Development and application of the WRFDA-Chem 3DVAR system: aiming to improve air quality forecast and diagnose model deficiencies

Wei Sun[1,2], Zhiquan Liu[2*], Dan Chen[3*], Pusheng Zhao[3], and Min Chen[3]

[1] National Space Science Center, Chinese Academy of Sciences, Beijing, 100190, China

[2] National Center for Atmospheric Research, Boulder, CO, 80301, USA

[3] Institute of Urban Meteorology, China Meteorology Administration, Beijing, 100089, China

* *Corresponding author: Dr. Zhiquan Liu (liuz@ucar.edu) and Dr. Dan Chen (dchen@ium.cn)*

Abstract

To improve the operational air quality forecasting over China, a new aerosol/gas phase pollutants assimilation capability is developed within the WRFDA system using 3DVAR algorithm. In this first application, the interface for MOSAIC aerosol scheme is built with flexible extending potentials. Based on the new WRFDA-Chem system, five experiments assimilating different surface observations, including $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO are conducted for January 2017 along with a control experiment without DA. Results exhibit that the WRFDA-Chem system evidently improves the air quality forecasting. On the analysis aspect, the assimilation of surface observations reduces the bias and RMSE in the initial condition (IC) remarkably; on the forecast aspect, better forecast performances are acquired up to 24-h, in which the experiment assimilating the six pollutants simultaneously displays the best forecast skill overall. With respect to the impact of DA cycling frequency, the responses toward IC updating are found out to be different among the pollutants. For $PM_{2.5}$, PM_{10}, SO_2, and CO, the forecast skills increase with the DA frequency. For O_3, although improvements are acquired at the 6-h cycling frequency, the advantage of more frequent DA could be consumed by the disadvantages of the unbalanced photochemistry (due to inaccurate precursor NO_x/VOC ratios) or the changed titration process (due to changed NO_2 concentrations but not NO) from assimilating the existing observations (only O_3 and NO_2, but no VOC and NO); yet the finding is based on the 00 UTC forecast for this winter season only and O_3 has strong diurnal and seasonal variations, more experiments should be conducted to draw further conclusions. In addition, considering after one aspect (IC) in the model is corrected by DA, the deficiencies from other aspects (e.g., chemical reactions) could be more evident, this study explores the model deficiencies by investigating the effects of assimilating gaseous precursors on the forecast of related aerosols. Results exhibit that the parameterization (uptake

coefficients) in the newly added Sulfate-Nitrate-Ammonium (SNA) relevant heterogeneous reactions in the model are not fully appropriate although it best simulates observed SNA aerosols without DA; since the uptake coefficients were originally tuned under the inaccurate gaseous precursor scenarios without DA, the biases from the two aspects (SNA reactions and IC DA) were just compensated. In the future chemistry development, parameterizations (such as uptake coefficients) for different gaseous precursor scenarios should be adjusted and verified with the help of DA technique. According to these results, DA ameliorates certain aspects by using observation as constraints, and thus provides an opportunity to identify and diagnose the model deficiencies; it is useful especially when the uncertainties of various aspects are mixed up and the reaction paths are not clearly revealed. In the future, besides being used to improve the forecast through updating IC, DA could be treated as another approach to explore necessary developments in the model.

1. Introduction

Air pollution is almost inevitable for all developed (historically) and developing (in present days) countries. From acid rain, haze to smog etc., the air pollution significantly impacts atmospheric visibility, human health, and climate. As one of the fastest-growing countries, China has been suffering from the extreme haze with high particulate matter (PM) national-wide and increasing tropospheric ozone (O_3) pollution in city clusters (Fu et al., 2019; Lu et al., 2019). To control the pollutions as well as to improve the air quality forecast, Chinese governments had enforced stricter air quality standards from 2012, and deployed monitoring network for six "criteria" air pollutants since 2013, which includes $PM_{2.5}$ and PM_{10} (aerosols/fine particulate matter with aerodynamic diameters less than 2.5 or 10 μm), SO_2 (sulfur dioxide), NO_2 (nitrogen dioxide), O_3 (ozone), and CO (carbon monoxide). Among

52 the six pollutants, the forecast on aerosols (especially $PM_{2.5}$) is of greatest research interest as the
53 severity of aerosol pollution and its negative effects on both health and climate. However, it's still
54 challenging to accurately simulate and forecast aerosols by pure air quality models due to some issues,
55 such as the large uncertainties in primary and precursor emissions processes, the incomplete
56 understanding and parameterization of secondary inorganic/organic reactions from precursors, and the
57 accumulation of meteorology simulation errors. In addition to aerosol forecast, the elevated O_3 levels
58 in city clusters over eastern China draw more and more attentions recently. Under this circumstance,
59 in the urban regions in China, where suffer from complex air pollution with both haze and smog, the
60 accurate forecast of air quality has been not only a challenge for operational centers, but also a common
61 concern for the scientific community.

62 To improve the forecast skill, data assimilation (DA), a combination of observations and numerical
63 model output, has been widely used in meteorology forecast since last century, and recently extended
64 to air pollutant forecasts. Based upon various techniques, DA is proven to be skillful at improving the
65 meteorology and aerosol forecasts (Bannister 2017; McHenry et al. 2015; Peng et al. 2018; Sandu and
66 Chai 2011; Schutgens et al. 2010; Sekiyama et al. 2010; Tang et al. 2011; Tang et al. 2013). Focusing
67 on aerosol assimilation, NCAR group had conducted a series of work. Using three-dimensional
68 variational (3DVAR) algorithm, Liu et al. (2011) implemented DA on aerosol optical depth estimates
69 within the Grid-point Statistical Interpolation (GSI) system. Schwartz et al. (2012), Jiang et al. (2013),
70 and Chen et al. (2019) further extended this system to assimilate surface $PM_{2.5}$ and PM_{10}. It should be
71 noted that the aerosols are complicated not merely from primary emissions but also secondary
72 reactions with gaseous precursors in the atmosphere (Huang et al. 2014; Nie et al. 2014; Xie et al.
73 2015). However, the assimilation of aerosols along with gas phase pollutants are seldom investigated.

Recently, it is encouraging that an Ensemble Kalman Filter (EnKF) DA system is developed to assimilate multi-species surface chemical observations (Peng et al. 2017), while the EnKF system may not be the favorite choice in operational applications due to its massive computational cost. In addition, at the Institute of Urban Meteorology (IUM), regional NWP system–RMAPS-ST (adapted from WRF) and regional air quality model–RMAPS-Chem (adapted from WRF-Chem) are applied operationally for the weather and air quality forecast over Northern China. RMAPS-ST provides the meteorology drivers for RMAPS-Chem, and WRFDA is utilized for the meteorology DA in RMAPS-ST (Fan et al. 2016; Yu et al. 2018). In result, to implement the assimilations of aerosols along with gas phase pollutants in the future air quality forecast operational system (e.g. the RMPAS-Chem), and to design an efficient and unified DA platform that satisfies the operational needs in both meteorology and air quality forecast, this study works on the WRFDA system with 3DVAR algorithm. To the authors' knowledge, this is the first attempt to assimilate hourly ground-based aerosols simultaneously with gas phase pollutants in the WRFDA system.

With regard to the aerosol data assimilation, the first and foremost challenge comes from the complex components related to the aerosol scheme. With different emphasis and applications, the chosen aerosol scheme in the model could be different, which will lead to various choices and treatments for the analysis variables in the DA system. For example, in the existed DA developments, many studies used the GOCART aerosol scheme to address the dust or the natural-source related events. However, the GOCART aerosol scheme is well known to underestimate the PM concentrations due to lack of secondary organic aerosol (SOA) formation, as well as aerosol species related to the anthropogenic emission, such as nitrate and ammonium (McKeen et al. 2009; Pang et al. 2018). Different from the GOCART scheme, the MOSAIC (Model for Simulating Aerosol Interactions and

96 Chemistry) aerosol scheme uses a sectional approach to represent the aerosol size distribution with
97 different size bins, and it takes black carbon, organic carbon, sulfate, nitrate, ammonium, sodium,
98 chloride, and other inorganic compounds that are related to anthropogenic emissions into consideration.
99 In result, the MOSAIC scheme exhibits a better performance in representing the complex $PM_{2.5}$
100 pollution over China (Chen et al. 2016; Chen et al. 2019). Therefore, to make the DA system suitable
101 for different emphasis and applications, a flexible aerosol assimilation capability is built within the
102 WRFDA system in this study, which will facilitate developments and applications for more chemistry
103 schemes in the future. Focusing on the air quality forecast over China, this study mainly analyses the
104 results of MOSAIC aerosol scheme.

105 It should be mentioned that the forecast performance with DA also relies on the air quality model
106 itself. Due to the limited observational information as constraint, the DA system uses large parts of
107 model mechanism and processes to derive the full analysis information (e.g. use total PM mass
108 observations to analyze all PM components). However, there are still potential deficiencies in the
109 model. For example, some reaction paths are missing in the heavily polluted events in China (e.g.
110 Wang et al., 2014), since the chemistry schemes are originally developed for relatively clean areas and
111 recently observed pathways haven't been timely reflected in the model. Moreover, the large
112 uncertainties of precursor and primary emissions could bring errors to the aerosol species partitioning
113 and size distribution in the model. Nevertheless, when it comes to DA, as one aspect (initial conditions
114 of aerosols and some precursors) in the model is corrected by using observation as constraints, the
115 deficiencies from other aspects, such as the above mentioned chemical reactions, could be more
116 evident. From this point of view, after investigating to what extent the DA technique can help to

improve the forecast of air quality, this study further explores the model deficiencies with the help of DA, aiming to provide helpful indications for future model development.

In the rest of the paper, an overview of the model description, observations, and methodology is presented in Section 2, followed by evaluations of the new WRFDA-Chem system in Section 3. Section 4 analyzes the DA experiments in consideration of potential issues in the model, aiming to provide beneficial references on further model development. Conclusions and discussions are given in section 5.

2. Model description, observations, and methodology

In this study, the interfaced air quality model is WRF-Chem. The WRF-Chem settings are very similar to those of Chen et al. (2016). Here, only a summary of the model configuration and observations is provided below. Descriptions of the most important development of this study, the WRFDA-Chem system, are presented in Section 2.3.

2.1 WRF-Chem model and emissions

As in Chen et al. (2016), version 3.6.1 of the WRF-Chem model is used in this study to simulate the aerosols and gas-phase chemistry processes. A summary of the used physical parameterizations is given in Table 1. Details of the WRF-Chem model have been described by Grell et al. (2005) and Fast et al. (2006). The Carbon Bond Mechanism version Z (CBMZ) and Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) schemes are used as the gas-phase and aerosol chemical mechanisms, respectively. The relative humidity (RH) dependent heterogeneous reactions added by Chen et al. (2016) are also applied in the simulations. The model computational domain covers most

of China and its surrounding regions. Figure 1 presents the horizontal range of the domain, which contains 121 x 121 horizontal grids at a 40.5-km resolution. Vertically, there are 57 levels extending from the surface to 10 hPa.

As in Chen et al. (2019), the emission input is based on the 2010 Multi-resolution Emission Inventory for China (MEIC) (He 2012; Lei et al. 2011; Li et al. 2014; Zhang et al. 2009), which has already been applied in many recent studies over China (Wang et al. 2016; Wang et al. 2013; Zheng et al. 2015). The emission inventory has also been processed to match the model grid spacing (40.5 km) from an original grid spacing of 0.25° × 0.25° (Chen et al. 2016). Admittedly, the difference between the emission base year and our simulation year and the spatial-temporal allocations may arise uncertainties in our simulation, this emission is the only publicly available emission inventory when the study is conducted. Meanwhile, the inhomogeneous spatial changes and large uncertainties in seasonal allocations of the emissions made it difficult to simply scale the original emission inventory for our study period (Chen et al. 2019).

The dust emission is the GOCART dust emission and the biogenic emission is calculated online by the Gunther scheme within the WRF-Chem model. Given the time period of this study (January) is not the period with massive fires (crop/biomass burning), the fire emission is not used in this study.

2.2 Observations

For the future application in RMAPS-Chem operational air quality forecast system, the WRFDA-Chem system is designed to assimilate the hourly surface observations of six major pollutants ($PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO) from the China National Environmental Monitoring Center (CNEMC). To verify the capability of the system, we use the data for the whole month of January 2017. As in

Chen et al. (2019), to perform statistical calculations, an observation dataset at 531 locations (Fig. 1) is acquired by averaging all the original observations (1600+ sites) that fall into the same model grid. Meanwhile, two steps of data quality control are conducted before DA. Firstly, observations lager than a threshold are treated as unrealistic and are not assimilated. Secondly, observations leading to innovations (observations minus the model-simulated values) higher than a maximum deviation are omitted. For $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO, the threshold in the first step is 500 μg m^{-3}, 700 μg m^{-3}, 200 μg m^{-3}, 200 μg m^{-3}, 200 μg m^{-3}, and 20 mg m^{-3}, respectively; the maximum deviation in the second step is 120 μg m^{-3}, 120 μg m^{-3}, 60 μg m^{-3}, 60 μg m^{-3}, 60 μg m^{-3}, and 6 mg m^{-3}, respectively.

To verify sulfate-nitrate-ammonium partitioning, a site observation of different chemical species is used in Section 4. The measurements were performed over January 14–20, 2017, and carried out on the roof of IUM in Beijing (green dot in Fig. 1). A detailed description for the features of the observation, including the quality assurance and quality control has been given by Su et al. (2018). This study mainly uses the sulfate (SO_4^{2-}) and nitrate (NO_3^-) in this dataset.

2.3 WRFDA-Chem system

In this study, an aerosol/chemical assimilation capability is built within the version 4.0.3 of the WRFDA system with 3DVAR algorithm. The WRFDA 3DVAR produces the analysis through the minimization of a scalar objective function $J(x)$ given by

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}[H(x) - y]^T R^{-1}[H(x) - y], \quad (1)$$

where x_b denotes the background vector, y is a vector of the observations, and B and R represent the background and observation error covariance matrices, respectively. The covariance matrices determine how close the analysis is weighted toward the background and observations. H is

179 the observation operator that interpolates model grid point values to observation space and converts
180 model-predicted variables to observed quantities.

181 Generally, the implementation of WRFDA-Chem 3DVAR includes several parts: WRF-Chem
182 model and surface air pollutants observation interface to WRFDA, the addition of aerosol/chemical
183 analysis variables, the surface air pollutants observation operators, the update of observation errors,
184 and the statistics of background error covariances for chemical analysis variables. Detailed
185 descriptions will be presented in the following parts. It's worth mentioning that the new WRFDA-
186 Chem system is designed with a flexible aerosol assimilation capability that can switch between
187 different aerosol schemes. Given the fact that WRF-Chem model predicts the PM concentrations in
188 the forms of different prognostic variables depending on the chosen aerosol scheme, the
189 aerosol/chemical prognostic variables are given in the registry file of the WRFDA-Chem, instead of
190 specifically defined in the code. With the help of the registry mechanism of WRF model, the prognostic
191 variables in the entire DA process can be easily adjusted by modifying the registry file. The WRFDA-
192 Chem system has been tested with GOCART and MOSAIC aerosol scheme, while this study focuses
193 on the MOSAIC scheme.

194 **2.3.1 Observation operators**

195 The WRFDA-Chem is designed to assimilate six types of surface aerosol/chemical observations,
196 including $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO. For aerosol assimilation, the aerosol species in the
197 MOSAIC scheme are defined as black carbon (BC), organic compounds (OCs), sulfate (SO_4^{2-}), nitrate
198 (NO_3^-), ammonium (NH_4^+), sodium (NA), chloride (CL), and other inorganic compounds (OIN). To
199 represent the aerosol size distribution, MOSAIC uses a sectional approach with different bins. This

study uses four size bins with aerosol diameters ranging from 0.039–0.1, 0.1–1.0, 1.0–2.5, and 2.5–10μm. The $PM_{2.5}$ total is controlled by the 24 variables in the first three bins (8 species multiplied by 3 bins), and the PM_{10} total is controlled by the 32 variables in the four bins (8 species multiplied by 4 bins). In result, the model-simulated $PM_{2.5}$ is computed by summing the 24 variables as

$$y^f_{PM_{2.5}} = \rho_d \sum_{i=1}^{3}[BC_i + OC_i + SO_{4_i} + NO_{3_i} + NH_{4_i} + CL_i + NA_i + OIN_i]. \quad (2)$$

The model-simulated PM_{10} observations are computed by summing the 32 variables as

$$y^f_{PM_{10}} = \rho_d \sum_{i=1}^{4}[BC_i + OC_i + SO_{4_i} + NO_{3_i} + NH_{4_i} + CL_i + NA_i + OIN_i]. \quad (3)$$

Correspondingly,

$$y^f_{PM_{10-2.5}} = \rho_d \sum_{i=4}^{4}[BC_i + OC_i + SO_{4_i} + NO_{3_i} + NH_{4_i} + CL_i + NA_i + OIN_i], \quad (4)$$

where ρ_d is the dry-air density, which is used to convert the unit of the analysis variable (μg/kg) to the observations (μg/m^3); i denotes the bin number in the MOSAIC aerosol scheme. In the experiment assimilating $PM_{2.5}$ alone, the $PM_{2.5}$ observations are used to analyze the species in the first three bins (Eq. 2). In the experiment assimilating $PM_{2.5}$ and PM_{10} simultaneously, the $PM_{2.5}$ observations are used to analyze the species in the first three bins (Eq. 2), and the $PM_{10\text{-}2.5}$ (PMcoarse, hereafter) in the observations is used to analyze the species in the 4th bin (Eq. 4). A similar approach has been adopted by Peng et al. (2018).

In the assimilation of the gas-phase pollutants, the model-simulated values are computed by

$$y^f_x = \rho_d \cdot \frac{M_x}{M_{dair}} \cdot R_x \cdot 10^3, \quad (5)$$

where x denotes the four gas-phases pollutants as in SO_2, NO_2, O_3, and CO, ρ_d is the dry-air density, M_x is the relative molecular mass for the four gas-phases pollutants, M_{dair} is the relative molecular mass for dry-air, and R_x is the mixing ratio for the four gas-phases pollutants. Since the gas-phase

pollutants observations are mass concentrations in μg/m^3 and the analysis variables are mixing ratios in ppmv, the Eq. 5 is used for the unit conversion.

2.3.2 Observation errors

Following Chen et al. (2019) and Peng et al. (2018), the observation error covariance matrix R in Eq. (1) is estimated from measurement error ε_0 and the representativeness error ε_r in this study. The measurement error ε_0 is defined as $\varepsilon_0 = 1.0 + 0.0075 \cdot M_i$, where M_i denotes the observation of the six major pollutants in unit μg/m^3; the representativeness error ε_r is defined as $\varepsilon_r = \gamma\varepsilon_0\sqrt{\frac{\Delta x}{L}}$. where γ is an adjustable parameter scaling (set as 0.5), Δx is the grid spacing (40.5 km in our case) and L is the radius of influence of the observation (set to 2 km). These parameter settings are based on the sensitivity tests by Chen et al. (2019). The total observation error (ε_x) is computed as $\varepsilon_x = \sqrt{\varepsilon_{0_x}^2 + \varepsilon_{r_x}^2}$, where x denotes the six major pollutants as in $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO.

2.3.3 Background error covariance

To implement the aerosol/chemical DA with the MOSAIC-4Bin scheme, this study expands the GEN_BE v2.0 (Descombes et al. 2015) to compute the **B** matrix in Eq. (1) for the 32 chemical variables as in Eq. 3 (BC, OC, SO_4^{2-}, NO_3^-, NH_4^+, NA, CL, and OIN in four bins), as well as the four gas-phase variables as in Eq. 5 (SO_2, NO_2, O_3, and CO). Since it is both technically and scientifically challenging to model the cross-correlations between different aerosol/chemical variables in a 3DVAR framework, they are not considered in this study. We plan to introduce the cross-variable correlations with the ensemble-variational approach in the future extension of the system. With the updated GEN_BE v2.0,

240 the statistics for background error covariance, such as standard deviation, vertical and horizontal length
241 scales, and vertical correlations, are computed for each of the aerosol/chemical variables. In this study,
242 the background error covariance is estimated using the National Meteorological Center (NMC) method
243 (Parrish and Derber, 1992) from one-month WRF-Chem forecasts over January 2017.

244 Following the analyses based on the GEN_BE v2.0 (Descombes et al. 2015), Figure 2 presents the
245 background error standard deviations of each species at different vertical levels. For the aerosols in the
246 first three size bins (Fig. 2a-2c), although the standard deviation errors vary across the species, the
247 errors of NO_3^-, SO_4^{2-}, NH_4^+, OC, and OIN are generally larger than that of the others (BC, Cl and NA)
248 in the three size bins. These results are consistent with the finding in Chen et al. (2019), which allows
249 inorganic compounds (NO_3^-, SO_4^{2-}, NH_4^+), OC and OIN to be adjusted more in corresponding to their
250 larger background errors. For the aerosols in the 4th size bin (Fig. 2d), the errors are unreasonably
251 much smaller than that in the first three bins due to model deficiency. Under this circumstance, to get
252 a reasonable bigger adjustment for the aerosols in the 4th size bin, it might need to enlarge their
253 background errors in the DA procedure. As for the gaseous pollutants (Fig. 2e), CO has the biggest
254 background errors in the middle and lower layers, followed by O_3, SO_2 and NO_2.

255 For the background error horizontal correlation length scales, the results are similar as in Liu et al.
256 (2011) (figure omitted). The length scales of aerosols are comparable in most of the species, which
257 generally span from 1.5 to 2.5 times the grid spacing, while the aerosol species NA exhibits a smaller
258 horizontal length scale than all the other species. For the background error vertical correlations (figure
259 omitted), the results are similar as in Descombes et al. (2015), in which the vertical correlations are
260 bigger in the lower levels (where they are emitted) in most of the species. According to Descombes et

al. (2015), the reactions with species emitted near the surface might create these strong correlations in the lower model levels.

2.3.4 Experimental design

To seek for the best forecast performance, six experiments were conducted for January 2017 in this study, including NODA, PM1, PM2, ALL, ALL_3h, and ALL_1h (detailed in Table 2). NODA is the control experiment without any data assimilation. The design of PM1, PM2, and ALL is to investigate the assimilation impacts of $PM_{2.5}$, PMcoarse, and gas-phase pollutants (SO_2, NO_2, O_3, CO) step-by-step.

The NODA experiment initialized a new WRF-Chem forecast every 6-h between 00:00 UTC, 20 December 2016 and 18:00 UTC 31 January 2017, in which the aerosol/chemical fields were simply carried over from cycle to cycle, and the meteorological initial condition/boundary conditions were updated from GFS data every 6-h. The first 10 days were treated as the spin up period, and only simulations in January were used in the following analyses. The PM1, PM2, and ALL experiments updated the chemical IC using the WRFDA-Chem system every 6-h starting from 00:00 UTC, 1 January. The background of the first cycle was obtained from the NODA experiment, and all subsequent cycles were derived from the 6-h forecast of the previous cycle. The only difference between PM1, PM2, and ALL experiments is that PM1 only assimilated $PM_{2.5}$ observations; PM2 assimilated PM2.5 and PMcoarse (PM10-2.5) simultaneously; ALL assimilated PM2.5, PM10-2.5, SO2, NO2, O3, and CO together.

In view of the cycling frequency is an important aspect in the DA strategy, especially for 3DVAR, two more experiments that assimilate all the six major pollutants with 3-h and 1-h cycling frequency are conducted respectively (experiment ALL_3h and ALL_1h). To investigate the forecast improvements, a 24-h forecast is initialized for all the experiments at 00:00 UTC of each day.

3. Performance of the WRFDA-Chem system

3.1 Impact on analyses

To evaluate the performance of the WRFDA-Chem system, the impact on analyses is firstly investigated. Figure 3 presents the domain-averaged bias and root-mean-square-error (RMSE) of the analysis at 00 UTC over January 1-31, 2017. For $PM_{2.5}$ (Fig. 3a), the NODA experiment displays a general overestimation of 36.60 $\mu g/m^3$, along with a large RMSE of 70.41 $\mu g/m^3$. After DA, in the PM1, PM2, and ALL experiments, the bias of $PM_{2.5}$ drops to 5.62 $\mu g/m^3$, 5.19 $\mu g/m^3$, and 5.98 $\mu g/m^3$, respectively; the RMSE drops to 22.10 $\mu g/m^3$, 22.84 $\mu g/m^3$, and 23.15 $\mu g/m^3$, respectively.

In the analyses of PM_{10}, it is noted that the PM1 experiment has a larger bias than the NODA run (Fig. 3b). To explain this phenomenon, Figure 4 presents the monthly mean difference between PM_{10} and $PM_{2.5}$ (PM_{10} minus $PM_{2.5}$, PMcoarse) in the analysis. In the observation, the PMcoarse generally increases from south to north, reaching above 50 $\mu g/m^3$ over northern China (Fig. 4a). However, the PMcoarse in the NODA experiment (with an average of 5.47 $\mu g/m^3$) is much smaller than that in the observation (with an average of 39.13 $\mu g/m^3$). This result suggests that the WRF-Chem model failed to reasonably represent the PMcoarse, which is actually the 4^{th} bin of the aerosol species in the MOSAIC scheme. Under this circumstance, when the assimilation of $PM_{2.5}$ trying to reduce its evident overestimation (Fig. 3a), components in the first three bins (within 2.5 μm) of PM_{10} decrease

301 dramatically. Meanwhile, since the simulated PMcoarse is too small, the PM_{10} variates are eventually
302 dominated by the adjustment of $PM_{2.5}$. In result, the assimilation of $PM_{2.5}$ causes a large negative bias
303 in the PM_{10} analysis (Fig. 3b). Correspondingly, compared to the NODA run, the PMcoarse in the PM1
304 experiment exhibit no significant changes (only slightly decrease) in the analysis (Figs. 4b and 4c) and
305 also in the forecast (Fig 4f).

306 To overcome this issue, several adjustments have been adapted in the PM_{10} assimilation: instead
307 of using the PM_{10} observations directly, the PMcoarse is used to analyze the species in the 4th bin (Eq.
308 4); to reflect the large uncertainty of the simulated PMcoarse and to appropriately weighting the model
309 and observation errors, the background error covariance of the PMcoarse (species in the 4th bin) is
310 arbitrarily inflated (inflation factor 1 is normally used and 90 is selected after tuning). By this means,
311 after assimilating the PM_{10} observations, the PM2 and ALL experiments exhibit similar distributions
312 in the PMcoarse (Figs. 4d-e, with an average of 34.58 $\mu g/m^3$ and 34.68 $\mu g/m^3$) as in the observation
313 (with an average of 39.13 $\mu g/m^3$). Correspondingly, compared to the NODA experiment, evident
314 improvements for PM_{10} analysis appear in the PM2 and ALL experiments, in which the bias and RMSE
315 drops evidently (Fig. 3b). Overall, the DA experiments exhibit strong contributions to the analyses of
316 $PM_{2.5}$ and PM_{10}, suggesting that the WRFDA-Chem system works effectively in updating the initial
317 conditions.

318 As for the analyses of gaseous pollutants (Figs. 3c-3f), large improvements can be seen in the ALL
319 experiment by further assimilating SO_2, NO_2, O_3, and CO. Compared to the PM2 experiment, although
320 the bias and RMSE for $PM_{2.5}$ and PM_{10} in the ALL experiment are slightly larger, the bias for the four
321 gaseous pollutants decrease from 4.74 $\mu g/m^3$, -4.59 $\mu g/m^3$, 4.92 $\mu g/m^3$, and -8.31 mg/m^3 (PM2
322 experiment) to -1.68 $\mu g/m^3$, -1.25 $\mu g/m^3$, -0.31 $\mu g/m^3$, and -0.18 mg/m^3 (ALL experiment),

323 respectively, and the corresponding RMSE drops from 37.87 μg/m^3, 15.39 μg/m^3, 21.04 μg/m^3, and
324 1.11 mg/m^3 (PM2 experiment) to 23.85 μg/m^3, 9.70 μg/m^3, 8.62 μg/m^3, and 0.43 mg/m^3 (ALL
325 experiment). In general, by assimilating all the six major pollutants, the ALL experiment displays the
326 largest improvement in the analyses of gaseous pollutants among all the experiments, along with a
327 comparable improvement in the analyses of the aerosols.

328 Due to the lack of vertical information within the observations, the common mathematical solution
329 to use the surface total mass observations to analyze multiple 3-D fields variables is to utilize prior
330 information in the background. As shown in Fig. 5, based on vertical correlations specified in the
331 background error covariance, the observation impact spreads to a certain height, even though the
332 analysis variables used in the observation operator (Eq. 2-5) are only at the lowest model level. It is
333 also noted that observations contribute differently to the analysis variables. Corresponding to the
334 strong overestimation of $PM_{2.5}$ (Fig. 3a), all the three DA experiments (PM1, PM2 and ALL) tend to
335 reduce the $PM_{2.5}$ below 6 km; corresponding to the distinct underestimation for CO (Fig. 3f), the
336 experiment assimilating CO (ALL experiment) increases the value below 9 km. Relative small analysis
337 increments are shown in the other three gas pollutants (SO_2, NO_2, and O_3).

338 **3.2 Forecast improvements**

339 After illustrating the effect of WRFDA-Chem on the analyses, this section further investigates the
340 forecast performances based on the new analyses. A 24-h forecast is performed at each 00 UTC from
341 1 to 31 January 2017. The forecast error statistics, including bias, RMSE, and correlation, are computed
342 by verifying against the surface observations at 531 stations over China.

As shown in Fig. 6, the model performs relative poorly in the forecast of aerosols without DA. For $PM_{2.5}$, the average bias, RMSE, and correlation over 0-24 h are 31.17 $\mu g/m^3$, 88.99 $\mu g/m^3$, and 0.41, respectively (Tab. 3). As expected, all the DA experiments improve the forecasts evidently. Along with the forecast range, distinct improvements on bias, RMSE and correlation last from 0 to 24 h. For example, in PM1 experiment, the average improvement percentages (over 0-24 h) for bias, RMSE and correlation reach up to 71.8%, 39.4%, and 43.9%, respectively. It is also noted that $PM_{2.5}$ observation is the dominant data source in improving $PM_{2.5}$ forecast. As for PM_{10}, distinct improvements on RMSE and correlation can be seen from 0 to 24 h. Especially after assimilating the PMcoarse ($PM_{10-2.5}$ in PM2 and All experiments), the averaged improvement percentage for RMSE and correlation reach up to about 27.0 % and 55.5%. For bias, since the statistics are averaged over the 531 stations, the offset of large positive and negative bias at different stations leads to the small averaged bias in the NODA run (see the spatial distribution of bias at the individual site in Section 1 of the supplementary material). Considering the DA experiments exhibit distinct improvements on RMSE and correlation, WRFDA-Chem still provides a generally positive contribution to the PM_{10} forecast.

Figure 7 presents the averaged forecast error statistics for SO_2, NO_2, O_3, and CO with respect to forecast range. In PM1 and PM2 experiments that do not assimilate the gas-phase observations, no significant changes appear in the forecasts of the gaseous pollutants compared to the NODA run; after assimilating the gas-phase observations, the ALL experiment shows evident improvements in all the four gaseous pollutants, in which the improvements for SO_2, NO_2, and O_3 are more significant in 0-10 h, and the improvements for CO last up to 24 h. According to the numbers shown in Table 3, for SO_2, NO_2, O_3, and CO, the average bias (RMSE) in the ALL experiment decreases by 43.3%, 42.2%, 73.9%,

and 74.0% (13.4%, 5.3%, 11.3%, and 33.7%), compared to the NODA run, and the average correlation increases by 37.9%, 8.3%, 41.4%, and 103.5%, respectively. It is worth noting that the WRFDA-Chem system has a positive impact on the forecast of NO_2 and O_3 by merely analyzing the IC. Since NO_2 and O_3 are related to complex photochemical reaction processes, the assimilation of NO_2 and O_3 usually does not work well as other gas-phase pollutants on the forecast aspect, even with both emission and IC analyzed (Peng et al. 2018). In result, the aerosol/chemical assimilation based on WRFDA-Chem could not only contribute to the conventional aerosol forecasts in operational applications but also provide valuable help in the emerging study demands for gaseous pollutants, especially O_3.

Air Quality Index (AQI), which is used for reporting daily air quality and issuing alarms, is one of the service products of RMAPS-Chem operational air quality model over Northern China. Generally, AQI is classified into six levels rating from good to hazardous: 0-50 (level 1), 51-100 (level 2), 101-150 (level 3), 151-200 (level 4), 201-300 (level 5), and 300+ (Level 6). Similar to previous studies (Kumar and Goyal 2011; Tao et al. 2015; Zheng et al. 2014), AQI is calculated for the six major pollutants. The pollutant with the highest AQI level is deemed as the "main pollutant" and its AQI determines the overall AQI level. Accordingly, the accurate forecast of AQI requires the overall good performances of the six pollutants. To reflect the integrated DA effect of aerosols and gas-phase pollutants, the threat score (TS), one of the most commonly used criterions in the verifications of meteorology forecast, is used for AQI at each AQI level. The threat score (TS) for air quality index (AQI) is calculated by

$$TS_i = \frac{H_i}{H_i+M_i+F_i} \quad (6)$$

where H, M, and F denotes the times of the hits, the misses, and the false alarms in the forecast of AQI, and i denotes the AQI levels from 1 to 6. In result, the TS is acquired at each AQI level ranging from 0 to 1, and the higher (lower) TS represents the better (worse) forecast performance.

As shown in Fig. 8, in the beginning of the forecast, DA experiments (PM1, PM2 and ALL) increase the TS remarkably at all AQI levels, and then gradually decrease (quickly drop) with the forecast range at AQI levels 2-6 (AQI level 1). Nevertheless, for the polluted situations with AQI levels 3-6, evident improvements can be seen from 0 to 24h in all the DA experiments, in which the average TS increase from 0.19, 0.09, 0.16, and 0.19 (NODA experiment) to about 0.27, 0.16, 0.27, and 0.26 (DA experiments), respectively. For heavily polluted situations with AQI levels 5-6 (Figs. 8e-f), compared to the PM1 case, TS experiences a further increase in the PM2 and ALL experiments after assimilating the PMcoarse ($PM_{10-2.5}$). This result indicates that for heavily polluted events during this period (January 2017), $PM_{2.5}$ and PM_{10} could be the "main pollutant" that contributes the most to the AQI.

In general, the new WRFDA-Chem evidently improves the aerosol/chemical forecasting. Based on the assimilation of the six major pollutants, the chemical ICs are improved distinctly and a better forecast performance is acquired up to 24 hours. Among different experiments, the ALL experiment displays the best forecast error statistics for most of the major pollutants along with the highest TS for AQI. In the following operational applications, it is recommended to assimilate the six major pollutants simultaneously, which will help to get better analyses and forecast skills on the whole.

405 **3.3 Response to DA cycling frequency**

406 Cycling frequency is an important aspect in the DA strategy. However, the responses toward IC
407 updating could be different among the pollutants. To figure out this issue and to provide helpful
408 references for future applications, DA experiments with different cycling frequencies were analyzed
409 in this section.

410 Figure 9 shows the domain-averaged bias and RMSE of the analysis as in Fig. 3, but for
411 experiments with different DA frequencies (ALL_6h, ALL_3h, and ALL_1h; the ALL_6h is the ALL
412 experiment in Tab.2). Except for O_3, most of the variables display a gradual improvement with the
413 increase of cycling frequency. For example, from NODA run to the 6-h cycling experiment, and then
414 to the 3-h and 1-h cycling experiment, the bias (RMSE) for $PM_{2.5}$ gradually decrease from 36.60 μg/m^3
415 (70.41 μg/m^3) to 5.98 μg/m^3 (23.15 μg/m^3), and then to 5.41 μg/m^3 (21.32 μg/m^3) and 4.30 μg/m^3
416 (18.54 μg/m^3). Similar results also exist in the bias for SO_2, NO_2, and CO, as well as the RMSE for
417 PM_{10}, SO_2, and CO. In accordance with the gradual improvements in the analyses, the forecast skills
418 increase with the cycling frequency in most of the variables except O_3 (Figs. 10-11). Especially for the
419 forecasts of aerosols, evident gradual improvements can be seen from 0 to 24 h. From the 6-h cycling
420 experiment to the 3-h and the 1-h cycling experiment, the averaged decrease percentage of RMSE for
421 $PM_{2.5}$ (PM_{10}) enlarges from 38.76% to 41.27% and 44.21% (27.31% to 30.17% and 32.97%); the
422 averaged increase percentage of correlation for $PM_{2.5}$ (PM_{10}) enlarges from 42.82% to 49.51% and
423 55.58% (57.71% to 66.39% and 74.89%). To further investigate the integrated DA effect of aerosols
424 and gas phase pollutants under different cycling frequency, the TS for AQI is shown in Fig. 12. The
425 forecast of air quality is improved step by step with the increase of cycling frequency. On AQI levels
426 2-6, the TS for the ALL_1h experiment situates above the ALL_3h experiment at most of the time,

427 and followed by the ALL_6h experiment. These results indicate that the frequent IC updating is helpful
428 to further improve the forecast for most of the pollutants.

429 However, the analysis and 24-hr forecast of O_3 become worse under higher cycling frequencies
430 for this winter season (Fig. 9e and 11c). Given the analysis is at 00 UTC, the worsen analysis in the
431 experiments with higher DA frequencies (1-h, 3-h) could be mainly due to the unfavorable changes in
432 the 1-h/3-h forecasts period (starting from 23 UTC, 21 UTC), which is different from the situation in
433 the 6-h cycling experiment. As for the forecasts, the 24-hr performances starting from 00 UTC show
434 complex changes along with the forecast range: compared to the 6-h cycling experiment, the biases in
435 the experiments with higher DA frequencies decrease at 09-14 UTC but increase for other hours; the
436 RMSE and correlations in the experiments with higher DA frequencies become worse in most of the
437 hours (Fig. 11c). It should be mentioned that O_3 is a relatively short-lived chemical reactive species,
438 and takes part in highly complex and photochemical reactions in association with NO_x and VOC (Peng
439 et al. 2018, Lu et al., 2019). From this perspective, the performances of O_3 could also rely on the
440 photochemistry and the NO_x titration, in addition to the IC. Although the winter month (January 2017)
441 is investigated here when ozone photochemistry is relatively weaker compared to other seasons, the
442 photochemistry and the NO_x titration still play their roles. Accordingly, when the assimilation of NO_2
443 changes the NO_2 concentration and leave the NO and VOC unadjusted due to the absence of NO and
444 VOC measurements, two results might occur: firstly, the NO_2/VOC ratio which determine the
445 photochemical reactions and even the regime might be changed (O_3 production/loss direction might
446 change); secondly, the NO_x titration process might be changed due to the NO_2 concentration updates
447 (but no change on NO). Considering the relevant NO_x-VOC-O_3 reactions take place quickly, changing
448 the O_3 concentration in a short period, the advantage of IC DA could compete with the disadvantages

449 of the disordered photochemistry (inaccurate NO_2/VOC ratios) or the changed titration (adjusted NO_2
450 concentrations but not NO) resulting from the DA. Under this circumstance, the more frequent the O_3
451 and NO_2 were assimilated, the more incompatibilities could be brought into the related
452 photochemical/titration reactions, resulting the model performs worse in the O_3 forecasts under higher
453 cycling frequencies. It is noted that these statistics were only for the analysis at 00UTC and the 24-hr
454 forecast starting from 00UTC for winter season. Since O_3 has strong diurnal and seasonal variations,
455 more experiments and statistics at different time of the day and different season of the year should be
456 conducted in the future.

457 According to the results above, it is better to assimilate $PM_{2.5}$, PM_{10}, SO_2, and CO every 1 h and
458 assimilate O_3 and NO_2 every 6 h in the future applications, given the fact that the 6-h cycling
459 experiment performs the best in the O_3 forecasting (Fig. 11c) and displays no significant differences
460 in the NO_2 forecasting with experiments under higher cycling frequencies (Fig. 11b). It could also be
461 helpful to assimilate the VOC along with O_3 and NO_2 after there are corresponding observations.

462 **4. Indications on further model development**

463 A higher forecast skill relies on not only better working of DA, but also better performance of the
464 forecast model. To further improve the forecast skill, a crucial task is to understand the deficiencies in
465 the model, while the challenge in chemistry model diagnostic is that uncertainties are from various
466 aspects and are mixed-up in the model simulations, and the situation becomes even more complex
467 when the reaction path is not yet revealed by laboratory. However, with the help of DA, as one aspect
468 (IC) in the model is corrected by using observation as constraints, the deficiencies from other aspects
469 (e.g. chemical reactions) could be more evident, and thus there could be a better chance to diagnose

the deficiencies in the model. Specifically, Sulfate-nitrate-ammonium (SNA) are the predominant inorganic aerosol species that contribute up to 50% of total $PM_{2.5}$ in heavily polluted events in northern China (Wang et al. 2014). In addition to the normal pathways in the MOSAIC scheme, we added SO_2-NO_2-NO_3 related heterogeneous reactions for high relative humidity case in WRF-Chem (Chen et al. 2016), which greatly improved the underestimated SNA simulations. Since the newly added reactions are calculated on both the concentration of precursors (SO_2, NO_2-NO_3) and the uptake coefficients in the model, after DA corrected the concentrations of the precursors (one aspect), the impacts of the uptake coefficients could be more evident (the other aspect than the one corrected). Ideally, if the newly added reactions depict the heterogeneous reaction processes properly, a forecast improvement on the aerosols could be expected by assimilating their gaseous precursors. Based on this notion, this section verifies the forecast of two specific aerosol species, sulfate (SO_4^{2-}) and nitrate (NO_3^-), against a size-resolved particle observation over Beijing IUM station (in view of the assimilated SO_2 and NO_2 are the corresponding gaseous precursors of these aerosol species), aiming to explore the deficiencies in the uptake coefficients in the newly added heterogeneous reactions, and to provide beneficial indications for future model development.

Figure 13 presents the time series of sulfate and nitrate over Beijing IUM station. In the ALL experiment, after assimilating both the PM concentrations and the gaseous precursors (SO_2, NO_2), the forecasts of sulfate and nitrate become even worse than the PM2 experiment which only assimilates the PM concentrations. In the ALL experiment, sulfate experiences a decrease, accompanied by the average RMSE grows from 4.32 to 4.88 $\mu g/m^3$; nitrate exhibits an increase, accompanied by the average RMSE grows from 8.74 to 10.12 $\mu g/m^3$. However, compared to the PM2 experiment, the precursors (SO_2 and NO_2) are indeed improved. Figure 14 displays the analysis statistics of SO_2 and

492 NO_2 in the ALL experiment around Beijing area (red dots in Fig. 1) on January 16, the period with the
493 largest changes of sulfate and nitrate (Fig. 13). To correct the overestimated SO_2 (underestimated NO_2)
494 in the background, the DA in reduces (enhances) the model value in the ALL experiment, making it
495 closer to the observations.

496 It should be mentioned that the heterogeneous reactions are added by using the sulfate-nitrate-
497 ammonium observations as constraints to tune the "observation-best-matched" uptake coefficients
498 under the scenario without DA, in which the precursor concentrations are from pure model thus not
499 very accurate. To best match the observation, when gaseous precursors are overestimated
500 (underestimated) in the model, the uptake coefficient is tuned to low-biased (high-biased) value. In
501 result, such a coefficient may no longer be suited for the cases with DA. For instance, after DA
502 reducing the overestimated SO_2, the uptake coefficient is still relatively low and thus the reaction from
503 SO_2 to sulfate will stay at a low rate (with both low value of SO_2 and low reaction coefficient). A
504 similar result also exists for the reaction from NO_2 to nitrate. From this perspective, the negative effects
505 on sulfate and nitrate in the ALL experiment may not be hard to understand (Fig. 13). Therefore, in
506 the future chemistry development, it is necessary to develop more appropriate coefficients for different
507 gaseous precursor scenarios, in which more constraints, such as precursor and species concentrations,
508 should be provided with the help of DA technique. Accordingly, further improvements on aerosol
509 forecast could be expected by assimilating their gaseous precursors.

510 According to the results above, the DA technique provides an opportunity to identify and diagnose
511 the deficiencies in the model. By correcting the precursor concentrations through DA (one aspect), the
512 deficiency of the uptake coefficients for the SNA heterogeneous reactions (the other aspect than the
513 one corrected) is revealed. In the future, besides being used to improve the forecast skill through

updating the IC, DA could be used as another approach to reveal the necessary developments in the model.

5. Conclusions and discussions

To improve the operational air quality forecasting over China, a flexible aerosol and gas phase pollutants assimilation capability that can switch between different aerosol schemes is developed based on the WRFDA system with 3DVAR algorithm. This flexibility is designed to address the complexity of current aerosol schemes and to facilitate future chemistry developments. In this first application, the assimilation capability of surface observations of six major pollutants, including $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO, is built with MOSAIC aerosol scheme.

Before application in the operational air quality model, capability of the WRFDA-Chem system is verified in terms of analysis and forecast performances. Using the updated system, five DA experiments (assimilate different combinations of pollutants in various frequencies) were conducted for January 2017, along with a control experiment without DA. Results exhibit that the WRFDA-Chem system evidently improves the forecast of aerosols and gas phase pollutants. On the aspect of analysis, the assimilation of different atmospheric-composition observation reduces the bias and RMSE in the IC remarkably (e.g. by about 68%, 61%, and 30-60% in the RMSE for $PM_{2.5}$, PM_{10}, and gas phase pollutants); on the aspect of forecast skill, better performances are acquired up to 24 hours with about 10-40% (30-50%) improvements in the RMSE (correlation) for different pollutants. Among different experiments, the one assimilating all the six pollutants displays the best forecast error statistics for most of the pollutants along with the highest TS for AQI. In future applications, to get a better analysis and forecast skill in general, it is recommended to assimilate the six major pollutants simultaneously.

As the cycling frequency is an important aspect in the DA strategy, DA experiments with various cycling frequencies are also analyzed. Results exhibit that the responses toward IC updating are different among the pollutants. For $PM_{2.5}$, PM_{10}, SO_2, and CO, the forecast skills increase with the DA frequency; for O_3, compared to a better performance at the 6-h cycling frequency, its analysis at 00 UTC and the following 24-hr forecast become generally worse under higher cycling frequencies for this winter season, although the biases did decrease at 09-14 UTC in the 24-hr forecast. Considering the relevant NO_x-VOC-O_3 reaction system changes the NO_2/O_3 concentration in a short period, the advantage of IC DA could compete with the disadvantages of the disordered photochemistry (inaccurate NO_2/VOC ratios) or the changed titration (adjusted NO_2 concentrations but not NO) resulting from the DA. In future applications, it is better to assimilate PM2.5, PM10, SO2, and CO every 1 h. For the frequency of O_3 and NO_2 assimilation, every 6 h is the best in this winter season in our study. Since O_3 has strong diurnal and seasonal variations, more experiments and statistics at different time of the day and different season of the year should be conducted in the future. Also, it might be helpful to assimilate NO/VOC simultaneously with O_3 and NO_2 after there are corresponding measurements.

By investigating the effect of assimilating gaseous precursors on the forecast of related aerosols, the deficiencies in the WRF-Chem model are further revealed. The uptake coefficients for Sulfate-Nitrate-Ammonium heterogeneous reactions in the model are found out to be not appropriate in the applications with gaseous precursors (SO_2 and NO_2) assimilations, since they were originally tuned under the gaseous precursor scenarios without DA and the biases from the two aspects (SNA reactions and IC DA) were just compensated. In the future chemistry development, it is necessary to develop

appropriate coefficients for different gaseous precursor scenarios, in which more constraints, such as precursor and species concentrations, should be provided with the help of DA technique.

As for the significantly underestimated PMcoarse in the model, the results might relate to the missing emissions under current situations. Different from the United states or European countries that national emission inventories are provided and updated frequently by the government (e.g. US National Emission Inventory NEI 05-08-11-14-17), the publicly available emission inventories for China are mainly established by several scientific research groups. In result, the uncertainties of the publicly available emission inventories in China are relatively larger compared with others (US, European countries), and it's a known problem that the fugitive dust emissions over the whole of China is still lack, which might cause the underestimated PMcoarse simulation in the model.

Contributed by the flexible aerosol assimilation capability of the WRFDA-Chem system, development for other aerosol schemes targeting different regions in Asia is undergoing. In the next step, a study will focus on assimilating chemical observations from different observing platforms, such as satellite AOD observations, which contain more information over the areas with sparse surface observations. In addition, more advanced DA techniques, such as 4DVAR and Hybrid DA, could be taken into consideration in further developing the aerosol/chemical DA system.

Code and data availability

The data used in the figures and the developed WRFDA-Chem codes are available from WS upon request.

Author contributions

WS and ZL conducted development of DA system. ZL, DC, WS, and MC designed research, WS performed experiments and analyzed results, PZ provided PM species observations, and WS and DC wrote the paper with contributions from all co-authors.

Acknowledgement

This work was supported by the National Key R&D Program on Monitoring, Early Warning and Prevention of Major Natural Disasters under grant (2017YFC1501406), and Basic R&D special fund for central level, scientific research institutes (IUMKYSZHJ201701, IUMKY201807) of China. NCAR is sponsored by the US National Science Foundation.

Competing interests

The authors declare that they have no conflict of interest.

References

Bannister, R., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143,** 607-633.

Chen, D., Z. Liu, J. Fast, and J. Ban, 2016: Simulations of sulfate–nitrate–ammonium (SNA) aerosols during the extreme haze events over northern China in October 2014. *Atmospheric Chemistry and Physics*, **16,** 10707-10724.

592 Chen, D., Z. Liu, J. Ban, P. Zhao, and M. Chen, 2019: Retrospective analysis of 2015–2017 wintertime
593 PM 2.5 in China: response to emission regulations and the role of meteorology. *Atmospheric*
594 *Chemistry and Physics*, **19,** 7409-7427.
595 Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn
596 State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly*
597 *Weather Review*, **129,** 569-585.
598 Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use
599 in general circulation models.
600 Descombes, G., T. Auligné, F. Vandenberghe, D. Barker, and J. Barre, 2015: Generalized background
601 error covariance matrix model (GEN_BE v2. 0). *Geoscientific Model Development*, **8,** 669-696.
602 Fan, S., and Coauthors, 2016: Introduction of Rapid-refresh Multi-scale Analysis and Prediction
603 System-short time (RMAP-ST) over Northern China (in Chinese). *33th Annual Meeting of*
604 *Chinese Meteorological Society*.
605 Fast, J. D., and Coauthors, 2006: Evolution of ozone, particulates, and aerosol direct radiative forcing
606 in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. *Journal*
607 *of Geophysical Research: Atmospheres*, **111**.
608 Fu, Y., H. Liao, and Y. Yang, 2019: Interannual and decadal changes in tropospheric ozone in China
609 and the associated chemistry-climate interactions: A review. Advances in Atmospheric Sciences,
610 **36**, 975-993.
611 Grell, G. A., and D. Dévényi, 2002: A generalized approach to parameterizing convection combining
612 ensemble and data assimilation techniques. *Geophysical Research Letters*, **29,** 38-31-38-34.

613 Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder,
614 2005: Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, **39,**
615 6957-6975.
616 He, K., 2012: Multi-resolution emission Inventory for China (MEIC): model framework and 1990–
617 2010 anthropogenic emissions, presented on the international Global Atmospheric Chemistry
618 Conference, 17–21 September 2012, Beijing, China.
619 Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment
620 of entrainment processes. *Monthly weather review*, **134,** 2318-2341.
621 Huang, X., Y. Song, C. Zhao, M. Li, T. Zhu, Q. Zhang, and X. Zhang, 2014: Pathways of sulfate
622 enhancement by natural and anthropogenic mineral aerosols in China. *Journal of Geophysical*
623 *Research: Atmospheres*, **119,** 14,165-114,179.
624 Jiang, Z., Z. Liu, T. Wang, C. S. Schwartz, H. C. Lin, and F. Jiang, 2013: Probing into the impact of
625 3DVAR assimilation of surface PM10 observations over China using process analysis. *Journal*
626 *of Geophysical Research: Atmospheres*, **118,** 6738-6749.
627 Kumar, A., and P. Goyal, 2011: Forecasting of daily air quality index in Delhi. *Science of the Total*
628 *Environment*, **409,** 5517-5523.
629 Lei, Y., Q. Zhang, K. He, and D. Streets, 2011: Primary anthropogenic aerosol emission trends for
630 China, 1990–2005. *Atmospheric Chemistry and Physics*, **11,** 931-954.
631 Li, M., and Coauthors, 2014: Mapping Asian anthropogenic emissions of non-methane volatile organic
632 compounds to multiple chemical mechanisms. *Atmos. Chem. Phys*, **14,** 5617-5638.

633 Liu, Z., Q. Liu, H. C. Lin, C. S. Schwartz, Y. H. Lee, and T. Wang, 2011: Three-dimensional
634 variational assimilation of MODIS aerosol optical depth: Implementation and application to a
635 dust storm over East Asia. Journal of Geophysical Research: Atmospheres, 116 (D23).
636 Lu, X., and Coauthors, 2019: Exploring 2016–2017 surface ozone pollution over China: source
637 contributions and meteorological influences. Atmospheric Chemistry and Physics, **19**, 8339-8361.
638 McHenry, J. N., J. M. Vukovich, and N. C. Hsu, 2015: Development and implementation of a remote-
639 sensing and in situ data-assimilating version of CMAQ for operational PM2. 5 forecasting. Part
640 1: MODIS aerosol optical depth (AOD) data-assimilation design and testing. *Journal of the Air*
641 *& Waste Management Association*, **65,** 1395-1412.
642 McKeen, S., and Coauthors, 2009: An evaluation of real-time air quality forecasts and their urban
643 emissions over eastern Texas during the summer of 2006 Second Texas Air Quality Study field
644 study. *Journal of Geophysical Research: Atmospheres*, **114**.
645 Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer
646 for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave.
647 *Journal of Geophysical Research: Atmospheres*, **102,** 16663-16682.
648 Nie, W., and Coauthors, 2014: Polluted dust promotes new particle formation and growth. *Scientific*
649 *reports*, **4,** 6634.
650 Pang, J., Z. Liu, X. Wang, J. Bresch, J. Ban, D. Chen, and J. Kim, 2018: Assimilating AOD retrievals
651 from GOCI and VIIRS to forecast surface PM2. 5 episodes over Eastern China. *Atmospheric*
652 *environment*, **179,** 288-304.

653 Peng, Z., Z. Liu, D. Chen, and J. Ban, 2017: Improving PM 2. 5 forecast over China by the joint
654 adjustment of initial conditions and source emissions with an ensemble Kalman filter.
655 *Atmospheric Chemistry and Physics*, **17,** 4837-4855.
656 Peng, Z., and Coauthors, 2018: The impact of multi-species surface chemical observation assimilation
657 on air quality forecasts in China. *Atmospheric Chemistry and Physics*, **18,** 17387-17404.
658 Sandu, A., and T. Chai, 2011: Chemical data assimilation—An overview. *Atmosphere*, **2,** 426-463.
659 Schutgens, N., T. Miyoshi, T. Takemura, and T. Nakajima, 2010: Applying an ensemble Kalman filter
660 to the assimilation of AERONET observations in a global aerosol transport model. *Atmospheric*
661 *Chemistry and Physics*, **10,** 2561-2576.
662 Schwartz, C. S., Z. Liu, H. C. Lin, and S. A. McKeen, 2012: Simultaneous three - dimensional
663 variational assimilation of surface fine particulate matter and MODIS aerosol optical depth.
664 *Journal of Geophysical Research: Atmospheres*, **117**.
665 Sekiyama, T., T. Tanaka, A. Shimizu, and T. Miyoshi, 2010: Data assimilation of CALIPSO aerosol
666 observations. *Atmospheric Chemistry and Physics*, **10,** 39-49.
667 Su, J., P. Zhao, and Q. Dong, 2018: Chemical compositions and liquid water content of size-resolved
668 aerosol in Beijing. *Aerosol Air Qual. Res*, **18,** 680-692.
669 Tang, X., J. Zhu, Z. Wang, and A. Gbaguidi, 2011: Improvement of ozone forecast over Beijing based
670 on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions.
671 *Atmospheric Chemistry and Physics*, **11,** 12901-12916.
672 Tang, X., and Coauthors, 2013: Inversion of CO emissions over Beijing and its surrounding areas with
673 ensemble Kalman filter. *Atmospheric environment*, **81,** 676-686.

674 Tao, J., L. Zhang, J. Gao, H. Wang, F. Chai, and S. Wang, 2015: Aerosol chemical composition and
675 light scattering during a winter season in Beijing. *Atmospheric Environment*, **110,** 36-44.
676 Wang, G., and Coauthors, 2016: Persistent sulfate formation from London Fog to Chinese haze.
677 *Proceedings of the National Academy of Sciences*, **113,** 13630-13635.
678 Wang, L., and Coauthors, 2013: The 2013 severe haze over the southern Hebei, China: model
679 evaluation, source apportionment, and policy implications. *Atmospheric Chemistry & Physics*
680 *Discussions*, **13**.
681 Wang, Y., and Coauthors, 2014: Enhanced sulfate formation during China's severe winter haze episode
682 in January 2013 missing from current models. *Journal of Geophysical Research: Atmospheres*,
683 **119,** 10,425-410,440.
684 Wild, O., X. Zhu, and M. J. Prather, 2000: Fast-J: Accurate simulation of in-and below-cloud
685 photolysis in tropospheric chemical models. *Journal of Atmospheric Chemistry*, **37,** 245-282.
686 Xie, Y., and Coauthors, 2015: Enhanced sulfate formation by nitrogen dioxide: Implications from in
687 situ observations at the SORPES station. *Journal of Geophysical Research: Atmospheres*, **120,**
688 12679-12694.
689 Yu, M., S. Miao, and H. Zhang, 2018: Uncertainties in the Impact of Urbanization on Heavy Rainfall:
690 Case Study of a Rainfall Event in Beijing on 7 August 2015. *Journal of Geophysical Research:*
691 *Atmospheres*, **123,** 6005-6021.
692 Zaveri, R. A., and L. K. Peters, 1999: A new lumped structure photochemical mechanism for large-
693 scale applications. *Journal of Geophysical Research: Atmospheres*, **104,** 30387-30415.
694 Zaveri, R. A., R. C. Easter, J. D. Fast, and L. K. Peters, 2008: Model for simulating aerosol interactions
695 and chemistry (MOSAIC). *Journal of Geophysical Research: Atmospheres*, **113**.

696 Zhang, Q., and Coauthors, 2009: Asian emissions in 2006 for the NASA INTEX-B mission.
697 *Atmospheric Chemistry and Physics*, **9,** 5131-5153.
698 Zheng, B., and Coauthors, 2015: Heterogeneous chemistry: a mechanism missing in current models to
699 explain secondary inorganic aerosol formation during the January 2013 haze episode in North
700 China. *Atmospheric Chemistry and Physics (Online)*, **15**.
701 Zheng, S., C.-X. Cao, and R. P. Singh, 2014: Comparison of ground based indices (API and AQI) with
702 satellite based aerosol products. *Science of the Total Environment*, **488,** 398-412.

703 **Tables and Figures**

704 **Table 1.** WRF-Chem model configurations.

705 **Table 2.** The detail setting of six experiments and the purposes.

706 **Table 3.** Averaged bias (units: μg/m^3), RMSE (units: μg/m^3), and correlation over forecast hour 0-24
707 h for different variables and different experiments. The statistics for gas phase pollutants in PM1 and
708 PM2 experiments are highly close to the results in NODA experiment, and thus leave with blank in
709 the table.

710 **Figure 1.** Computation domain. Dots depict surface observations with 531 stations spreading over
711 China. The red dots indicate the observations around Beijing. The green dot indicates the IUM station.

712 **Figure 2**. Background error standard deviations of aerosol species in the (a) 1st size bin, (b) 2nd size
713 bin, (c) 3rd size bin, (d) 4th size bin, and of (e) gas pollutants. The units for the x-axis are μg m^{-3} for
714 (a)-(d) and ppm for (e). The left y-axis denotes the model level, and the right y-axis denotes the vertical
715 height (units: km).

Figure 3. Averaged bias (color bar, left y-axis) and RMSE (hallow bar, right y-axis) of the analysis at 00 UTC over January 1-31, 2017 for (a) $PM_{2.5}$, (b) PM_{10}, (c) SO_2, (d) NO_2, (e) O_3 and (f) CO in different experiments, verified against the surface observations of 531 stations in China. The blue, red, green and gray shaded bars denote the bias of the experiment NODA, PM1, PM2, ALL, respectively; the corresponding hallow bars denote the RMSE of these experiments. Units of the y-axis are $\mu g/m^3$ in Figs. 3a-e and mg/m^3 in Fig. 3f.

Figure 4. Averaged PMcoarse ($PM_{10-2.5}$, units: $\mu g/m^3$) at 00 UTC over January 1-31, 2017 in (a) observation and four experiments (b) NODA, (c) PM1, (d) PM2, (e) ALL, and (f) averaged bias (units: $\mu g/m^3$) for PMcoarse in different experiments as a function of forecast range (the blue, red, green and gray lines denote the results of experiment NODA, PM1, PM2, ALL, respectively), verified against the surface observations of 531 stations in China. The numbers on the top of each panel denote the average PMcoarse concentrations over 531 stations (units: $\mu g/m^3$).

Figure 5. Vertical profile of the analysis at 00 UTC over January 1-31, 2017 for (a) $PM_{2.5}$, (b) PM_{10}, (c) SO_2, (d) NO_2, (e) O_3, and (f) CO in different experiments, averaged over the 531 surface stations in China. The blue, red, green and gray lines denote the results of experiment NODA, PM1, PM2, and ALL, respectively. Units of the y-axis are $\mu g/m^3$ in Figs. 5a-e and mg/m^3 in Fig. 5f.

Figure 6. Averaged bias (units: $\mu g/m^3$), RMSE (units: $\mu g/m^3$), and correlation for (a) $PM_{2.5}$ and (b) PM_{10} in different experiments as a function of forecast range, verified against the surface observations of 531 stations in China. The blue, red, green and gray lines denote the results of experiment NODA, PM1, PM2, ALL, respectively.

Figure 7. Same as Fig. 6, but for the forecast of (a) SO_2, (b) NO_2, (c) O_3 (units: $\mu g/m^3$), and (d) CO (units: mg/m^3).

738 **Figure 8.** Averaged threat score (TS) for Air Quality Index (AQI) from AQI level 1 to level 6 (a-f) in
739 different experiments as a function of forecast range, verified against the surface observations of 531
740 stations in China. The blue, red, green and gray lines denote the results of experiment NODA, PM1,
741 PM2, and ALL, respectively. The numbers on the right of each panel denote the averaged TS from 0
742 to 24 h for different experiments.
743 **Figure 9.** Same as Fig. 3, but for the experiments of NODA, ALL_6h, ALL-3h, ALL_1h, respectively.
744 Units of the y-axis are $\mu g/m^3$ in Figs. 9a-e and mg/m^3 in Fig. 9f.
745 **Figure 10.** Averaged bias (units: $\mu g/m^3$), RMSE (units: $\mu g/m^3$), and correlation for (a) $PM_{2.5}$ and (b)
746 PM_{10} in different experiments as a function of forecast range, verified against the surface observations
747 of 531 stations in China. The blue, red, green and gray lines denote the results of experiment NODA,
748 ALL_6h, ALL_3h, and ALL_1h, respectively.
749 **Figure 11.** Same as Fig. 10, but for the forecast of (a) SO_2, (b) NO_2, (c) O_3 (units: $\mu g/m^3$), and (d) CO
750 (units: mg/m^3).
751 **Figure 12**. Same as Fig. 8, but for the experiments of NODA, ALL_6h, ALL-3h, ALL_1h, respectively.
752 **Figure 13**. Time series of (a) sulfate, (b) nitrate over January 14-20, verified against the size-resolved
753 particle observation at IUM station (units: $\mu g/m^3$). The gray, blue and red lines denote the observation
754 and the results of experiment PM2 and ALL, respectively. The numbers on the right of each panel
755 denote the averaged RMSE over January 14-20 for different experiments.
756 **Figure 14**. Averaged scatter plot of (a, c) observation versus background and (b, d) observation versus
757 analysis for (a, b) SO2 and (c, d) NO2 around Beijing area (red dots in Fig. 1) on January 16. The
758 numbers on the title denote the accumulated numbers of the used observations around Beijing area
759 during January 16 (1600 UTC, 1606 UTC, 1612 UTC, and 1618 UTC).

760 **Table 1.** WRF-Chem model configurations.

Aerosol scheme	MOSAIC (four bins, Zaveri et al. (2008))
Photolysis scheme	Fast-J (Wild et al. 2000)
Gas-phase chemistry	CBM-Z (Zaveri and Peters 1999)
Cumulus parameterization	Grell 3-D scheme
Short-wave radiation	Goddard Space Flight Center short-wave radiation scheme (Chou and Suarez 1994)
Long-wave radiation	RRTM (Mlawer et al. 1997)
Microphysics	Single-moment 6-class scheme (Grell and Dévényi 2002)
Land-surface model (LSM)	NOAH LSM (Chen and Dudhia 2001)
Boundary-layer scheme	YSU (Hong et al. 2006)
Meteorology initial and boundary conditions	GFS analysis and forecast every 6 h
Initial condition for chemical species	11-day spin-up
Boundary conditions for chemical species	Averages of mid-latitude aircraft profiles
Dust and sea salt emissions	GOCART

761

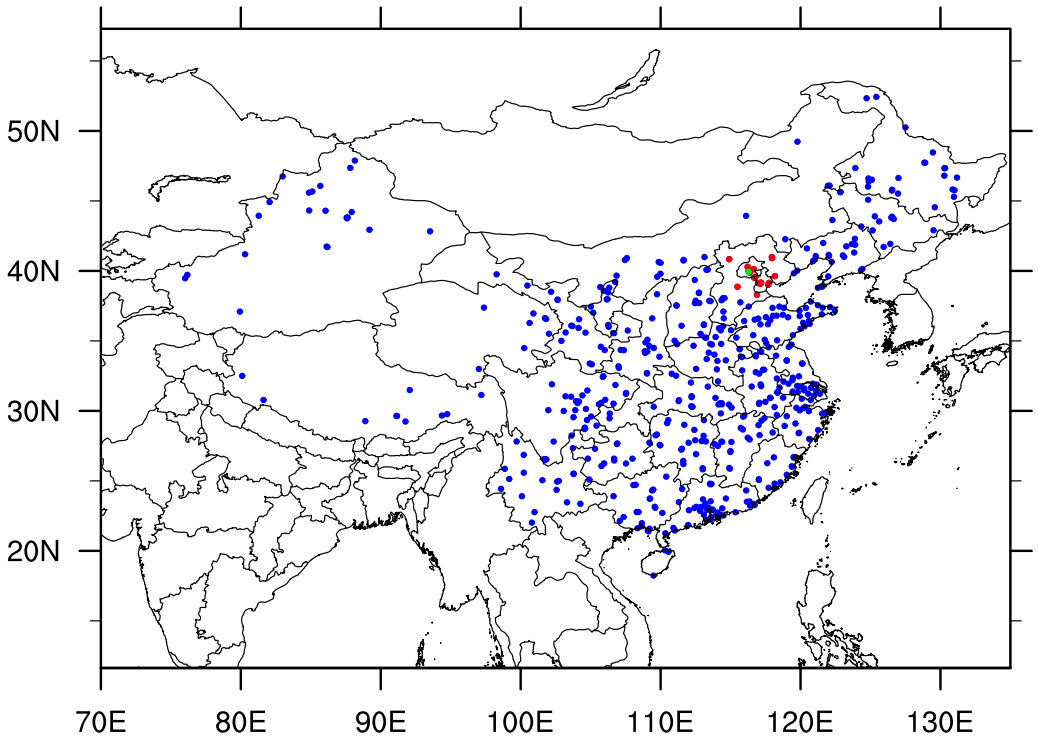
762 **Table 2.** The detail setting of six experiments and the purposes

Experiments	$PM_{2.5}$ assimilation	$PM_{10-2.5}$ assimilation	Gas phase (SO_2, NO_2, O_3, CO) assimilation	Assimilated time (UTC)	Purposes for forecast performances
NODA	No	No	No	--	Control simulation
PM1	Yes	No	No	00, 06, 12, 18	Basic $PM_{2.5}$ assimilation
PM2	Yes	Yes	No	00, 06, 12, 18	$PM_{2.5}$ and $PM_{10-2.5}$ assimilation
ALL	Yes	Yes	Yes	00, 06, 12, 18	Aerosol and precursor simultaneously assimilation
ALL_3h	Yes	Yes	Yes	00, 03, 06, 09, 12, 15, 18, 21	Different assimilation frequencies on forecast performances
ALL_1h	Yes	Yes	Yes	0-23, every hour	

763

Table 3. Averaged bias (units: μg/m^3), RMSE (units: μg/m^3), and correlation over forecast hour 0-24 h for different variables and different experiments. The statistics for gas phase pollutants in PM1 and PM2 experiments are highly close to the results in NODA experiment, and thus leave with blank in the table.

		NODA	PM1	PM2	ALL
$PM_{2.5}$	Bias	31.17	8.78	8.39	9.36
	RMSE	88.99	53.93	54.35	54.49
	Correlation	0.41	0.59	0.58	0.59
PM_{10}	Bias	-1.13	-22.73	-15.43	-14.41
	RMSE	98.5	74.41	71.9	71.6
	Correlation	0.36	0.54	0.56	0.56
SO_2	Bias	6.67	-	-	3.78
	RMSE	44.11	-	-	38.18
	Correlation	0.29	-	-	0.4
NO_2	Bias	-2.87	-	-	-1.66
	RMSE	25.61	-	-	24.26
	Correlation	0.48	-	-	0.52
O_3	Bias	-3.22	-	-	-0.84
	RMSE	31.96	-	-	28.36
	Correlation	0.29	-	-	0.41
CO	Bias	-0.73	-	-	-0.19
	RMSE	1.13	-	-	0.75
	Correlation	0.28	-	-	0.57

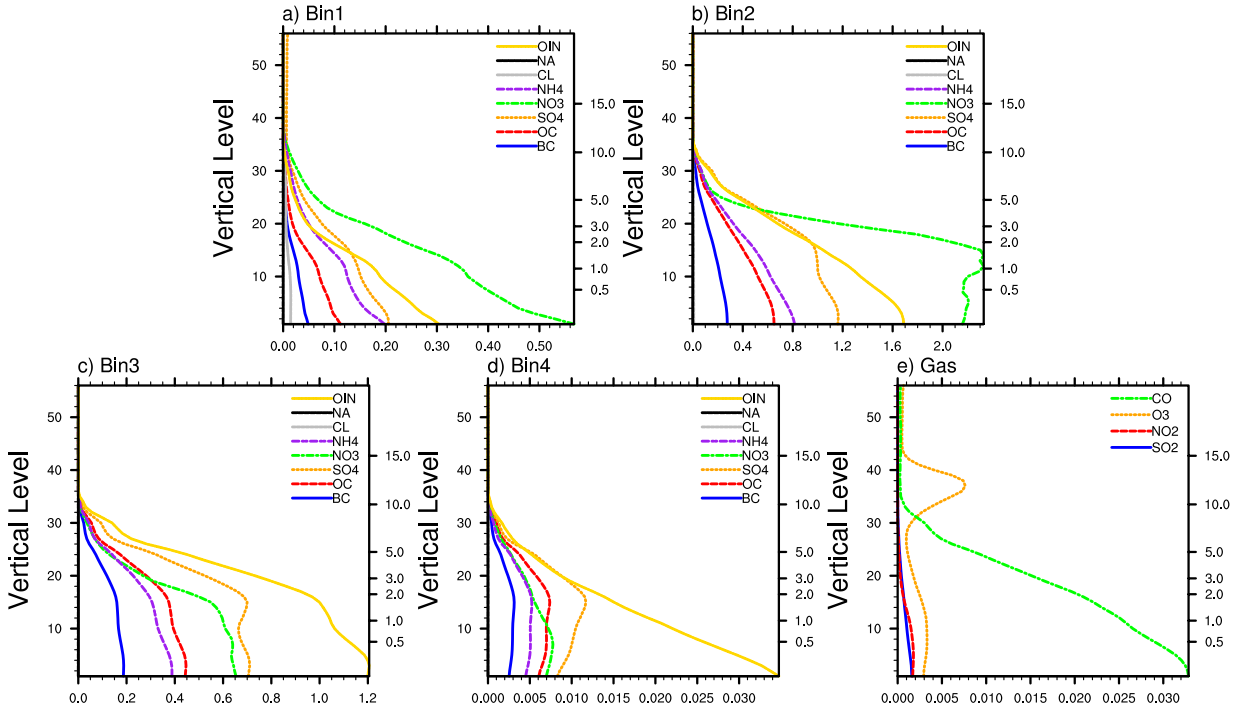


769

770 **Figure 1.** Computation domain. Dots depict surface observations with 531 stations spreading over

771 China. The red dots indicate the observations around Beijing. The green dot indicates the IUM station.

772



773

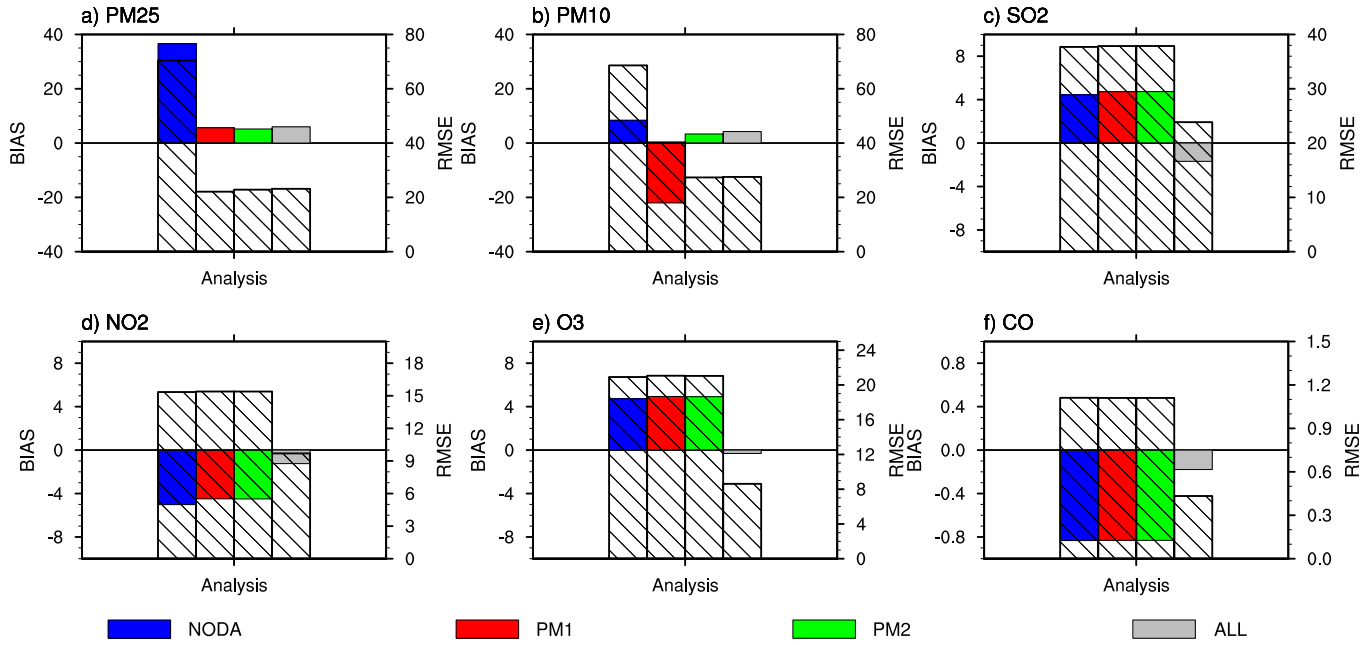
774 **Figure 2**. Background error standard deviations of aerosol species in the (a) 1st size bin, (b) 2nd size

775 bin, (c) 3rd size bin, (d) 4th size bin, and of (e) gas pollutants. The units for the x-axis are μg m^{-3} for

776 (a)-(d) and ppm for (e). The left y-axis denotes the model level, and the right y-axis denotes the vertical

777 height (units: km).

778



779

780 **Figure 3.** Averaged bias (color bar, left y-axis) and RMSE (hallow bar, right y-axis) of the analysis at

781 00 UTC over January 1-31, 2017 for (a) $PM_{2.5}$, (b) PM_{10}, (c) SO_2, (d) NO_2, (e) O_3 and (f) CO in different

782 experiments, verified against the surface observations of 531 stations in China. The blue, red, green

783 and gray shaded bars denote the bias of the experiment NODA, PM1, PM2, ALL, respectively; the

784 corresponding hallow bars denote the RMSE of these experiments. Units of the y-axis are $\mu g/m^3$ in

785 Figs. 3a-e and mg/m^3 in Fig. 3f.

786

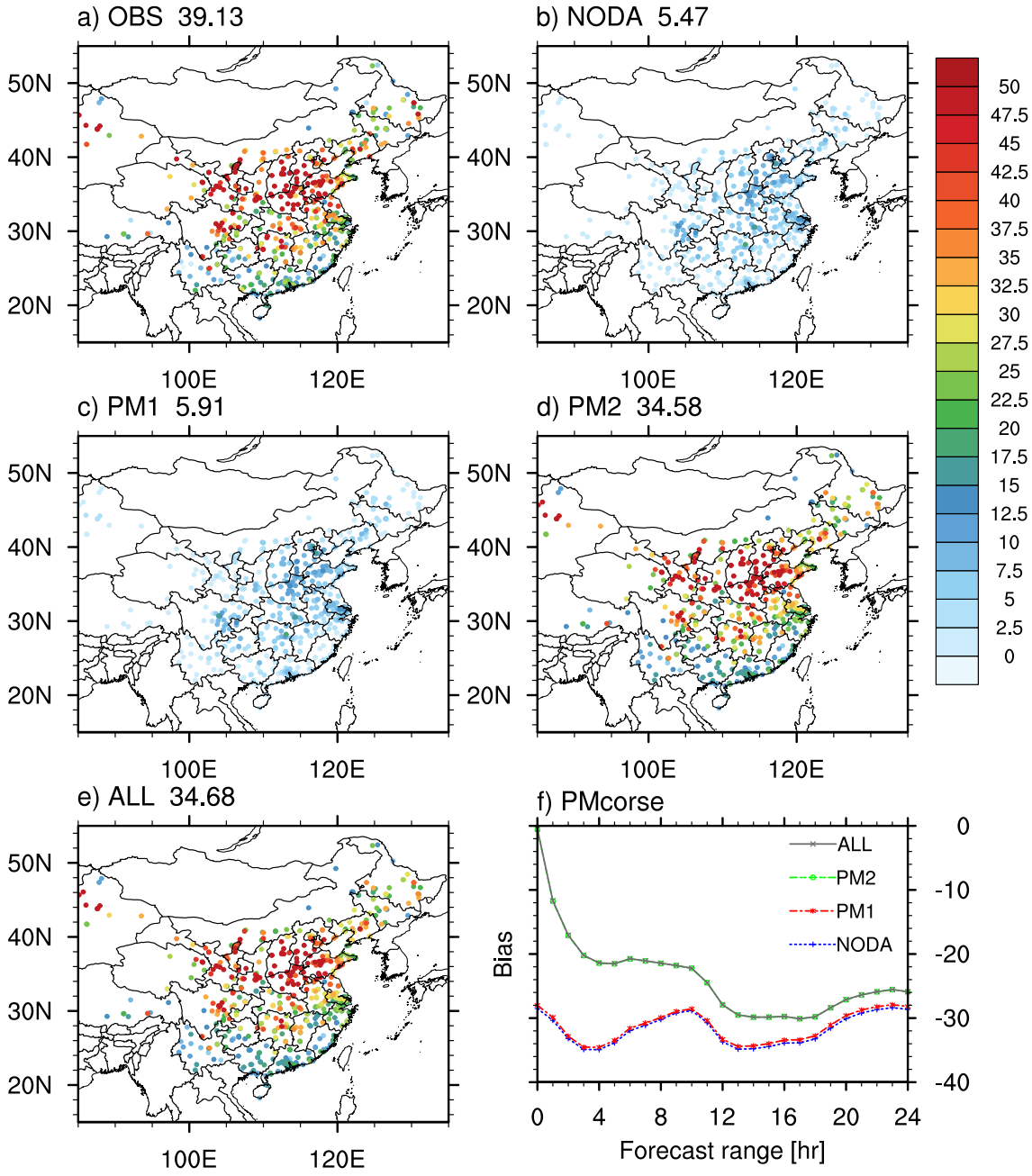
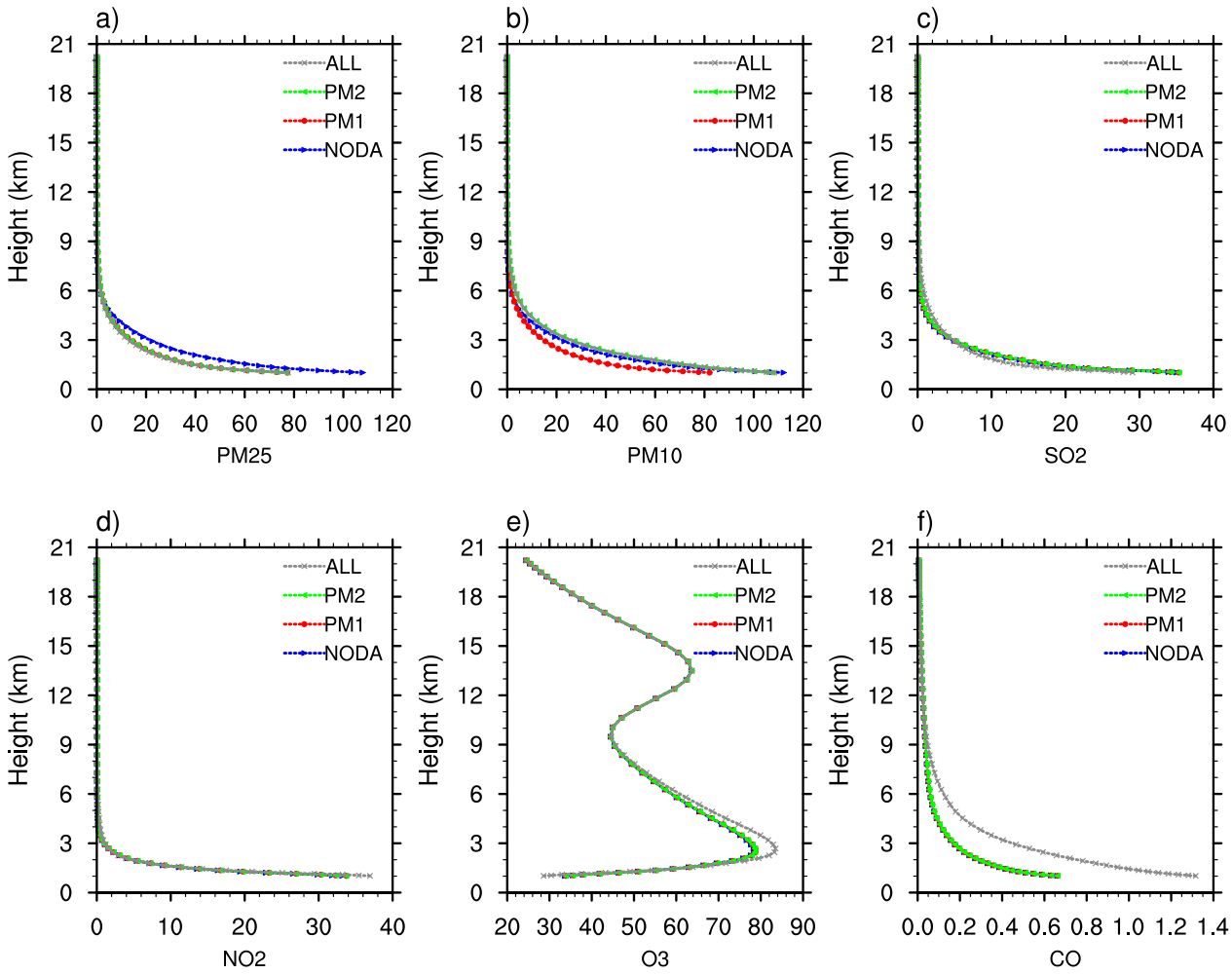


Figure 4. Averaged PMcoarse ($PM_{10-2.5}$, units: $\mu g/m^3$) at 00 UTC over January 1-31, 2017 in (a) observation and four experiments (b) NODA, (c) PM1, (d) PM2, (e) ALL, and (f) averaged bias (units: $\mu g/m^3$) for PMcoarse in different experiments as a function of forecast range (the blue, red, green and gray lines denote the results of experiment NODA, PM1, PM2, ALL, respectively), verified against the surface observations of 531 stations in China. The numbers on the top of each panel denote the average PMcoarse concentrations over 531 stations (units: $\mu g/m^3$).



794

795 **Figure 5.** Vertical profile of the analysis at 00 UTC over January 1-31, 2017 for (a) $PM_{2.5}$, (b) PM_{10},
796 (c) SO_2, (d) NO_2, (e) O_3, and (f) CO in different experiments, averaged over the 531 surface stations
797 in China. The blue, red, green and gray lines denote the results of experiment NODA, PM1, PM2, and
798 ALL, respectively. Units of the y-axis are $\mu g/m^3$ in Figs. 5a-e and mg/m^3 in Fig. 5f.

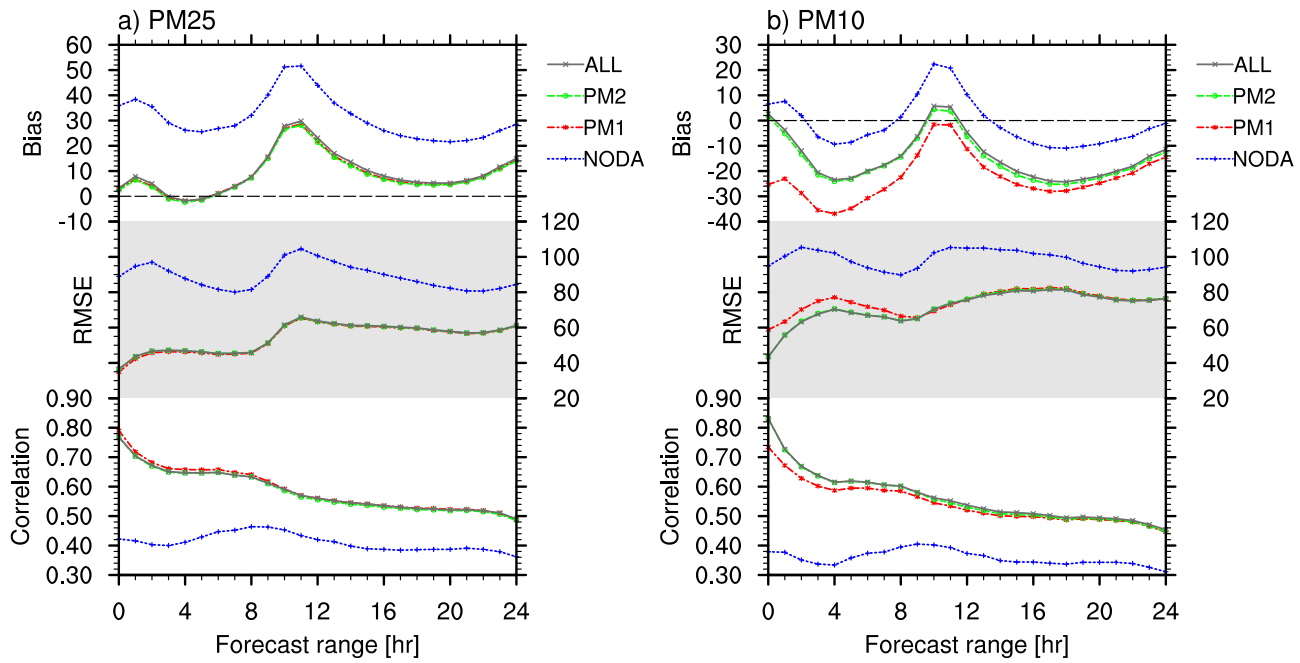
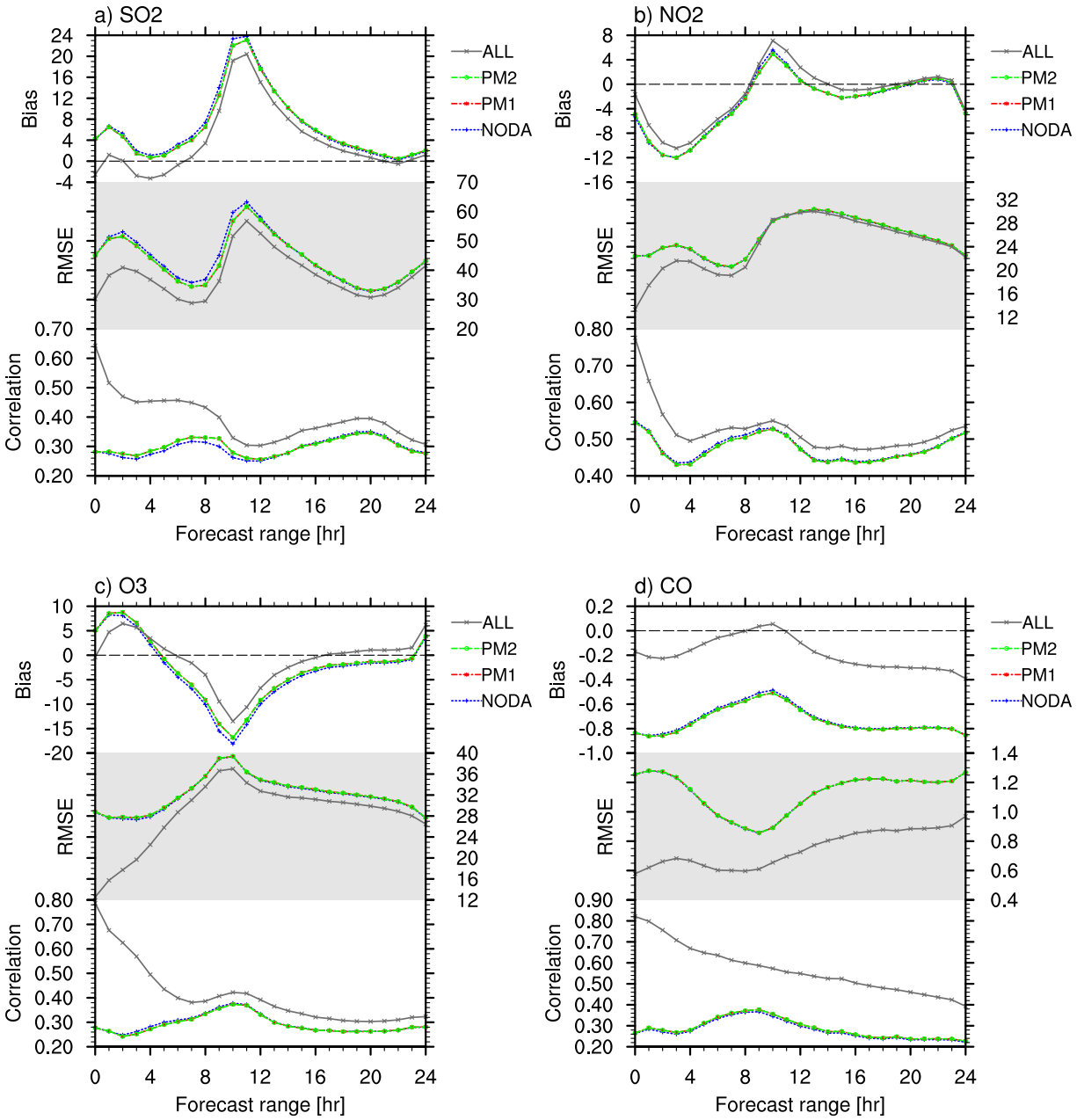


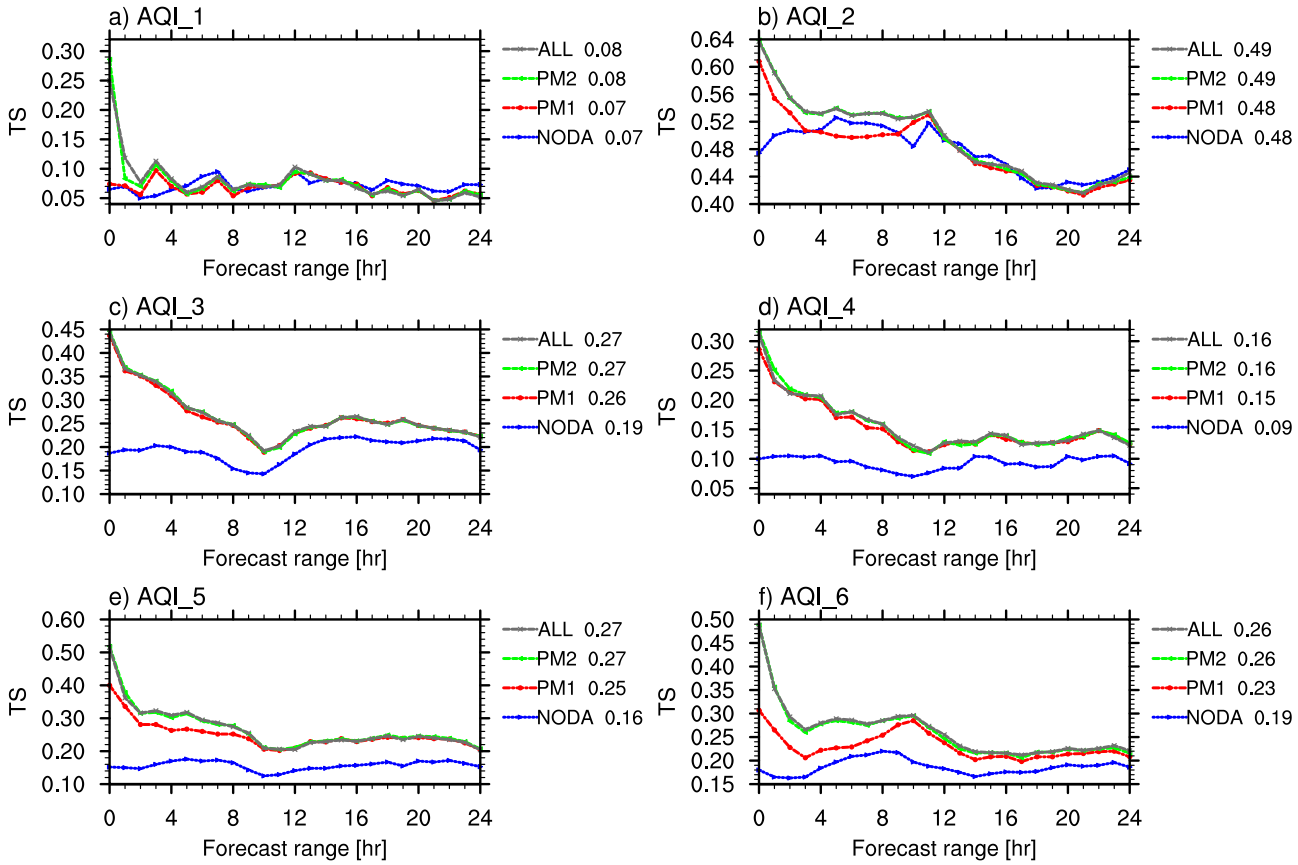
Figure 6. Averaged bias (units: μg/m^3), RMSE (units: μg/m^3), and correlation for (a) $PM_{2.5}$ and (b) PM_{10} in different experiments as a function of forecast range, verified against the surface observations of 531 stations in China. The blue, red, green and gray lines denote the results of experiment NODA, PM1, PM2, ALL, respectively.



804

805 **Figure 7.** Same as Fig. 6, but for the forecast of (a) SO_2, (b) NO_2, (c) O_3 (units: $\mu g/m^3$), and (d) CO

806 (units: mg/m^3).



807

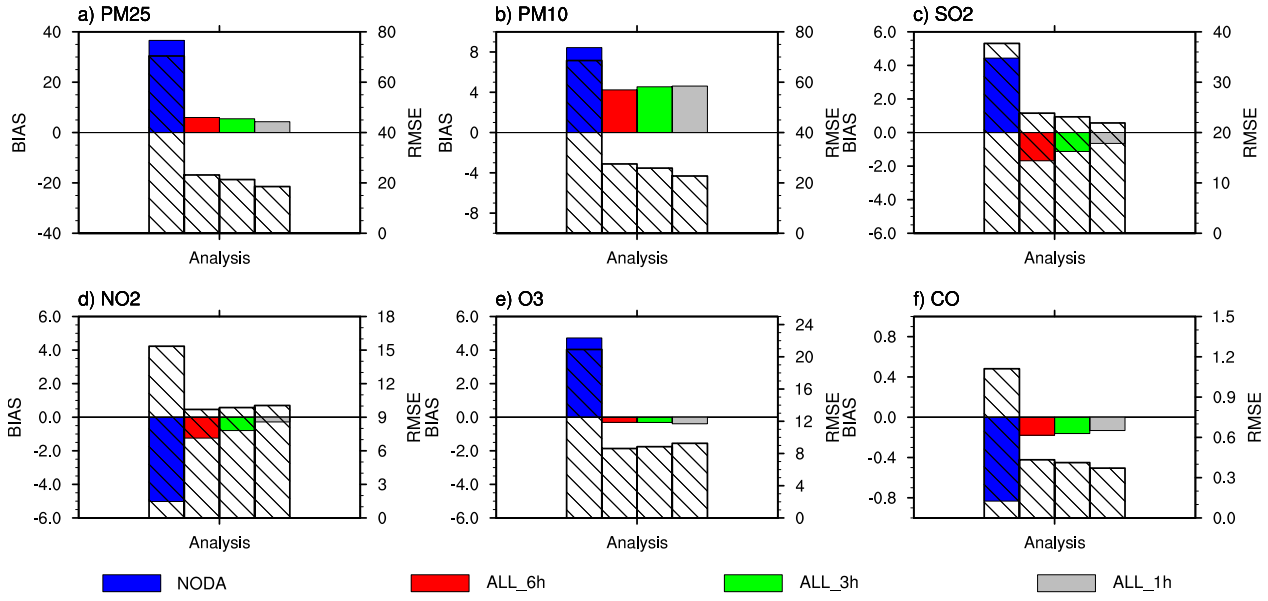
808 **Figure 8.** Averaged threat score (TS) for Air Quality Index (AQI) from AQI level 1 to level 6 (a-f) in

809 different experiments as a function of forecast range, verified against the surface observations of 531

810 stations in China. The blue, red, green and gray lines denote the results of experiment NODA, PM1,

811 PM2, and ALL, respectively. The numbers on the right of each panel denote the averaged TS from 0

812 to 24 h for different experiments.

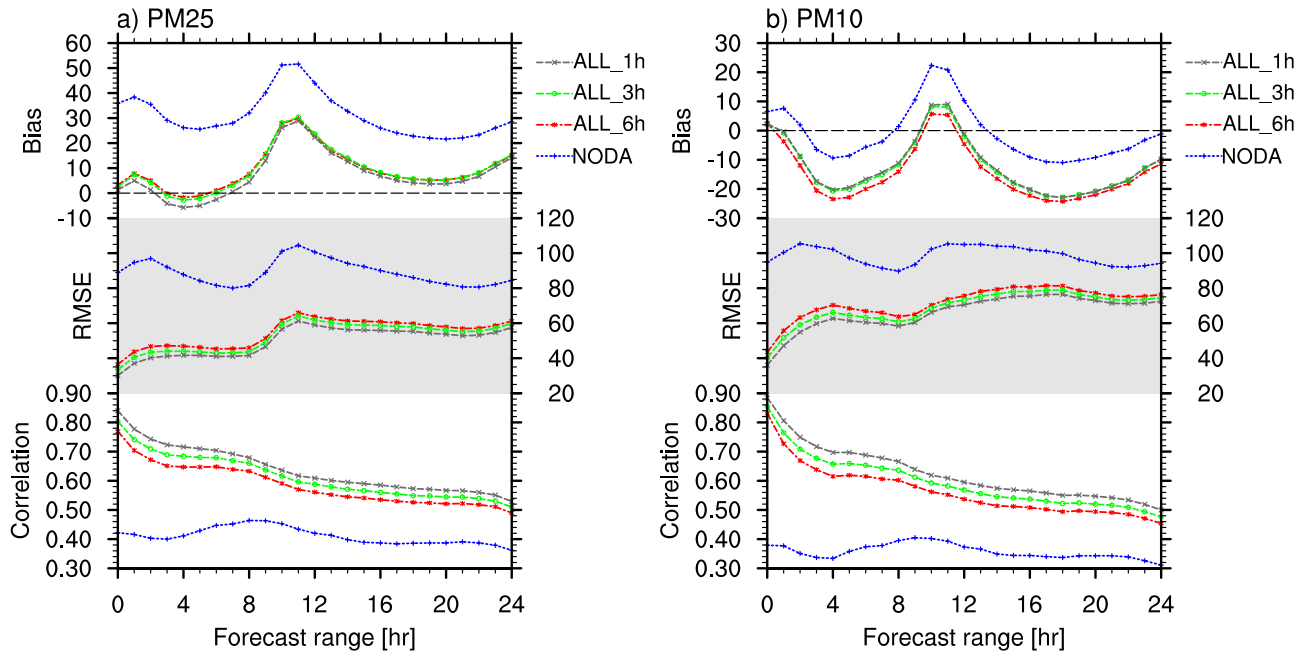


813

814 **Figure 9.** Same as Fig. 3, but for the experiments of NODA, ALL_6h, ALL-3h, ALL_1h, respectively.

815 Units of the y-axis are μg/m^3 in Figs. 9a-e and mg/m^3 in Fig. 9f.

816



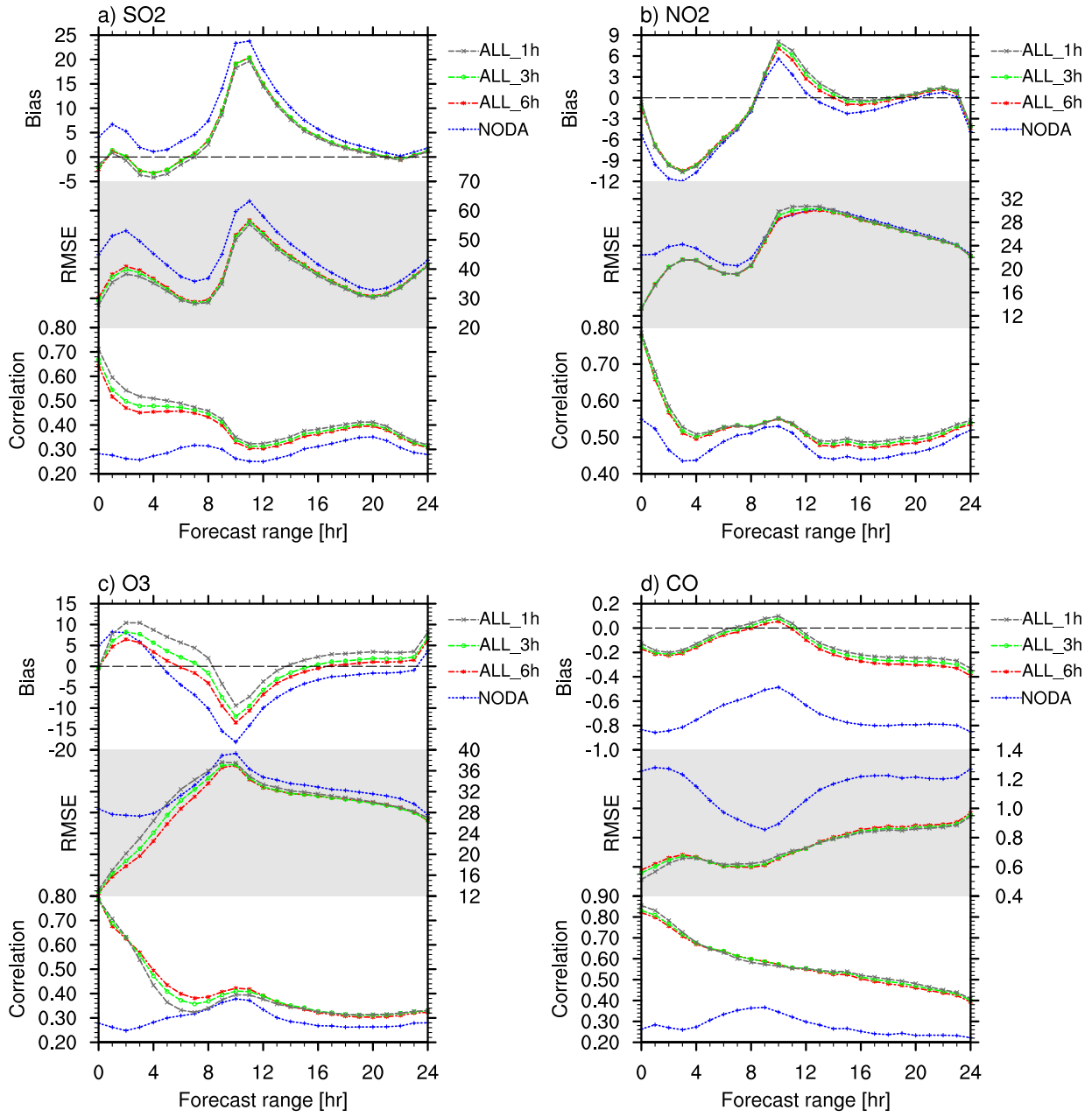
817

818 **Figure 10.** Averaged bias (units: μg/m^3), RMSE (units: μg/m^3), and correlation for (a) $PM_{2.5}$ and (b)

819 PM_{10} in different experiments as a function of forecast range, verified against the surface observations

820 of 531 stations in China. The blue, red, green and gray lines denote the results of experiment NODA,

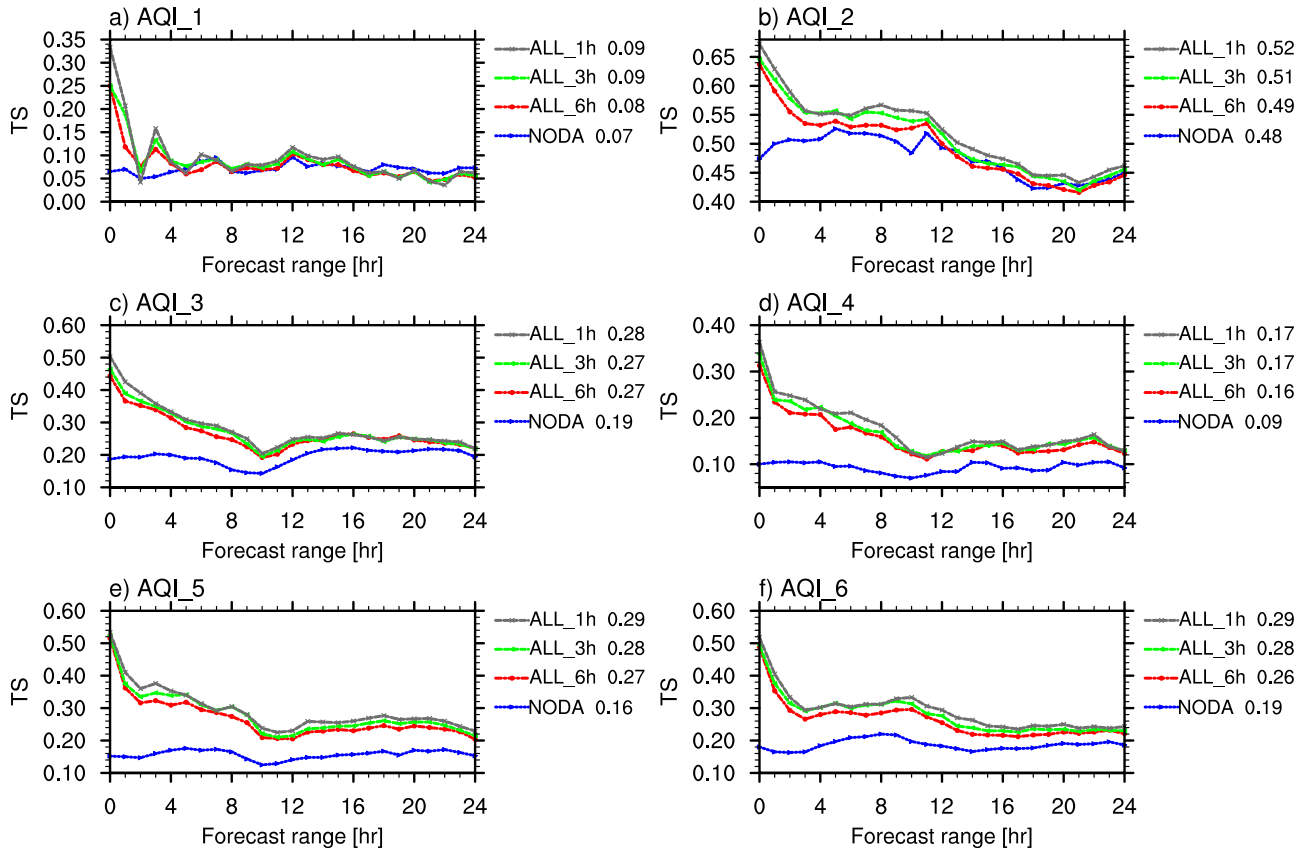
821 ALL_6h, ALL_3h, and ALL_1h, respectively.



822

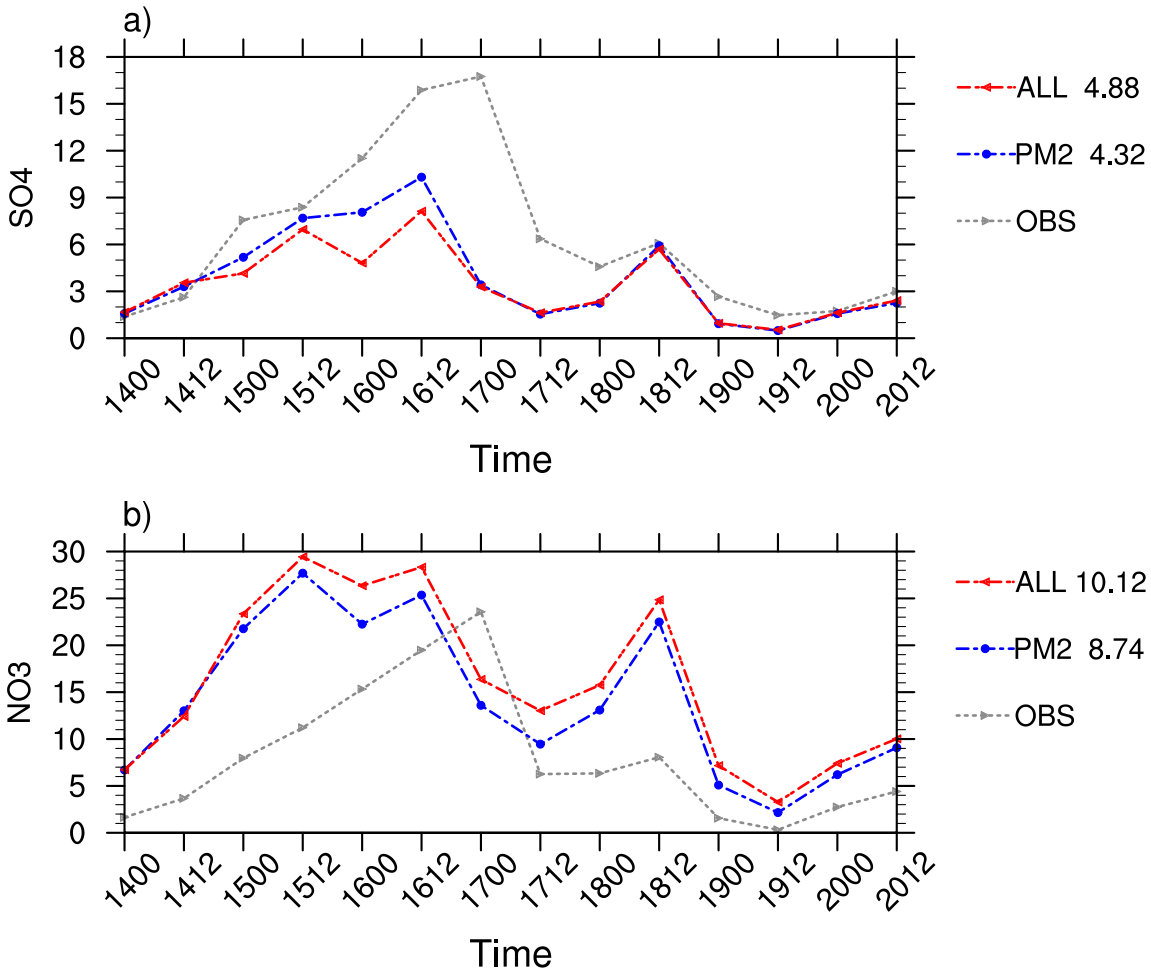
823 **Figure 11.** Same as Fig. 10, but for the forecast of (a) SO_2, (b) NO_2, (c) O_3 (units: μg/m^3), and (d) CO

824 (units: mg/m^3).



825

826 **Figure 12.** Same as Fig. 8, but for the experiments of NODA, ALL_6h, ALL-3h, ALL_1h, respectively.



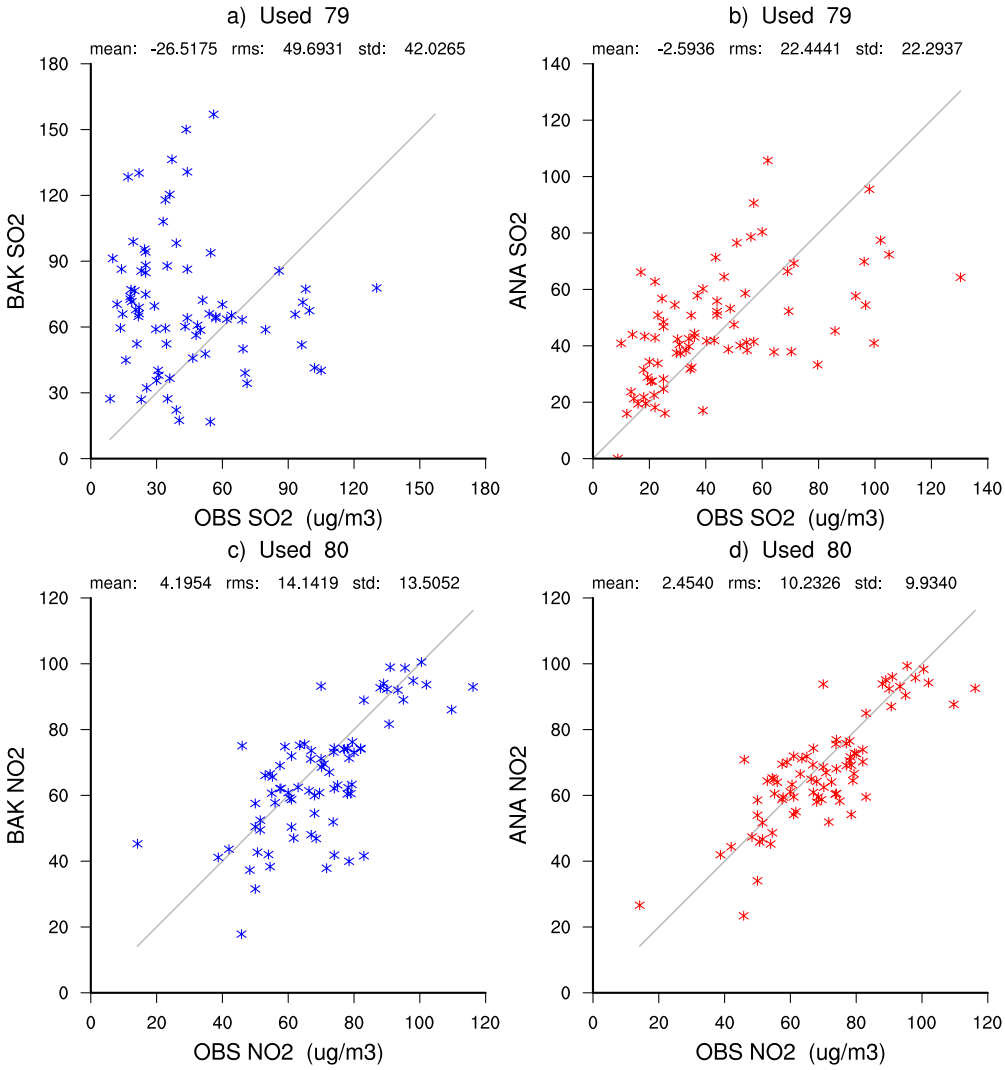
827

828 **Figure 13.** Time series of (a) sulfate, (b) nitrate over January 14-20, verified against the size-resolved

829 particle observation at IUM station (units: μg/m^3). The gray, blue and red lines denote the observation

830 and the results of experiment PM2 and ALL, respectively. The numbers on the right of each panel

831 denote the averaged RMSE over January 14-20 for different experiments.



832

833 **Figure 14.** Averaged scatter plot of (a, c) observation versus background and (b, d) observation versus

834 analysis for (a, b) SO_2 and (c, d) NO_2 around Beijing area (red dots in Fig. 1) on January 16. The

835 numbers on the title denote the accumulated numbers of the used observations around Beijing area

836 during January 16 (1600 UTC, 1606 UTC, 1612 UTC, and 1618 UTC).