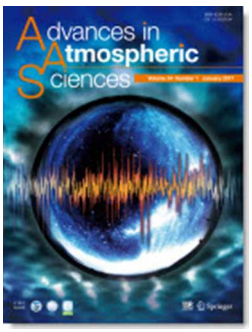


Comparisons of three-dimensional variational data assimilation and Model Output Statistics in improving atmospheric chemistry forecast

Journal:	*Advances in Atmospheric Sciences*
Manuscript ID	AAS-2017-0179
Manuscript Type:	Original Article
Date Submitted by the Author:	26-Jul-2017
Complete List of Authors:	Ma, Chaoqun; Nanjing University Wang, Tijian; Nanjing University Zang, Zengliang Li, Zhijin; Joint Institute for Regional Earth System Science and Engineering

Comparisons of three-dimensional variational data assimilation and Model Output Statistics in improving atmospheric chemistry forecast

Chaoqun Ma [1], Tijian Wang [1,*1], Zengliang Zang [2], and Zhijin Li [3]

[1]*School of Atmospheric Sciences, CMA-NJU Joint Laboratory for Climate Prediction Studies, Jiangsu Collaborative Innovation Center for Climate Change, Nanjing University, Nanjing 210000, China*

[2]*Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 210000, China*

[3]*Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles 90001, California, USA*

ABSTRACT

Atmospheric chemistry models usually perform badly in forecasting wintertime air pollution because of their uncertainties. Generally, such uncertainties could be decreased effectively by techniques such as data assimilation (DA) and Model Output Statistics (MOS). However, the relative importance and combined effects of the two techniques have not been clarified. One-month air quality forecast with the Weather Research and Forecasting-Chemistry (WRF-Chem) model was carried out here in virtually operational manner focusing on Hebei province, China. Meanwhile, 3D variational DA (3Dvar DA) and MOS based on one-dimensional Kalman filtering were implemented separately and simultaneously to investigate their relative performance in model improvement. Comparison with observations shows that chemical forecast with MOS outperforms that with 3Dvar DA. Such superiority of MOS in improving the forecast accuracy could be

*Corresponding author : Tijian Wang
Email: tjwang@nju.edu.cn

seen in all the species tested and extend to 48 h and 72 h forecast in which 3Dvar DA has almost no or negligible effect. Combined use of both techniques does not guarantee better forecast than MOS only—improvements and degradations are both small and appear rather randomly. Results probably indicate that the implementation of MOS is more suitable than 3Dvar DA in improving operational forecast ability of chemistry model.

Key words: Data assimilation; model output statistics; WRF-Chem; operational forecast

1. Introduction

In recent years, an unexpected outbreak of severe air pollution events engulfed China during autumn and winter. These air pollution episodes arouse a deep concern and panic in the public and has since then been the top priority of local government. To efficiently control the atmospheric pollution, it's a must to achieve accurate forecast for atmospheric chemical constituents.

Since the 21st century, the air quality forecast systems, such as Weather Research and Forecasting-Chemistry (WRF-Chem), have been gradually put into operation in key cities across China by many organizations and institution. However, without any additional measures, those numerical forecast systems are often incompetent for the application of air quality forecast due to the uncertainties within the parameterization schemes and input data of models (van Loon et al., 2007; Zhang et al., 2016). During the last decades, scientists have developed multiple techniques in either pre- or post-processing manners for model improvement in the operational prediction of meteorological and chemical fields. For example, data assimilation (DA)—a measure applied before the model run—is

an effective approach in improving the model forecast skill of air pollution via reducing the uncertainty of chemical initial conditions (CICs) or other parameters. For instance, Barbu et al. (2009) achieved better forecast by assimilating the measurements of SO_2 and SO_4 to adjust the emission and conversion rates of SO_2 in model; the research of Liu et al. (2011) and Yin et al. (2016) showed improvement of aerosol analysis and forecast from assimilating the Moderate Resolution Imaging Spectroradiometer (MODIS) total aerosol optical depth (AOD) retrieval products. Different models and observation data had been tried and results showed similar conclusions (Wang et al., 2014; Zhang et al., 2015; Mizzi et al., 2016; Tang et al., 2016). Readers who are interested could refer to Bocquet et al. (2015) for more details. On the other hand, through processing the model output, the forecast error could also be corrected effectively by another approach called Model Output Statistics (MOS) (Glahn and Lowry, 1972). MOS works through statistically relating the historical model output with the corresponding observations and then applying it to the model forecasts. This approach has been widely used in the post-processing of operational numerical weather prediction and most forecast biases could be corrected (Wilson and Vallee, 2003) especially for temperature (Taylor and Leslie, 2005; Libonati et al., 2008) and humidity (Anadranistakis et al., 2004).

Although it has been proved that both DA and MOS are effective in improving the forecast performance, little attention, however, was paid to the comparison or the combined use of the two methods, especially concerning the atmospheric chemistry model. Two problems made it relatively hard to fairly compare the two methods using the same dataset of observation. Firstly, early (even recent) research concerning the DA in atmospheric chemistry models focused on assimilating observation of satellite-derived

products to generate analysis skillful in improving the forecasts for variables like CO (Barret et al., 2008), CO_2, O_3, NO_2 (Inness et al., 2015; Wargan et al., 2015), CH_4 (Alexe et al., 2015) and aerosols (Benedetti et al., 2009; Yumimoto et al., 2016). However, MOS works only with in-situ observations from surface stations. Secondly, unlike numerical weather prediction, MOS still remains in its infancy for the operational numerical forecast of atmospheric chemical variables. Such works, if any, were usually based on regression approaches (Denby et al., 2008; Honore et al., 2008; Struzewska et al., 2016) which are effective in improving the air quality forecast for all analyzed species but too dependent on local pollution conditions, which makes them inconvenient to be applied as widely as DA approaches.

This study aims at comparing the potentials of the two approaches in improving the atmospheric chemistry forecast with WRF-Chem modeling system in the operational context. To overcome problems mentioned above, we firstly adapted a 3Dvar DA system based on Li et al. (2013) (L13 henceforth) and Jiang et al. (2013) to the assimilation of observation data from surface stations. The authors then modified a MOS scheme from Galanis and Anadranistakis (2002) (hereafter G02) which was used in adjusting the meteorological forecast, to make it able to correct the chemical output from atmospheric model. The paper is organized as follows: model, setup and experimental designs will be described in Section 2; evaluation of model improvement with the two methods are discussed in Section 3; a summary and conclusion are given in Section 4.

2. Method and data

2.1 Model setup

The WRF-Chem model is an online three-dimensional, Eulerian chemical transport model that considers the complex physical and chemical processes in the troposphere (Grell et al., 2005). It has been applied in various research settings, especially those concerning feedbacks of air pollution on weather and chemical data assimilation (Saide et al., 2012; Makar et al., 2015; Saide et al., 2015; Mizzi et al., 2016). In this study, WRF-Chem 3.7.1 was used to simulate air quality in Hebei province, China. Two nested domain was set as Fig. 1. The outer domain covered East-Asia with a horizontal resolution of 75 km×75 km and 106×81 grids, while the inner domain covered the Hebei province with a horizontal resolution of 15 km×15 km and 76×81 grids. Model vertical resolution was defined as 24 vertical levels with 100 hpa as model top. 0.5°×0.5° data from NCEP Global Forecast System (GFS) was used to provide the meteorological initial conditions and lateral boundary meteorological conditions every 12 hours. Atmospheric gaseous chemistry and aerosol was simulated using RADM2/MADE/SORGAM (Stockwell et al., 1990; Ackermann et al., 1998; Schell et al., 2001) scheme. Anthropogenic emission inventory was provided by Multi-resolution Emission Inventory for China (MEIC) in 2012 (http://www.meicmodel.org). Detailed configuration of WRF-Chem model was listed in Table 1.

2.2 MOS method

What MOS does is to find statistic relationship from the training samples which would be then applied on model forecast outputs. By doing so, it is expected to correct the model errors and generate forecast that fit the observation better. In this paper, one-dimensional Kalman filter was chosen as the algorithm to realize the MOS process for its situation-independence and stable effects. This paper formulated the algorithm mostly

resembling G02, so detailed theory and formulations will not be introduced here except some modifications described in the next paragraph.

Firstly, in this work, measurement *y(t)* as well as real value *x(t)* could be both the difference and ratio between forecast and observation while they only denoted difference in equation (1) and (2) of G02. Furthermore, hourly concentrations of 5 species from the 3 day model output were split into 3×5×24 independent daily concentration series. At last, given Kalman filtering could predict only one time step ahead (one day ahead in this context), the correction could only work on 24 h forecast while leaving 24 – 48 h and 48 – 72 h (48 h and 72 h hereafter) forecast uncorrected for lack of future observation. So, in order to extend the algorithm to the further, the corrected results from 24 h (48 h) forecast would be used as a low-grade proxy of observation at corresponding time to correct the 48 h (72 h) model output. Detailed steps could refer to Appendix A.

2.3 Configuration of DA

In this paper, 3Dvar data assimilation was implemented to optimize the CICs for the inner model domain. The data assimilation system and formulation used here is based on L13 with the following modifications.

In addition to $PM_{2.5}$ assimilated in L13, particulate mass with diameter between 2.5 μm and 10 μm ($PM_{2.5}$–PM_{10} hereafter) was also assimilated and the analysis increment was added to the corresponding model variables following L13. The gaseous species, including SO_2, NO_2 and O_3, also joined the assimilation to decrease the uncertainty of their concentration in the model CICs.

For background error, The National Meteorological Center (NMC) method (Parrish, 1992) was adopted here to estimate the background root mean square error (RMSE) and the three Kronecker product members of the background error correlation matrix. The NMC method utilized the difference between 12 and 24 h WRF-Chem forecasts valid at the same time of 12:00 UTC for a whole month. No cross correlation between different species was assigned for background error.

Considering that all the stations were built and maintained under the same standards, no difference the measurement and representativeness error between different stations was assumed. In addition, cross correlation between different species and stations were also set to zero for the lack of information. Therefore, observation error consisted of merely 5 numbers, one for each species. The measurement error was assigned as 1.0 μg/m3, 1.0 μg/m3, 1.0 ppb, 1.0 ppb and 1.0 ppb for $PM_{2.5}$, $PM_{2.5-10}$, SO_2, NO_2 and O_3 respectively. Representativeness error was estimated following Elbern et al. (2007) and Schwartz et al. (2012) using the formula

$$\varepsilon_r = \gamma \varepsilon_0 \sqrt{\frac{\Delta x}{L}} \quad \text{Eq. (1)}$$

where ε_0 and ε_r are the measurement error and representativeness error, γ means a adjustable parameter that accounts for the lifetime of the species (0.5 for $PM_{2.5}$, $PM_{2.5-10}$ and O_3, 1 for SO_2 and 2 for NO_2), Δx is the grid spacing (here 15km) and L is the radius of influence determined according to the location of stations (here 4.0km for suburban stations assumed for all sites). If the total observation error was defined as the sum of measurement error and representativeness error, the standard deviation of observation error ends out to be 2.0 μg/m^3, 2.0 μg/m^3, 3.0 ppb, 4.9 ppb and 2.0 ppb for

$PM_{2.5}$, $PM_{2.5-10}$, SO_2, NO_2 and O_3 respectively. Though the observation error was determined fairly arbitrarily and empirically here, the uncertainty relating to it should not have significant influence on the conclusion. That is because the results are usually not very sensitive to the specification of error (Geer et al., 2006) and similar analysis were obtained from our experiments with observation error increased or reduced 2–3 times.

2.4 Experiment design

To compare the relative importance of MOS and DA, 4 parallel experiments were designed: Sim_base, Sim_DA, Sim_MOS and Sim_DM. Sim_base worked as the base simulation without applying both DA and MOS.; Sim_DA was an experiment with only DA being employed to optimize the model CICs; Sim_MOS was the same to Sim_base except the model output had been corrected by one-dimensional Kalman filtering; in Sim_DM, both DA and MOS techniques were used.

To simulate the operational forecast scenes, as Fig. 2 shows, all experiments initiated a new daily WRF-Chem forecast at 12:00 UTC between 12:00 UTC 30 November 2014 and 12:00 UTC 31 December 2014. Each forecast was integrated 84 h to generate 72 hours' forecasts of each day with the earliest 12 h being discarded as the spin-up time. CICs for each initiation came from the 24 h forecasts of the previous cycle, which would be the background fields to be assimilated with valid observations for experiments with DA before initializing the WRF-Chem. The very first CICs at 12:00 UTC 30 November 2014 were the spun-up over 2 days beginning from the climatological background chemical profile. For all experiments, the chemical boundary conditions (CBCs) were the default climatological chemical profile for the outer domain who supplied CBCs through interpolation for the inner domain.

2.5 Observation data

Hourly concentration of SO_2, NO_2, PM_{10}, O_3 and $PM_{2.5}$ at surface level from 207 sites was provided by Ministry of Environmental Protection (MEP) of China. Data covers the whole month of December 2014 and has been subjected to routine quality control. As shown in Fig. 3, only 155 stations were randomly selected from the 207 stations to be assimilated with the 52 stations' data left to verify the assimilation process. Furthermore, it should be noted that only the 155 sites who provide their data into the 3Dvar data assimilation will participate into the MOS process.

3 Results

3.1 Model evaluation

Table 2 presents the mean bias (MB), relative bias (RB), RMSE and correlation coefficient (Corr) for the 24 h, 48 h and 72 h forecast of Sim_base. In general, the base model simulation could give fairly good result especially for NO_2 whose bias is small and correlation is high. When it comes to particulate matter and SO_2, model tends to systematically underestimate the concentration of SO_2 as well as that of $PM_{2.5}$ and PM_{10}. Even so, the model well reproduces the temporal variations of particulate matter, with correlation coefficient higher than 0.47 for PM_{10} and 0.54 for $PM_{2.5}$. For O_3, the model may seem to have some problem—simulated concentrations (20 – 80 $\mu g/m^3$) seriously overestimate the observed values (5 – 45 $\mu g/m^3$), which leads to positive bias (about 40 $\mu g/m^3$) and lower correlation coefficient (0.44) than other species. Fortunately, when

viewed the RMSE of O_3 from the aspect of MB, it is apparent that error is mainly caused by bias and therefore the model is still able to reproduce the variation of the O_3. The biases mentioned above could usually be attributed to the uncertainties from emission inventory (Tang et al., 2011) and model schemes (Yerramilli et al., 2010). Though 24 h forecast performs the best for all species, 48 h and 72 h simulation are also good enough to yield fairly reliable results, which is critical to the success of MOS in the whole 72 hours' forecast. In a word, the model shows forecast skill that is enough to make itself competent for the success of DA and MOS process.

3.2 Validation of MOS

Fig. 4 depicts the site averaged hourly concentration simulated by Sim_MOS plotted versus ground observations. Note that, although the hourly concentration are averaged over 155 stations, the MB and RMSE attached in Fig. 4 are generated by calculating the 155 stations' own errors at the first before averaging.

From Fig. 4, it could be concluded that forecast from the Sim_MOS fits the observation fairly close especially for SO_2, NO_2 and O_3. However, when it comes to $PM_{2.5}$ and PM_{10}, the points locate within a wider space and those extremely high observations are hard for MOS to forecast. Even so, when compared with Table 2, $PM_{2.5}$ and PM_{10} together with the other 3 species could see a clear correction obtained for all the forecast time. Excluding the 48 h forecast of NO_2, MOS could decrease the MB to a large extent, which means MOS is able to remove the majority of model systematic bias. Thanks to the reduction of MB, RMSE were also decreased for all the cases except the 24 h forecast of SO_2. What's more, the effect on reducing the error is unlikely to get poor as the forecast time advances. That is to say, no matter for 72 h and 48 h forecast or 48 h

and 24 h forecast, the effect MOS has on the former could rival or even exceed that on the latter, e.g. the RMSE reduction of $PM_{2.5}$ is even larger for 72 h than 48 h forecast. Among all the 5 species, O_3 seems to benefit the most from the MOS process. That is not surprising because O_3 usually follows a very regular daily variation which makes the hourly-split but daily-linked concentration series almost perfect for the assumptions of one-dimensional kalman filtering.

It must be admitted that MOS degrades the forecast in few cases (48 h forecast of NO_2 and 24 h forecast of SO_2 as mentioned above). Those error increase usually will not concern the users and may well be accepted, in that they are extremely small and only appears at times when model outputs to be corrected are already fairly close to the observation. However, when viewed from correlation, such degradation becomes more obvious. Except NO_2 and O_3, the correlation all experience a decrease of about 0.1 – 0.2. It seems that the MOS approach tends to reduce the bias and error at the expense of correlation.

3.3 Validation of DA

Fig. 5 and Fig. 6 shows the RMSE and Corr change over the integration time from forecast −12 h (right after the DA) to forecast 10 h (already integrated 22 h) for experiments Sim_base and Sim_DA. RMSE and Corr are averaged over the 52 stations who have not provided their observation data into the 3Dvar DA process, so any improvement in RMSE and Corr should be attributed to the success of DA.

It could be seen from the forecast −12 h of Fig. 5 and Fig. 6 that DA lead to better initial condition for the simulation, especially for NO_2, PM_{10}, $PM_{2.5}$ and O_3 whose RMSE experienced large decrease at almost all the sites. For example, particulate matter such as PM_{10} and $PM_{2.5}$ has a RMSE reduction of about 50–100 μg/m^3, which is about half the RMSE of Sim_base. Such results are as good as those obtained by L13 and Jiang et al. (2013) who also worked on assimilating ground observation using 3Dvar. For SO_2, the RMSE decrease seems not so apparent (though RMSE change is still negative when 52 sites are averaged) but correlation after DA is still obviously larger than before. The marginal RMSE decrease of SO_2 could be accepted considering correlation increase is rather obvious and the data representativeness of some stations is dubious (Zhang et al., 2016).

However, as expected, the effect of DA slowly diminishes when the integration goes on, which has also been observed in other works (Jiang et al., 2013; Li et al., 2013). After the model has been integrated over 14 hours, RMSE after DA minus that before DA (RMSE change henceforth) are still negative but their absolute values are apparently smaller when compared to the earlier. The effect of DA keeps longer for O_3, PM_{10} and $PM_{2.5}$ (RMSE change keeps negative for the 14 hours), mainly benefiting from their relatively long lifetime. For example, $PM_{2.5}$ and PM_{10} still maintain a RMSE reduction of about 10–20 μg/m^3, which is even better than results from L13 and Jiang et al. (2013). However, when it comes to NO_2 whose lifetime is short, the two experiment shows almost no difference in RMSE after 4 hours' integration. Because the initial improvement from DA is relatively small, the forecast of SO_2 soon loses its improvement from DA and shows little RMSE change almost immediately after the run of model. For SO_2 and NO_2,

the RMSE change are positive in some cases, but such positive values are usually very small compared with the original RMSEs and therefore unlikely to be a problem. Conclusions from the view of Correlation are similar to that from RMSE, except the effect of DA seems more obvious and long-lasting.

Overall, for most cases, DA successfully produce better chemical initial condition for the model and could help to improve the forecast skill in the following half to one day.

3.4 Effect of MOS and DA

Forecast error could be seen to vary with forecast hours from Fig. 7 which plots the RMSE of the 4 experiments and 5 species at different forecast hours from −12 h to 72 h. Considering the RMSE is calculated from the statistics of one month, Fig. 7 could well represent a relatively general fact of the four experiments.

From comparing experiments using MOS (solid lines) and those without MOS (dotted lines), all species see large decrease in the 72 hours forecast span and such decrease is much larger than that DA could provide (solid lines are below the dotted lines a larger distance than blue lines are below the red). Take SO_2 as an example, the averaged RMSE decrease of the 72 hours for Sim_MOS to Sim_base is 4.34 μg/m^3 while Sim_DA could only provide a decrease of 0.48 μg/m^3. What is worse, when forecast runs to its second or third day, the effect of DA would inevitably diminish (it could be seen from the overlapped red dotted and blue dotted lines after forecast 24 h) while MOS could still work during this period (solid lines do not overlap with the dotted lines even after forecast 24 h).

The blue dotted lines represent simulation RMSE corrected only with MOS while red dotted lines are results processed by both MOS and DA. Overlap of the two lines could

be seen at almost all the times from all the species, which means that DA of the initial condition provides little help to the MOS effect though providing better initial condition for the model to generate better forecast. However, sometimes the two lines do not overlap and do show some differences, which is common for all the species but most obvious for NO_2 and most unobvious for PM_{10} and $PM_{2.5}$. In time when DA still could improve the forecast or say the red dotted line is below the blue dotted line, the red solid line could be either above (forecast 1 h in SO_2) or below (forecast 10 h in O_3) the red blue line, which means better forecast from DA may either improve or degrade the MOS effect. Because in this work, MOS correct the one day's forecast using the correction results and forecast from previous days, it is not strange that Sim_DM and Sim_MOS show discrepancy when Sim_DA and Sim_base coincide after forecast 24 h. However, like what has been stated, such discrepancy could be either improvement or degradation.

3.5 Discussion

This section will give additional discussion on two facts. First, MOS could improve forecast far more than 3Dvar DA of CICs. That is reasonable because MOS could keep effective through the whole forecast period while the effect of 3Dvar via optimized initial condition usually diminishes after 24 hours of model integration. The loss of benefit from 3Dvar DA is unavoidable because atmospheric chemistry is less sensitivity to chemical initial condition rather than other driving factors like meteorological conditions and emission (Henze et al., 2009; Semane et al., 2009; Tang et al., 2011). What is worse, the earliest 12 hours' forecast, which benefits the most from 3Dvar DA, usually makes no sense in real operational forecast environment and is excluded from evaluation as spin-up. In fact, when compared with Sim_base, Sim_DA could account for 43.85% O_3

RMSE decrease in the first 12 hours after initialization (forecast −12 h to forecast −1 h), well acting as a rival of Sim_MOS who contributes 55.94% O_3 RMSE decrease in its first 12 hours (forecast 0 h to forecast 11h). However, when discussed within the same forecast time period, say forecast 12 h to forecast 24 h, Sim_DA could produce only 3.93% O_3 RMSE decrease which is far less than Sim_MOS (61.26%), despite the following hours during which DA has no effect at all.

The second fact to be explained is that Sim_DM does not always outperform Sim_MOS. This result is somehow against the experience that, when corrected with the same MOS algorithm, better input should lead to better or at least not worse output. However, it should be noted that the error's temporal consistency, rather than its magnitude, decides the effects of MOS on the model outputs. When reducing the magnitude of the error of model outputs, the 3Dvar DA process may at the same time violate or increase its consitency to degrade or improve the effects of MOS from case to case. Therefore, such phenomenon is uncorrelated with the inherent or necessary nature of the model, DA as well as MOS processes and will be changed randomly whenever the three changes their setup. The assumption is supported by the evidence that things get very different when the whole experiment is replicated except the space resolution of original anthropogenic emission is changed from 0.1°×0.1° to 0.25°×0.25°. (See Fig. 8, $PM_{2.5}$ and PM_{10} are not plotted here considering the problem was not obvious for them in Fig. 7.) For example, in Fig. 7, SO_2 is predicted better by Sim_DA than Sim_base at forecast 2 h but Sim_DM is beaten by Sim_MOS. When it comes to Fig. 8, however the same comparison lead to an inverse result to see Sim_DM is better than Sim_MOS. Given the fact that experiment only covers a period of only 1 month, it is possible that the

forecast ability of Sim_MOS is slightly worse than or almost the same to Sim_DM if experiment is carried with longer time. However, results from short time experiment, which contains random just like everyday forecast, could still reveal that using MOS and DA together does not gurantee better output than MOS only, to which people should be careful.

4. Conclusion

A comparison between the effect of MOS and DA on improving forecast skill of atmospheric chemistry model was performed in near real operational context. Evaluation according to observations shows that both 3Dvar DA and MOS based on one kalman filtering are effective measures to decrease errors in model forecast.

Forecast with solely MOS (Sim_MOS) performs better than that with solely 3Dvar DA of CICs (Sim_DA). Such superiority of MOS could also be seen in all five species and even extend to 48 h and 72 h forecast where 3Dvar DA of CICs lose its effect in forecast improvement. That is to say, the implementation of MOS rather than 3Dvar DA on CICs is more suitable for the aim of improving operational forecast ability.

Considering the randomness of DA's influence on error consistency, it is not impossible that combined use of both techniques sometimes yields worse forecast than MOS only. The potential degradation, which is expected to be mitigated by long term average, should be paid attention to but is not likely to concern the forecasters for the relatively limited difference yielded.

Given the indications that the refinement of model grid is promising for additional forecast skills (Elbern et al., 2007), future work will concentrate on trying finer model grid as well as different 3Dvar observation error setup in order to improve the effect of

DA on forecast. Of particular interest are species like SO_2 and NO_2 which are unsatisfactorily forecasted by model with 3Dvar. Also, to see if conclusions will be different, we are going to try different DA methods, like 4Dvar and inverse modeling, which are able to adjust model parameters and emissions who work as control parameters in a successful forecast (Schmidt and Martin, 2003; Dubovik et al., 2008). Another work concerns itself with extending the experiment time to explore whether any long‐term statistic properties exists when using DA and MOS techniques together.

Acknowledgements. This work was supported by the State Key Research and Development Program (2016YFC0208504, 2016YFC0203303) and National Natural Science Foundation of China (91544230, 41575145, 41621005 and 41275128).

REFERENCES

Ackermann, I. J., H. Hass, M. Memmesheimer, A. Ebel, F. S. Binkowski, and U. Shankar, 1998: Modal aerosol dynamics model for Europe: Development and first applications. *Atmos. Environ.*, **32**, 2981-2999.

Alexe, M., and Coauthors, 2015: Inverse modelling of CH4 emissions for 2010-2011 using different satellite retrieval products from GOSAT and SCIAMACHY. *Atmos. Chem. Phys.*, **15**, 113-133.

Anadranistakis, M., K. Lagouvardos, V. Kotroni, and H. Elefteriadis, 2004: Correcting temperature and humidity forecasts using Kalman filtering: potential for agricultural protection in Northern Greece. *Atmos. Res.*, **71**, 115-125.

Barbu, A. L., A. J. Segers, M. Schaap, A. W. Heemink, and P. J. H. Builtjes, 2009: A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe. *Atmos. Environ.*, **43**, 1622-1631.

Barret, B., and Coauthors, 2008: Transport pathways of CO in the African upper troposphere during the monsoon season: a study based upon the assimilation of spaceborne observations. *Atmos. Chem. Phys.*, **8**, 3231-3246.

Benedetti, A., and Coauthors, 2009: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation. *J. Geophys. Res.*, **114**.

Bocquet, M., and Coauthors, 2015: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmos. Chem. Phys.*, **15**, 5325-5358.

Denby, B., M. Schaap, A. Segers, P. Builtjes, and J. Horalek, 2008: Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale. *Atmos. Environ.*, **42**, 7122-7134.

Dubovik, O., T. Lapyonok, Y. J. Kaufman, M. Chin, P. Ginoux, R. A. Kahn, and A. Sinyuk, 2008: Retrieving global aerosol sources from satellites using inverse modeling. *Atmos. Chem. Phys.*, **8**, 209-250.

Elbern, H., A. Strunk, H. Schmidt, and O. Talagrand, 2007: Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys.*, **7**, 3749-3769.

Galanis, G., and M. Anadranistakis, 2002: A one-dimensional Kalman filter for the correction of near surface temperature forecasts. *Meteorol. Appl.*, **9**, 437-441.

Geer, A. J., and Coauthors, 2006: The ASSET intercomparison of ozone analyses: method and first results. *Atmos. Chem. Phys.*, **6**, 5445-5474.

Glahn, H. R., and D. A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *Journal of Applied Meteorology*, **11**, 1203-1211.

Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, 2005: Fully coupled "online" chemistry within the WRF model. *Atmos. Environ.*, **39**, 6957-6975.

Henze, D. K., J. H. Seinfeld, and D. T. Shindell, 2009: Inverse modeling and mapping US air quality influences of inorganic PM2.5 precursor emissions using the adjoint of GEOS-Chem. *Atmos. Chem. Phys.*, **9**, 5877-5903.

Honore, C., and Coauthors, 2008: Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV'AIR system. *J. Geophys. Res.*, **113**.

Inness, A., and Coauthors, 2015: Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's Composition-IFS. *Atmos. Chem. Phys.*, **15**, 5275-5303.

Jiang, Z. Q., Z. Q. Liu, T. J. Wang, C. S. Schwartz, H. C. Lin, and F. Jiang, 2013: Probing into the impact of 3DVAR assimilation of surface PM10 observations over China using process analysis. *J. Geophys. Res.*, **118**, 6738-6749.

Li, Z., and Coauthors, 2013: A three-dimensional variational data assimilation system for multiple aerosol species with WRF/Chem and an application to PM2.5 prediction. *Atmos. Chem. Phys.*, **13**, 4265-4278.

Libonati, R., I. Trigo, and C. C. Dacamara, 2008: Correction of 2m-temperature forecasts using Kalman Filtering technique. *Atmos. Res.*, **87**, 183-197.

Liu, Z. Q., Q. H. Liu, H. C. Lin, C. S. Schwartz, Y. H. Lee, and T. J. Wang, 2011: Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia. *J. Geophys. Res.*, **116**.

Makar, P. A., and Coauthors, 2015: Feedbacks between air pollution and weather, Part 1: Effects on weather. *Atmos. Environ.*, **115**, 442-469.

Mizzi, A. P., A. F. Arellano, D. P. Edwards, J. L. Anderson, and G. G. Pfister, 2016: Assimilating compact phase space retrievals of atmospheric composition with WRF-Chem/DART: a regional chemical transport/ensemble Kalman filter data assimilation system. *Geosci. Model. Dev.*, **9**, 965-978.

Parrish, D. F., 1992: The National Meteorological Center's spectral-statistical interpolation analysis system. *Mon. Weather. Rev.*, **120**, 62.

Saide, P. E., G. R. Carmichael, S. N. Spak, P. Minnis, and J. K. Ayers, 2012: Improving aerosol distributions below clouds by assimilating satellite-retrieved cloud droplet number. *P. Natl. Acad. Sci. USA.*, **109**, 11939-11943.

Saide, P. E., and Coauthors, 2015: Central American biomass burning smoke can increase tornado severity in the US. *Geophys. Res. Lett.*, **42**, 956-965.

Schell, B., I. J. Ackermann, H. Hass, F. S. Binkowski, and A. Ebel, 2001: Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. *J. Geophys. Res.*, **106**, 28275-28293.

Schmidt, H., and D. Martin, 2003: Adjoint sensitivity of episodic ozone in the Paris area to emissions on the continental scale. *J. Geophys. Res.*, **108**.

Schwartz, C. S., Z. Q. Liu, H. C. Lin, and S. A. McKeen, 2012: Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. *J. Geophys. Res.*, **117**.

Semane, N., and Coauthors, 2009: On the extraction of wind information from the assimilation of ozone profiles in M,t,o-France 4-D-Var operational NWP suite. *Atmos. Chem. Phys.*, **9**, 4855-4867.

Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang, 1990: The second generation regional acid deposition model chemical mechanism for regional air quality modeling. *J. Geophys. Res.*, **951**, 16343-16367.

Struzewska, J., J. W. Kaminski, and M. Jefimow, 2016: Application of model output statistics to the GEM-AQ high resolution air quality forecast. *Atmos. Res.*, **181**, 186-199.

Tang, X., J. Zhu, Z. F. Wang, and A. Gbaguidi, 2011: Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions. *Atmos. Chem. Phys.*, **11**, 12901-12916.

Tang, X., and Coauthors, 2016: Limitations of ozone data assimilation with adjustment of NOx emissions: mixed effects on NO2 forecasts over Beijing and surrounding areas. *Atmos. Chem. Phys.*, **16**, 6395-6405.

Taylor, A. A., and L. M. Leslie, 2005: A single-station approach to model output statistics temperature forecast error assessment. *Weather Forecast.*, **20**, 1006-1020.

van Loon, M., and Coauthors, 2007: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble. *Atmos. Environ.*, **41**, 2083-2097.

Wang, Y., K. N. Sartelet, M. Bocquet, and P. Chazette, 2014: Modelling and assimilation of lidar signals over Greater Paris during the MEGAPOLI summer campaign. *Atmos. Chem. Phys.*, **14**, 3511-3532.

Wargan, K., and Coauthors, 2015: The global structure of upper troposphere-lower stratosphere ozone in GEOS-5: A multiyear assimilation of EOS Aura data. *J. Geophys. Res.*, **120**, 2013-2036.

Wilson, L. J., and M. Vallee, 2003: The Canadian Updateable Model Output Statistics (UMOS) system: Validation against perfect prog. *Weather Forecast.*, **18**, 288-302.

Yerramilli, A., and Coauthors, 2010: Simulation of Surface Ozone Pollution in the Central Gulf Coast Region Using WRF/Chem Model: Sensitivity to PBL and Land Surface Physics. *Adv. Meteorol.*, 24.

Yin, X., T. Dai, N. A. J. Schutgens, D. Goto, T. Nakajima, and G. Shi, 2016: Effects of data assimilation on the global aerosol key optical properties simulations. *Atmos. Res.*, **178-179**, 175-186.

Yumimoto, K., and Coauthors, 2016: Aerosol data assimilation using data from Himawari-8, a next-generation geostationary meteorological satellite. *Geophys. Res. Lett.*, **43**, 5886-5894.

Zhang, L., and Coauthors, 2016: Sources and Processes Affecting Fine Particulate Matter Pollution over North China: An Adjoint Analysis of the Beijing APEC Period. *Environ. Sci. Technol.*, **50**, 8731-8740.

Zhang, L., and Coauthors, 2015: Source attribution of particulate matter pollution over North China with the adjoint method. *Environ. Res. Lett.*, **10**.

Table 1. Configuration of the model's physical and chemical schemes

	Scheme
Cloud microphysics scheme	WSM 5-Class
Long-wave radiation scheme	RRTM
Short-wave radiation scheme	Goddard
Urban canopy scheme	Off
Surface layer scheme	Modified MM5 Monin-Obukhov
Land surface scheme	Unified Noah
Planetary boundary layer scheme	YSU
Cumulus parameterization scheme	Grell-Devenyi
Chemistry scheme	RADM2/MADE/SORGAM
Photolysis scheme	Madronich F-TUV

Table 2. Site averaged MB, RB, RMSE and Corr for 24 h, 48 h and 72 h forecast of 5 species in Sim_base. Statistics is calculated according to hourly concentration from each of the 207 sites before being averaged. Unit is μg/m^3 for MB and RMSE.

	24 h forecast				48 h forecast				72 h forecast			
	MB	RB(%)	RMSE	Corr	MB	RB(%)	RMSE	Corr	MB	RB(%)	RMSE	Corr
SO_2	−33.49	−35.2	73.39	0.50	−36.54	−38.4	75.42	0.48	−41.78	−43.9	77.82	0.47
NO_2	2.12	3.8	35.45	0.56	0.05	0.1	36.13	0.53	−4.47	−8.1	35.84	0.51
PM_{10}	−136.92	−72.3	173.60	0.52	−140.65	−74.3	177.60	0.50	−145.27	−76.7	182.68	0.47
O_3	39.99	160.8	43.47	0.44	40.08	161.2	43.61	0.44	41.34	166.2	44.75	0.43
$PM_{2.5}$	−72.10	−62.3	104.30	0.60	−74.74	−64.6	107.26	0.57	−78.64	−67.9	111.48	0.54

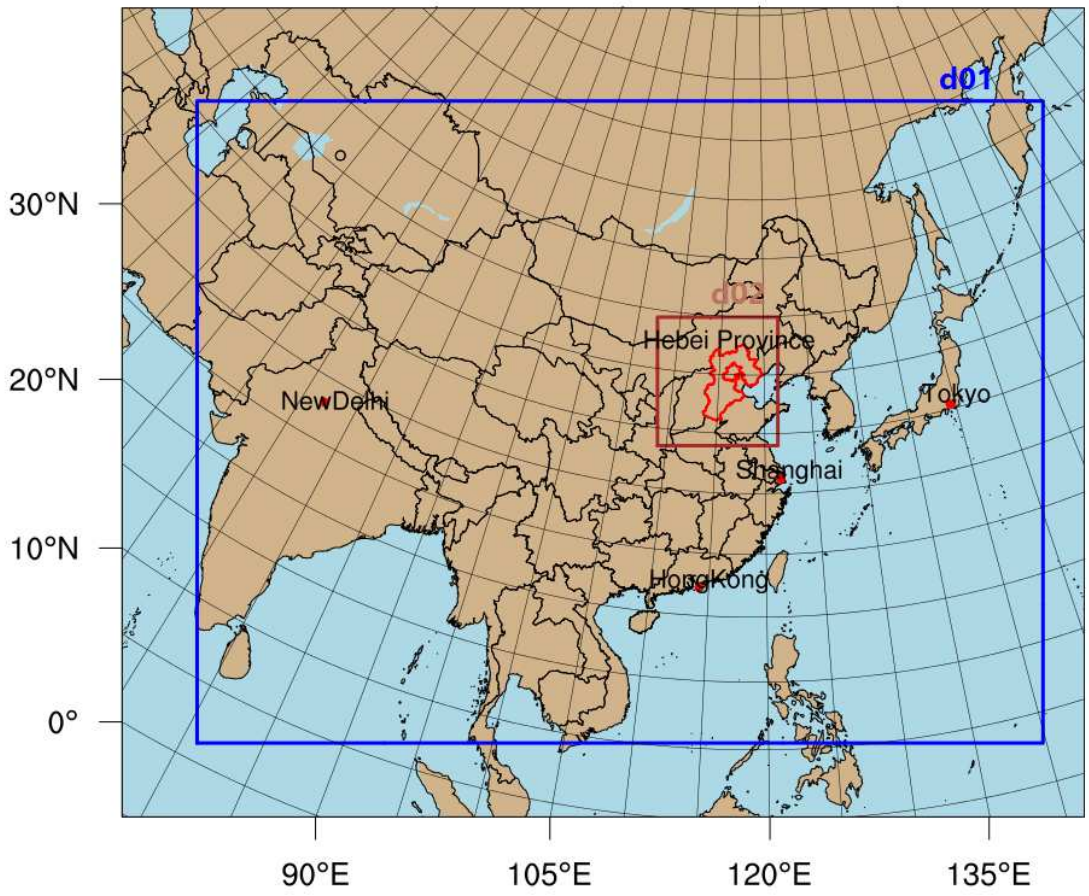


Fig. 1. The nested domains of the WRF-Chem model.

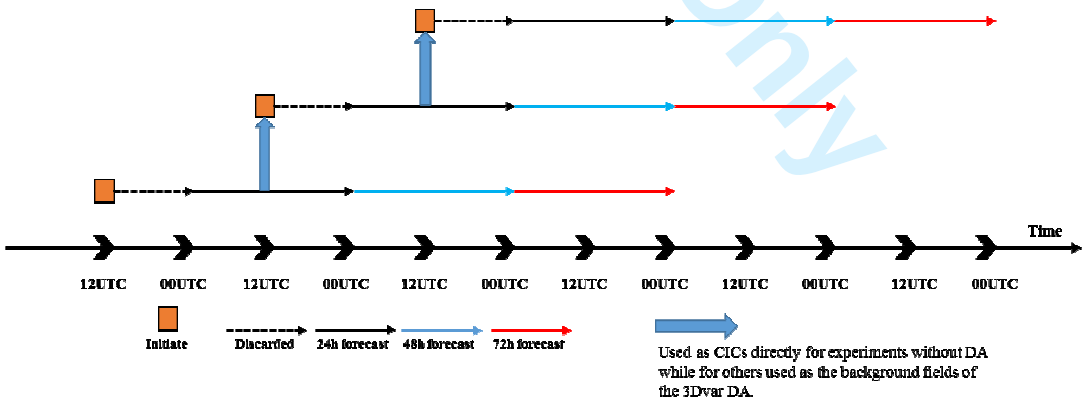


Fig. 2. Time setting of the model in the four experiments.

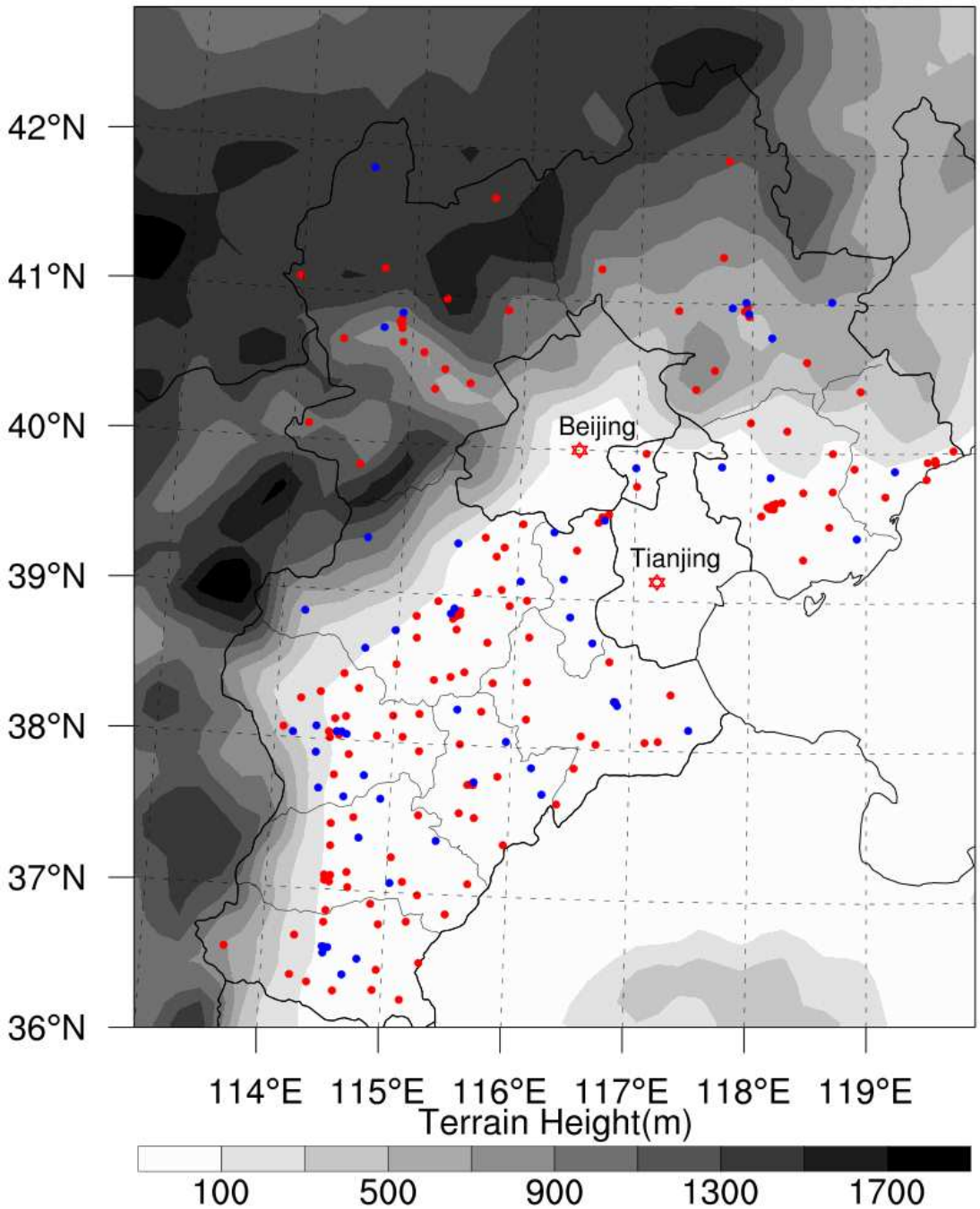
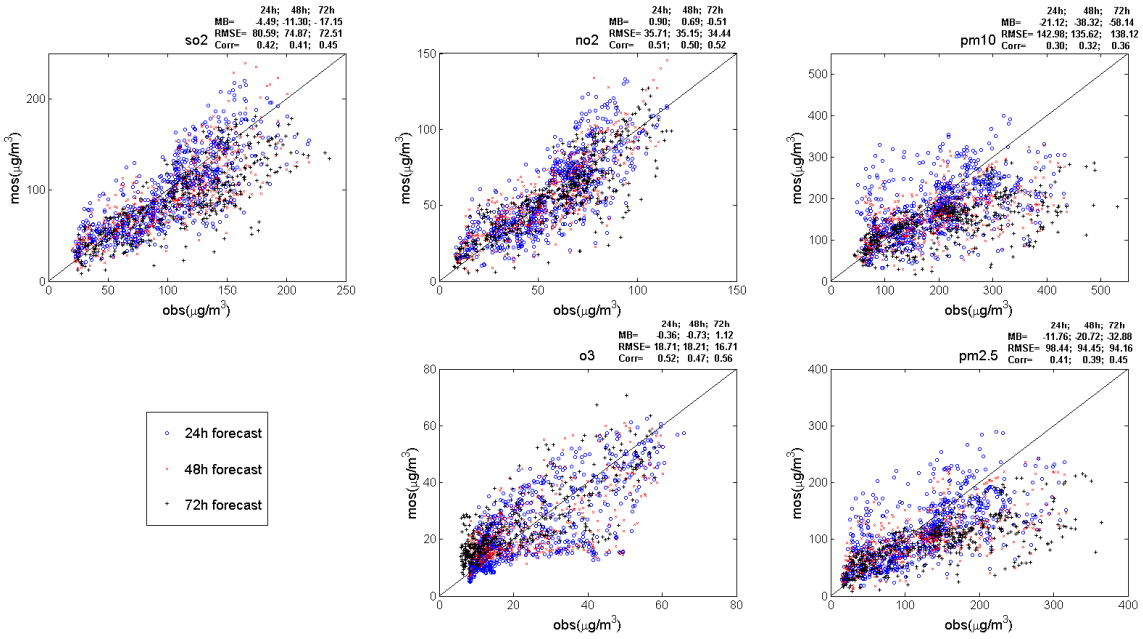


Fig. 3. The terrain height of the Hebei province with monitoring sites plotted as filled dots. Red dots means sites who participated into 3Dvar and MOS process while blue ones were only used in validation of DA effect.



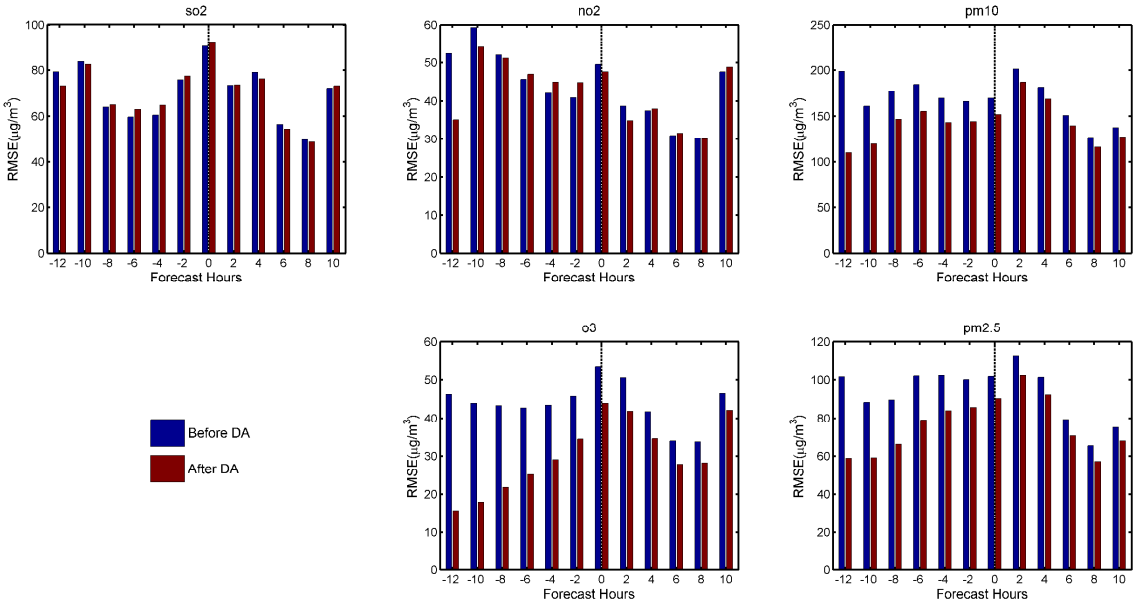
514

515 **Fig. 4.** Hourly concentration simulated by Sim_MOS plotted versus ground station

516 observations averaged over 155 stations. Mean bias and RMSE each forecast period are

517 attached with unit of μg/m^3.

518



519

520 **Fig. 5.** The RMSE change over the integration time from forecast −12 h (right after the

521 DA) to forecast 10 h (already integrated 22 hours) for the five species. All the RMSE

values have been averaged over the 52 stations who have not provided observation into DA.

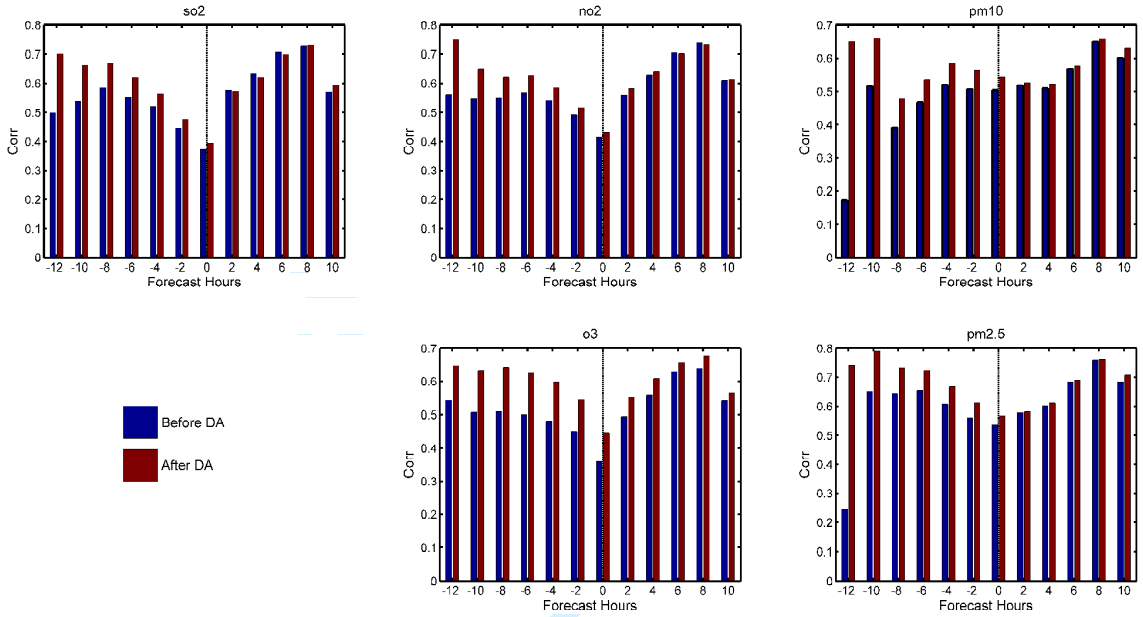


Fig. 6. Same as Fig. 5 except correlation coefficient is plotted instead of RMSE.

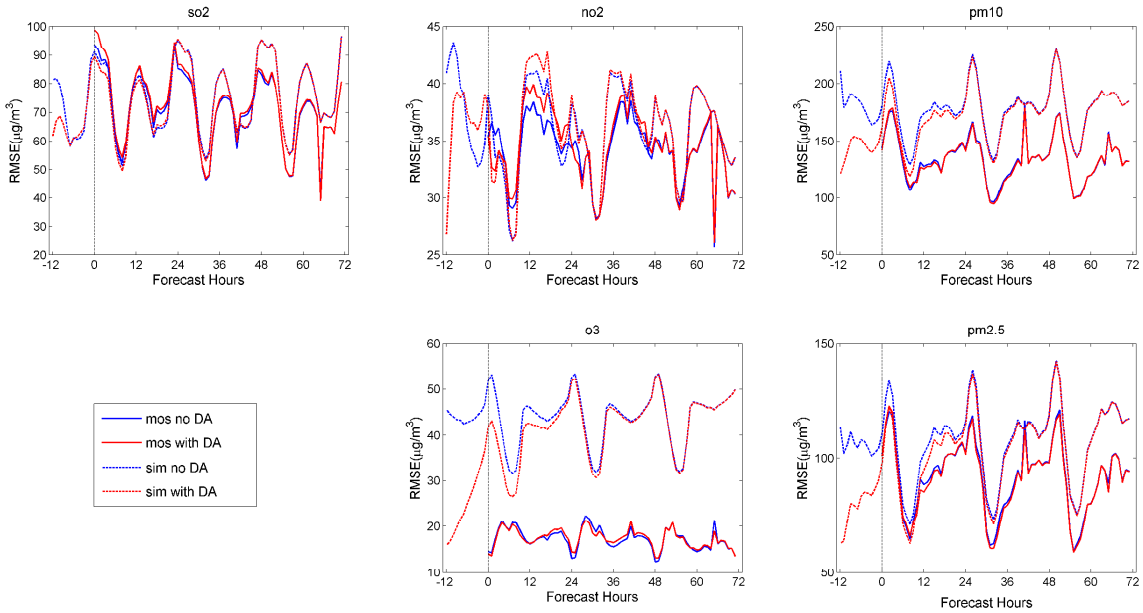


Fig. 7. The RMSE of the 4 experiments and 5 species at different forecast hours from −12 h to 72 h.

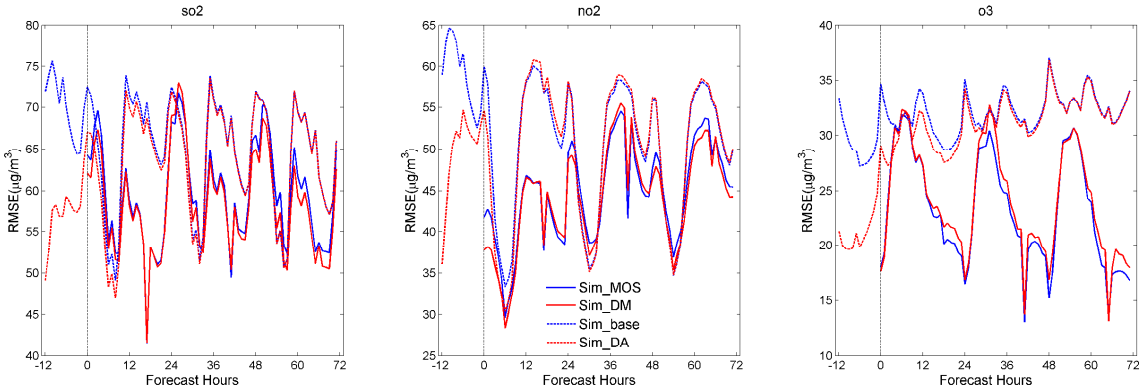


Fig. 8. The RMSE of 3 species at different forecast hours from −12 h to 72 h from the same 4 experiments with solely emission is perturbed.

Appendix A:

For convenience, the full equations of one - dimension Kalman filtering were given below as *k1~k4*:

$$\begin{aligned}
\tilde{x}(t) &= \tilde{x}(t-1) + K(t)\,[\mathrm{y}(t) - \tilde{x}(t-1)\,] \cdots\cdots k1 \\
K(t) &= P_1(t)\,[P_1(t) + V(t)\,]^{-1} \cdots\cdots k2 \\
P_1(t) &= P(t-1) + W(t) \cdots\cdots k3 \\
P(t) &= P_1(t) - K(t)P_1(t) \cdots\cdots k4
\end{aligned} \qquad \text{Eq. (A.1)}$$

These equations will give the best estimate of a sequence of $\mathrm{x}(t)$ when the system equation writes as

$$\mathrm{x}(t) = x(t-1) + w(t) \qquad \text{Eq. (A. 2)}$$

and measurement equation is

$$y(t) = \mathrm{x}(t) + v(t) \qquad \text{Eq. (A. 3)}$$

To complete a single one - dimension kalman filtering iteration, $P(t-1)$ and $\tilde{x}(t-1)$ from the previous calculation are input the equations to be updated by a newly

obtained measurement $y(t)$. Then the result $P(t)$ and $\tilde{x}(t)$ will be saved for the next iteration.

Here there sequences of O_3 concentrations at 12:00 UTC from 3 days' model output will be taken as an example to show the detailed steps of completing one iteration of MOS. It is assumed that, at the day of *t* (here I use *t* to represent "today"), the following variables had been prepared: observation sequence $O_i : i = 1,2 \ldots t$ from ground station, model 24 h forecast sequence and its correction $f_i^{24} : i = 1,2 \ldots t+1$ and $\widetilde{f}_i^{24} : i = 1,2 \ldots t$, model 48 h forecast sequence and its correction $f_i^{48} : i = 1,2 \ldots t+2$ and $\widetilde{f}_i^{48} : i = 1,2 \ldots t+1$, model 72 h forecast sequence and its correction $f_i^{72} : i = 1,2 \ldots t+3$ and $\widetilde{f}_i^{72} : i = 1,2 \ldots t+2$ (model forecast sequence could be get from interpolating the 3D field of model output to the position of site) as well as all the P and $\tilde{x}$ from previous iteration. What we want is to generate $\widetilde{f}_{t+1}^{24}$, $\widetilde{f}_{t+2}^{48}$ and $\widetilde{f}_{t+3}^{72}$ from all variables above.

To get $\widetilde{f}_{t+1}^{24}$, two different approach will be applied simultaneously. In one approach, $x(i) = O_i - f_i^{24} : i = 1,2, \cdots t$ and $y(i) = O_i - f_i^{24} : i = 1,2, \cdots t$ in Eq. (A. 2) and (A. 3), so this approach would be called difference approach hereafter. $v(i)$ in Eq. (A. 3) was set to zero to assume no measurement error, which means V in *k2* equals zero and the estimation of $\mathbb{W}$ would be unnecessary here. The calculation of *k1* to *k2* gives the $\tilde{x}(t)$, and $\widetilde{f}_{t+1,d}^{24} = f_{t+1}^{24} + \tilde{x}(t+1)$ if we assume $\tilde{x}(t+1) = \tilde{x}(t)$. Subscript *d* means the corrected result from difference approach. In another approach called ratio approach, all the things

is the same except $x(i) = O_i / f_i^{24} : i = 1,2,\cdots t$, $y(i) = O_i / f_i^{24} : i = 1,2,\cdots t$ and at last $\widetilde{f}_{t+1,\mathrm{r}}^{24} = f_{t+1}^{24} \times \tilde{x}(t+1)$, where subscript r means ratio approach. The final result $\widetilde{f}_{t+1}^{24}$ will be chosen from $\widetilde{f}_{t+1,d}^{24}$ and $\widetilde{f}_{t+1,r}^{24}$ according to the method described in the last of the appendix.

With $\widetilde{f}_{t+1}^{24}$ obtained, it is now possible to calculate $\widetilde{f}_{t+2}^{48}$. Similar to $\widetilde{f}_{t+1}^{24}$, we still carry the calculation in two approaches: difference and ratio approach. Now, $x(i) = O_i - f_i^{48}$ *or* $O_i / f_i^{48} : i = 1,2,\cdots t$ and $y(i) = \widetilde{f}_i^{24} - f_i^{48}$ *or* $\widetilde{f}_i^{24} / f_i^{48} : i = 1,2,\cdots t+1$. At this time, $v(\mathrm{i})$ could not be set to zero in Eq. (A. 3) and the evaluation of V and $\mathbb{W}$ become necessary in completing the iteration. The estimation of $\mathbb{W}$ and V resembles G02 in which

$$\mathbb{W} = \frac{1}{\mathrm{n}-1}\sum_{1=0}^{n-1}\left[\left(x(t-1) - x(t-1-1)\right) - \left(\frac{\sum_{m=0}^{n-1}\left(x(t-m) - x(t-m-1)\right)}{n}\right)\right]^2 \quad \text{Eq. (A. 4)}$$

$$\mathrm{V} = \frac{1}{\mathrm{n}-1}\sum_{1=0}^{n-1}\left[\left(y(t-1+1) - x(t-1+1)\right) - \left(\frac{\sum_{m=0}^{n-1}\left(y(t-m+1) - x(t-m+1)\right)}{n}\right)\right]^2 \quad \text{Eq. (A. 5)}$$

The variable n in Eq. (A. 4) and Eq. (A. 5) is the number of sequence members participate into the statistics. A value from 7 to 9 is enough for n to generate fairly good MOS effect, which means history data from the past 7~9 days' observation and model output works as the virtual training sample for MOS. After completing the calculation of

k1~k4, $\tilde{x}(t+1)$ will be obtained and then $\tilde{x}(t+2)$, $\tilde{f}^{48}_{t+2,d}$ could be produced replicating the process from the paragraph above. Another ratio approach will also be done to generate $\tilde{f}^{48}_{t+2,\mathrm{r}}$ and $\tilde{f}^{48}_{t+2}$ is chosen from $\tilde{f}^{48}_{t+2,d}$ and $\tilde{f}^{48}_{t+2,\mathrm{r}}$ in the way like determining $\tilde{f}^{24}_{t+1}$.

From $\tilde{f}^{48}_{t+2}$ to obtain $\tilde{f}^{72}_{t+3}$, what have to be done is almost the same as $\tilde{f}^{24}_{t+1}$ to $\tilde{f}^{48}_{t+2}$. Differences mainly lie at the form of some formulations, taking difference approach as example, $x(i) = O_i - f_i^{72} : i = 1,2,\cdots t$, $y(i) = \tilde{f}_i^{48} - f_i^{72} : i = 1,2,\cdots t+2$ and the estimation of V is changed to

$$\mathrm{V} = \frac{1}{\mathrm{n}-1}\sum_{1=0}^{n-1}\left[\left(y(t-1+2)-x(t-1+2)\right)-\left(\frac{\sum_{m=0}^{n-1}\left(y(t-m+2)-x(t-m+2)\right)}{n}\right)\right]^2$$

Eq. (A. 6)

Fig. A.1 is a schematic diagram that shows how the algorithm described above is done step by step.

The final output should be chosen from difference and ratio approach according to their reasonability. From our experience, the difference approach tends to occasionally yield unreasonably low value while the ratio approach sometimes gives result too high. Fortunately the two conditions never happen simultaneously and therefore the final output is from ratio approach when its result is not too high (lower than the yearly averaged value of the species, for example). In time when ratio approach appears too high, the difference approach take the place.

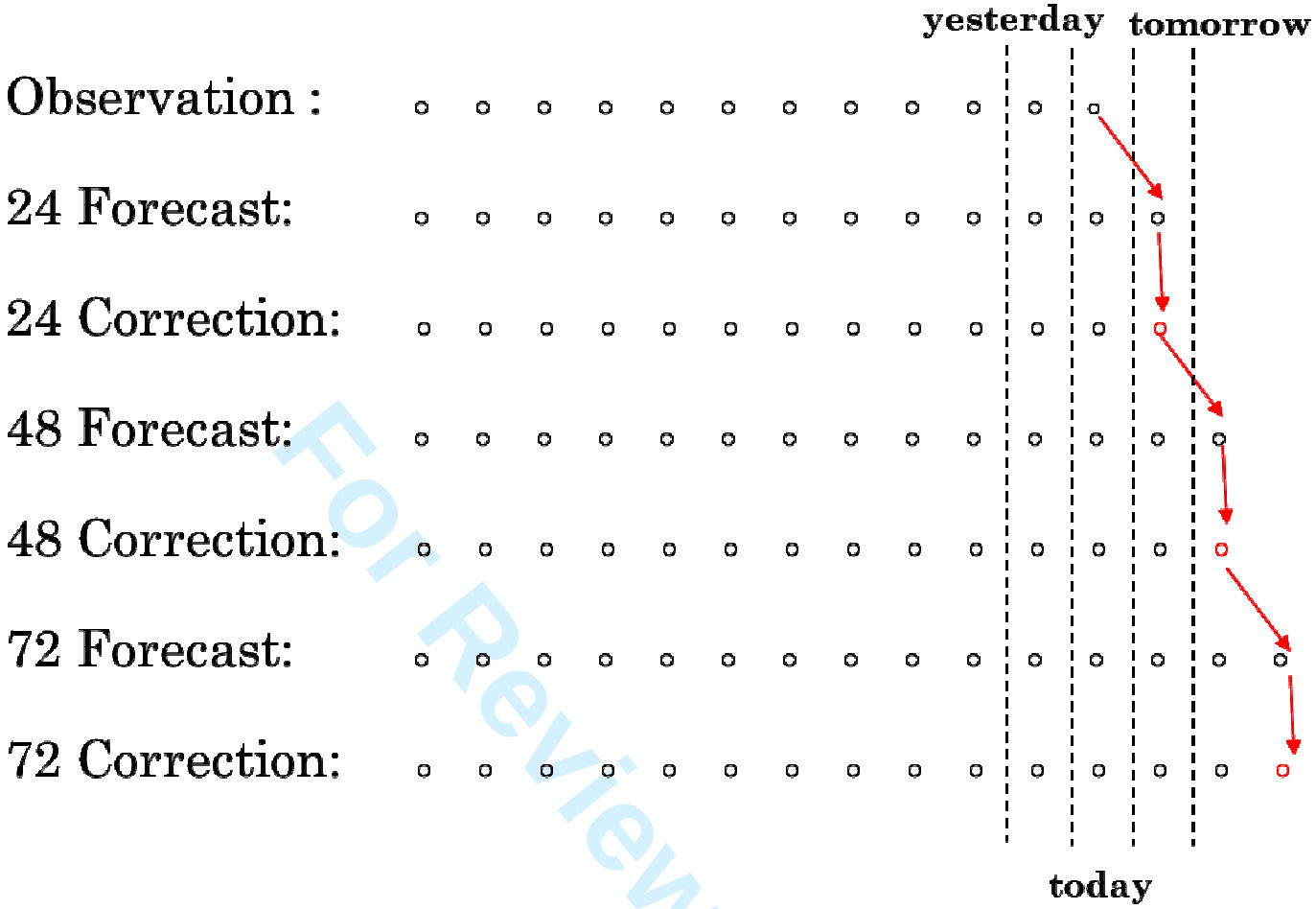


Fig. A. 1. The steps of one-dimensional kalman filtering to apply on 72 h correction. Note that circles from left to right stands for values at the same hour from a day to the next day and those in the same column are valid at the same time and day. Black circles are those exist before today's iteration while red ones are those generated in the calculation of today.