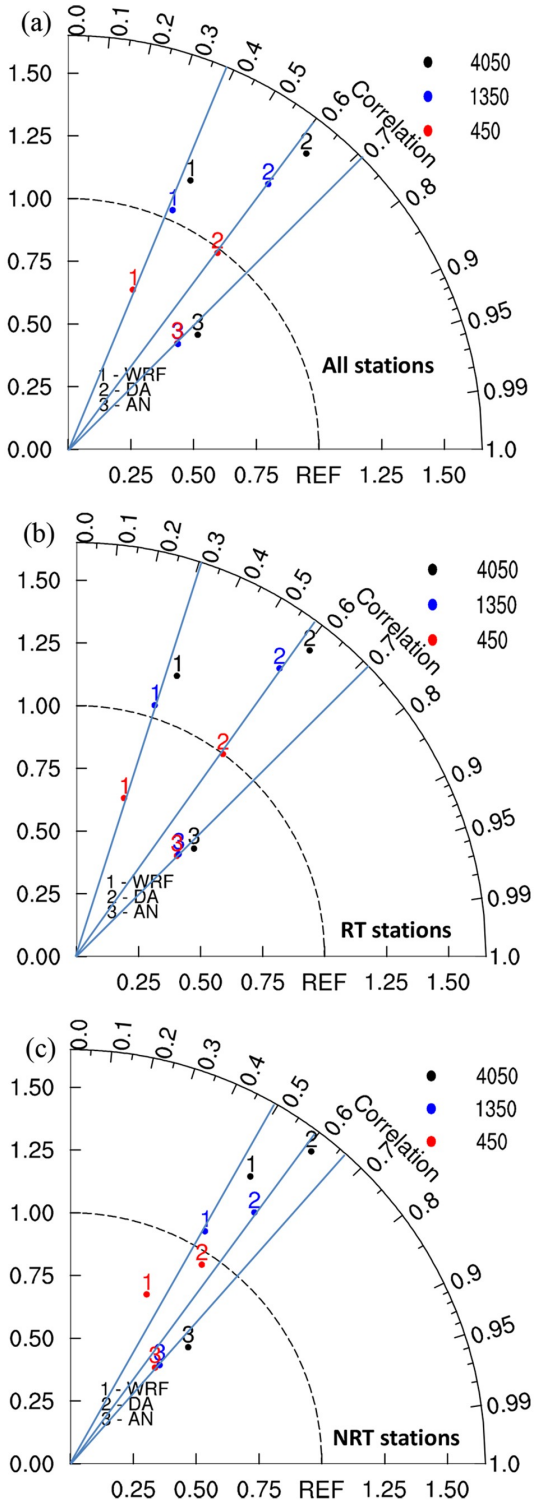
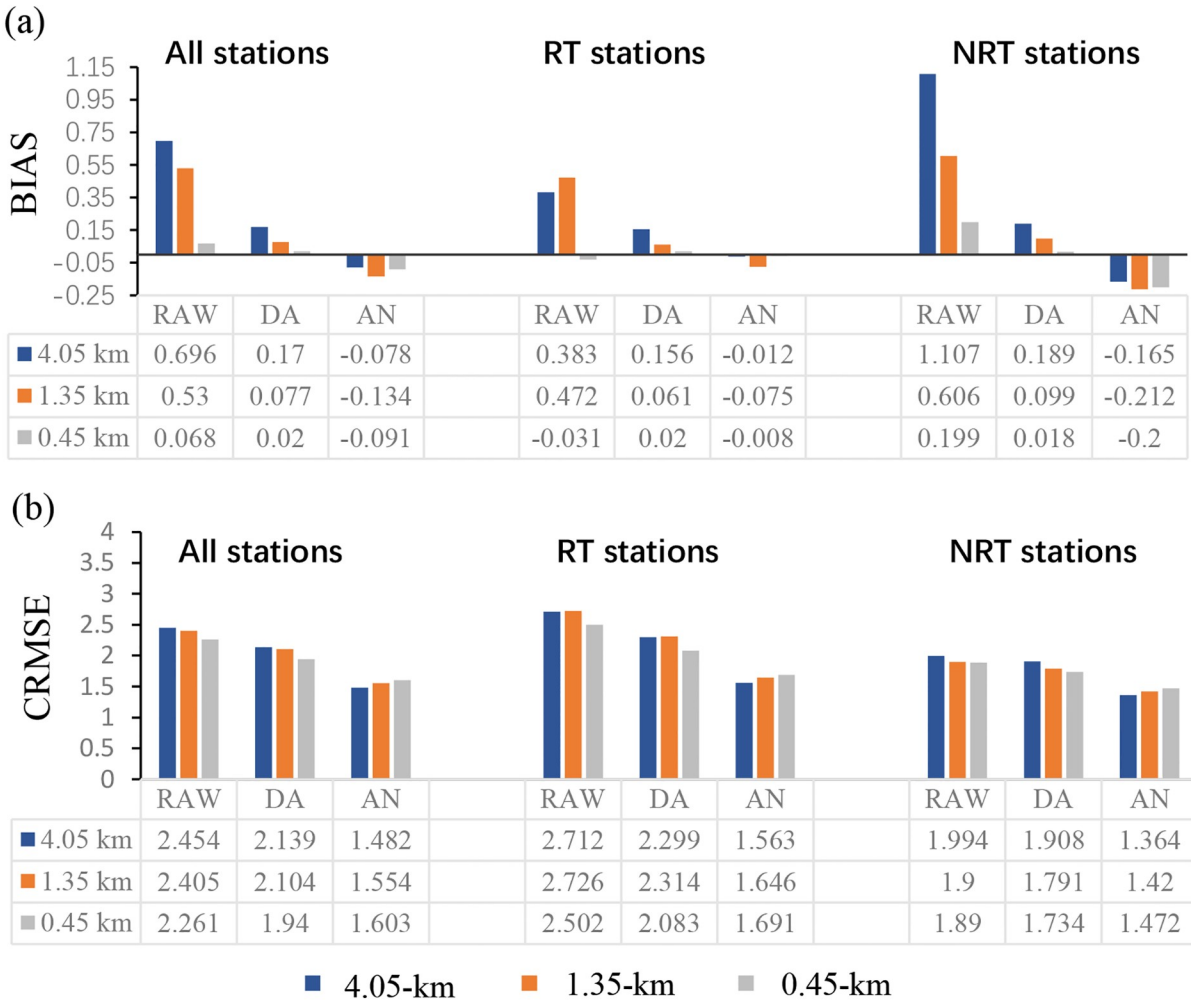


Fig. 2. Taylor diagrams based on COR and NSD of the raw WRF forecasts and the forecasts obtained with the DA, AN, and ANE methods of the 10-m wind speed for the different domains at (a) all stations; (b) RT stations; (c) NRT stations. The black dot indicates the 4.05-km domain, the blue dot indicates the 1.35-km domain, the red dot indicates the 0.45-km domain, 1 indicates the WRF raw forecast, 2 indicates the DA forecast, 3 indicates the AN forecast, and 4 indicates the ANE forecast.



	RAW	DA	AN	RAW	DA	AN	RAW	DA	AN
4.05 km	0.696	0.17	-0.078	0.383	0.156	-0.012	1.107	0.189	-0.165
1.35 km	0.53	0.077	-0.134	0.472	0.061	-0.075	0.606	0.099	-0.212
0.45 km	0.068	0.02	-0.091	-0.031	0.02	-0.008	0.199	0.018	-0.2

	RAW	DA	AN	RAW	DA	AN	RAW	DA	AN
4.05 km	2.454	2.139	1.482	2.712	2.299	1.563	1.994	1.908	1.364
1.35 km	2.405	2.104	1.554	2.726	2.314	1.646	1.9	1.791	1.42
0.45 km	2.261	1.94	1.603	2.502	2.083	1.691	1.89	1.734	1.472

Fig. 3. Histogram of (a) BIAS and (b) CRMSE of the raw WRF forecasts and the forecasts with the DA, and AN methods of the 10-m wind speed for the different domains.

still greater than that at the NRT stations (5.2%). This result is similar to that reported in Vionnet et al. (2015), in which a greater improvement of the near-surface wind speed forecasts at high-altitude stations is obtained with increasing horizontal resolution.

A preliminary result from the above global evaluations is that decreasing the model grid spacing does improve the near-surface wind speed forecasts, especially at the RT stations, although the model performance at the RT stations may not be better than that at the NRT stations (based on COR and CRMSE).

3.1.2. *The performance of the postprocessing methods*

Figures 2 and 3 are revisited here to examine the performance of the various postprocessing methods in the calibration of the 10-m wind speed forecasts via a comparison to the performance attained by refining the horizontal resolution of the WRF model. According to Fig. 2a, both postprocessing methods generate larger COR values than those of the raw WRF forecasts. Compared to the model raw forecasts, the DA method increases COR to approximately 0.6, and the AN method can improve this value to approximately 0.7, indicating a better performance of the AN method in forecasting the wind speed change pattern. This result is also applicable to the RT and NRT stations, as shown in Figs. 2b–c. However, a noteworthy point of the evaluations is that COR at the RT stations can be improved to approximately 0.7 after calibration with the AN method. This value is even larger than the COR value at the NRT stations, although the COR value of the raw forecasts at the RT stations is much smaller than that at the NRT stations. This indicates that calibration by the postprocessing methods is an effective way to generate more accurate near-surface wind speed forecasts in the competition areas in Chongli. This is also observed for BIAS and CRMSE, as shown in Fig. 3.

Figure 3 shows that compared to the raw WRF forecasts, the postprocessing methods are able to reduce BIAS and CRMSE in each of the domains, except for the 0.45-km-grid BIAS values. Regarding the BIAS values in the 0.45-km domain, the raw WRF forecasts attain a lower BIAS value than that of the AN method. However, as indicated above, the BIAS results require further examination due to the possible canceling of any positive and negative BIAS values. The relative performance of the DA and AN

methods in reducing the bias of the model raw forecasts is not systematic (Fig. 3a). For example, at the RT stations, the AN method attains better performance. At all stations, the DA method performs better than the AN method. However, in terms of reducing CRMSE (Fig. 3b), it is observed that the AN method performs better. This is consistent with the results of Delle Monache et al. (2011) whereby the AN method reduces the random model errors and thus improves the predictive skills on near-surface wind speed forecasts.

3.1.3. *Evaluations according to the wind speed intensity at the RT stations*

The evaluations in this subsection are similar to the above evaluations except that the wind speed intensity is considered. The wind speed is separated into three bins (0–5, 5–10, and >10 m s^{-1}), and BIAS, CRMSE, and COR values are calculated for each of the bins across the available RT station observations and times (Fig. 4).

In Fig. 4a, at all model resolutions, a positive bias is observed at wind speeds below 5 m s^{-1}, with the 0.45-km grid exhibiting the lowest bias. The bias is observed to decrease with increasing wind speed, resulting in a negative bias at high wind speeds. The negative bias at wind speeds of 5–10 m s^{-1} occurs at all resolutions, with a finer resolution resulting in a larger underestimation. At wind speeds higher than 10 m s^{-1}, all model resolutions exhibit the highest negative bias, at approximately –5.5 m s^{-1}, indicating the inability of the WRF model to forecast the high wind speeds along the tracks of the competition areas in Chongli. The relationship between the bias and wind speed is similar to that reported by Jiménez and Dudhia (2012) and Frediani et al. (2016) for the Yonsei University (YSU) PBL parameterization scheme, which has been applied in this study. In their study, it was argued that the high negative bias related to high wind speeds occurs due to the higher drag, which is described by the drag expression in the Monin-Obukhov similarity theory.

The random error component measured by CRMSE also exhibits an increase with increasing wind speed (Fig. 4b) for the WRF raw forecast. At wind speeds below 5 m s^{-1} and at a moderate wind speeds of 5–10 m s^{-1}, the increase in horizontal resolution from the kilometer to subkilometer scale results in a decrease in CRMSE. However, at high wind speeds, i.e., higher than 10 m s^{-1}, this decrease is not evident. The observed COR trend at each wind speed range is similar to that depicted in Fig. 2, which shows that COR does not change much with increasing horizontal resolution. Moreover, at each wind speed range, COR further decreases, with the moderate wind speed range attaining the smallest COR value.

These analyses indicate that increasing the horizontal resolution may generate better forecasts at low wind speeds, but at high wind speeds, retrogression may occur.

Figure 4 reveals that the postprocessing method plays an important role in reducing both (positive or negative) BIAS and CRMSE and improving COR, especially at high wind speeds. To reduce the high negative bias at wind speeds exceeding 5 m s^{-1}, the DA method performs better than the AN method. At wind speeds higher than 10 m s^{-1}, the DA method can reduce the negative bias by approximately 65%. The AN method yields better performance in reducing the random errors at both high and low wind speeds. However, it seems that the postprocessing method is not able to improve COR at wind speeds higher than 10 m s^{-1}, with the COR values of the AN method even smaller than those of the raw forecasts.

3.2. *Spatial analysis*

In this subsection, we analyze how BIAS, CRMSE, and COR behave according to terrain variability. These parameters are computed with all available N_p pairs of the predictions and observations of the 10-m wind speed at each station (N_p = forecast time × times). The relation between the terrain and 10-m wind speed forecast BIAS, CRMSE, and COR at each of the different resolutions at the RT and NRT stations is shown in Fig. 5. In Fig. 5, the stations are arranged according to their actual terrain from lowest to highest elevation, with the 4.05-, 1.35-, and 0.45-km model terrain elevations of the corresponding observation sites also shown to examine the terrain bias in each of the domains.

By comparing Figs. 5a, c, and e to Figs. 5b, d, and f, the elevation of the RT stations ranges from 1625 to 2100 m, and most of the NRT stations are lower than the RT stations, corresponding to Fig. 2a. Generally, the 0.45-km-grid model terrain (the green solid line) matches the actual terrain (the black solid line) the best, followed by the 1.35-km grid (the red solid line) and 4.05-km grid (the green solid line). However, the terrain biases at the RT and NRT stations are quite different. At the RT stations, the model exhibits great limitations in resolving the terrain fluctuations in the competition areas in Chongli. Compared to the smoothly changing actual terrain from the lower RT stations to the higher RT stations (the black solid line in Figs. 5a, c, and e), the model terrain elevation fluctuations in each of the domains are notable (the colored solid lines in Figs. 5a, c, and e). However, at the NRT stations, the trend of the actual terrain elevation is successfully resolved by the 0.45-km grid and mostly represented by the 4.05- and 1.35-km grids, although a higher bias occurs in the coarser grids.

Although many factors can induce a bias in the near-surface wind speed forecasts, the mismatch between the actual and model terrains is always an important factor. As indicated by Jiménez and Dudhia (2012), an unresolved topography produces an additional drag, which is a possible reason for the frequent overestimation of the near-surface wind speed by the WRF model. Vionnet et al. (2015) indicated that the details of complex terrains may notably influence near-surface wind speed forecasts, such as wind deceleration in sheltered areas or wind increase due to the presence of crests and mountains (Raderschall et al., 2008).

In Figs. 5a–b, the BIAS change trend with the terrain elevation is not very evident, but the model tends to have a

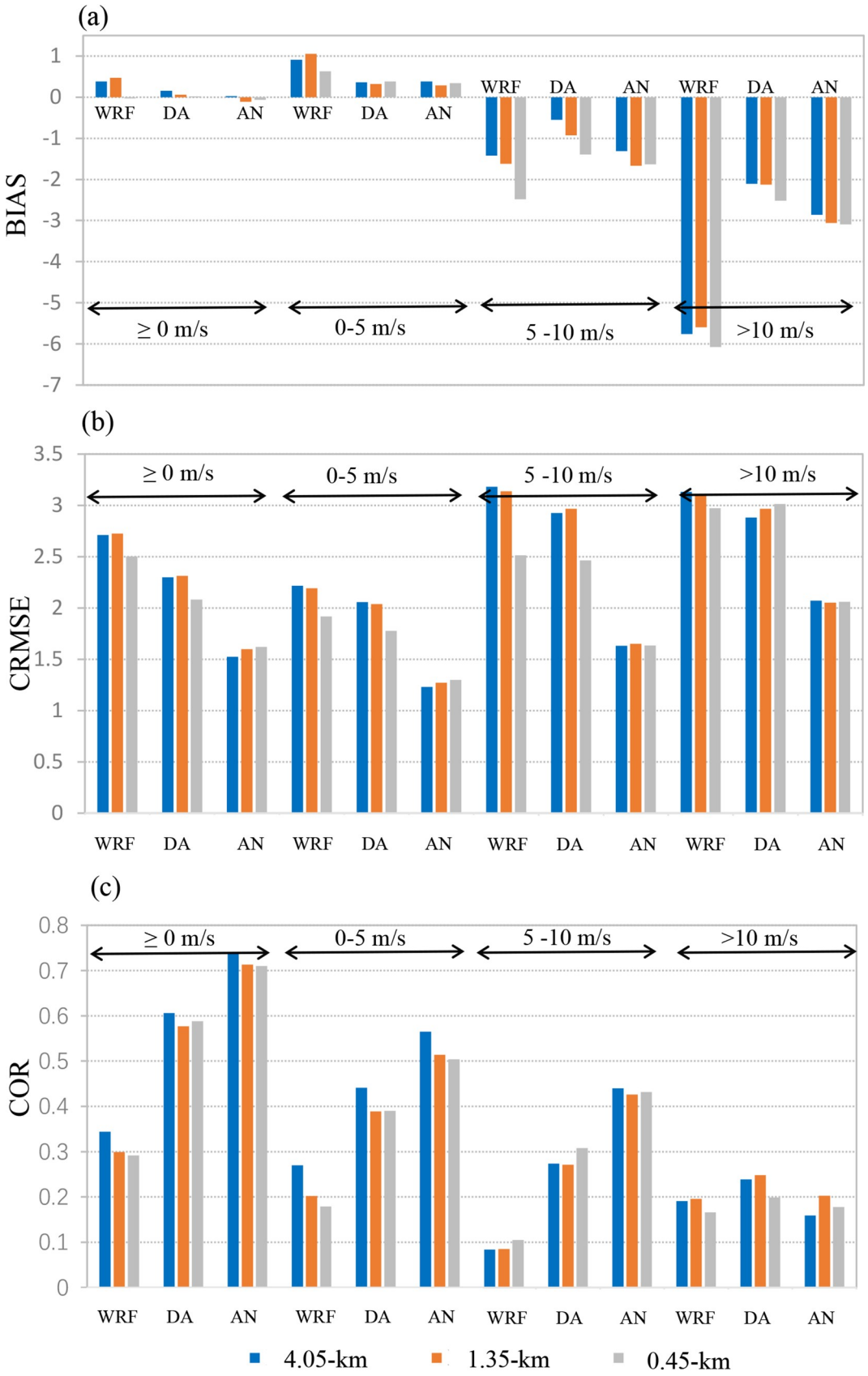


Fig. 4. (a) BIAS, (b) CRMSE and (c) COR of the raw WRF forecasts and the forecasts obtained with the DA and AN methods at the different wind speed intensities for the various model domains.

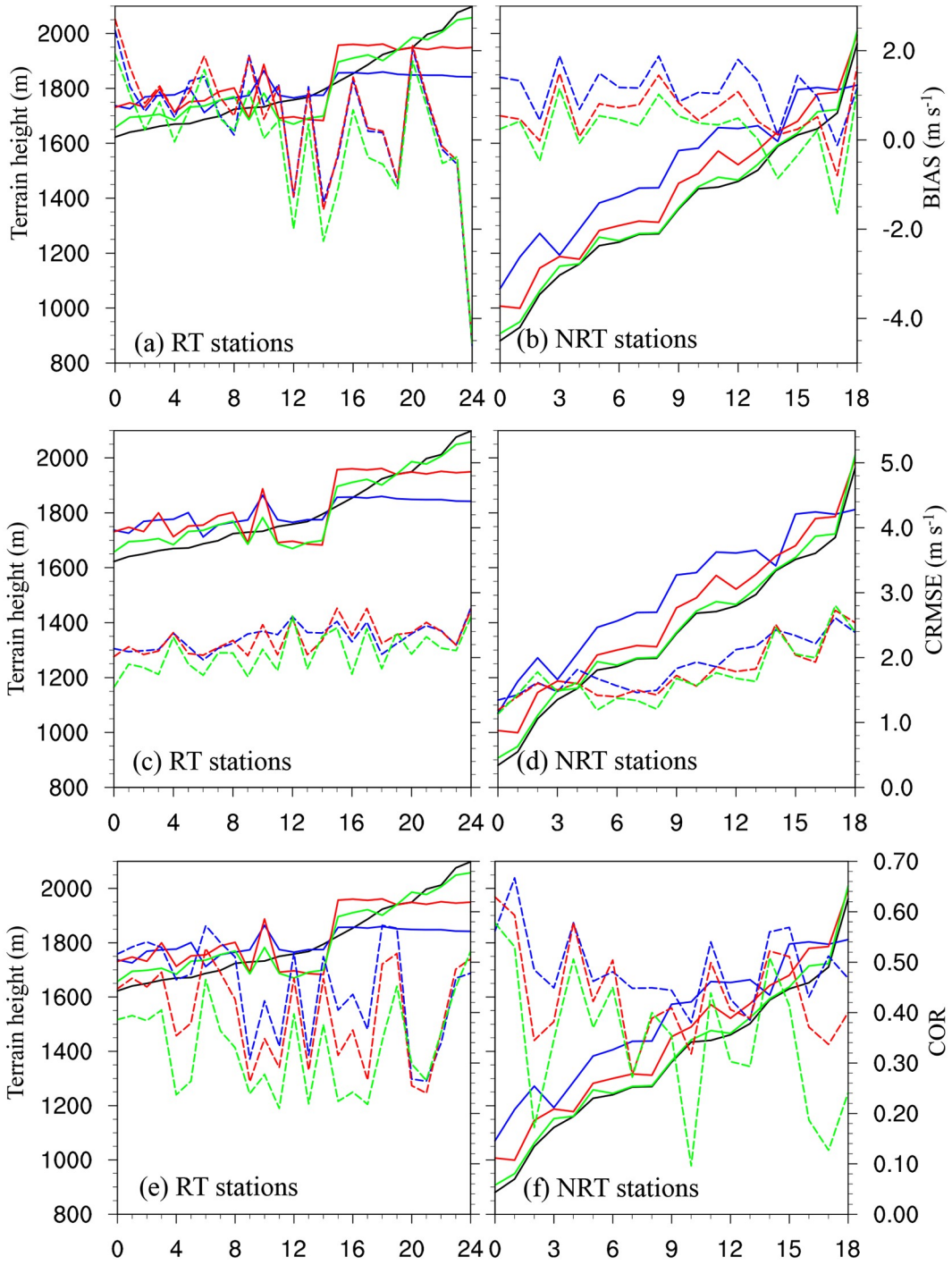


Fig. 5. Change in (a, b) BIAS, (c, d) CRMSE and (e, f) COR at the RT and NRT stations (*x*-axis), which are arranged from the lowest to highest elevation according to the actual terrain elevation. The blue color indicates the 4.05-km grid, the red color indicates the 1.35-km grid and the green color indicates the 0.45-km grid. The solid lines are the station elevations with the actual terrain elevation in black. The dashed lines indicate BIAS, CRMSE and COR.

lower bias at low altitudes (the NRT stations) than at high altitudes (the RT stations). At most NRT stations, the model exhibits an overestimation of the near-surface wind speed. Underestimation occurs at some of the NRT stations with a terrain elevation higher than 1600 m. At the RT stations, overestimation also occurs at low altitudes, but a very large underestimation is observed at the RT stations above 1700 m. The CRMSE positive trend with increasing terrain elevation is more evident, as shown in Figs. 5c–d, and larger CRMSE values are observed at the RT stations. Compared to BIAS and CRMSE, the fluctuations in COR with increasing station elevation are much larger, ranging from 0.2 to 0.6 at the RT sta-

tions and from 0.1 to 0.68 at the NRT stations.

Generally, increasing the horizontal resolution from the kilometer to subkilometer scale results in lower BIAS and CRMSE values (Figs. 5a–d), with the exception of the high-altitude RT and NRT stations. At these stations, a higher (positive or negative) BIAS is observed at the finer resolution (in the 0.45-km grid). This is different from the results of Duan et al. (2018), in which resolution refinement reduces the wind speed forecast error, or of Vionnet et al. (2015), where the underestimation is reduced at the finer resolution due to its ability to capture a crest. The underestimation at the high altitudes in the coarser domains is likely due to the negative terrain bias (the solid lines in Fig. 5), namely, the model is unable to discriminate the details of the high peaks of the terrain (Vionnet et al., 2015). However, a better resolution of terrain details may produce more drag in the PBL parameterization scheme and thus result in a larger underestimation in the finer domain (the 0.45-km grid) (Jiménez and Dudhia, 2012). These uncertainties make it difficult to accurately forecast the near-surface wind speeds in Chongli under such complex terrain conditions.

The performance of the two postprocessing methods in calibrating the 10-m wind speed forecasts at the RT and NRT stations is shown in Figs. 6 and 7, respectively. As shown in Figs. 6a–c and 7a–c, both methods can successfully reduce the notable bias in the raw WRF forecasts. The various postprocessing methods result in lower spatial variabilities than that of the raw WRF forecasts. In particular, the high negative bias at the high-altitude RT stations is also evidently reduced. The highest bias of –4.5 m s^{-1} at the highest station (2100 m, Figs. 6a–c) is successfully reduced to –0.5–0.5 m s^{-1}. This result is similar to the global analysis results in the above subsection.

On reducing the random errors measured by CRMSE,

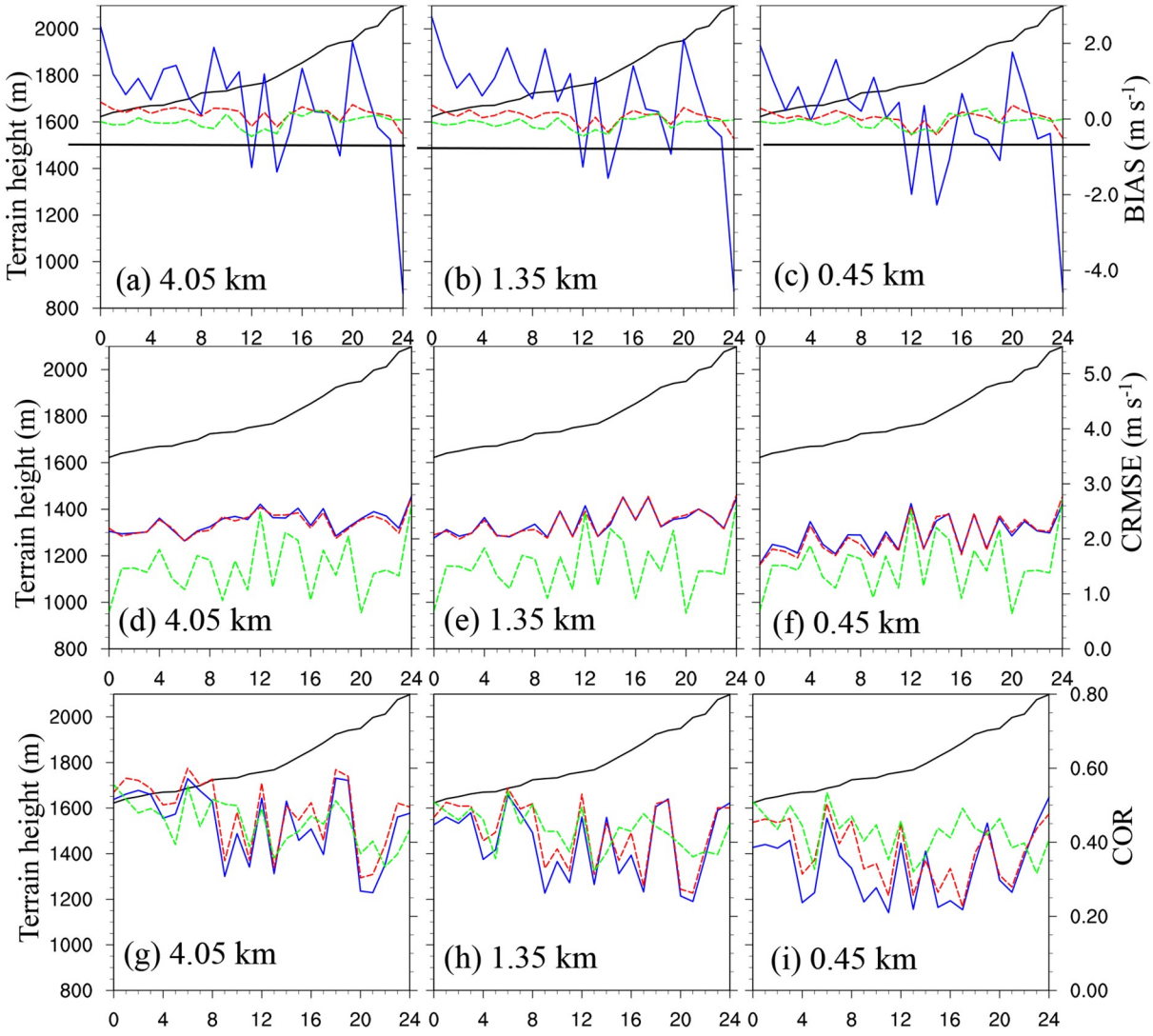


Fig. 6. Change in BIAS, CRMSE, and COR of the raw model forecasts (blue solid lines) and the forecasts obtained with the DA (red dashed lines) and AN (green dashed lines) methods of the 10-m wind speed at the RT stations (*x*-axis), which are arranged from the lowest to highest elevation according to the actual terrain elevation: (a, d, g) 4.05-km grid; (b, e, h) 1.35-km grid; and (c, f, i) 0.45-km grid. The black solid lines are the actual terrain elevations.

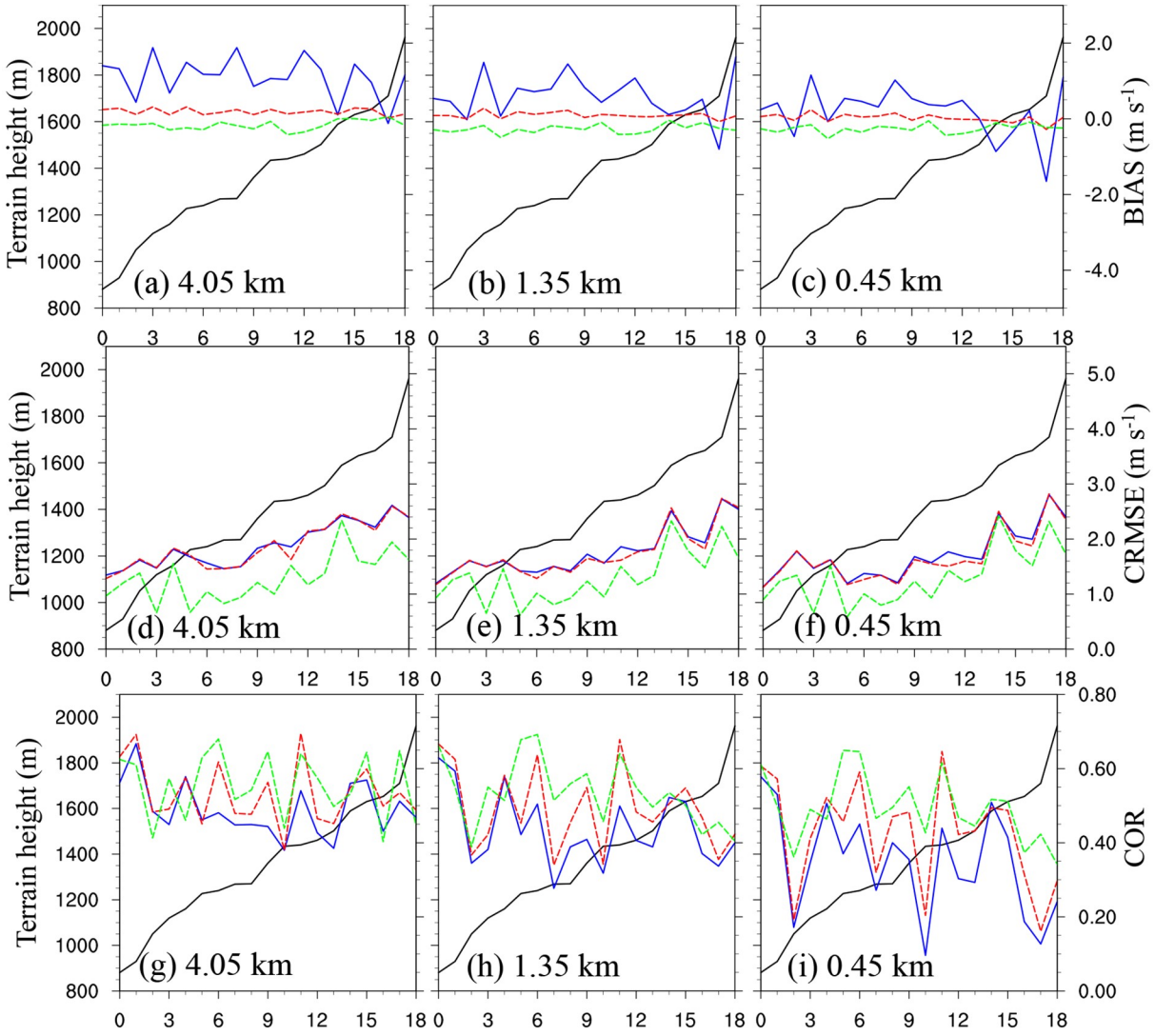


Fig. 7. Same as Fig. 6 except for the NRT stations (*x*-axis).

the DA method is almost unable to reduce the random errors, as the red lines nearly coincide with the blue lines in Figs. 6d–f and 7d–f. Nevertheless, the AN method exhibits a clear potential for CRMSE reduction. In addition, the AN method also performs better in improving COR (Figs. 6g–i). However, it seems that the COR improvement at the RT stations is not as large as that at the NRT stations (Figs. 7g–i).

3.3. *Diurnal variations in the errors*

The diurnal cycle is a basic characteristic of the surface variables in the PBL due to the influence of surface heat and momentum fluxes. The accurate simulation of the diurnal variations in near-surface variables is one of the most difficult tasks in NWP models. Although the various PBL parameterization schemes can basically reproduce the diurnal variations in the surface wind speed, the resulting errors are often large. Zhang et al. (2004) found that PBL parameterization schemes tend to underestimate the surface wind speed during the day and overestimate it at night. However, with the presence of nocturnal jets, positive bias errors of 10-m wind speed forecasts were observed in the analysis of Zhang et al. (2013). In our evaluations, the diurnal variations in both the systematic (BIAS) and random (CRMSE) errors of the WRF raw 10-m wind speed forecast are significant.

Figures 8 and 9 show the temporal variations in BIAS, CRMSE, and COR at the different horizontal resolutions and for the various forecast methods at the RT and NRT stations. Here, BIAS, CRMSE, and COR are computed with all available observations and predictions across all the stations and times at a given forecast time (N_p = stations × times). As shown in Figs. 8a–c and 9a–c, the bias of the raw model forecasts exhibits a distinct diurnal cycle at both the RT and NRT stations. In the 4.05-km domain (Figs. 8a and 9a), the bias exhibits a trough after sunrise in the early morning, followed by a ridge in the afternoon before sunset, a small trough during sunset and a wide ridge throughout the night over one cycle in the first 24 hours of the forecast, and a similar cycle is observed over the next 24–48 hours of the forecast. Corresponding to this diurnal pattern, the WRF

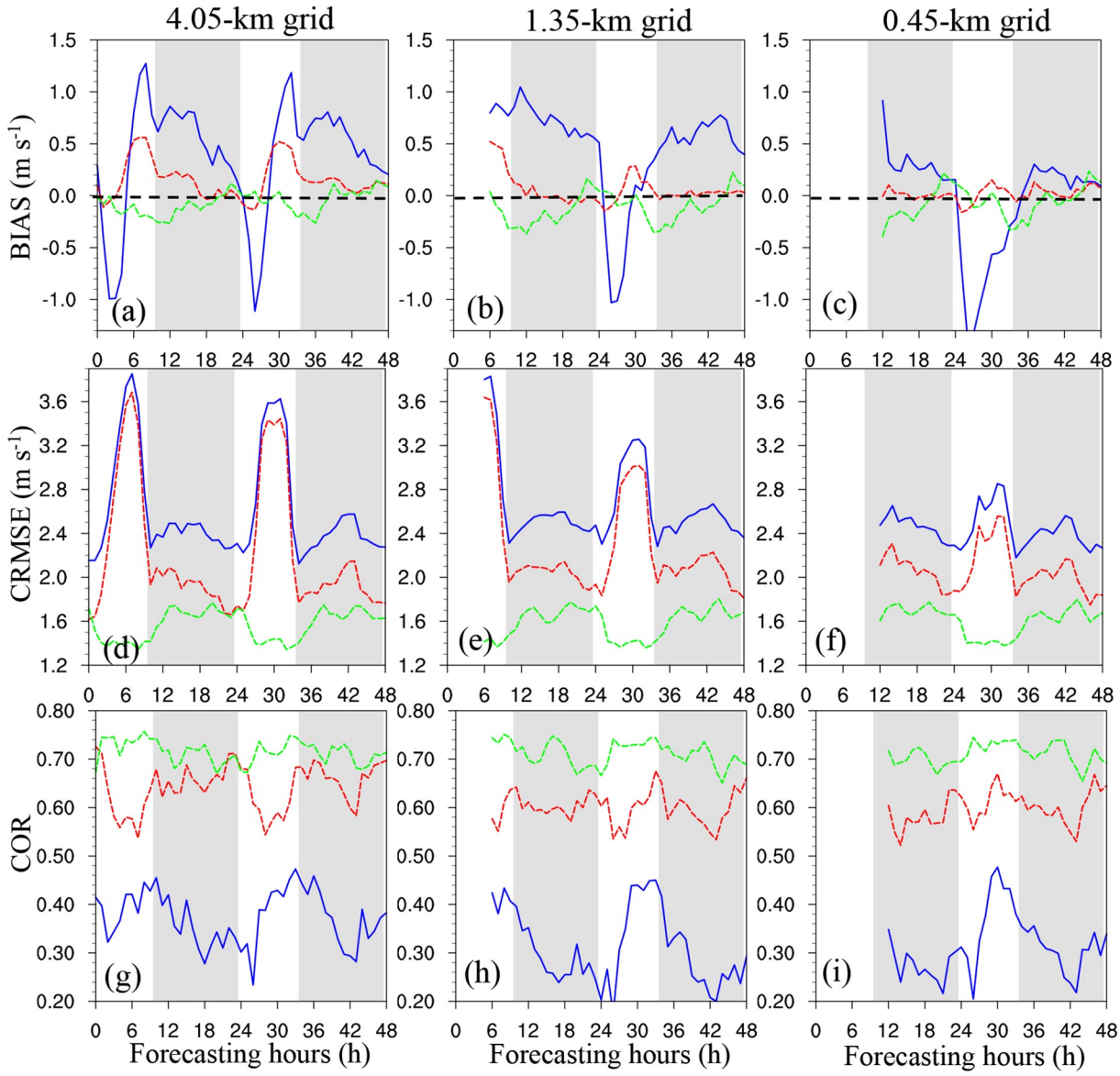


Fig. 8. Diurnal variations in BIAS, CRMSE, and COR at the different horizontal resolutions and for the various forecast methods at the RT stations. The *x*-axis is forecast hours. The blue solid line indicates the WRF raw forecast, the red dashed line indicates the DA method and the green dashed line indicates the AN method. The gray shaded areas indicate nighttime data, while the white areas indicate daytime data. This figure applies to sunrise and sunset in Chongli in January.

model tends to strongly underestimate the 10-m wind speed in the early morning, which is then greatly overestimated in the afternoon, and the notable overestimation persists throughout the night until sunrise at the RT stations (Fig. 8a). At the NRT stations, systematic overestimations are observed throughout the diurnal cycle (Fig. 9a). In particular, over one cycle, a minimum bias occurs at the NRT stations in the early morning, but this situation does not last long in the finer domains. In the 1.35- and 0.45-km domains (Figs. 8b–c and 9b–c), the diurnal pattern is slightly different from that in the 4.05-km domain with the absence of the small-amplitude ridge and trough at sunset. In addition, it is observed that the whole cycle moves from a positive bias toward a negative bias with increasing model resolution. This results in a persistent decrease in the positive bias during the nighttime and an increase in the negative bias during the daytime. In particular, with increasing model resolution, the model performance improves at night but deteriorates during the day. This phenomenon occurs at both the RT and NRT stations, and the daytime negative bias at the RT stations is higher than that at the NRT stations (Figs. 8b–c and 9b–c). The finding whereby large underestimation occurs in the daytime surface wind speed forecasts at the RT stations in the finer-resolution domains is critical as most races only occur during the day.

In Figs. 8a–c and 9a–c, both methods greatly reduce the bias of the raw forecasts, with the diurnal variations in the bias notably decreasing after calibration. At the RT stations,

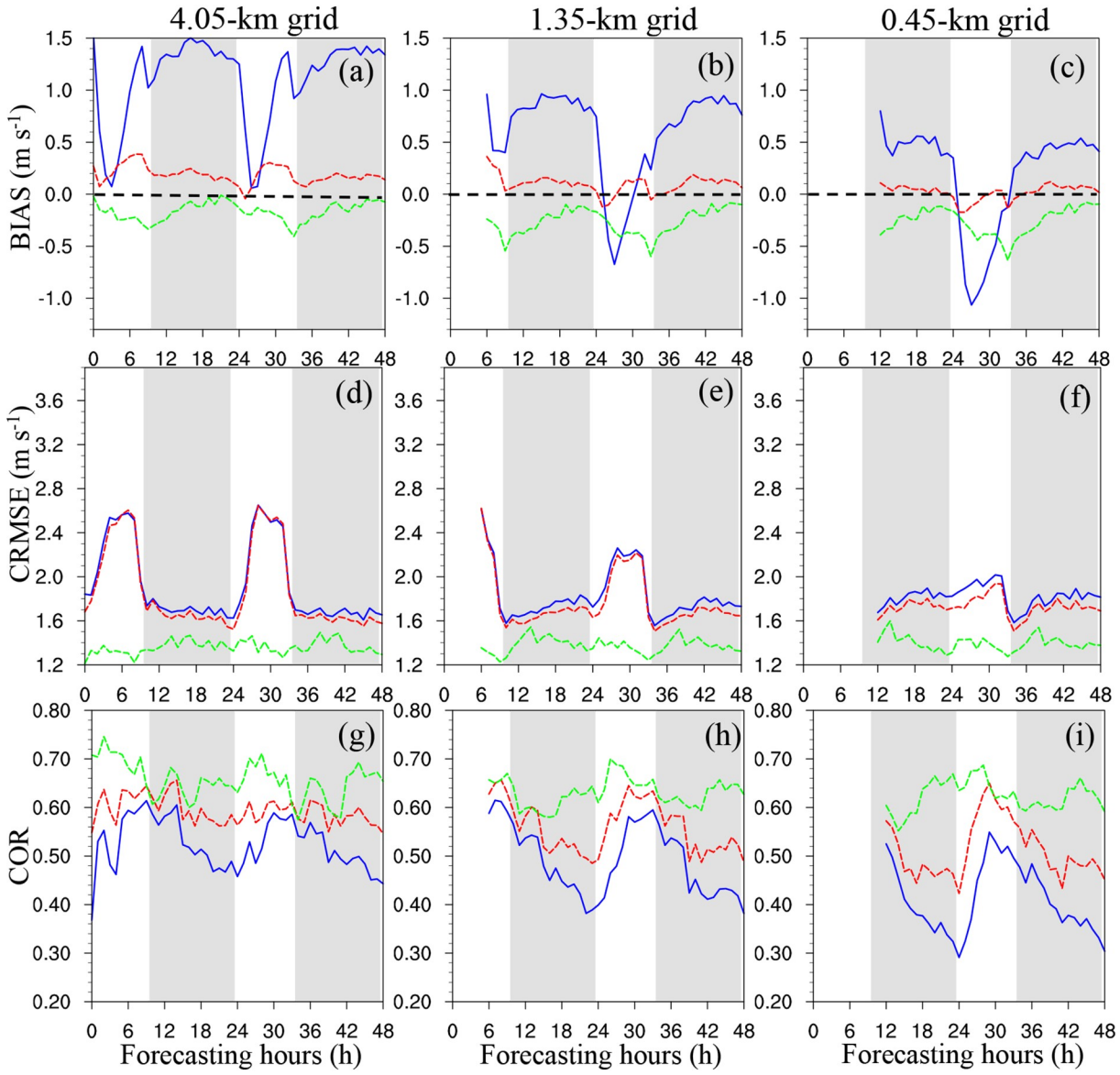


Fig. 9. Same as Fig. 8, except for the NRT stations.

the AN method outperforms the DA method in reducing the daytime negative bias to nearly zero (Figs. 8a–c), but its performance during the nighttime is not as good as that during the daytime. In the 1.35- and 0.45-km domains, the DA method attains better performance in reducing the nighttime positive bias. At the NRT stations, the performance of the AN method is not as good as that at the RT stations, where an evident negative bias is observed (Figs. 9a–c). The DA method performs better than the AN method either during the daytime or the nighttime.

Similar to BIAS, CRMSE of the raw WRF forecast also exhibits notable diurnal variations with high CRMSE values occurring in the daytime and low CRMSE values occurring in the nighttime (Figs. 8d–f and 9d–f). The first CRMSE peak is observed in the daytime approximately 6 h after sunrise, which then rapidly decreases until sunset. Right after sunset, CRMSE begins to increase again and exhibits a second peak at midnight or it remains constant throughout the night. Despite the similar pattern, CRMSE at the RT stations is much larger than that at the NRT stations, further reflecting the poor predictive skill of the WRF model in predicting the surface wind speed at the RT stations, especially during the daytime. With increasing horizontal model resolution, CRMSE during the daytime clearly decreases, but CRMSE does not change much during the nighttime for the raw WRF forecasts. CRMSE during the daytime decreases from approximately 3.6 to 2.8 m s^{-1} at the RT stations (Figs. 8d–f) and decreases from 2.6 to 2.0 m s^{-1} at the NRT stations (Figs. 9d–f) with the horizontal resolution increasing from 4.05 to 0.45 km. However, at night, there is little change.

Similar to the previous results, the random errors represented by CRMSE can be effectively reduced by the AN method, with the diurnal variations in CRMSE fluctuating

less than those in the raw forecasts. At the RT stations, CRMSE of the AN method even exhibits a trough during the daytime, and the daytime CRMSE decreases to approximately 1.4 m s^{-1} (Figs. 8d–f).

The diurnal variations in COR of the raw forecasts are slightly different from those in BIAS and CRMSE, which reveals the highest correlation at sunset and the lowest correlation at sunrise. COR at the RT stations (Figs. 8g–i) is much smaller than that at the NRT stations (Figs. 9g–i). However, it is clear that the postprocessing methods attain a better performance in improving COR at the RT stations (Figs. 8g–i). Figures 8g–i and 9g–i show that the AN method can improve COR to almost 0.7, while the DA method yields a lower COR value.

4. Conclusions

The ability of the WRF model to generate accurate near-surface wind speed forecasts for the 2022 Winter Olympic and Paralympic Games in Chongli is evaluated in this study. Two postprocessing methods, including the DA and AN methods, are applied to calibrate the raw model forecasts of the near-surface wind speed to improve the forecast accuracy. As one of the host locations of the Winter Olympic Games, accurate forecasts of near-surface variables in Chongli are of great importance for the success of these games. However, the very complex terrain in Chongli results in large uncertainties in near-surface wind speed forecasts. To acquire a better model performance, these uncertainties need to be understood in advance, in addition to the effectiveness of various postprocessing methods in improving raw model forecasts.

Several questions, as described in the Introduction section, are addressed, and they are mostly answered through this study. A global evaluation across all the stations and times showed that increasing the horizontal resolution could improve the model's 10-m wind speed forecasts by reducing both BIAS and CRMSE, especially at the RT stations. However, a more detailed analysis considering the wind speed intensity, the spatial variability, and the diurnal cycle revealed several problems at the RT stations, particularly those at altitudes higher than 1700 m. In terms of the 10-m wind speed forecasts at wind speeds higher than 5 m s^{-1}, the WRF model exhibits a high negative bias, which even increases when the horizontal resolution is refined from the kilometer to subkilometer scale. Although the finer-resolution domain contains a much better description of the actual terrain, at certain high-altitude RT locations, a high positive or negative bias is observed, and the negative bias even increases with increasing horizontal resolution. The diurnal cycles of the model errors in forecasting the surface wind speed are quite evident at these RT stations, with a clear negative bias during the daytime. Moreover, the negative bias is also exaggerated in the subkilometer domain. The possible reasons for these problems are the mismatch between the actual and model terrains and misrepresentation of the near-surface physical processes in the PBL and LSM parameterization schemes. These uncertainties pose a great challenge to the near-surface wind speed forecasting in the competition areas in Chongli.

An evaluation of the use of the various postprocessing methods to calibrate the near-surface wind speed forecasts indicates that these methods improve the prediction accuracy of the near-surface wind forecasts and partly resolve the above problems. Both tested methods clearly reduce the bias, and the AN method is better able to reduce the random errors represented by CRMSE. In addition, the notable bias at high wind speeds and the forecast accuracy issues at the high-altitude stations and during the daytime are also notably reduced. These results indicate the necessity of applying postprocessing methods.

Acknowledgements. This study was supported by the Strategic Pilot Science and Technology Special Program of the Chinese Academy of Sciences (Grant No. XDA17010105), the National Key Research and Development Project (Grant No. 2018YFC1507104), the Key Scientific and Technology Research and Development Program of Jilin Province (Grant No. 20180201035SF), and the National Natural Science Foundation of China (Grant Nos. 41875056, 41775140, 42075013 and 41575065).

REFERENCES

Alapaty, K., D. Niyogi, F. Chen, P. Pyle, A. Chandrasekar, and N. Seaman, 2008: Development of the flux-adjusting surface data assimilation system for mesoscale models. *J. Appl. Meteorol. Climatol.*, **47**, 2331–2350, https://doi.org/10.1175/2008JAMC1831.1.

Bakhshaii, A., and R. Stull, 2009: Deterministic ensemble forecasts using gene-expression programming. *Wea. Forecasting*, **24**, 1431–1451, https://doi.org/10.1175/2009WAF2222192.1.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1.

Cheng, W. Y. Y., and W. J. Steenburgh, 2005: Evaluation of surface sensible weather forecasts by the WRF and the Eta Models over the western United States. *Wea. Forecasting*, **20**, 812–821, https://doi.org/10.1175/WAF885.1.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, https://doi.org/10.1175/BAMS-D-11-00040.1.

Cui, B., Z. Toth, Y. J. Zhu, and D. C. Hou, 2012: Bias correction for global ensemble forecast. *Wea. Forecasing*, **27**, 396–410, https://doi.org/10.1175/WAF-D-11-00011.1.

Delle Monache, L., and Coauthors, 2008: A Kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone. *Tellus*, **60B**, 239–249, https://doi.org/10.1111/j.1600-0889.2007.00332.x.

Delle Monache, L. D., T. Nipen, Y. B. Liu, G. Roux, and R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon. Wea. Rev.*, **139**,

3554–3570, https://doi.org/10.1175/2011MWR3653.1.

Duan, H. X., Y. H. Li, T. J. Zhang, Z. X. Pu, C. L. Zhao, and Y. P. Liu, 2018: Evaluation of the forecast accuracy of near-surface temperature and wind in Northwest China based on the WRF model. *Journal of Meteorological Research*, **32**, 469–490, https://doi.org/10.1007/s13351-018-7115-9.

Frediani, M. E. B., J. P. Hacker, E. N. Anagnostou, and T. Hopson, 2016: Evaluation of PBL parameterizations for modeling surface wind speed during storms in the Northeast United States. *Wea. Forecasing*, **31**, 1511–1528, https://doi.org/10.1175/WAF-D-15-0139.1.

Frogner, I. L., T. Nipen, A. Singleton, J. B. Bremnes, and O. Vignes, 2016: Ensemble prediction with different spatial resolutions for the 2014 Sochi Winter Olympic Games: The effects of calibration and multimodel Approaches. *Wea. Forecasing*, **31**, 1833–1851, https://doi.org/10.1175/WAF-D-16-0048.1.

Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, https://doi.org/10.1175/MWR2904.1.

Hacker, J. P., and D. L. Rife, 2007: A practical approach to sequential estimation of systematic error on near-surface mesoscale grids. *Wea. Forecasting*, **22**, 1257–1273, https://doi.org/10.1175/2007WAF2006102.1.

Han, K., J. T. Choi, and C. Kim, 2016: Comparison of prediction performance using statistical postprocessing methods. *Asia-Pacific Journal of Atmospheric Sciences*, **52**, 495–507, https://doi.org/10.1007/s13143-016-0034-8.

Han, K., J. T. Choi, and C. Kim, 2018: Comparison of statistical post-processing methods for probabilistic wind speed forecasting. *Asia-Pacific Journal of Atmospheric Sciences*, **54**, 91–101, https://doi.org/10.1007/s13143-017-0062-z.

Hanna, S. R., and R. X. Yang, 2001: Evaluations of mesoscale models' simulations of near-surface winds, temperature gradients, and mixing depths. *J. Appl. Meteor.*, **40**, 1095–1104, https://doi.org/10.1175/1520-0450(2001)040<1095:EOMMSO>2.0.CO;2.

Hong, S. Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Huang, L. P., D. H. Chen, L. T. Deng, Z. F. Xu, F. Yu, Y. Jiang, and F. F. Zhou, 2017: Main technical improvements of GRAPES Meso V4.0 and verification. *Journal of Applied Meteorological Science*, **28**(1), 25–37, https://doi.org/10.11898/1001-7313.20170103. (in Chinese with English abstract)

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Isaac, G. A., and Coauthors, 2014: Science of nowcasting Olympic weather for Vancouver 2010 (SNOW-V10): A world weather research programme project. *Pure Appl. Geophys.*, **171**, 1–24, https://doi.org/10.1007/s00024-012-0579-0.

Jiménez, P. A., and J. Dudhia, 2012: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. *J. Appl. Meteorol. Climatol.*, **51**, 300–316, https://doi.org/10.1175/JAMC-D-11-084.1.

Joe, P., and Coauthors, 2010: Weather services, science advances, and the Vancouver 2010 Olympic and Paralympic winter games. *Bull. Amer. Meteor. Soc.*, **91**, 31–36, https://doi.org/10.1175/2009BAMS2998.1.

Kiktev, D., and Coauthors, 2017: FROST-2014: The Sochi winter Olympics international project. *Bull. Amer. Meteor. Soc.*, **98**, 1908–1929, https://doi.org/10.1175/BAMS-D-15-00307.1.

Lee, Y. H., G. Lee, S. Joo, and K. D. Ahn, 2018: Observational study of surface wind along a sloping surface over mountainous terrain during Winter. *Adv. Atmos. Sci.*, **35**(3), 276–284, https://doi.org/10.1007/s00376-017-7075-5.

Leroyer, S., S. Bélair, S. Z. Husain, and J. Mailhot, 2014: Subkilometer numerical weather prediction in an urban coastal area: A case study over the Vancouver metropolitan area. *J. Appl. Meteorol. Climatol.*, **53**, 1433–1453, https://doi.org/10.1175/JAMC-D-13-0202.1.

Liu, Y. J., S. G. Miao, L. Liu, and F. Hu, 2019: Effects of a modified sub-grid-scale terrain parameterization scheme on the simulation of low-layer wind over complex terrain. *Journal of Applied Meteorological Science*, **30**, 70–81, https://doi.org/10.11898/1001-7313.20190107. (in Chinese with English abstract)

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest *Bull. Amer. Meteor. Soc.*, **83**, 407–430, https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2.

Min, J. J, 2014: Evaluation on surface meteorological element forecast by Beijing Rapid Update Cycle System. *Journal of Applied Meteorological Science*, **25**, 265–273, https://doi.org/10.3969/j.issn.1001-7313.2014.03.002. (in Chinese with English abstract)

Raderschall, N., M. Lehning, and C. Schär, 2008: Fine-scale modeling of the boundary layer wind field over steep topography. *Water Resour. Res.*, **44**, W09425, https://doi.org/10.1029/2007WR006544.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, https://doi.org/10.1175/MWR2906.1.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, https://doi.org/10.1175/2010MWR3425.1.

Shafran, P. C., N. L. Seaman, and G. A. Gayno, 2000: Evaluation of numerical predictions of boundary layer structure during the Lake Michigan Ozone Study. *J. Appl. Meteorol. Climatol.*, **39**, 412–426, https://doi.org/10.1175/1520-0450(2000)039<0412:EONPOB>2.0.CO;2.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

Thompson, G., and T. Eidhammer, 2014: A study of Aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved Bulk Microphysics Scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Vasil'ev, E. V., and T. G. Dmitrieva, 2015: Forecasting extreme weather phenomena and processes during the test events and Sochi-2014 Olympic and Paralympic Games. *Russian Meteorology and Hydrology*, **40**(8), 513–522, https://doi.org/10.3103/S1068373915080038.

Vionnet, V., S. Belair, C. Girard, and A. Plante, 2015: Wintertime subkilometer numerical forecasts of near-surface variables in the Canadian Rocky Mountains. *Mon. Wea. Rev.*, **143**, 666–686, https://doi.org/10.1175/MWR-D-14-00128.1.

Whiteman, C. D., S. Zhong, W. J. Shaw, J. M. Hubbe, X. Bian, and J. Mittelstadt, 2001: Cold pools in the Columbia Basin. *Wea. Forecasting*, **16**, 432–447, https://doi.org/10.1175/1520-0434(2001)016<0432:CPITCB>2.0.CO;2.

Zhang, D. L., and W. Z. Zheng, 2004: Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. *J. Appl. Meteorol. Climatol.*, **43**, 157–169, https://doi.org/10.1175/1520-0450(2004)043<0157:DCOSWA>2.0.CO;2.

Zhang, H. L., Z. X. Pu, and X. B. Zhang, 2013: Examination of errors in near-surface temperature and wind from WRF numerical simulations in regions of complex terrain. *Wea. Forecasting*, **28**, 893–914, https://doi.org/10.1175/WAF-D-12-00109.1.

Zhang, Y. T., H. Tong, and J. Sun, 2020: Application of a bias correction method to meteorological forecast for the Pyeongchang Winter Olympic Games. *Journal of Applied Meteorological Science*, **31**(1), 27–41. (in Chinese with English abstract)

Zhong, S. Y., and J. Fast, 2003: An evaluation of the MM5, RAMS, and Meso-Eta models at subkilometer resolution using VTMX field campaign data in the Salt Lake Valley. *Mon. Wea. Rev.*, **131**, 1301–1322, https://doi.org/10.1175/1520-0493(2003)131<1301:AEOTMR>2.0.CO;2.