Article

Comparative Study of O$_3$ Forecast Performance Using Multiple Models in Beijing–Tianjin–Hebei and Surrounding Regions

Lili Zhu [1,2,3], **Wei Wang** [3], **Huihui Zheng** [3], **Xiaoyan Wang** [3,*], **Yonghai Huang** [3] **and Bing Liu** [3]

1   Aerospace Information Research Institute, Chinese Academy of Sciences, No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China; zhull@cnemc.cn
2   University of Chinese Academy of Sciences, No.1 Yanqihu East Rd, Huairou District, Beijing 101408, China
3   China National Environmental Monitoring Centre, No.8-2 Anwai Dayangfang, Chaoyang District, Beijing 100012, China; wangwei2015@cnemc.cn (W.W.); zhenghh@3clear.com (H.Z.); huangyh@cnemc.cn (Y.H.); liubing@cnemc.cn (B.L.)
*   Correspondence: wangxy@cnemc.cn

Abstract: In order to systematically understand the operational forecast performance of current numerical, statistical, and ensemble models for O$_3$ in Beijing–Tianjin–Hebei and surrounding regions, a comprehensive evaluation was conducted for the 30 model sets regarding O$_3$ forecasts in June–July 2023. The evaluation parameters for O$_3$ forecasts in the next 1–3 days were found to be more reasonable and practically meaningful than those for longer lead times. When the daily maximum 8 h average concentration of O$_3$ was below 100 μg/m^3 or above 200 μg/m^3, a significant decrease in the percentage of accurate models was observed. As the number of polluted days in cities increased, the overall percentage of accurate models exhibited a decreasing trend. Statistical models demonstrated better overall performance in terms of metrics such as root mean square error, standard mean bias, and correlation coefficient compared to numerical and ensemble models. Numerical models exhibited significant performance variations, with the best-performing numerical model reaching a level comparable to that of statistical models. This finding suggests that the continuous tuning of operational numerical models has a more pronounced practical effect. Although the best statistical model had higher accuracy than numerical and ensemble models, it showed a significant overestimation when O$_3$ concentrations were low and a significant underestimation when concentrations were high. In particular, the underestimation rate for heavy polluted days was significantly higher than that for numerical and ensemble models. This implies that statistical models may be more prone to missing high-concentration O$_3$ pollution events.

Keywords: O$_3$ forecast; forecast performance assessment; multi-model comparison

1. Introduction

Near-surface ozone has become one of the primary atmospheric pollutants. High ozone concentrations can be harmful to human beings and the environment, potentially causing respiratory issues and damaging plant growth [1,2]. Fluctuations in O$_3$ concentrations are primarily associated with temperature, relative humidity, wind speed, and wind direction. Different meteorological conditions have varying impacts on O$_3$ concentrations. In general, higher surface temperatures, lower relative humidity, and stronger solar radiation favor the photochemical reactions of O$_3$, leading to an increase in O$_3$ concentrations [3–5]. Since the issuance of the "Action Plan for Air Pollution Prevention and Control" in 2013, China has achieved significant success in controlling PM$_{2.5}$ pollution. However, with the adjustment of industrial structure and changes in meteorological conditions, the issue of O$_3$ pollution has become increasingly prominent [6–10], particularly in Beijing–Tianjin–Hebei and surrounding regions [11]. Urban air quality forecast provides crucial information and acts as a basis for decision making, allowing local governments to initiate emergency responses to heavy polluted days, formulate pollution control measures,

and identify technical means for air pollution control and precise regulation [12]. Thus, in the face of a complex pollution situation, the accurate forecast of O_3 is the key for precise control decision making.

Currently, models used for O_3 forecasts mainly include numerical models, statistical models, and ensemble models. O_3 numerical models [13], based on atmospheric dynamics and physical–chemical processes, predict future O_3 concentrations by simulating O_3 production and consumption processes in the real atmosphere. Their advantage lies in providing a physical interpretation of the simulation process, aiding in understanding the fundamental mechanisms of the atmospheric system. However, they require significant computational resources and detailed input data, including meteorological fields, topographic conditions, and emission inventories. O_3 statistical models [14] employ traditional statistical methods (like regression analysis and time series analysis) or machine learning algorithms (such as non-linear or deep learning), which have been widely used in energy and medicine fields [15–17]. They are used to analyze historical O_3 monitoring data and meteorological conditions, as well as to establish mathematical relationships for predicting future O_3 concentrations, since meteorological factors such as temperature, relative humidity, and solar radiation exhibit a strong correlation with O_3 concentration [18]. The advantage of statistical models is the fast execution speed, but they lack detailed a physical–chemical interpretation of results and are considered black box models [19,20]. O_3 ensemble models [21] utilize statistical methods to analyze and compute results based on numerical model forecasts combined with observed data, ultimately providing adjusted O_3 forecast results. The ensemble models can enhance the system's resilience to noise and outliers, improving the overall model robustness [22,23]. However, the ensemble models depend on the outputs of individual sub-models. This may lead to difficulties in providing a clear model interpretation, which is a similar shortcoming for statistical models.

Yang et al. [24], in a review of over 200 O_3 numerical model applications in China since 2010, found that O_3 numerical models are more widely adopted and perform well in the Yangtze River Delta and Pearl River Delta regions, but their applications are relatively limited in Beijing–Tianjin–Hebei and surrounding areas. In recent years, numerous researchers have conducted extensive research on the establishment and performance evaluation of O_3 statistical models [25–27] in Beijing–Tianjin–Hebei and surrounding regions. However, the forecast lead time generally does not exceed 3 days, and the evaluation indicators used vary from model to model. The applications of O_3 ensemble models in China are relatively scarce. There is limited research on the evaluation of O_3 forecast model performance in the current operational processes in China, and the referential quality of various O_3 models in operational processes has not been investigated. This study conducted a comprehensive evaluation of O_3 forecast results in Beijing–Tianjin–Hebei and surrounding regions for June and July 2023 based on 30 models of numerical, statistical, and ensemble types. Multiple evaluation metrics were employed to provide a comprehensive assessment of their forecast performance, aiming to enhance the supportive role of model forecast in environmental management.

2. Materials and Methods

2.1. Models and Setting

The forecast results evaluated in this study consisted of a total of 30 model sets, including numerical, statistical, and ensemble models. The forecast period for each set spanned from 1 June to 31 July 2023, covering 57 cities at the prefecture level and above in six provinces (or municipalities), namely Beijing, Tianjin, Hebei, Shanxi, Shandong, and Henan. The forecast content included hourly time series of O_3 concentrations for the next 7 days in each city.

Numerical Models: The numerical model forecast results evaluated in this study consisted of a total of 20 sets. Among them, six sets utilized the Nested Air Quality Prediction Modeling System (NAQPMS) developed by the Institute of Atmospheric Physics, the Chinese Academy of Sciences. Another six sets used the Community Multiscale Air Quality

(CMAQ) Model developed by the United States Environmental Protection Agency. Additionally, six sets employed the Comprehensive Air Quality Model with Extension (CAMx). One set was based on the Weather Research and Forecast—Chemistry (WRF-chem) Model developed by the National Oceanic and Atmospheric Administration (NOAA), and the last one set utilized the RuiTu Map (RMAPS)—Chemistry subsystem developed by the Beijing Urban Meteorological Research Institute based on WRF-chem. Meteorological driving data were sourced from National Centers for Environmental Prediction—Global Forecast System (NCEP-GFS) or China Meteorological Administration—Global Assimilation Forecast System (CMA-GFS).

Statistical Models: The statistical model forecast results evaluated in this study consisted of a total of 6 sets, all constructed using machine learning methods based on historical O_3 concentrations and meteorological conditions over a period of time. The nonlinear machine learning methods employed included support vector regression (SVR), random forest regression (RFR), and other algorithms. Additionally, deep learning methods, such as long short-term memory (LSTM), recurrent neural networks (RNNs), and deep neural networks (DNNs) were utilized. One of the models considered the effects of pollutant emissions, and two models considered the spatial transport of pollutants. Multiple models incorporated advanced algorithms, such as extreme gradient boosting, time series analysis, and attention mechanisms, to enhance the predictive performance based on machine learning. The statistical models used in this study were not categorized by the applied machine learning methods because several different machine learning methods were applied to one single statistical model.

Ensemble Models: The ensemble model forecast results evaluated in this study consisted of a total of 4 sets. Two sets were constructed by merging multiple-source observation data based on the forecast results from the Community Multiscale Air Quality (CMAQ) Model. The remaining two sets were constructed by merging observation data based on forecast results from multiple numerical models, including NAQPMS, CMAQ, CAMx, and WRF-chem. The ensemble models used in this study were based on numerical forecast and observed data and generated new forecast results by establishing the relationship between them with machine learning methods.

2.2. Evaluation Methods

This study employed three categories of forecast performance evaluation indicators, including general statistical metrics, evaluation metrics for pollution events, and comprehensive assessment metrics based on the Individual Air Quality Index (IAQI) [28] for O_3. The observation data sourced from the China National Environmental Monitoring Center, specifically data from 287 national-level ground monitoring stations in Beijing–Tianjin–Hebei and surrounding regions, were used for verification purposes. The O_3 concentration monitoring data had a temporal resolution of 1 h. Considering the timeliness of operational forecasting, only the forecast results generated before 8:00 AM (local time) each day were considered valid.

General Statistical Metrics: This category includes four common statistical indicators, namely the correlation coefficient (R), the root mean square error (RMSE), normalized mean bias (NMB), and mean bias (MB). These indicators are used to assess the forecast performance of O_3 numerical models.

Pollution Event Evaluation Metrics: The O_3 daily assessment indicator is the 8 h sliding average maximum value (denoted as O_{3-8h} max). The limit value for O_{3-8h} max in cities was set at 160 μg/m^3, and exceeding this limit indicated the occurrence of a pollution event. The distribution of observed and model-predicted values was classified into four scenarios, where cities were designated as follows: "*a*" for days when both observed and predicted values were below the limit; "*b*" for days when observed values were below the limit but predicted values exceeded it; "*c*" for days when observed values exceeded the limit but predicted values were below it; and "*d*" for days when both observed and predicted values exceeded the limit. The threshold forecast accuracy (FC) is used to assess the overall

performance of the model's ability to predict the occurrence of O_3 pollution events and is calculated using Formula (1).

$$FC = \frac{a+d}{a+b+c+d} \times 100\% \tag{1}$$

The probability of detection (POD) is defined as the proportion of forecasted exceedance days among all observed O_3 exceedance days. It examines how well the model predicts O_3 pollution events, with a higher POD indicating better performance. The POD is calculated using Formula (2):

$$POD = \frac{d}{c+d} \times 100\% \tag{2}$$

The false alarm rate (FAR) is defined as the proportion of forecasted exceedance days among all days where observed values do not exceed the limit. It assesses the model's tendency to issue false alarms for O_3 pollution events, with a lower FAR indicating better performance. The FAR is calculated using Formula (3):

$$FAR = \frac{b}{b+d} \times 100\% \tag{3}$$

IAQI Comprehensive Evaluation: The model-predicted O_{3-8h} max values for each city were converted into corresponding IAQI values. The IAQI forecast range was obtained by allowing a 25% fluctuation above and below the model-predicted IAQI values. If the observed IAQI value for the O_{3-8h} max value in a city fell within this forecast range, it was considered accurate; otherwise, it was determined as a high or low IAQI forecast. Based on the daily IAQI forecast range, corresponding IAQI forecast level ranges were determined. If the observed IAQI level for the daily O_{3-8h} max concentration fell within this forecast level range, it was considered an accurate forecast; otherwise, it was categorized as a high or low forecast. The IAQI comprehensive accuracy rate (S) is calculated using Formula (4):

$$S = \frac{n}{N} \times 100\% \tag{4}$$

n represents the number of days with an accurate IAQI comprehensive forecast and N represents the total number of valid days during the evaluation period.

3. Results and Discussion

For the 30 sets of forecast results from cities in Beijing–Tianjin–Hebei and surrounding regions, evaluation metrics for different forecast lead times were calculated based on each city. The regional assessment results were represented by the average values of these metrics across all cities. In Figure 1a, the distribution of regional O_3 IAQI comprehensive accuracy levels for different forecast lead times is presented (where the *x*-axis represents 1d, 2d ...7d, corresponding to forecasts for the next 1–7 days). The IAQI comprehensive accuracy could reach up to 83%, 81%, 79%, and 78% for lead times of 1–4 days, respectively, with average accuracy levels not falling below 60%. For lead times of 1–3 days, most models achieved accuracy levels above 65%. Figure 1b–d present statistical results for pollution event assessment indicators. For lead times of 1–6 days, the POD of the top-performing forecast models could reach around 90%, with most models above 70% for lead times of 1–2 days. For lead times of 1–3 days, the FC could exceed 80%, with average values surpassing 70%, and the FAR for a lead time of 3 days could be as low as 4%, with most models having a FAR below 26%. Figure 1e–h provide regional average statistical results for general statistical indicators. The correlation of O_3 forecasts significantly decreased with increasing lead times. For lead times of 1–4 days, the models with the highest correlation coefficients could reach 0.83, 0.82, 0.79, and 0.77, respectively, while for lead times of 5–7 days, the highest correlation coefficients were in the range of 0.67–0.68. For lead times of 1–3 days, most models had correlation coefficients exceeding 0.60. Smaller RMSE values

indicated better forecasting performance. For lead times of 1–4 days, the models with the smallest RMSE were 27, 28, 28, and 30 $\mu g/m^3$, respectively, while most models did not exceed 40 $\mu g/m^3$. The MB and NMB were similar indicators, and when they approached 0, better model performance was achieved. Positive values indicated overestimation, while negative values indicated underestimation. The majority of models in the statistical period showed O_3 forecast results leaning toward underestimation. In summary, the comprehensive statistical results of all evaluation metrics for the 30 sets of O_3 forecasts in Beijing–Tianjin–Hebei and surrounding regions indicated a general deterioration trend with an increasing lead time. The results for forecasts within the next 1–3 day range showed significant reference significance.
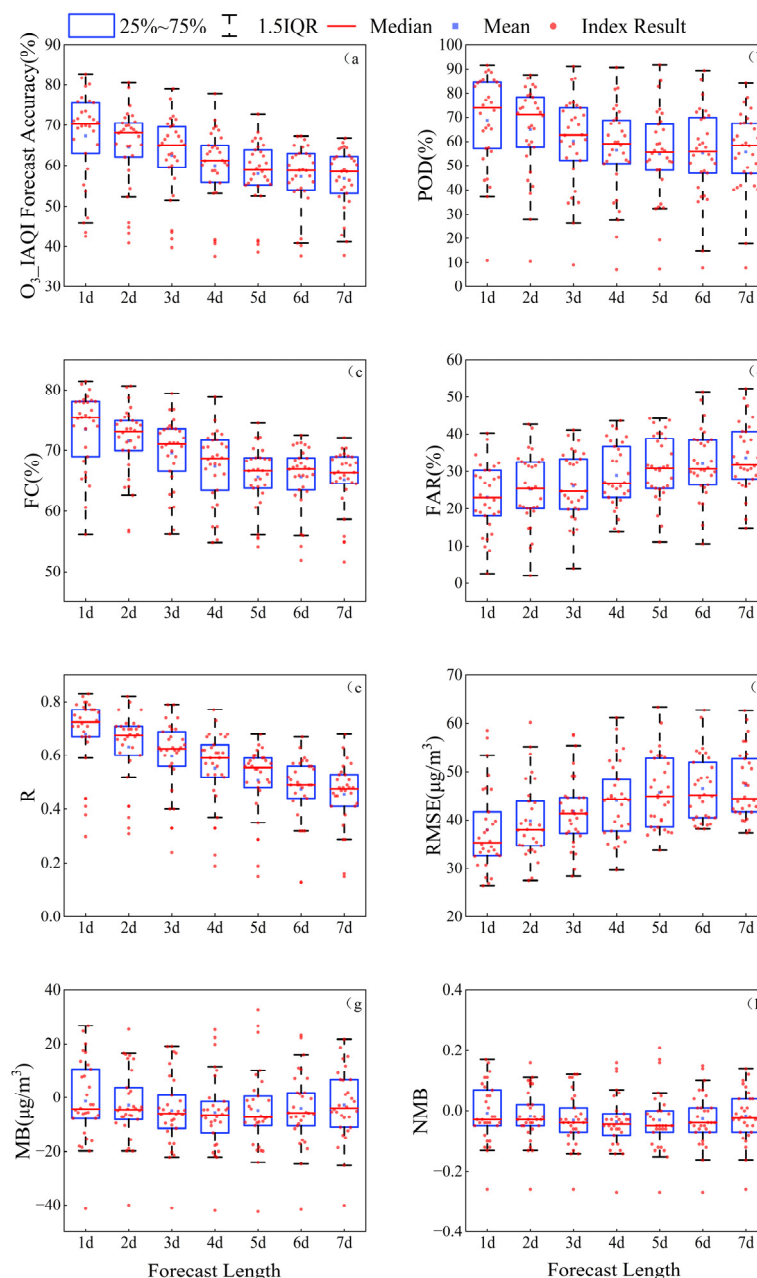

Figure 1. The distribution of (**a**) O_3 IAQI forecast accuracy; (**b**) POD; (**c**) FC; (**d**) FAR; (**e**) R; (**f**) RMSE; (**g**) MB; (**h**) NMB of various lead times.

The comprehensive analysis of the forecast results for three types of models (numerical, statistical, and ensemble) for a lead time of 3 days is presented in Figure 2. The figure provides a synthesis of the regional average RMSE, NMB, and R values for each set

of forecasts. In the plot, each circle represents a set of forecast results, with the *y*-axis corresponding to the RMSE, the *x*-axis corresponding to the NMB, and the size of the circle indicating the magnitude of R. The figure shows that the statistical models exhibited better overall performance in terms of RMSE and R values compared to the numerical and ensemble models. The majority of numerical and ensemble models showed small biases, while the statistical models demonstrated good consistency in all three evaluation metrics. The statistical models with the smallest RMSE also exhibited good NMB and correlation coefficients. For the four ensemble forecasts, R values were very close, ranging from 0.6 to 0.7, with RMSE differences not exceeding 8 μg/m^3. However, there was a notable disparity in the distribution of NMB, ranging from -0.26 to 0.12. The results from the twenty numerical models showed significant variability, but the best-performing numerical model was comparable to the statistical models and the best ensemble model. Overall, the statistical models demonstrated a closer resemblance to the observed O$_{3-8h}$ max concentrations, and the best-performing numerical and ensemble models achieved a similar forecasting level.
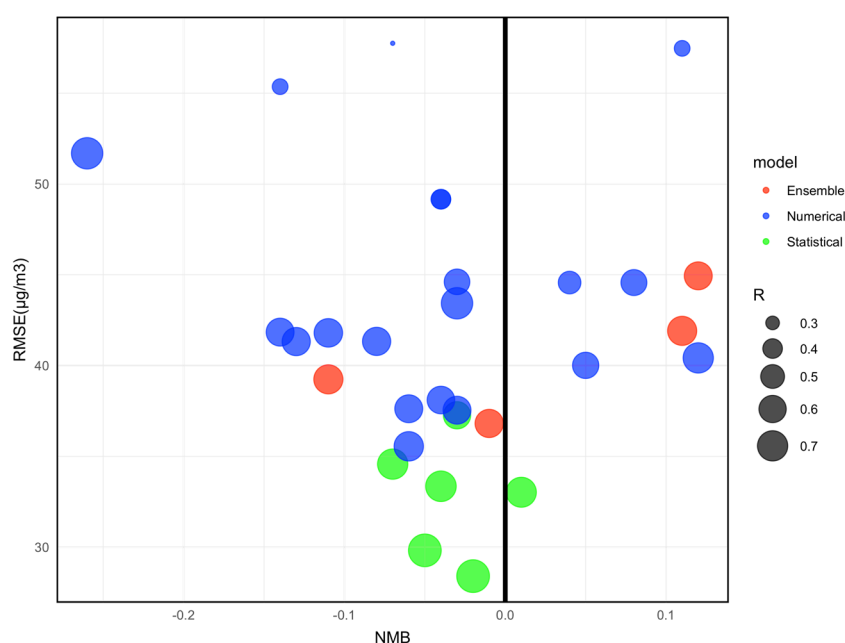
Figure 2. Comprehensive chart of multiple evaluation metrics for different types of forecast models for a lead time of 3 days.

In terms of IAQI comprehensive accuracy with a lead time of 3 days, the optimal model forecast results were selected as the top model. The models with the best IAQI comprehensive accuracy in numerical, statistical, and ensemble categories were denoted as top numerical, top statistical, and top ensemble, respectively. Figure 3a displays the regional average forecast performance for the top numerical, top statistical, and top ensemble models, which are abbreviated in the figure as Top_n, Top_s, and Top_e, respectively. The average IAQI comprehensive accuracy is ranked as follows: top statistical (65%) > top numerical (59%) > top ensemble (58%). Both numerical and ensemble models tended to overestimate, while statistical models tended to underestimate the O$_3$ concentration observation value. When a day had a O$_{3-8h}$ max concentration over 160 μg/m^3, this day was defined as a polluted day. Figure 3b focuses on evaluating the model's forecast performance on polluted days during the evaluation period. The average IAQI comprehensive accuracy was ranked as follows: top ensemble (65%) > top numerical (64%) > top statistical (62%). The forecast accuracy of numerical and ensemble models improved on polluted days and surpassed that of statistical models. Additionally, it was observed that the top statistical bias rate is significantly higher than that of top numerical and top ensemble. This may be

attributed to the limited predictive ability of statistical models concerning atypical events due to the quality and quantity of historical data [29,30]. Moreover, the lack of a description of long-distance transport and terrain for O_3 and its precursors can lead to significant deviations in pollution forecasting from complex environments [31,32]. Considering the varying performance of each model in different cities, Figure 3c provides a breakdown of the forecast results for top models in each city. For all days in June and 25 July 2023, out of 57 cities, the top model had an accuracy exceeding 80%, with three cities reaching 90%. Additionally, 30 cities had top model accuracy ranging between 70% and 80%. Figure 3d evaluates the forecast performance on polluted days, where black dots represent polluted days (i.e., the sample size for IAQI forecast performance assessment). It was observed that three cities had a top model accuracy rate of 100%, indicating accurate forecasts for all polluted days. Furthermore, 34 cities achieved accuracy rates exceeding 90%, with only three cities having accuracy rates below 80%. The forecast bias in all cities was mainly characterized by underestimation, demonstrating that the models tended to underestimate pollution events.
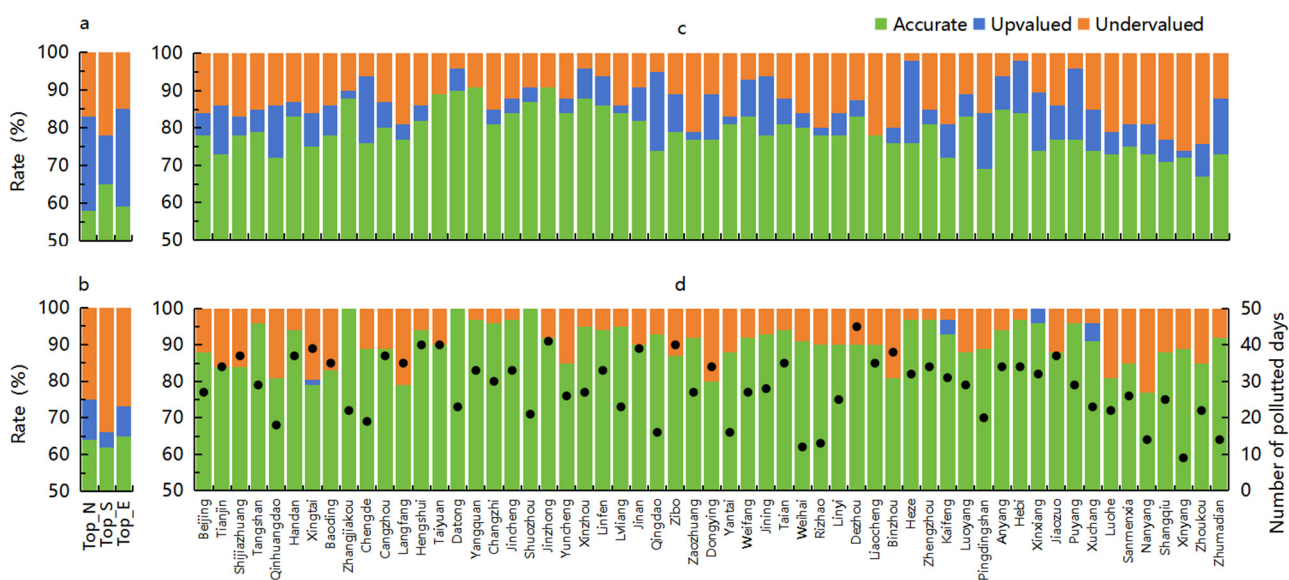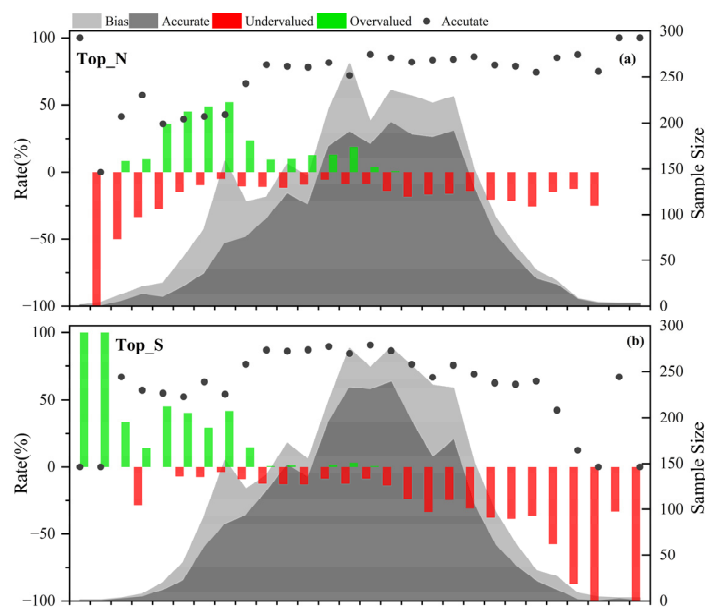

Figure 3. The IAQI comprehensive forecast evaluation for a lead time of 3 days, considering (**a**) the regional average for all days, (**b**) the regional average for polluted days, (**c**) the top model for all days in each city, and (**d**) the top model for polluted days in each city (black dots: polluted days, corresponding to the axis on the right).

Figure 4 illustrates the relationship between O_3 concentration and the accuracy of model forecasts. The shaded areas represent the percentage of models with accurate IAQI comprehensive forecasts within the corresponding O_3 concentration range for each city. The *x*-axis arranges cities in descending order based on the number of polluted days with daily mean O_{3-8h} max concentration above 200 μg/m^3, which is used to represent the pollution level. As the pollution level in cities decreased, there was an overall upward trend in the percentage of accurate models. When O_{3-8h} max concentrations ranged from 100 to 200 μg/m^3, over 60% of models could make accurate forecasts. Specifically, in the concentration range [100, 160), corresponding to the "Good" IAQI level, 63% of models provided accurate forecasts. In the concentration range of "Light Pollution" [160, 215), an average of 57% of models could accurately predict O_3 levels. However, when O_{3-8h} max was less than 100 μg/m^3 or greater than 215 μg/m^3, the percentage of models with accurate forecasts significantly dropped below 40%, especially when O_{3-8h} max reached the "Heavy Pollution" level, with the accuracy rate dropping to 28%.
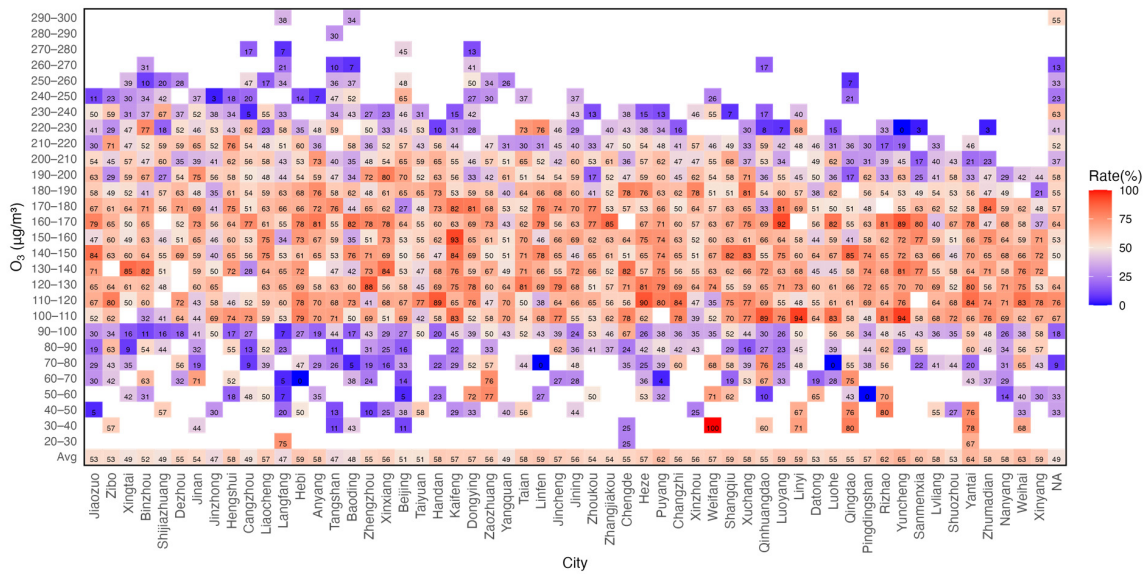
Figure 4. Heatmap depicting the relationship between the percentage of models with the next 3-day accurate forecasts and O_3 concentrations.

To analyze the forecast performance of different models across different O_3 concentration ranges, the IAQI comprehensive accuracy of the next 3-day was used as the indicator to pick the top numerical, statistical, and ensemble models for each city. The forecast performance for all cities against the variations in the observed O_3 was then analyzed in Figure 5. When O_{3-8h} max was in the concentration range of [110, 190), the accuracy of all three models was relatively high, i.e., around 80%. When the O_{3-8h} max value was low, both overestimation and underestimation could be observed in the numerical and ensemble models, while the statistical models tended to overestimate only. When the O_{3-8h} max value was high, the forecast bias for all three types of models was mainly underestimation. The numerical models had significantly more accuracy than the statistical models and slightly more accuracy than the ensemble models. The statistical models showed a significantly higher underestimation rate than the numerical and ensemble models, indicating that the statistical models have higher possibilities in failing to report high-concentration O_3 pollution events.

Figure 5. *Cont.*
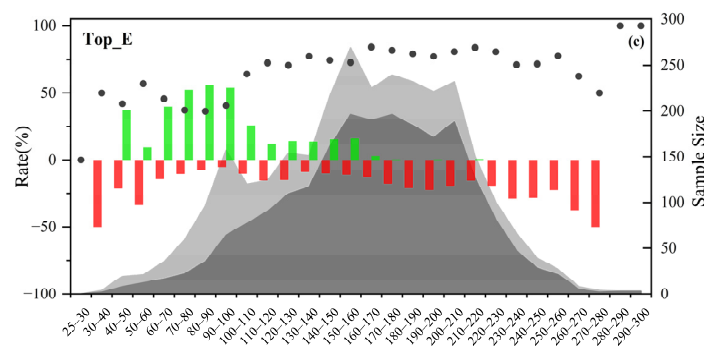
Figure 5. Comparison of the next 3–day IAQI comprehensive forecast results across various O_3 concentrations for (**a**) top numerical, (**b**) top statistical, and (**c**) top ensemble models (red and green bars represent underestimation and overestimation rates; black dots represent the accuracy, corresponding to the *y*–axis on the right).

4. Conclusions

This study conducted O_3 forecast experiments for all cities in Beijing–Tianjin–Hebei and surrounding regions, generating 30 sets of data from different numerical, statistical, and ensemble forecasting models. A comprehensive evaluation of 30 sets of forecasts results was explored and the conclusions are listed below:

When the lead time increases, the declining trend in O_3 concentration forecast performance becomes more evident. The forecasts for lead times of 1–3 days had higher reference significance, while the forecasts for 5–7 days had larger errors. At a lead time of 3 days, most models achieved an IAQI forecast accuracy rate exceeding 65%, with the highest reaching 79%. For most models, the POD values surpassed 65%, with the highest being 90%. The lowest FAR did not exceed 4% and most models had a FAR below 26%. However, at a lead time of 5–7 days, forecast performance significantly declined, with an average IAQI forecast accuracy rate below 60%, POD averaging around 60%, a FAR over at least 10%, an average RMSE exceeding 40 $\mu g/m^3$, and an average R dropping below 0.6.

An evaluation of different models across cities with varying pollution levels revealed that when O_{3-8h} max is in the concentration range of (100, 200), over 60% of models can accurately predict it. However, beyond this concentration range, the proportion of models with accurate predictions significantly dropped to less than 40%. Overall, as the pollution level in cities increased, the proportion of models with accurate predictions showed a decreasing trend, indicating that there are still considerable shortcomings in the ability of various O_3 models that can be used to accurately predict polluted days. Future efforts should focus on optimizing models to enhance the forecasting capability of O_3 pollution processes.

From the perspective of statistical metrics, such as RMSE, NMB, and R, statistical models outperform numerical models and ensemble models in general. Numerical models exhibited significant performance variations, with only the best-performing numerical and ensemble models being comparable to statistical models, suggesting that well-designed numerical models have the potential to achieve high forecast accuracy. The continuous optimization of numerical models has significant importance. In terms of IAQI forecast-related metrics, statistical models exhibited significantly higher rates of underestimation compared to numerical and ensemble models. For the forecast of polluted days, numerical models and ensemble models could achieve accuracy rates of 65% and 64%, respectively, surpassing statistical models at 62%. The underestimation rate of numerical models was 25%, lower than that of other models. As underestimation may lead to missed opportunities for implementing control measures in advance, the results from best-performing numerical and ensemble forecasts provide more meaningful insight into pollution process. The best-performing numerical model showed significantly better forecasting performance in the high O_3 concentration range compared to the best-performing statistical model, and slightly outperformed the best-performing ensemble model. In general, the overall forecast

performance of numerical models was less accurate than that of statistical models, but the statistical model was less effective in predicting the O$_3$ pollution process.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the size of the dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, T.; Yan, M.; Ma, W.; Ban, J.; Liu, T.; Lin, H.; Liu, Z. Short-term effects of multiple ozone metrics on daily mortality in a megacity of China. *Environ. Sci. Pollut. Res.* **2015**, *22*, 8738–8746. [CrossRef] [PubMed]
2. Avnery, S.; Mauzerall, D.L.; Liu, J.; Horowitz, L.W. Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage. *Atmos. Environ.* **2011**, *45*, 2284–2296. [CrossRef]
3. Tang, G.; Wang, Y.; Li, X.; Ji, D.; Hsu, S.; Gao, X. Spatial-temporal variations in surface ozone in Northern China as observed during 2009–2010 and possible implications for future air quality control strategies. *Atmos. Chem. Phys.* **2012**, *12*, 2757–2776. [CrossRef]
4. Chen, Z.; Zhuang, Y.; Xie, X.; Chen, D.; Cheng, N.; Yang, L.; Li, R. Understanding long-term variations of meteorological influences on ground ozone concentra-tions in Beijing during 2006–2016. *Environ. Pollut.* **2019**, *245*, 29–37. [CrossRef]
5. Zeng, S.; Zhang, Y. The Effect of Meteorological Elements on Continuing Heavy Air Pollution: A Case Study in the Chengdu Area during the 2014 Spring Festival. *Atmosphere* **2017**, *8*, 71. [CrossRef]
6. Wang, T.; Xue, L.; Brimblecombe, P.; Lam, Y.F.; Li, L.; Zhang, L. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total. Environ.* **2017**, *575*, 1582–1596. [CrossRef]
7. Zeng, Y.; Cao, Y.; Qiao, X.; Seyler, B.C.; Tang, Y. Air pollution reduction in China: Recent success but great challenge for the future. *Sci. Total Environ.* **2019**, *663*, 329–337. [CrossRef]
8. Li, K.; Jacob, D.J.; Shen, L.; Lu, X.; De Smedt, I.; Liao, H. Increases in surface ozone pollution in China from 2013 to 2019: Anthropogenic and meteoro-logical influences. *Atmos. Chem. Phys.* **2020**, *20*, 11423–11433. [CrossRef]
9. Mousavinezhad, S.; Choi, Y.; Pouyaei, A.; Ghahremanloo, M.; Nelson, D.L. A comprehensive investigation of surface ozone pollution in China, 2015–2019: Separating the contributions from meteorology and precursor emissions. *Atmos. Res.* **2021**, *257*, 105599. [CrossRef]
10. Liu, R.; Ma, Z.; Liu, Y.; Shao, Y.; Zhao, W.; Bi, J. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach. *Environ. Int.* **2020**, *142*, 105823. [CrossRef] [PubMed]
11. Wang, P.; Yang, Y.; Li, H.; Chen, L.; Dang, R.; Xue, D.; Li, B.; Tang, J.; Leung, L.R.; Liao, H. North China Plain as a hot spot of ozone pollution exacerbated by extreme high tempera-tures. *Atmos. Chem. Phys.* **2022**, *22*, 4705–4719. [CrossRef]
12. Zifa, W.; Chengming, P.; Jiang, Z. IAP Progress in atmospheric environment modeling research. *Chin. J. Atmos. Sci.* **2008**, *32*, 987–995. (In Chinese)
13. An, J.; Huang, M.; Wang, Z.; Zhang, X.; Ueda, H.; Cheng, X. Numerical Regional Air Quality Forecast Tests over the Mainland of China. *Water Air Soil Pollut.* **2001**, *130*, 1781–1786. [CrossRef]
14. Yafouz, A.; Ahmed, A.N.; Zaini, N.A.; El-Shafie, A. Ozone concentration forecasting based on artificial intelligence techniques: A systematic review. *Water Air Soil Pollut.* **2021**, *232*, 1–29. [CrossRef]
15. Krzywanski, J.; Blaszczuk, A.; Czakiert, T.; Rajczyk, R.; Nowak, W. Artificial intelligence treatment of NOX emissions from CFBC in air and oxy-fuel conditions. In Proceedings of the CFB-11—11th International Conference on Fluidized Bed Technology, Beijing, China, 14–17 May 2014; pp. 619–624.
16. Ahmed, I.; Ahmad, M.; Chehri, A.; Jeon, G. A heterogeneous network embedded medicine recommendation system based on LSTM. *Future Gener. Comput. Syst.* **2023**, *149*, 1–11. [CrossRef]
17. Permanasari, A.E.; Zaky, A.M.; Fauziati, S.; Fitriana, I. Predicting the Amount of Digestive Enzymes Medicine Usage with LSTM. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2018**, *8*, 1845–1849. [CrossRef]
18. Tang, G.; Zhu, X.; Xin, J.; Hu, B.; Song, T.; Sun, Y.; Zhang, J.; Wang, L.; Cheng, M.; Chao, N.; et al. Modelling study of boundary-layer ozone over northern China—Part I: Ozone budget in summer. *Atmos. Res.* **2017**, *187*, 128–137. [CrossRef]
19. Shahraiyni, H.T.; Sodoudi, S. Statistical Modeling Approaches for PM$_{10}$ Prediction in Urban Areas; A Review of 21st-Century Studies. *Atmosphere* **2016**, *7*, 15. [CrossRef]

20. Li, X.; Peng, L.; Hu, Y.; Shao, J.; Chi, T. Deep learning architecture for air quality predictions. *Environ. Sci. Pollut. Res.* **2016**, *23*, 22408–22417. [CrossRef]
21. Sayeed, A.; Choi, Y.; Eslami, E.; Jung, J.; Lops, Y.; Salman, A.K.; Lee, J.-B.; Park, H.-J.; Choi, M.-H. A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. *Sci. Rep.* **2021**, *11*, 10891. [CrossRef]
22. Tangang, F.T.; Tang, B.; Monahan, A.H.; Hsieh, W.W. Forecasting ENSO events: A neural network–extended EOF approach. *J. Clim.* **1998**, *11*, 29–41. [CrossRef]
23. Wu, X.; Xie, K.; Liu, J.; Liu, D.; Zhou, J.; Tang, L. Regional forecasting of fine particulate matter concentrations: A novel hybrid model based on principal component regression and EOF. *Earth Space Sci.* **2021**, *8*, e2021EA001694. [CrossRef]
24. Yang, J.; Zhao, Y. Performance and application of air quality models on ozone simulation in China–A review. *Atmos. Environ.* **2023**, *293*, 119446. [CrossRef]
25. Lv, B.; Cobourn, W.G.; Bai, Y. Development of nonlinear empirical models to forecast daily PM2.5 and ozone levels in three large Chinese cities. *Atmos. Environ.* **2016**, *147*, 209–223. [CrossRef]
26. Lyu, Y.; Ju, Q.; Lv, F.; Feng, J.; Pang, X.; Li, X. Spatiotemporal variations of air pollutants and ozone prediction using machine learning algorithms in the Beijing-Tianjin-Hebei region from 2014 to 2021. *Environ. Pollut.* **2022**, *306*, 119420. [CrossRef] [PubMed]
27. Cheng, M.; Fang, F.; Navon, I.M.; Zheng, J.; Zhu, J.; Pain, C. Assessing uncertainty and heterogeneity in machine learning-based spatiotemporal ozone prediction in Beijing-Tianjin-Hebei region in China. *Sci. Total Environ.* **2023**, *881*, 163146. [CrossRef] [PubMed]
28. *HJ 633—2012*; Technical Regulation on Ambient Air Quality Index (AQI). China Environmental Science Press: Beijing, China, 2012.
29. Markou, M.; Singh, S. Novelty detection: A review—Part 1: Statistical approaches. *Signal Process.* **2003**, *83*, 2481–2497. [CrossRef]
30. Ruff, L.; Vandermeulen, R.; Goernitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep one-class classification. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 4393–4402.
31. Soja, G.; Soja, A.M. Ozone indices based on simple meteorological parameters: Potentials and limitations of regression and neural network models. *Atmos. Environ.* **1999**, *33*, 4299–4307. [CrossRef]
32. Ma, R.; Ban, J.; Wang, Q.; Li, T. Statistical spatial-temporal modeling of ambient ozone exposure for environmental epidemiology studies: A review. *Sci. Total Environ.* **2020**, *701*, 134463. [CrossRef]