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

Wei Sun, Zhiquan Liu, Dan Chen, Pusheng Zhao, and Min Chen

1 National Center for Atmospheric Research, Boulder, CO, 80301, USA
2 National Space Science Center, Chinese Academy of Sciences, Beijing, 100190, China
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 disadvantage of unbalanced photochemistry (due to inaccurate precursor NOx/VOC ratios) from assimilating the existing observations (only O$_3$ and NO$_2$, but no VOC). 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 further 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 extreme haze with high particulate matter (PM) national-wide and increasing tropospheric ozone (O₃) 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 PM₂.₅ and PM₁₀ (aerosols/fine particulate matter with aerodynamic diameters less than 2.5 or 10 μm), SO₂ (sulfur dioxide), NO₂ (nitrogen dioxide), O₃ (ozone), and CO (carbon monoxide). Among the six pollutants, the forecast on aerosols (especially PM₂.₅) is of greatest research interest as the severity of aerosol pollution and its negative effects on both health and climate. However, it’s still challenging to accurately simulate and forecast aerosols by pure air quality models due to some issues, such as the...
large uncertainties in primary and precursor emissions processes, the incomplete understanding and parameterization of secondary inorganic/organic reactions from precursors, and the accumulation of meteorology simulation errors. In addition to aerosol forecast, the elevated $O_3$ levels in city clusters over eastern China draw more and more attentions recently. Under this circumstance, in the urban regions in China, where suffer from complex air pollution with both haze and smog, the accurate forecast of air quality has been not only a challenge for operational centers, but also a common concern for scientific community.

To improve the forecast skill, data assimilation (DA), a combination of observations and numerical model output, has been widely used in meteorology forecast since last century, and recently extended to air pollutant forecasts. Based upon various techniques, DA is proven to be skillful at improving the meteorology and aerosol forecasts (Bannister 2017; McHenry et al. 2015; Peng et al. 2018; Sandu and Chai 2011; Schutgens et al. 2010; Sekiyama et al. 2010; Tang et al. 2011; Tang et al. 2013). Focusing on aerosol assimilation, NCAR group had conducted a series of work. Using three-dimensional variational (3DVAR) algorithm, Liu et al. (2011) implemented DA on aerosol optical depth estimates within the Grid-point Statistical Interpolation (GSI) system. Schwartz et al. (2012), Jiang et al. (2013), and Chen et al. (2019) further extended this system to assimilate surface PM$_{2.5}$ and PM$_{10}$. It should be noted that the aerosols are complicated not merely from primary emissions but also secondary reactions with gaseous precursors in the atmosphere (Huang et al. 2014; Nie et al. 2014; Xie et al. 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 treatment 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 Chemistry) aerosol scheme uses a sectional approach to represent the aerosol size distribution with different size bins, and it takes black carbon, organic carbon, sulfate, nitrate, ammonium, sodium, chloride, and other inorganic compounds that are related to anthropogenic emissions into consideration.
In result, the MOSAIC scheme exhibits a better performance in representing the complex PM$_{2.5}$ pollution over China (Chen et al. 2016; Chen et al. 2019). Therefore, to make the DA system suitable for different emphasis and applications, a flexible aerosol assimilation capability is built within the WRFDA system in this study, which will facilitate developments and applications for more chemistry schemes in the future. Focusing on the air quality forecast over China, this study mainly analyses the results of MOSAIC aerosol scheme.

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

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 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 larger 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
9

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 ($\text{SO}_4^{2-}$) and nitrate ($\text{NO}_3^-$) 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

$$ J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + \frac{1}{2} [H(x) - y]^T R^{-1} [H(x) - y], $$ (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 the observation operator that interpolates model grid point values to observation space and converts model-predicted variables to observed quantities.

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

2.3.1 Observation operators

The WRFDA-Chem is designed to assimilate six types of surface aerosol/chemical observations, including PM$_{2.5}$, PM$_{10}$, SO$_2$, NO$_2$, O$_3$, and CO. For aerosol assimilation, the aerosol species in the MOSAIC scheme are defined as black carbon (BC), organic compounds (OCs), sulfate (SO$_4^{2-}$), nitrate (NO$_3^-$), ammonium (NH$_4^+$), sodium (NA), chloride (CL), and other inorganic compounds (OIN). To 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$_{2.5}$ are computed by summing the 24 variables as

$$y_{PM_{2.5}} = \rho_d \sum_{i=1}^{24} [BC_i + OC_i + SO_{4i} + NO_{3i} + NH_{4i} + CL_i + NA_i + OIN_i]. \quad (2)$$
The model-simulated PM$_{10}$ observations are computed by summing the 32 variables as

$$y^f_{PM_{10}} = \rho_d \sum_{i=1}^{8} [BC_i + OC_i + SO_{4_i} + NO_{3_i} + NH_4_i + CL_i + NA_i + OI\text{Ni}_i].$$  \hspace{1cm} (3)

Correspondingly,

$$y^f_{PM_{10-2.5}} = \rho_d \sum_{i=4}^{8} [BC_i + OC_i + SO_{4_i} + NO_{3_i} + NH_4_i + CL_i + NA_i + OI\text{Ni}_i].$$  \hspace{1cm} (4)

where $\rho_d$ 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 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 first three bins (Eq. 2), and the PM$_{10-2.5}$ (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

$$y^f_x = \rho_d \cdot \frac{M_x}{M_{\text{air}}} \cdot R_x \cdot 10^3,$$  \hspace{1cm} (5)

where $x$ denotes the four gas-phase pollutants as in SO$_2$, NO$_2$, O$_3$, and CO, $\rho_d$ is the dry-air density, $M_x$ is the relative molecular mass for the four gas-phase pollutants, $M_{\text{air}}$ is the relative molecular mass for dry-air, and $R_x$ is the mixing ratio for the four gas-phase 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 $\varepsilon_0$ and the representativeness error $\varepsilon_r$ in this study. The measurement error $\varepsilon_0$ 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 $\mu g/m^3$; the representativeness error $\varepsilon_r$ is defined as $\varepsilon_r = \gamma \varepsilon_0 \frac{\Delta x}{L}$, where $\gamma$ is an adjustable parameter scaling (set as 0.5), $\Delta x$ 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 ($\varepsilon_x$) is computed as $\varepsilon_x = \sqrt{\varepsilon_0^2 + \varepsilon_r^2}$, 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). Cross-correlations between different aerosol/chemical variables were not considered. With the updated GEN_BE v2.0, the statistics for background error covariance, such as standard deviation, vertical and horizontal length scales, and vertical correlations, are computed for each of the aerosol/chemical variable. In this study, the background error covariance is estimated using the National Meteorological Center (NMC) method (Parrish and Derber, 1992) from one-month WRF-Chem forecasts over January 2017.
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 PM$_{2.5}$ and PMcoarse (PM$_{10-2.5}$) simultaneously; ALL assimilated PM$_{2.5}$, PM$_{10-2.5}$, SO$_2$, NO$_2$, O$_3$, and CO together.

In view of the cycling frequency is an important aspect in the DA strategy, especially for 3DVAR, two more experiments assimilating all the six major pollutants with different cycling frequencies are further conducted, in which the ALL_3h and ALL_1h experiments assimilate the data with 3-h and 1-
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 2 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. 2a), the NODA experiment displays a general overestimation of 36.60 μg/m$^3$, along with a large RMSE of 70.41 μg/m$^3$. After DA, in the PM1, PM2, and ALL experiments, the bias of PM$_{2.5}$ drops to 5.62 μg/m$^3$, 5.19 μg/m$^3$, and 5.98 μg/m$^3$, respectively; the RMSE drops to 22.10 μg/m$^3$, 22.84 μg/m$^3$, and 23.15 μg/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. 2b). To explain this phenomenon, Figure 3 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 μg/m$^3$ over northern China (Fig. 3a). However, the PMcoarse in the NODA experiment (with an average of 5.47 μg/m$^3$) is much smaller than that in the observation (with an average of 39.13 μg/m$^3$). This result suggests that the WRF-Chem model failed to reasonably represent the PMcoarse, which is actually the 4+ 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. 2a), components in the first three bins (within 2.5 μm) of PM$_{10}$ decrease dramatically. Meanwhile, since the simulated PMcoarse is too small, the PM$_{10}$ variates are eventually
dominated by the adjustment of PM2.5. In result, the assimilation of PM2.5 causes a large negative bias in the PM10 analysis (Fig. 2b). Correspondingly, compared to the NODA run, the PMcoarse in the PM1 experiment exhibit no significant changes (only slightly decrease) in the analysis (Figs. 3b and 3c) and also in the forecast (Fig 3f).

To overcome this issue, several adjustments have been adapted in the PM10 assimilation: instead of using the PM10 observations directly, the PMcoarse is used to analyze the species in the 4+ bin (Eq. 4); to reflect the large uncertainty of the simulated PMcoarse and to appropriately weighting the model and observation errors, the background error covariance of the PMcoarse (species in the 4+ bin) is arbitrarily inflated (inflation factor 1 is normally used and 90 is selected after tuning). By this means, after assimilating the PM10 observations, the PM2 and ALL experiments exhibit similar distributions in the PMcoarse (Figs. 3d-e, with an average of 34.58 μg/m3 and 34.68 μg/m3) as in the observation (with an average of 39.13 μg/m3). Correspondingly, compared to the NODA experiment, evident improvements for PM10 analysis appear in the PM2 and ALL experiments, in which the bias and RMSE drops evidently (Fig. 2b). Overall, the DA experiments exhibit strong contributions to the analyses of PM2.5 and PM10, suggesting that the WRFDA-Chem system works effectively in updating the initial conditions.

As for the analyses of gaseous pollutants (Figs. 2c-2f), large improvements can be seen in the ALL experiment by further assimilating SO2, NO2, O3, and CO. Compared to the PM2 experiment, although the bias and RMSE for PM2.5 and PM10 in the ALL experiment is slightly larger, the bias for the four gaseous pollutants decrease from 4.74 μg/m3, -4.59 μg/m3, 4.92 μg/m3, and -8.31 mg/m3 (PM2 experiment) to -1.68 μg/m3, -1.25 μg/m3, -0.31 μg/m3, and -0.18 mg/m3 (ALL experiment), respectively, and the corresponding RMSE drops from 37.87 μg/m3, 15.39 μg/m3, 21.04 μg/m3, and
1.11 mg/m³ (PM2 experiment) to 23.85 μg/m³, 9.70 μg/m³, 8.62 μg/m³, and 0.43 mg/m³ (ALL experiment). In general, by assimilating all the six major pollutants, the ALL experiment displays the largest improvement in the analyses of gaseous pollutants among all the experiments, along with a comparable improvement in the analyses of the aerosols.

Due to lack of vertical information within the observations, the common mathematical solution to use the surface total mass observations to analyze multiple 3-D fields variables is to utilize prior information in the background. As show in Fig. 4, based on vertical correlations specified in the background error covariance, the observation impact spreads to a certain height, even though the analysis variables used in the observation operator (Eq. 2-5) are only at the lowest model level. It is also noted that observations contribute differently among the analysis variables. Corresponding to the strong overestimation of PM$_{2.5}$ (Fig. 2a), all the three DA experiments (PM1, PM2 and ALL) tend to reduce the PM$_{2.5}$ below 6 km; corresponding to the distinct underestimation for CO (Fig. 2f), the experiment assimilating CO (ALL experiment) increases the value below 9 km. Relative small analysis increments are shown in the other three gas pollutants (SO$_2$, NO$_2$, and O$_3$).

3.2 Forecast improvements

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

As shown in Fig. 5, model performs relatively poor in the forecast of aerosols without DA. For PM$_{2.5}$, the average bias, RMSE, and correlation over 0-24 h are 31.17 μg/m³, 88.99 μg/m³, 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. Averaging over 0-24 h, the improvement percentage for bias, RMSE and correlation reach up to 72.4%, 39.0%, 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 26.2% 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 lead to the small averaged bias in the NODA run (see the spatial distribution of bias at 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 general positive contribution to the PM$_{10}$ forecast.

Figure 6 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 13.4%, 42.3%, 74.0%, and 74.5% (13.4%, 5.3%, 11.3%, and 33.7%), compared to the NODA run, and the average correlation increases by 34.8%, 9.6%, 40.0%, 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 O₃ are related to complex photochemical reaction processes, the assimilation of NO₂ and O₃ 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 O₃.

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, calculation methodology in Section 2 of the supplemental material) for AQI at each AQI level is further analyzed. As shown in Fig. 7, 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 heavy polluted situations with AQI levels 5-6 (Figs. 7e-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 heavy 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.

### 3.3 Response to DA cycling frequency

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

Figure 8 shows the domain-averaged bias and RMSE of the analysis as in Fig. 2, but for experiments with different DA frequencies (ALL_6h, ALL_3h and ALL_1h). Except for O$_3$, most of the variables display a gradual improvement with the increase of cycling frequency. For example, from NODA run to the 6-h cycling experiment, and then to the 3-h and 1-h cycling experiment, the bias (RMSE) for PM$_{2.5}$ gradually decrease from 36.60 µg/m$^3$ (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$^3$ (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 PM$_{10}$, SO$_2$, and CO. In accordance with the
gradual improvements in the analyses, the forecast skills increase with the cycling frequency in most of the variables except O₃ (Figs. 9-10). Especially for the forecasts of aerosols, evident gradual improvements can be seen from 0 to 24 h. From the 6-h cycling experiment to the 3-h and the 1-h cycling experiment, the averaged decrease percentage of RMSE for PM₂.₅ (PM₁₀) enlarges from 38.76% to 41.27% and 44.21% (27.31% to 30.17% and 32.97%); the averaged increase percentage of correlation for PM₂.₅ (PM₁₀) enlarges from 42.82% to 49.51% and 55.58% (57.71% to 66.39% and 74.89%). To further investigate the integrated DA effect of aerosols and gas phase pollutants under different cycling frequency, the TS for AQI is shown in Fig. 11. The forecast of air quality is improved step by step with the increase of cycling frequency. On AQI levels 2-6, the TS for the ALL_1h experiment situates above the ALL_3h experiment at most of the time, and followed by the ALL_6h experiment. These results indicate that the frequent IC updating is helpful to further improve the forecast for most of the pollutants.

However, the analysis and forecast of O₃ becomes worse under higher cycling frequencies (Fig. 8e and 10c). As a short-lived chemical reactive species, O₃ takes part in highly complex and rapid photochemical reactions in association with NOₓ and VOC (Peng et al. 2018, Lu et al., 2019). From this perspective, the performances of O₃ could mostly rely on the rapid photochemistry, in addition to the IC. In the DA experiments, the assimilation of NOₓ changes the NOₓ concentration and leave the VOC unadjusted due to the absence of VOC measurements. In result, the NOₓ/VOC ratio which determine the photochemical reactions and even regime might be changed (O₃ production/loss direction might change). Since the relevant NOₓ-VOC-O₃ reactions take place so quickly, changing the O₃ concentration in minutes, the advantage of IC DA is competing with the disadvantage of the disordered photochemistry (inaccurate NOₓ/VOC ratios) from the unadjusted VOC and the updated O₃.
and NO₂, and thus the improvement of IC DA could be consumed quickly. Under this circumstance, the more frequent the O₃ and NO₂ were assimilated, the more incompatibilities could be brought into the related photochemical reactions, resulting the model performs worse in the forecast of O₃ under higher cycling frequencies.

According to the results above, it is better to assimilate PM₂.₅, PM₁₀, SO₂ and CO every 1 h and assimilate O₃ and NO₂ every 6 h in the future applications, given the fact that the 6-h cycling experiment performs the best in the O₃ forecasting (Fig. 10c) and displays no significant differences in the NO₂ forecasting with experiments under higher cycling frequencies (Fig. 10b). It could also be helpful to assimilate the VOC along with O₃ and NO₂ after there are corresponding observations.

4. Indications on further model development

A higher forecast skill relies on not only better working of DA, but also better performance of the forecast model. To further improve the forecast skill, a crucial task is to understand the deficiencies in the model, while the challenge in chemistry model diagnostic is that uncertainties are from various aspects and are mixed-up in the model simulations, and the situation becomes even more complex when the reaction path is not yet revealed by laboratory. However, with the help of DA, as one aspect (IC) in the model is corrected by using observation as constraints, the deficiencies from other aspects (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 PM₂.₅ in heavy polluted events in northern China (Wang et al. 2014). In addition to the normal pathways in the MOSAIC scheme, we added SO₂-NO₂-NO₃ 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\textsubscript{2}, NO\textsubscript{2}-NO\textsubscript{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\textsubscript{4}\textsuperscript{2-}) and nitrate (NO\textsubscript{3}⁻), against a size-resolved particle observation over Beijing IUM station (in view of the assimilated SO\textsubscript{2} and NO\textsubscript{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 12 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\textsubscript{2}, NO\textsubscript{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 μg/m\textsuperscript{3}; nitrate exhibits an increase, accompanied by the average RMSE grows from 8.74 to 10.12 μg/m\textsuperscript{3}. However, compared to the PM2 experiment, the precursors (SO\textsubscript{2} and NO\textsubscript{2}) are indeed improved. Figure 13 displays the analysis statistics of SO\textsubscript{2} and NO\textsubscript{2} in the ALL experiment around Beijing area (red dots in Fig. 1) on January 16, the period with the largest changes of sulfate and nitrate (Fig. 12). To correct the overestimated SO\textsubscript{2} (underestimated NO\textsubscript{2}) in the background, the DA in reduces (enhances) the model value in the ALL experiment, making it closer to the observations.
It should be mentioned that the heterogeneous reactions are added by using the sulfate-nitrate-ammonium observations as constraints to tune the “observation-best-matched” uptake coefficients under the scenario without DA, in which the precursor concentrations are from pure model thus not very accurate. To best match the observation, when gaseous precursors are overestimated (underestimated) in the model, the uptake coefficient is tuned to low-biased (high-biased) value. In result, such a coefficient may no longer be suited for the cases with DA. For instance, after DA reducing the overestimated \( \text{SO}_2 \), the uptake coefficient is still relatively low and thus the reaction from \( \text{SO}_2 \) to sulfate will stay at a low rate (with both low value of \( \text{SO}_2 \) and low reaction coefficient). A similar result also exists for the reaction from \( \text{NO}_2 \) to nitrate. From this perspective, the negative effects on sulfate and nitrate in the ALL experiment may not be hard to understand (Fig. 12). Therefore, in the future chemistry development, it is necessary to develop more 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. Accordingly, further improvements on aerosol forecast could be expected by assimilating their gaseous precursors. According to the results above, the DA technique provides an opportunity to identify and diagnose the deficiencies in the model. By correcting the precursor concentrations through DA (one aspect), the deficiency of the uptake coefficients for the SNA heterogeneous reactions (the other aspect than the 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 38%, 26%, and 10-30% 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$_3$, although improvements are acquired at the 6-h cycling frequency, the advantage of more frequent IC DA could be consumed by the disordered photochemistry (inaccurate NO$_2$/VOC ratios) due to the unadjusted VOC and the updated O$_3$ and NO$_2$ from DA. In future applications, it is better to assimilate PM$_{2.5}$, PM$_{10}$, SO$_2$, and CO every 1 h and assimilate O$_3$ and NO$_2$ every 6 h. It might also be helpful to assimilate 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.

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, study will focus on assimilating chemical observations from different observing platforms, such as satellite AOD observations, which contains 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.
512 **Code and data availability**

513 The WRF-Chem code used in this study is the public release version available at https://www2.mmm.ucar.edu/wrf/users/download/get_sources.html#WRF-Chem. The WRFDA code in this study is developed based on the public release version available at https://www2.mmm.ucar.edu/wrf/users/download/get_sources.html#WRFDA, and the developed WRFDA-Chem code could be incorporated into the public release version at appropriate time. The example run directory and dataset for this paper are available upon request from the corresponding authors (liuz@ucar.edu and dchen@ium.cn) and Wei Sun (weisun0416@gmail.com).

520 **Author contributions**

521 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.

524 **Acknowledgement**

525 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.
529 Competing interests

530 The authors declare that they have no conflict of interest.

531 References

532 Bannister, R., 2017: A review of operational methods of variational and ensemble-variational data
535 during the extreme haze events over northern China in October 2014. *Atmospheric Chemistry and
536 Physics*, 16, 10707-10724.
538 PM 2.5 in China: response to emission regulations and the role of meteorology. *Atmospheric
539 Chemistry and Physics*, 19, 7409-7427.
540 Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn
543 Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use
544 in general circulation models.
545 Descombes, G., T. Auligné, F. Vandenbergh, D. Barker, and J. Barre, 2015: Generalized background


Zheng, B., and Coauthors, 2015: Heterogeneous chemistry: a mechanism missing in current models to explain secondary inorganic aerosol formation during the January 2013 haze episode in North China. *Atmospheric Chemistry and Physics (Online)*, **15**.


### Tables and Figures

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

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

**Table 3.** Averaged bias (units: $\mu g/m^3$), RMSE (units: $\mu g/m^3$), 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.

**Figure 1.** Computation domain. Dots depict surface observations with 531 stations spreading over China. The red dots indicate the observations around Beijing.

**Figure 2.** 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. 2a-e and mg/m$^3$ in Fig. 2f.

**Figure 3.** 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 4.** 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. 4a-e and mg/m$^3$ in Fig. 4f.

**Figure 5.** 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 6.** Same as Fig. 5, but for the forecast of (a) SO$_2$, (b) NO$_2$, (c) O$_3$ (units: µg/m$^3$), and (d) CO (units: mg/m$^3$).

**Figure 7.** Averaged threat score (TS) for Air Quality Index (AQI) from AQI level 1 to level 6 (a-f) 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, and ALL, respectively. The numbers on the right of each panel denote the averaged TS from 0 to 24 h for different experiments.

**Figure 8.** Same as Fig. 2, but for the experiments of NODA, ALL_6h, ALL-3h, ALL_1h, respectively.

Units of the y-axis are µg/m$^3$ in Figs. 8a-e and mg/m$^3$ in Fig. 8f.

**Figure 9.** 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, ALL_6h, ALL_3h, and ALL_1h, respectively.

**Figure 10.** Same as Fig. 9, but for the forecast of (a) SO$_2$, (b) NO$_2$, (c) O$_3$ (units: µg/m$^3$), and (d) CO (units: mg/m$^3$).

**Figure 11.** Same as Fig. 7, but for the experiments of NODA, ALL_6h, ALL-3h, ALL_1h, respectively.

**Figure 12.** Time series of (a) sulfate, (b) nitrate over January 14-20, verified against the size-resolved particle observation at IUM station. The gray, blue and red lines denote the observation and the results of experiment PM2 and ALL, respectively. The numbers on the right of each panel denote the averaged RMSE over January 14-20 for different experiments.
Figure 13. Averaged scatter plot of (a, c) observation versus background and (b, d) observation versus analysis for (a, b) SO$_2$ and (c, d) NO$_2$ around Beijing area (red dots in Fig. 1) on January 16.
Table 1. WRF-Chem model configurations.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerosol scheme</td>
<td>MOSAIC (four bins, Zaveri et al. (2008))</td>
</tr>
<tr>
<td>Photolysis scheme</td>
<td>Fast-J (Wild et al. 2000)</td>
</tr>
<tr>
<td>Gas-phase chemistry</td>
<td>CBM-Z (Zaveri and Peters 1999)</td>
</tr>
<tr>
<td>Cumulus parameterization</td>
<td>Grell 3-D scheme</td>
</tr>
<tr>
<td></td>
<td>Goddard Space Flight Center short-wave radiation scheme</td>
</tr>
<tr>
<td>Short-wave radiation</td>
<td>Chou and Suarez (1994)</td>
</tr>
<tr>
<td>Long-wave radiation</td>
<td>RRTM (Mlawer et al. 1997)</td>
</tr>
<tr>
<td>Microphysics</td>
<td>Single-moment 6-class scheme (Grell and Dévényi 2002)</td>
</tr>
<tr>
<td>Land-surface model (LSM)</td>
<td>NOAH LSM (Chen and Dudhia 2001)</td>
</tr>
<tr>
<td>Boundary-layer scheme</td>
<td>YSU (Hong et al. 2006)</td>
</tr>
<tr>
<td>Meteorology initial and boundary</td>
<td>GFS analysis and forecast every 6 h</td>
</tr>
<tr>
<td>conditions</td>
<td></td>
</tr>
<tr>
<td>Initial condition for chemical</td>
<td>11-day spin-up</td>
</tr>
<tr>
<td>species</td>
<td></td>
</tr>
<tr>
<td>Boundary conditions for</td>
<td></td>
</tr>
<tr>
<td>chemical species</td>
<td></td>
</tr>
<tr>
<td>Dust and sea salt emissions</td>
<td>GOCART</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. The detail setting of six experiments and the purposes

<table>
<thead>
<tr>
<th>Experiments</th>
<th>PM2.5 assimilation</th>
<th>PM10-2.5 assimilation</th>
<th>Gas phase (SO(_2), NO(_2), O(_3), CO) assimilation</th>
<th>Assimilated time (UTC)</th>
<th>Purposes for forecast performances</th>
</tr>
</thead>
<tbody>
<tr>
<td>NODA</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>--</td>
<td>Control simulation</td>
</tr>
<tr>
<td>PM1</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>00, 06, 12, 18</td>
<td>Basic PM2.5 assimilation</td>
</tr>
<tr>
<td>PM2</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>00, 06, 12, 18</td>
<td>PM2.5 and PM10-2.5 assimilation</td>
</tr>
<tr>
<td>ALL</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>00, 06, 12, 18</td>
<td>Aerosol and precursor simultaneously assimilation</td>
</tr>
<tr>
<td>ALL_3h</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>00, 03, 06, 09, 12, 15, 18, 21</td>
<td>Different assimilation frequencies on forecast performances</td>
</tr>
<tr>
<td>ALL_1h</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0-23, every hour</td>
<td></td>
</tr>
</tbody>
</table>

https://doi.org/10.5194/acp-2019-969
Preprint. Discussion started: 23 January 2020
© Author(s) 2020. CC BY 4.0 License.
Table 3. Averaged bias (units: $\mu g/m^3$), RMSE (units: $\mu g/m^3$), 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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>NODA</th>
<th>PM1</th>
<th>PM2</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM$_{2.5}$</td>
<td>Bias</td>
<td>31.17</td>
<td>8.78</td>
<td>8.39</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>88.99</td>
<td>53.93</td>
<td>54.35</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.41</td>
<td>0.59</td>
<td>0.58</td>
</tr>
<tr>
<td>PM$_{10}$</td>
<td>Bias</td>
<td>-1.13</td>
<td>-22.73</td>
<td>-15.43</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>98.5</td>
<td>74.41</td>
<td>71.9</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.36</td>
<td>0.54</td>
<td>0.56</td>
</tr>
<tr>
<td>SO$_2$</td>
<td>Bias</td>
<td>6.67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>44.11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>Bias</td>
<td>-2.87</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>25.61</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.48</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>O$_3$</td>
<td>Bias</td>
<td>-3.22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>31.96</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.29</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CO</td>
<td>Bias</td>
<td>-0.73</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>1.13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
<td>0.28</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 1. Computation domain. Dots depict surface observations with 531 stations spreading over China. The red dots indicate the observations around Beijing. The green dot indicates the IUM station.
Figure 2. 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 μg/m$^3$ in Figs. 2a-e and mg/m$^3$ in Fig. 2f.
Figure 3. Averaged PMcoarse (PM_{10-2.5}, units: μg/m³) 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: μg/m³) 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: μg/m³).
Figure 4. 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. 4a-e and mg/m$^3$ in Fig. 4f.
Figure 5. 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 6. Same as Fig. 5, 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$^3$).
Figure 7. Averaged threat score (TS) for Air Quality Index (AQI) from AQI level 1 to level 6 (a-f) 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, and ALL, respectively. The numbers on the right of each panel denote the averaged TS from 0 to 24 h for different experiments.
Figure 8. Same as Fig. 2, but for the experiments of NODA, ALL_6h, ALL_3h, ALL_1h, respectively.

Units of the y-axis are μg/m³ in Figs. 8a-e and mg/m³ in Fig. 8f.
Figure 9. Averaged bias (units: μg/m³), RMSE (units: μg/m³), 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, ALL$_{6h}$, ALL$_{3h}$, and ALL$_{1h}$, respectively.
Figure 10. Same as Fig. 9, 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$^3$).
Figure 11. Same as Fig. 7, but for the experiments of NODA, ALL_6h, ALL_3h, ALL_1h, respectively.
Figure 12. Time series of (a) sulfate, (b) nitrate over January 14-20, verified against the size-resolved particle observation at IUM station. The gray, blue and red lines denote the observation and the results of experiment PM2 and ALL, respectively. The numbers on the right of each panel denote the averaged RMSE over January 14-20 for different experiments.
Figure 13. Averaged scatter plot of (a, c) observation versus background and (b, d) observation versus analysis for (a, b) SO$_2$ and (c, d) NO$_2$ around Beijing area (red dots in Fig. 1) on January 16.