Supplementary information for

Evaluating the EPICC-Model for Regional Air Quality Simulation: A Comparative Study with CAMx and CMAQ

Mengjie Lou[1], Qizhong Wu[1], Wending Wang[2,3,4], Huansheng Chen[2,3,4], Kai Cao[2], Xiaohan Fan[5], Dingyue Liang[5], Fenfen Yu[5], Jiating Zhang[1], Wei Wang[6], Zifa Wang[2,3,4]

[1]Institute of Earth System Science, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China
[2]State Key Laboratory of Atmospheric Environment and Extreme Meteorology, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[3]State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[4]University of Chinese Academy of Sciences, Beijing, 100049, China
[5]3Clear Technology Co., Ltd., Beijing, 100029, China
[6]China National Environmental Monitoring Centre, Beijing, 100012, China

Correspondence: Qizhong Wu (wqizhong@bnu.edu.cn) and Wending Wang (wangwending@mail.iap.ac.cn)

Contents of this file:

S1 Evaluation Metrics

The evaluation metrics used in the model assessment include the correlation coefficient (R), the root mean square error (RMSE), the mean bias (MB), the normalized mean bias (NMB), the fraction of predictions within a factor of two of observations (FAC2), and the index of agreement (IOA). R represents the similarity in the temporal variation between predicted and observed values; RMSE quantifies the average magnitude of the difference between predictions and observations; MB indicates the average absolute difference between predictions and observations; NMB indicates the average relative bias of predictions compared to observations; FAC2 reflects the fraction of predictions that fall within a factor of two of the observations; IOA measures the degree of agreement between predictions and observations (Willmott, 1981). The calculation formulas are as follows:

$$R = \frac{\sum_{i=1}^{n}(M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \quad (S1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - O_i)^2} \quad (S2)$$

$$MB = \frac{1}{n}\sum_{i=1}^{n}(M_i - O_i) \quad (S3)$$

$$NMB = \frac{\sum_{i=1}^{n}(M_i - O_i)}{\sum_{i=1}^{n} O_i} \times 100\% \quad (S4)$$

$$FAC2 = fraction\ of\ predictions\ satisfying\ 0.5 \le \frac{M_i}{O_i} \le 2 \quad (S5)$$

$$IOA = 1 - \frac{\sum_{j=1}^{n}\left(P_j - Q_j\right)^2}{\sum_{j=1}^{n}\left(\left|P_j - \overline{Q_j}\right| + \left|Q_j - \overline{Q_j}\right|\right)^2} \quad (S6)$$

In the formula, M_i denotes the predicted value on day i, O_i denotes the observed value on day i, and n represents the number of days with valid simultaneous model predictions and observations during the evaluation period. $\bar{M}$ and $\bar{O}$ are the mean predicted and observed values over the evaluated days, respectively. For the calculation of FAC2, a value of 1 is assigned if the criterion is met; otherwise, 0 is assigned. The sum of these binary results is then divided by n to obtain the FAC2 score. Here, P_j represents the simulated value at site j, Q_j the observed value at site j, $\bar{Q}$ the mean observed value across all sites, and N the total number of sites. IOA ranges from 0 to 1, where values closer to 1 indicate higher spatial consistency between the model simulation and observations, while values closer to 0 indicate larger discrepancies.

S2 Evaluation of WRF Meteorological Field Simulations

S2.1 January

As shown in Fig. S1, the WRF model reproduced the diurnal variation of 2 m temperature in January 2021 reasonably well over most parts of China, with correlation coefficients (R) generally exceeding 0.7 except for parts of central China. In Northeast China, temperatures were generally underestimated (mean bias, MB, mostly -1 to -5 °C), with local RMSE values exceeding 6 °C. In North China, distinct north-south differences were observed: the northern part exhibited larger cold biases and stronger fluctuations (RMSE > 3 °C), whereas the southern part showed warm biases with smaller fluctuations (RMSE around 3-4 °C). The model performance was generally good in East and South China, except for cold biases in parts of Fujian and Zhejiang (MB = -1 to -3 °C, RMSE = 4-5 °C). In Northwest China, warm biases were more pronounced, with MB of 1-3 °C at most locations and exceeding 3 °C at some Xinjiang stations, while RMSE values were mostly above 5℃. In Southwest China and the southern part of Central China, the model showed larger deviations, with the former generally underestimated (MB < -3 °C, RMSE > 5 °C) and the latter, such as in Hunan, exhibiting low correlations (R = 0.3-0.4) and RMSE exceeding 7 °C.

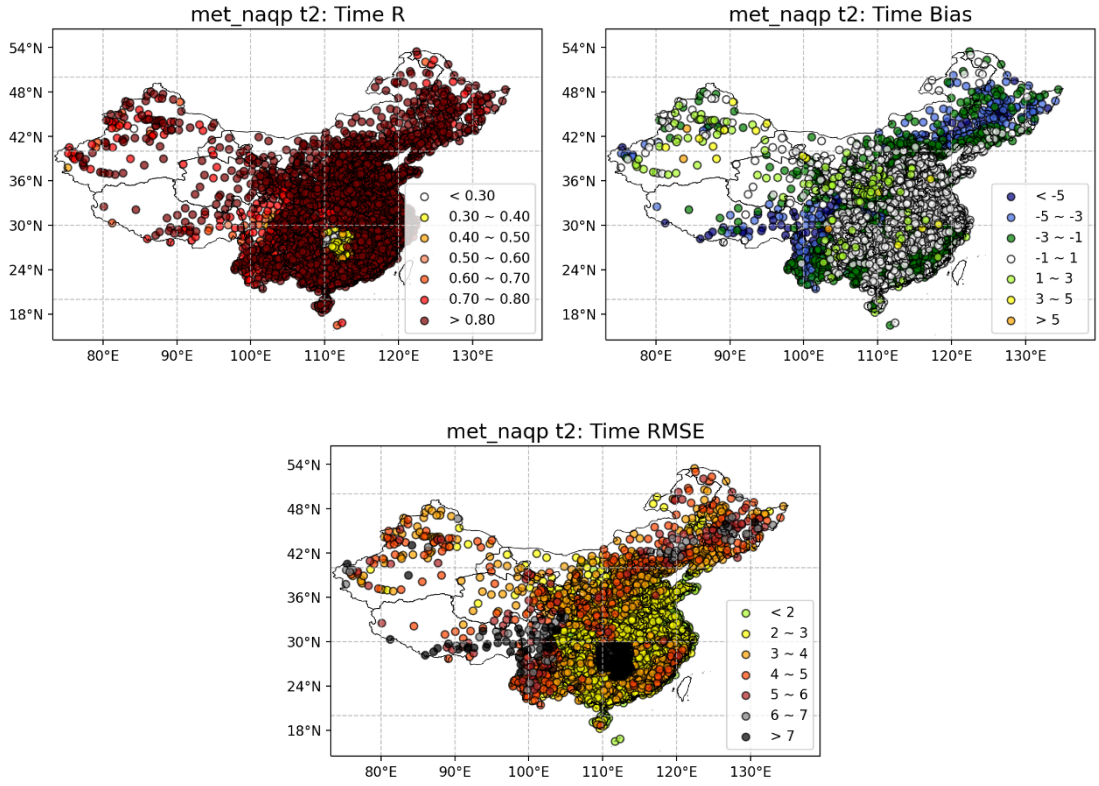


Fig. S1 Spatial distribution of hourly evaluation metrics for 2 m air temperature simulated by the WRF model across China from 1 to 31 January 2021.

As shown in Fig. S2, the WRF model well reproduces the diurnal variations of 2 m relative humidity across most regions of China, with R generally exceeding 0.7. Overall, substantial overestimations are observed in southern North China (e.g., Shanxi) and Southwest China, with MB exceeding 15% and RMSE mostly above 23%. Additionally,

moderate overestimations with considerable variability are present at some stations in Central China, South China, and parts of Northwest China.

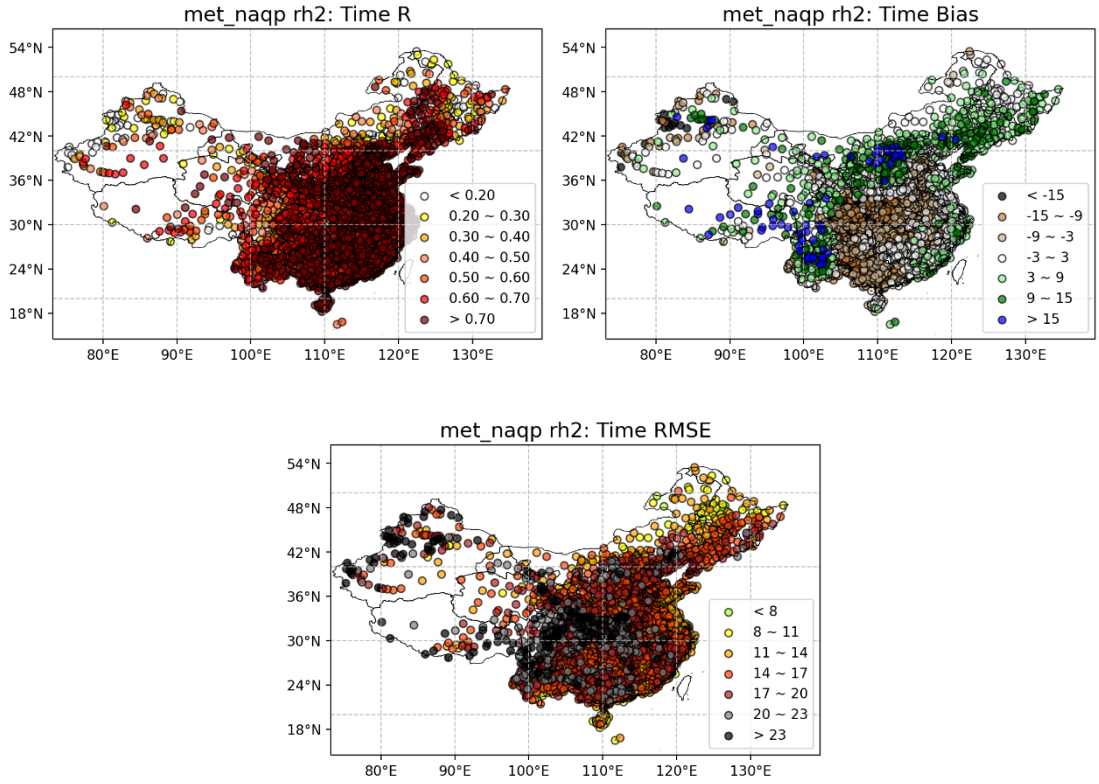
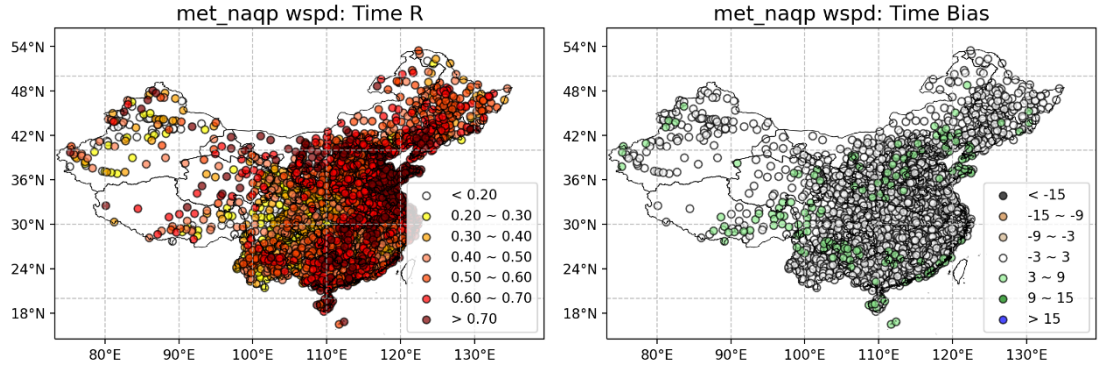


Fig. S2 Spatial distribution of hourly evaluation metrics for 2 m relative humidity simulated by the WRF model across China from 1 to 31 January 2021.

As shown in Fig. S3, the WRF model reproduces the diurnal variations of 10 m wind speed well across most regions of China, except for the Sichuan Basin and northwestern Xinjiang. The deviations are mainly characterized by overestimation, with RMSE generally below 8 m/s, indicating relatively small fluctuations and good overall consistency in wind speed simulations nationwide.



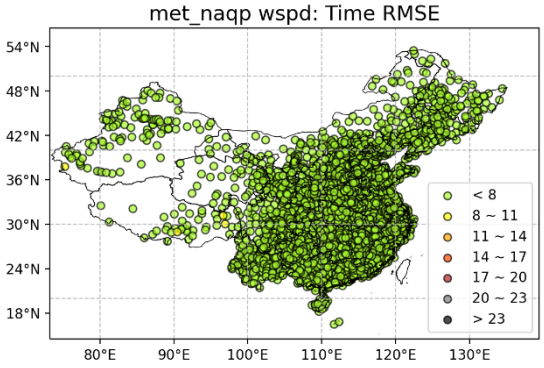
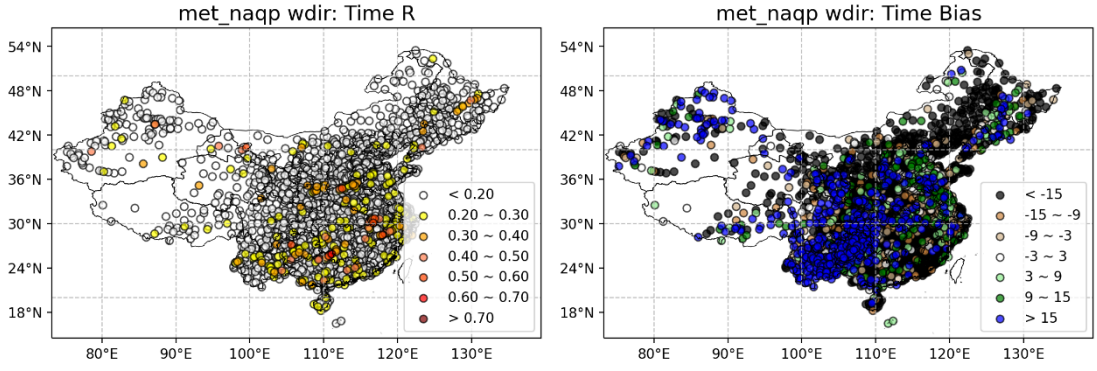


Fig. S3 Spatial distribution of hourly evaluation metrics for 10 m wind speed simulated by the WRF model across China from 1 to 31 January 2021.

As shown in Fig. S4, the WRF model exhibits overall poor performance in simulating 10 m wind direction, with only a few stations showing satisfactory results.



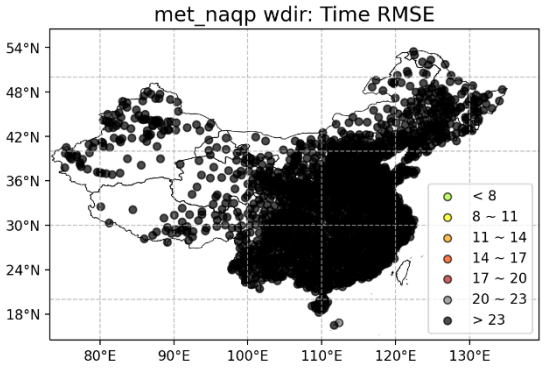


Fig. S4 Spatial distribution of hourly evaluation metrics for 10 m wind direction simulated by the WRF model across China from 1 to 31 January 2021.

As shown in Fig. S5, the WRF model reproduces the precipitation processes across most regions of China in January 2021 reasonably well ($R > 0.4$), although performance was relatively poor in Southwest China.

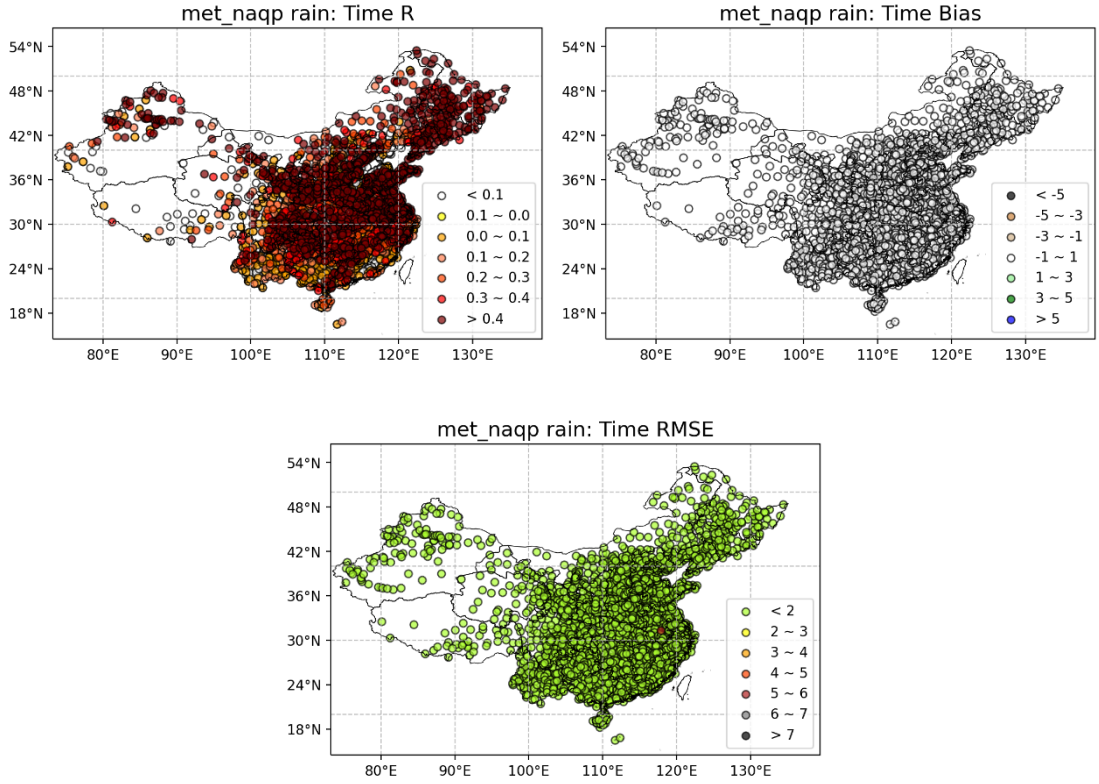
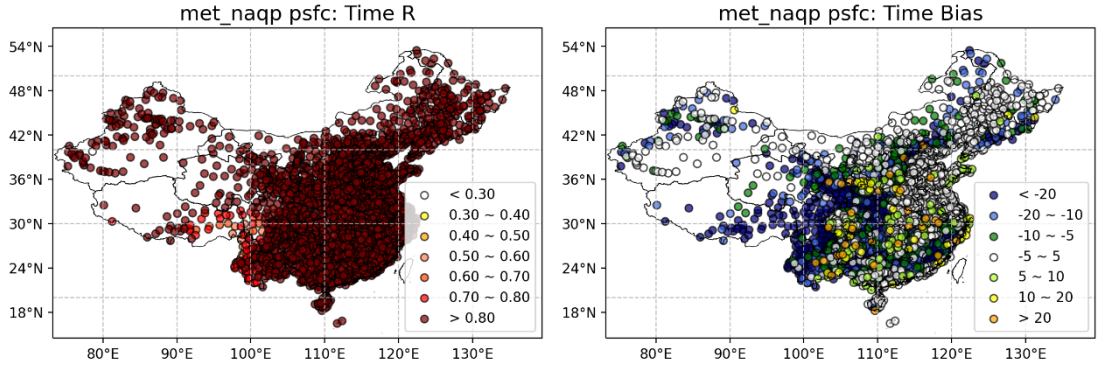


Fig. S5 Spatial distribution of hourly evaluation metrics for precipitation simulated by the WRF model across China from 1 to 31 January 2021.

As shown in Fig. S6, the WRF model reasonably reproduces the diurnal variations of surface pressure across mainland China ($R > 0.7$). Overall, simulations tended to underestimate pressure with larger fluctuations in parts of Southwest and Central China, with mean biases below 20 hPa, whereas slight overestimations were observed in regions such as Guizhou and Chongqing.



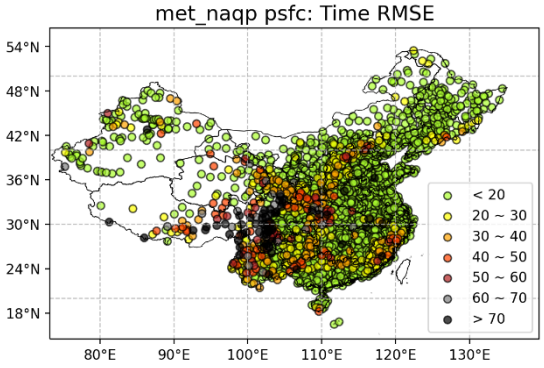


Fig. S6 Spatial distribution of hourly evaluation metrics for surface pressure simulated by the WRF model across China from 1 to 31 January 2021.

S2.2 April

As shown in Fig. S7, the WRF model generally captures the diurnal variation of 2 m temperature across most regions of China in April 2021 ($R > 0.7$). Overall, simulations were biased low in Southwest and Northwest China, with MB below -1 °C and relatively large fluctuations, and RMSE exceeding 6 °C in some areas; in South China, temperatures were overestimated ($1 < MB < 3$ °C) with smaller variability (RMSE between 3 and 4 °C). Only a few stations in other regions exhibited substantial temperature deviations.

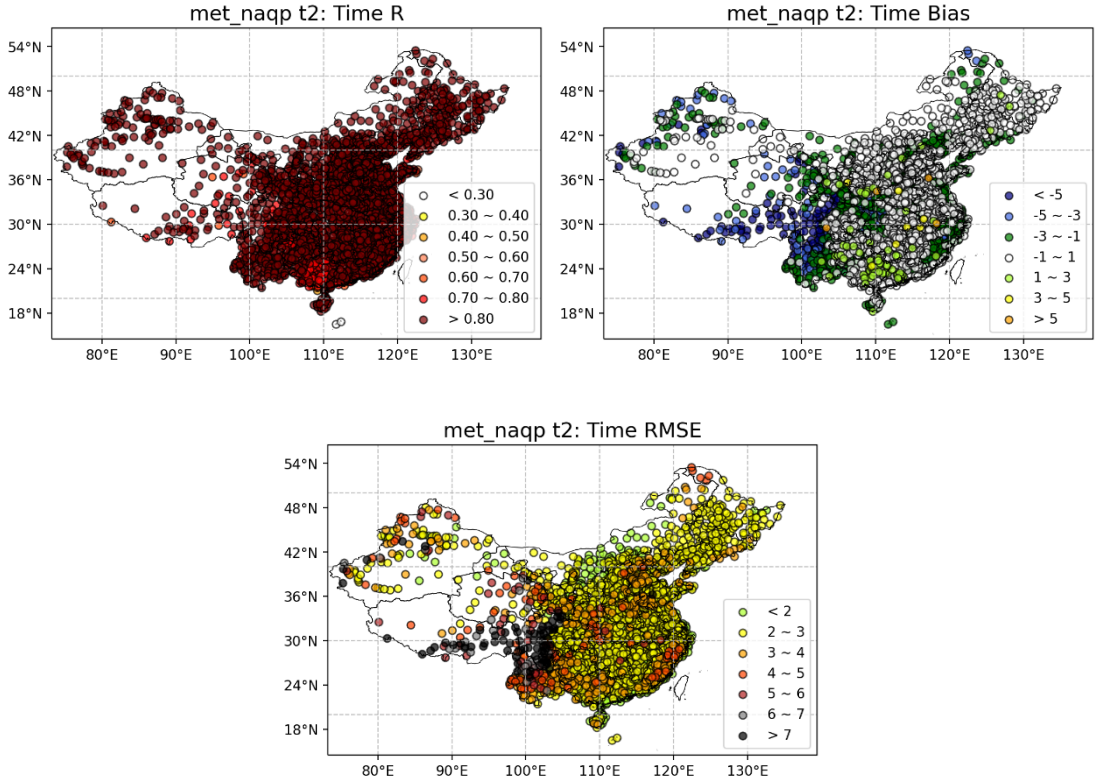


Fig. S7 Spatial distribution of hourly evaluation metrics for 2 m temperature simulated by the WRF model across China from 1 to 30 April 2021.

As shown in Fig. S8, the WRF model reasonably reproduces the diurnal variation

of 2 m relative humidity across most regions of China in April 2021 (R > 0.7). Overall, performance was poorer in Southwest, Northwest, and Northeast China. In the Southwest, 2 m humidity was overestimated with large variability (MB > 15%, RMSE > 23%). In the Northwest, results were polarized due to complex terrain, with underestimation in western Gansu and overestimation in northern Xinjiang. In the Northeast, humidity was generally underestimated, while coastal provinces exhibited overestimation.

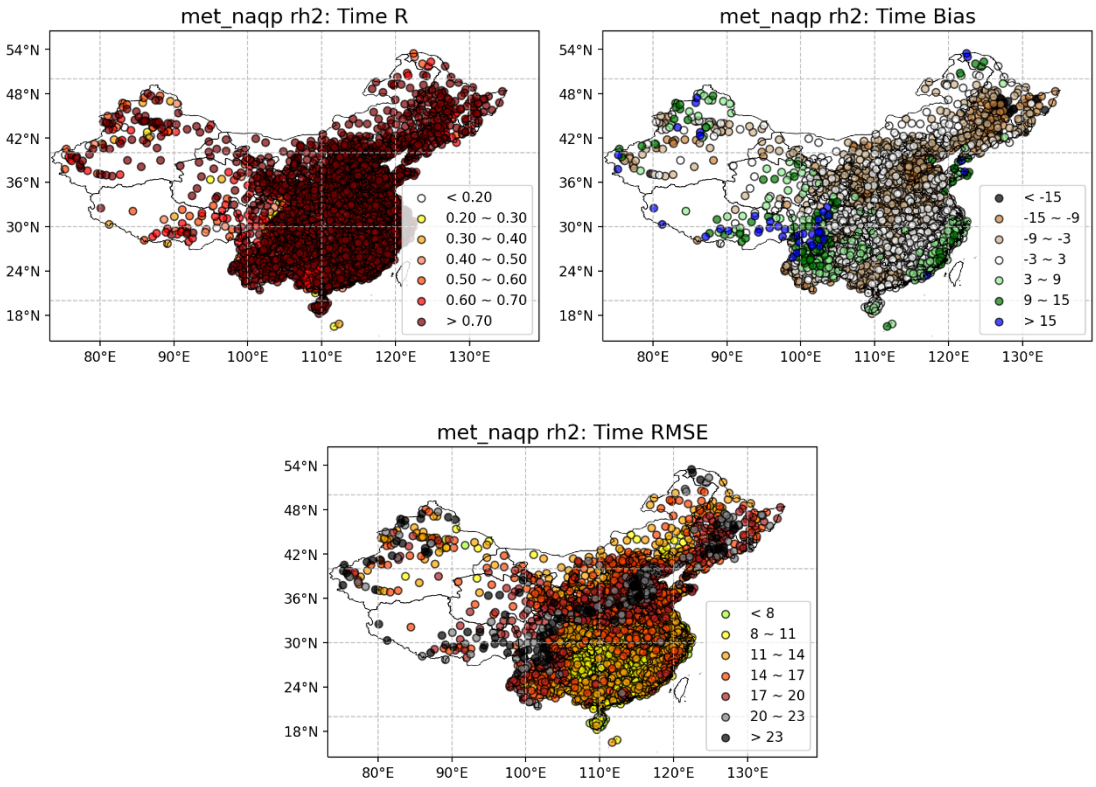
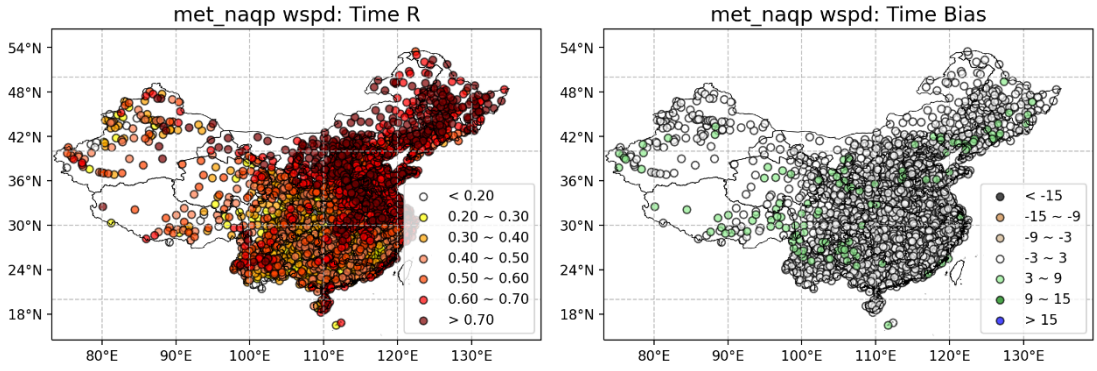
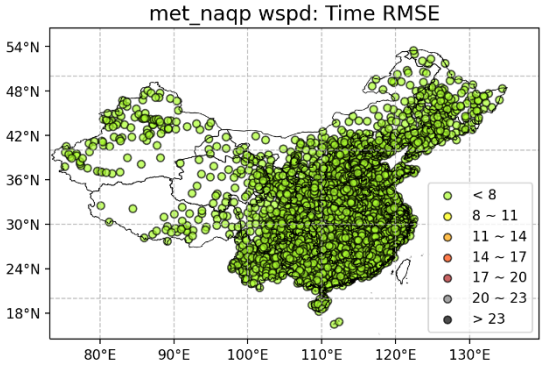


Fig. S8 Spatial distribution of hourly evaluation metrics for 2 m relative humidity simulated by WRF across China from 1 to 30 April 2021.

As shown in Fig. S9, the WRF model reproduces the diurnal variations of 10 m wind speed across most regions of China in April 2021, with deviations primarily manifesting as overestimations. Overall, RMSE remains below 8 m/s, indicating relatively small variability and good consistency in the wind speed simulation nationwide.

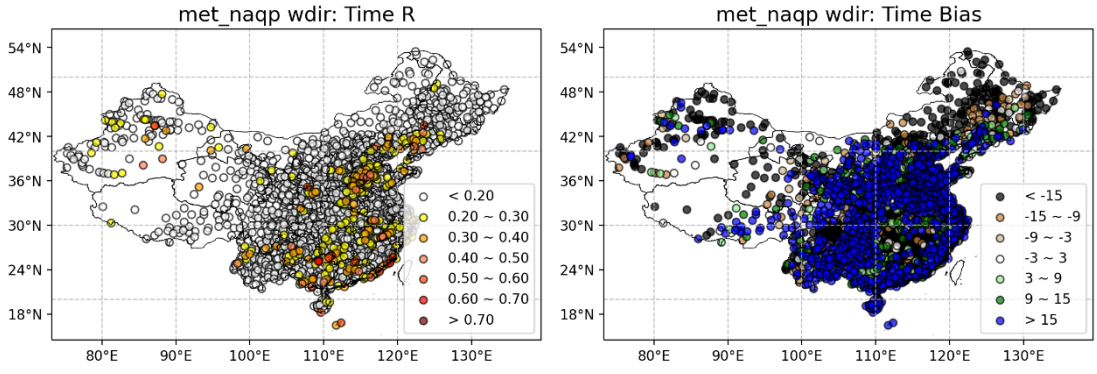




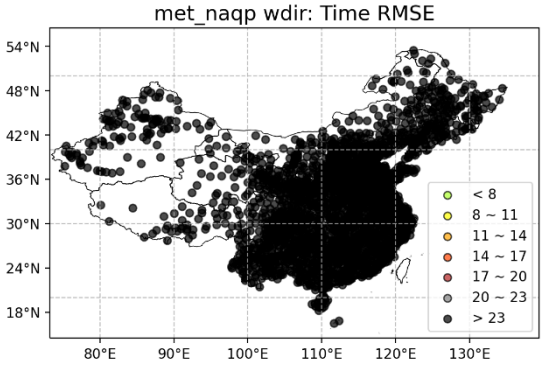
158

159 **Fig. S9** Spatial distribution of hourly evaluation metrics for 10 m wind speed simulated by WRF
160 across China from 1 to 30 April 2021.

161 As shown in Fig. S10, the WRF model generally reproduces 10 m wind direction
162 poorly across China, with R performing relatively well only in most areas of the
163 Beijing-Tianjin-Hebei region, Central China, South China, and at some other stations.



164



165

166 **Fig. S10** Spatial distribution of hourly evaluation metrics for 10 m wind direction simulated by
167 WRF across China from 1 to 30 April 2021.

168 As shown in Fig. S11, the WRF model reproduces the precipitation patterns across
169 most regions of China in April 2021 ($R > 0.4$), although performance is relatively poor
170 at some stations in Southwest and Northwest China.

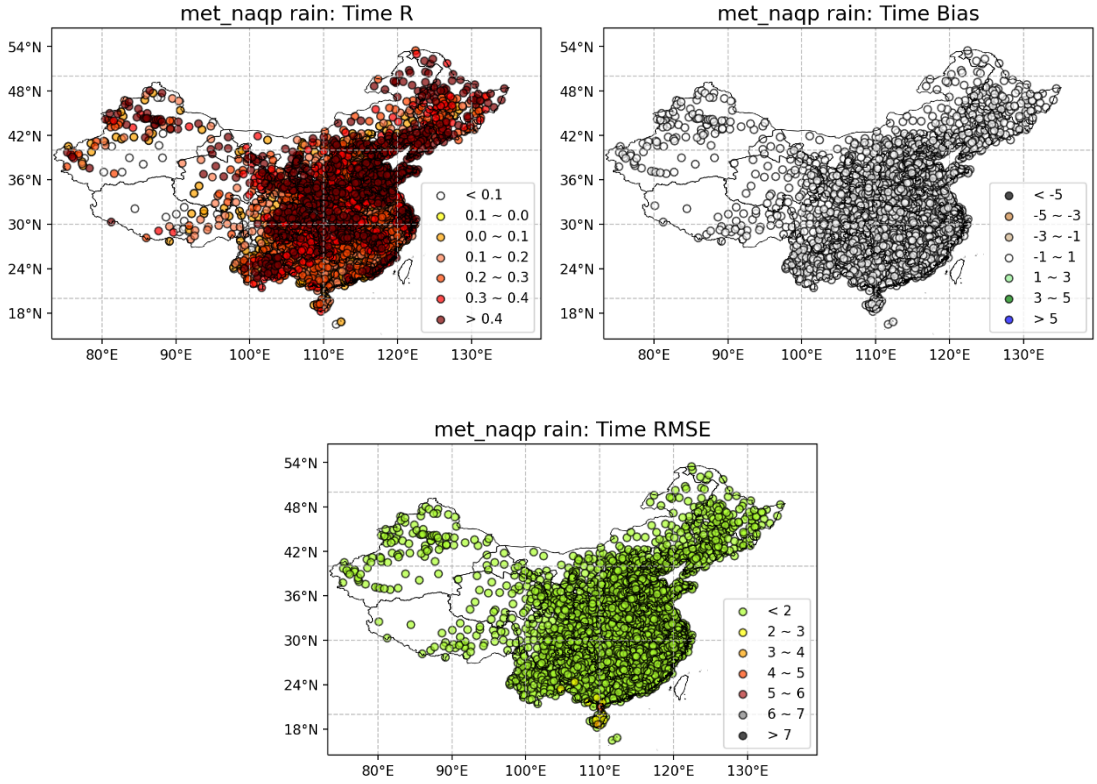
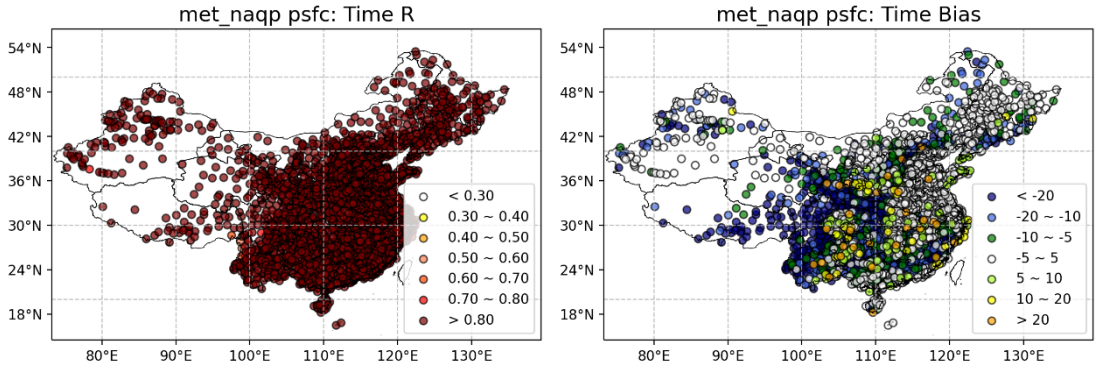


Fig. S11 Spatial distribution of hourly evaluation metrics for precipitation simulated by WRF across China from 1 to 30 April 2021.

As shown in Fig. S12, the WRF model reproduces the diurnal variations of surface pressure across China ($R > 0.7$). Overall, simulated values are lower and more variable in parts of Southwest and Central China (mean bias < 20 hPa), whereas over Guizhou and Chongqing, the simulation tends to be higher.



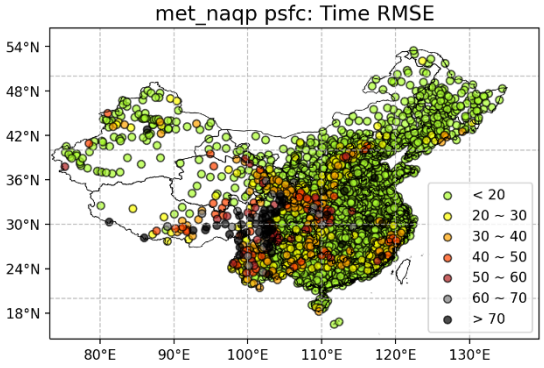


Fig. S12 Spatial distribution of hourly evaluation metrics for surface pressure simulated by WRF across China from 1 to 30 April 2021.

S2.3 July

As shown in Fig. S13, the WRF model reproduces the diurnal variations of 2 m temperature across most regions of China in July 2021 ($R > 0.7$). In Northeast China, simulated values deviate by 1-3 °C, with inland areas overestimated and coastal areas underestimated; overall, North China and Xinjiang are overestimated. In the southeastern coastal regions of East China, temperatures are generally underestimated with larger variability (MB: -3 to -1 °C, RMSE: 4-5 °C). The largest deviations occur in Southwest China, where temperatures are underestimated by more than 3 °C and RMSE exceeds 5 °C.

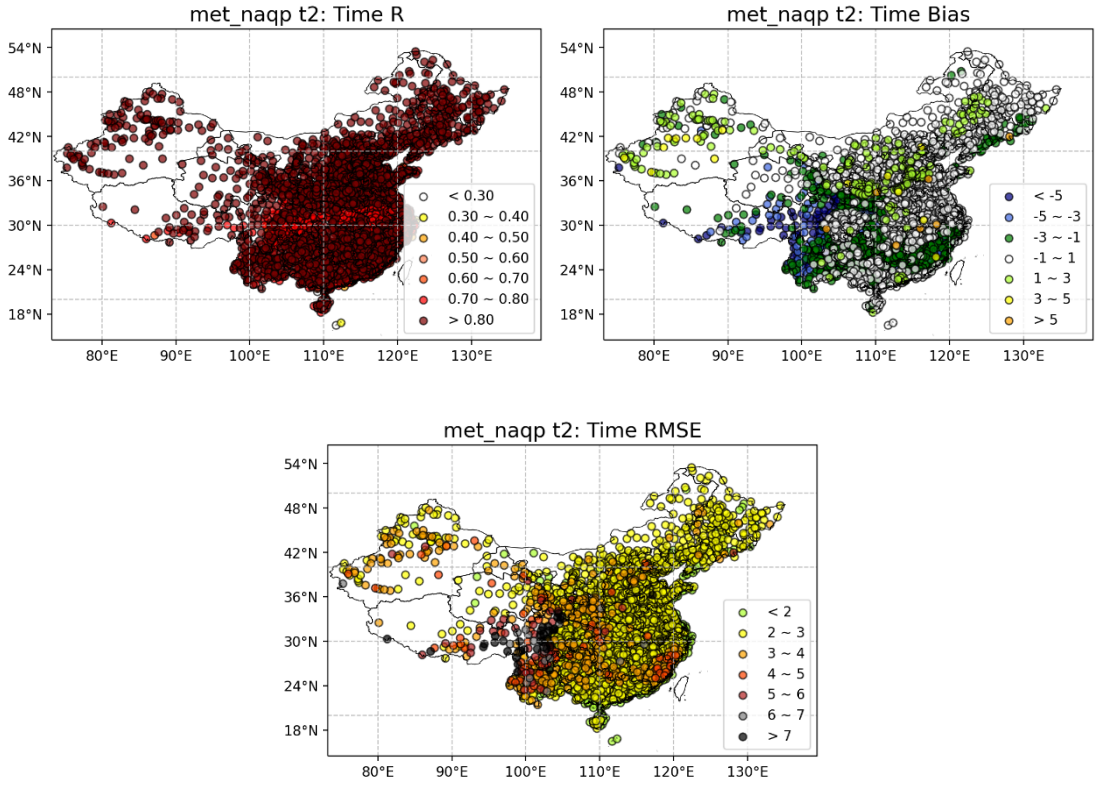


Fig. S13 Spatial distribution of hourly evaluation metrics for 2 m temperature simulated by WRF across China from 1 to 31 July 2021.

As shown in Fig. S14, the WRF model reproduces the diurnal variations of 2 m relative humidity across most regions of China in July 2021 ($R > 0.7$). Deviations mainly occur in Southwest, Northwest, and Northeast China: humidity is generally overestimated in the Southwest with large fluctuations (MB > 9%, RMSE > 11%); the Northwest is mostly underestimated, particularly in Xinjiang; and most of Northeast China is underestimated. In addition, coastal regions generally exhibit overestimation of humidity.

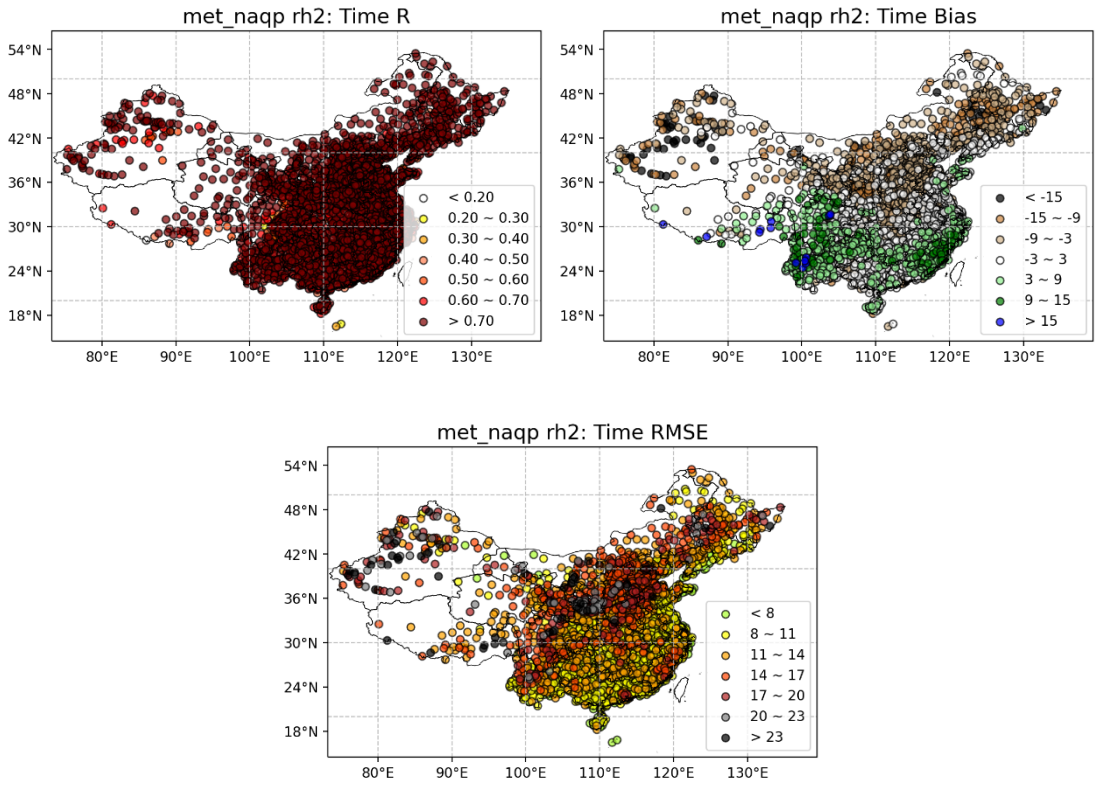
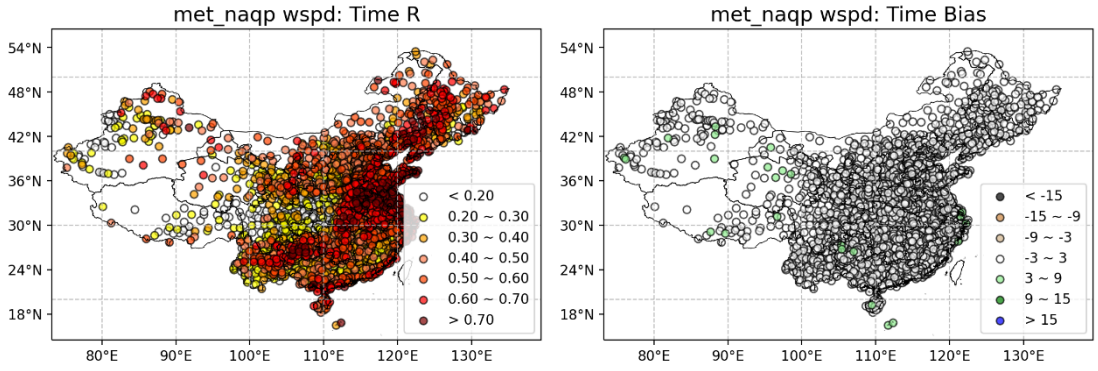


Fig. S14 Spatial distribution of hourly evaluation metrics for 2 m relative humidity simulated by WRF across China from 1 to 31 July 2021.

As shown in Fig. S15, the WRF model reproduces the diurnal variations of 10 m wind speed across most regions of China in July 2021, with larger deviations mainly manifesting as overestimation; poorer performance is observed in the Sichuan Basin, Tibet, and parts of Xinjiang. Overall, the RMSE of wind speed is generally below 8 m/s, indicating small fluctuations and good spatial consistency in the simulations.



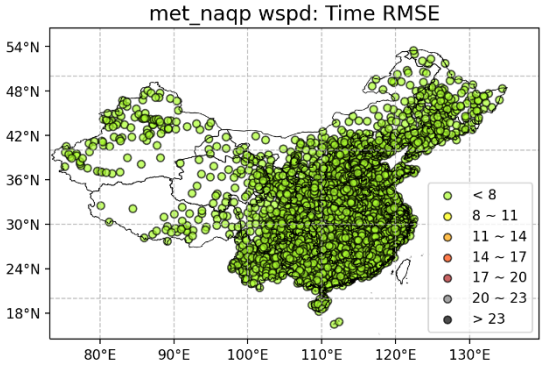
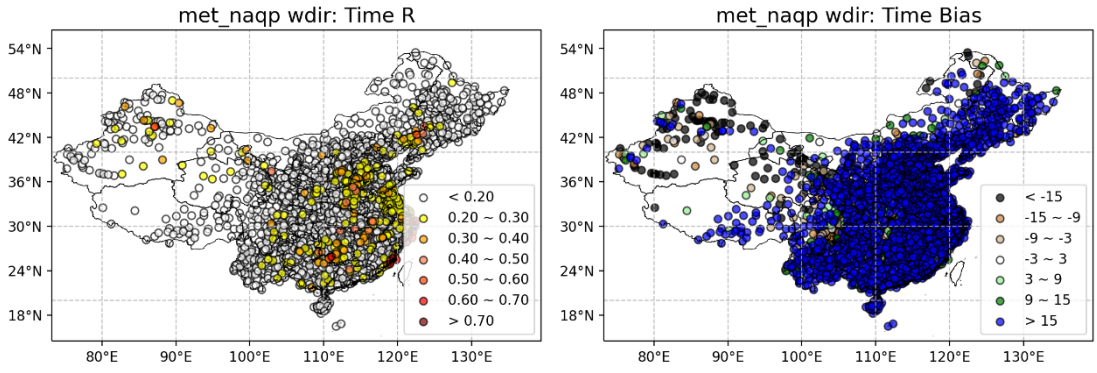


Fig. S15 Spatial distribution of hourly evaluation metrics for 10 m wind speed simulated by WRF across China from 1 to 31 July 2021.

As shown in Fig. S16, the WRF simulation of 10 m wind direction generally reproduces the observed patterns poorly, with low correlations across most regions of China, except for moderate performance in parts of the Beijing-Tianjin-Hebei region, East China, Central China, South China, and a few isolated stations.



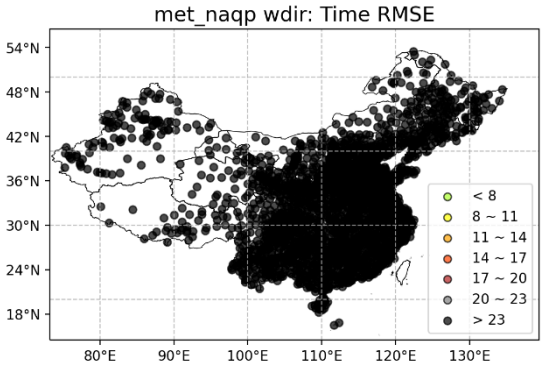


Fig. S16 Spatial distribution of hourly evaluation metrics for 10 m wind direction simulated by WRF across China from 1 to 31 July 2021.

As shown in Fig. S17, the WRF model reproduces the precipitation patterns in July 2021 relatively well mainly in the Beijing-Tianjin-Hebei region ($R > 0.4$), while correlations in other regions are generally poor, with only a few stations exhibiting R values above 0.4.

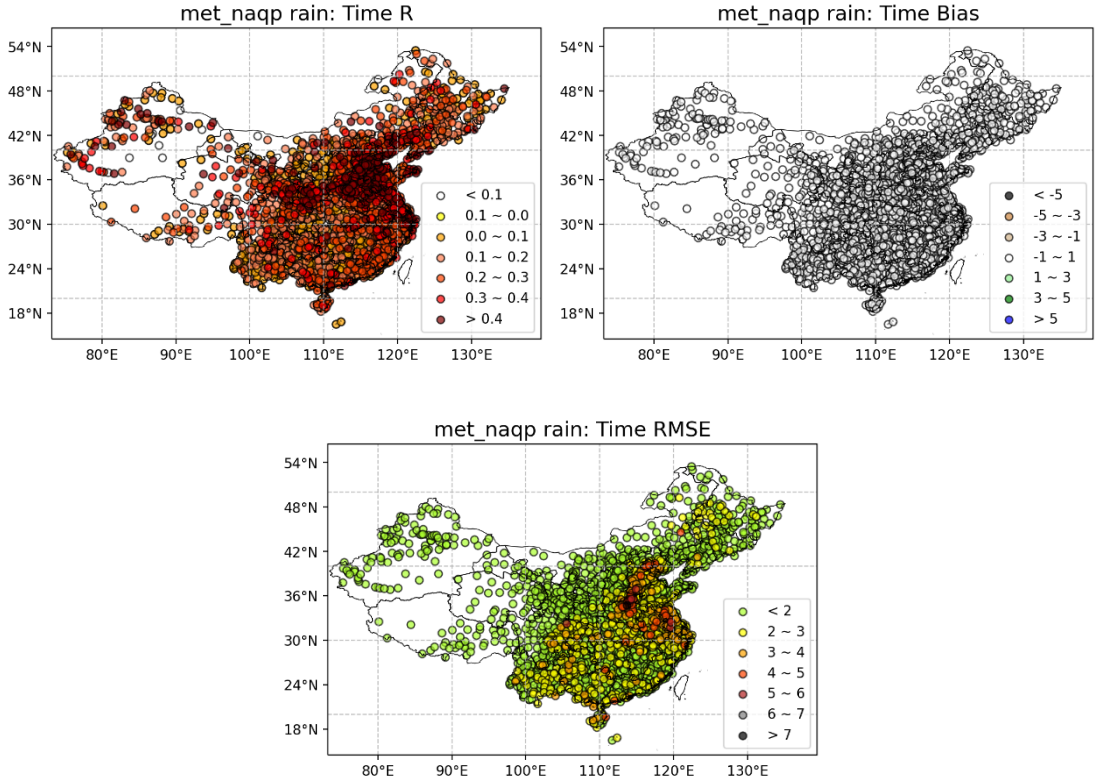
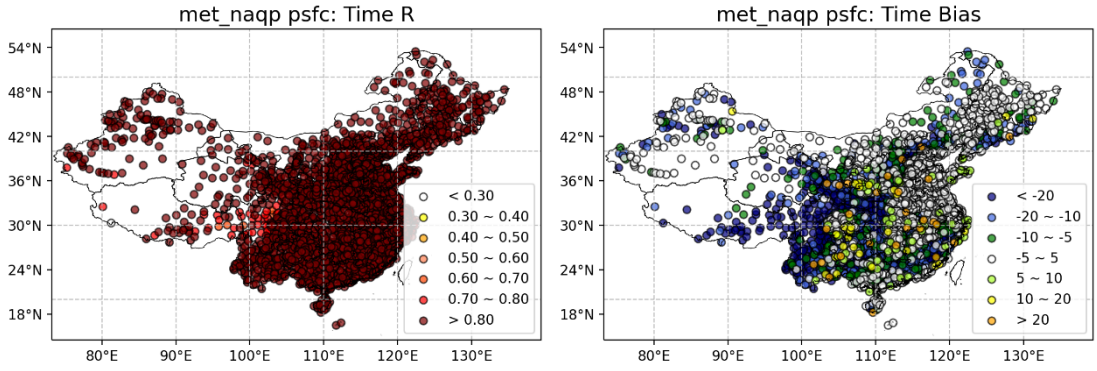


Fig. S17 Spatial distribution of hourly evaluation metrics for precipitation simulated by WRF across China from 1 to 31 July 2021.

As shown in Fig. S18, the WRF model reproduces the diurnal variations of surface pressure across China well ($R > 0.7$); however, simulated values are lower and more variable in parts of Southwest and Central China (mean bias < 20 hPa), whereas slight overestimations are observed in the Guizhou and Chongqing regions.



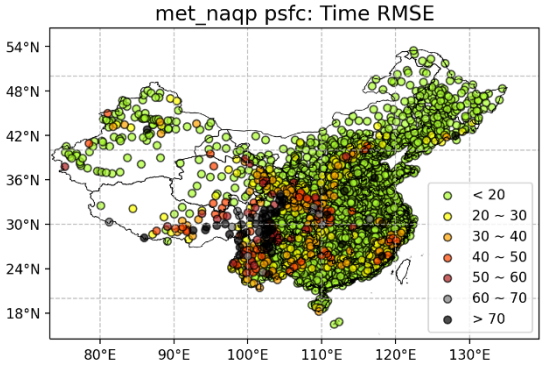
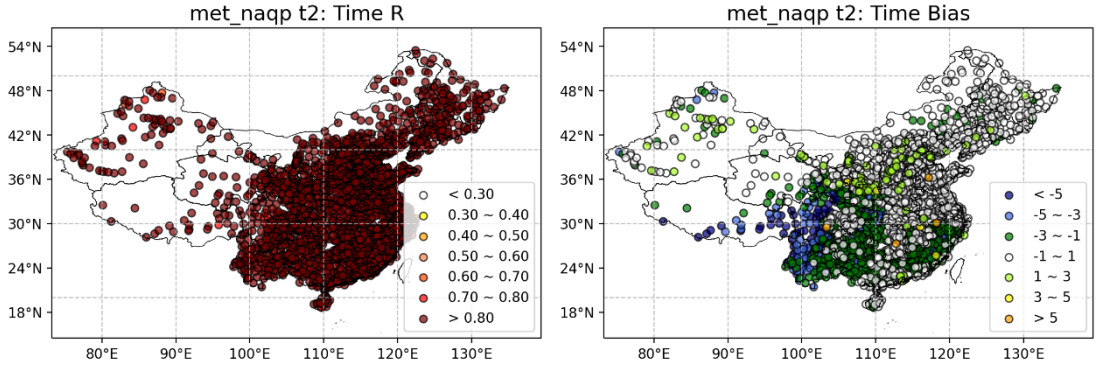


Fig. S18 Spatial distribution of hourly evaluation metrics for surface pressure simulated by WRF across China from 1 to 31 July 2021.

S2.4 October

As shown in Fig. S19, the current WRF simulation reproduces the diurnal variations of 2 m temperature in October 2021 well ($R > 0.7$); simulated values are slightly overestimated in southern North China (1-3 °C) and underestimated along the East China coast (1-3 °C); the Southwest region exhibits poorer performance, with temperatures generally underestimated by more than 3 °C and larger variability (RMSE > 7 °C).



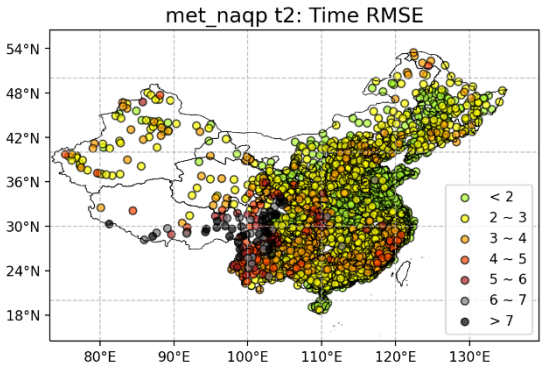


Fig. S19 Spatial distribution of hourly evaluation metrics for 2 m temperature simulated by WRF across China from 1 to 31 October 2021.

As shown in Fig. S20, the WRF model reproduces the diurnal variation of 2 m relative humidity across most regions of China in October 2021 ($R > 0.7$). Overall,

humidity is overestimated in Southwest China, with MB exceeding 15% and relatively large fluctuations (RMSE > 17%). In East and South China, the simulation slightly overestimates humidity with smaller variability, whereas most areas in Central, North, Northeast, and Northwest China show underestimation.

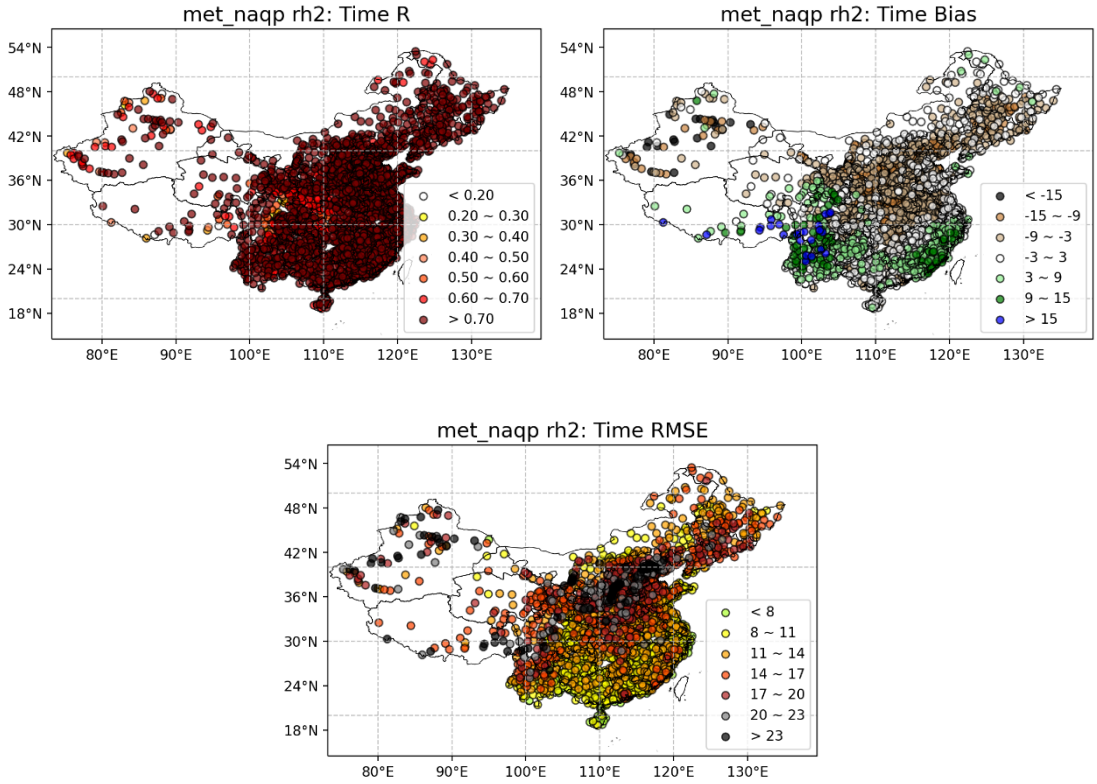
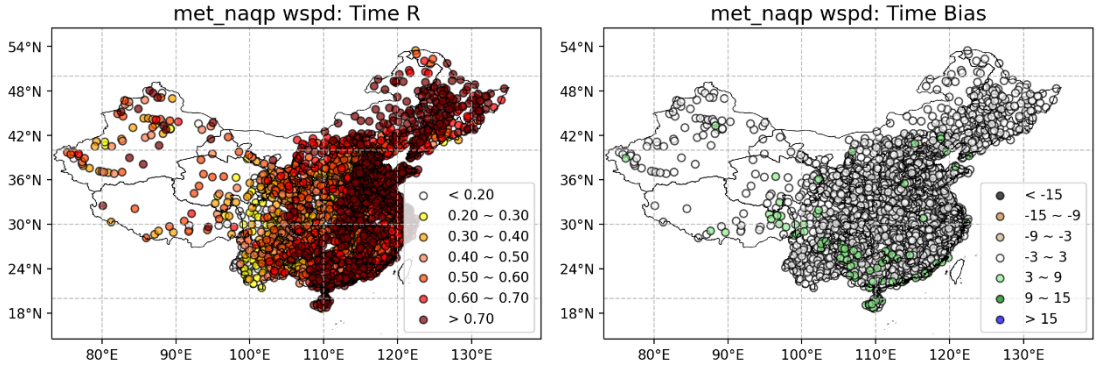


Fig. S20 Spatial distribution of hourly evaluation metrics for 2 m relative humidity simulated by WRF across China from 1 to 31 October 2021.

As shown in Fig. S21, the WRF model reproduces the diurnal variation of 10 m wind speed across most regions of China, except for certain areas in the Sichuan Basin, Xinjiang, and Yunnan. When deviations are relatively large, wind speed is generally overestimated. RMSE values are below 8 m/s nationwide, indicating small variability and good consistency of wind speed simulations across the country.



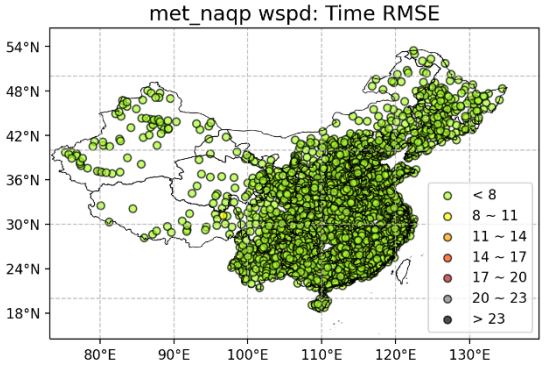


Fig. S21 Spatial distribution of hourly evaluation metrics for 10 m wind speed simulated by WRF across China from 1 to 31 October 2021.

As shown in Fig. S22, the WRF model generally performs poorly in simulating 10 m wind direction across China, although better agreement is observed in coastal provinces, most areas of the Beijing-Tianjin-Hebei region, and selected sites in Yunnan, Guizhou, and Xinjiang.

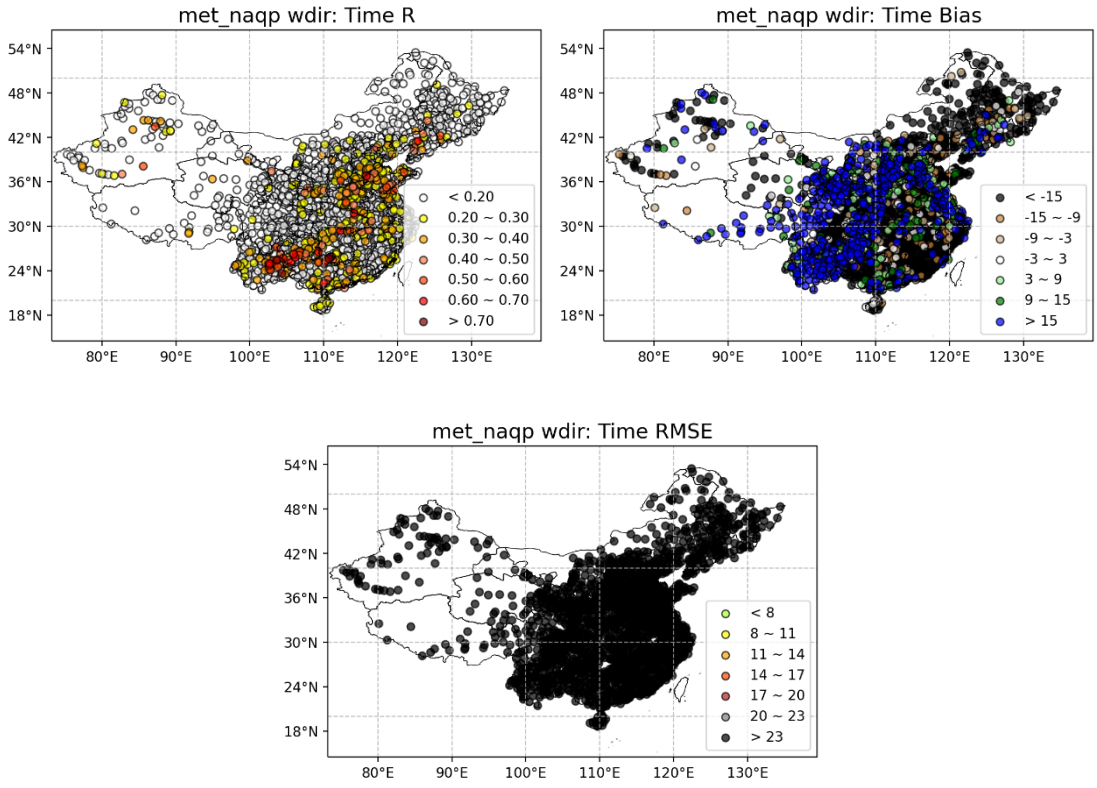


Fig. S22 Spatial distribution of hourly evaluation metrics for 10 m wind direction simulated by WRF across China from 1 to 31 October 2021.

As shown in Fig. S23, the WRF model reproduces the precipitation process well across most regions of China in October 2021 ($R > 0.4$), with poorer performance observed only in parts of Southwest, Northwest, and South China.

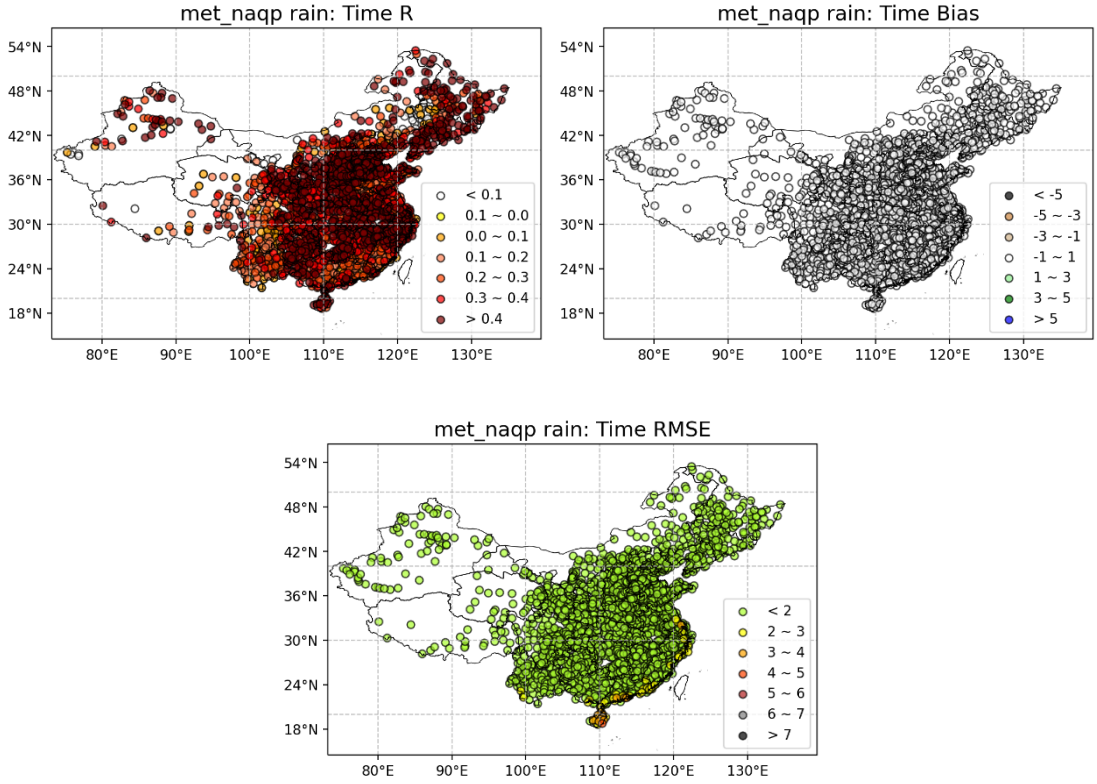
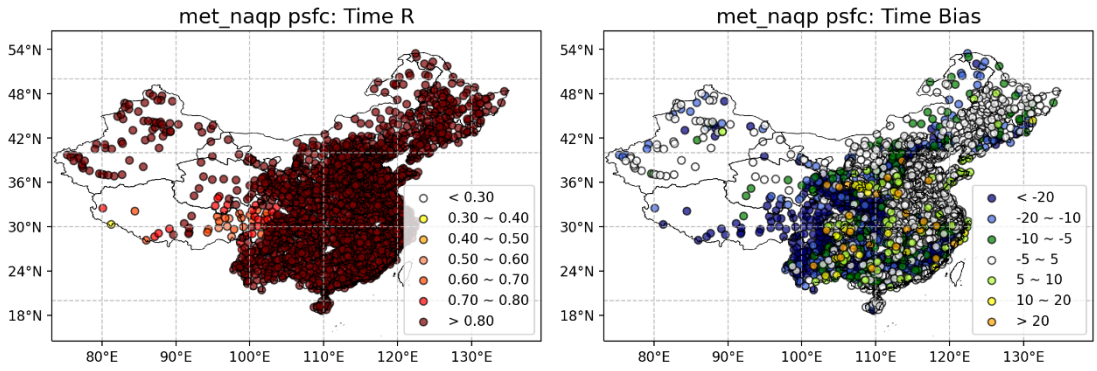


Fig. S23 Spatial distribution of hourly evaluation metrics for precipitation simulated by WRF across China from 1 to 31 October 2021.

As shown in Fig. S24, the WRF model reproduces the diurnal variation of surface pressure well across China in October 2021 ($R > 0.7$). Overall, simulated pressure is underestimated with relatively large fluctuations in parts of Southwest and Central China ($MB < 20$ hPa), whereas overestimation is observed in regions such as Guizhou and Chongqing.



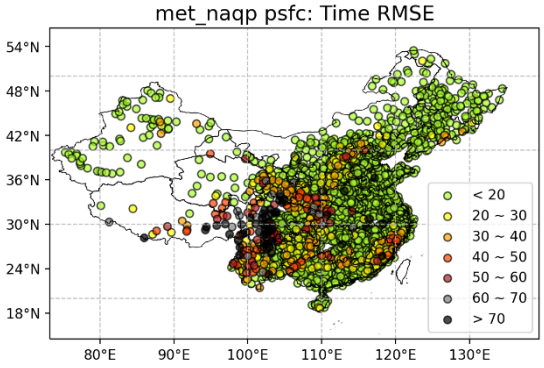


Fig. S24 Spatial distribution of hourly evaluation metrics for surface pressure simulated by WRF across China from 1 to 31 October 2021.

S3 Distribution of 1,644 National Control Stations

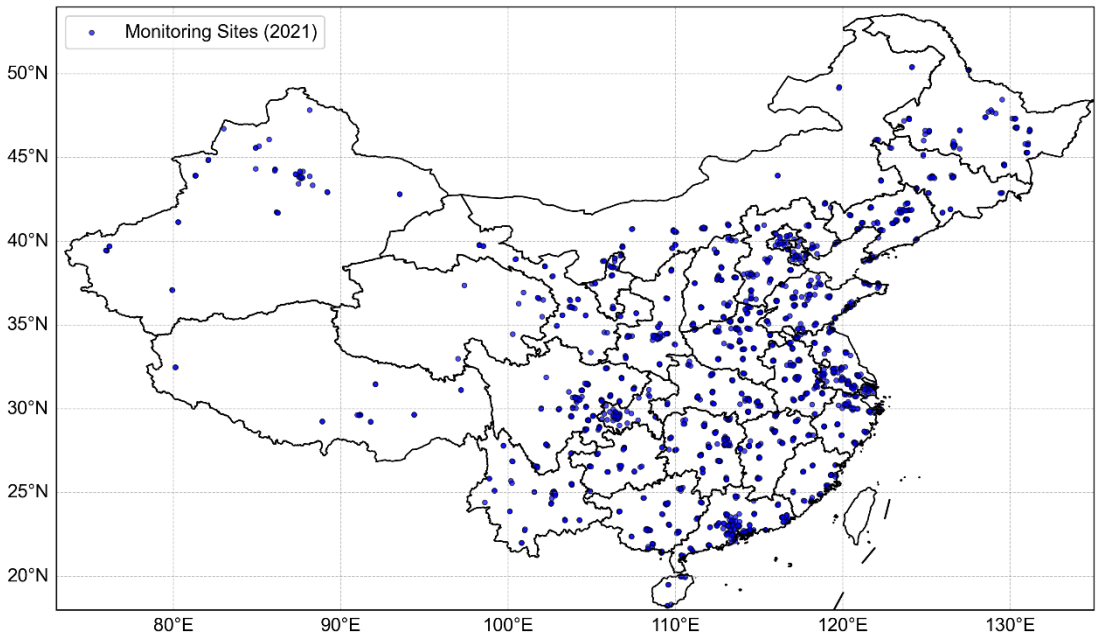


Fig. S25 Spatial distribution of the 1,644 national control monitoring stations across China in 2021, based on data from the China National Environmental Monitoring Center (CNEMC).

S4 Air Quality Index (AQI) and Individual Air Quality Index (IAQI)

S4.1 Definition

AQI (Air Quality Index): A standardized index used to communicate overall air quality to the public. It integrates multiple pollutant concentrations to reflect the potential health risks associated with air pollution.

IAQI (Individual Air Quality Index): Represents the contribution of a single pollutant to the overall AQI. It is calculated based on the observed concentration of the pollutant and the breakpoint concentrations defined in the air quality standards.

S4.2 Calculation

IAQI for pollutant i:

$$IAQI_i = \frac{IAQI_{Hi} - IAQI_{Lo}}{C_{Hi} - C_{Lo}} \times (C_i - C_{Lo}) + IAQI_{Lo} \quad (S7)$$

In the formula, C_i is the observed concentration of pollutant i. C_{Hi} and C_{Lo} are the upper and lower breakpoint concentrations corresponding to C_i. $IAQI_{Hi}$ and $IAQI_{Lo}$ are the IAQI values corresponding to C_{Hi} and C_{Lo}.

AQI:

$$AQI = \max(IAQI_1, IAQI_2, \ldots, IAQI_n) \quad (S8)$$

That is, the AQI is equal to the maximum IAQI among all monitored pollutants.

S5 Hit Rate, False Alarm Rate, and the Distance from the Random Operating Characteristic (DROC)

In the evaluation of graded air quality forecasts, we adopt Hit Rate, False Alarm Rate, and the distance from the random operating characteristic (DROC) as verification metrics. Their definitions and formulas are as follows:

S5.1 Hit Rate

Definition: The proportion of observed pollution categories that are correctly forecasted by the model.

Formula:

$$Hit\ Rate = \frac{Hits}{Hits + Misses} \quad (S9)$$

Where $Hits$ refers to the number of cases where the model prediction matches the observation, and $Misses$ denotes the cases that were observed but not predicted.

S5.2 False Alarm Rate

Definition: The proportion of pollution categories that were forecasted by the model but did not occur in the observations.

Formula:

$$False\ Alarm\ Rate = \frac{FalseAlarms}{Hits + FalseAlarms} \quad (S10)$$

Where $FalseAlarms$ refers to the number of cases that were forecasted but not observed.

S5.3 The Distance from the Random Operating Characteristic (DROC)

Definition: DROC measures the distance of forecast-observation points from the diagonal in the scatter plot of observed versus predicted pollution events. A larger distance indicates higher forecasting skill. It combines hit rate and false alarm rate to evaluate overall performance.

Formula:

$$DROC = \frac{Hit\ Rate - False\ Alarm\ Rate}{\sqrt{2}} \quad (S11)$$

Reference

Willmott, C.J., 1981. ON THE VALIDATION OF MODELS. Phys. Geogr. 2, 184-194. https://doi.org/10.1080/02723646.1981.10642213