Atmos. Chem. Phys. Discuss., 15, 35693–35718, 2015
www.atmos-chem-phys-discuss.net/15/35693/2015/
doi:10.5194/acpd-15-35693-2015

Atmospheric Chemistry and Physics Discussions

Limitations of ozone data assimilation with adjustment of NO_x emissions: mixed effects on NO_2 forecast over Beijing and surrounding areas

X. Tang[1], J. Zhu[1], Z. F. Wang[1], A. Gbaguidi[2], C. Y. Lin[3], J. Y. Xin[1], T. Song[1], and B. Hu[1]

[1]LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[2]AECOM Asia, Hong Kong, China
[3]Aviation Meteorological Center, Air Traffic Management Bureau, Civil Aviation Administration of China, Beijing, China

Received: 6 September 2015 – Accepted: 13 December 2015 – Published: 18 December 2015

Correspondence to: X. Tang (tangxiao@mail.iap.ac.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This study investigates a cross-variable ozone data assimilation (DA) method based on an ensemble Kalman filter (EnKF) that has been validated as an efficient approach for improving ozone forecasts. The main purpose is to delve into the impacts of the cross-variable adjustment of nitrogen oxides (NO_x) emissions on the nitrogen dioxide (NO_2) forecasts over Beijing and surrounding regions during the 2008 Beijing Olympic Games. A mixed effect on the NO_2 forecasts was observed during the application of the cross-variable assimilation approach in real-data assimilation (RDA) experiments. The method improved the NO_2 forecast over almost half of the urban sites with reductions of the root mean square errors (RMSEs) by 15–36 % in contrast to big increases of the RMSEs over other urban stations by 56–239 %. Over the urban stations with negative DA impacts, improvement of the NO_2 forecasts with 7 % reduction of the RMSEs was noticed during the night and the morning vs. significant deterioration of the forecasts during daytime with 190 % increase of the RMSEs, suggesting the negative DA impacts mainly occurred during daytime. Ideal data assimilation (IDA) experiments with a box model and the same cross-variable assimilation method, as a further investigation, confirmed the mixed effects found in the RDA experiments. An improvement of the NO_x emission estimation was obtained from the cross-variable assimilation under relatively small errors in the prior estimation of NO_x emissions during daytime, while deterioration of the NO_x emission estimation was found under large biases in the prior estimation of NO_x emissions during daytime. However, the cross-variable assimilation improved the NO_x emission estimations during the night and the morning even with large biases in the prior estimations. The mixed effects observed in the cross-variable assimilation, i.e., positive DA impacts on NO_2 forecast over some urban sites, negative DA impacts over the other urban sites and weak DA impacts over suburban sites, were found to be strongly associated with the linearization of the EnKF at the analysis step and the fast variability of the relationship between ozone concentrations and NO_x emissions. When the uncertainties of the daytime ozone concentrations were strongly nonlinearly related

ACPD
15, 35693–35718, 2015

to those of the NO_x emissions, the linearization analysis of the EnKF brought out an inefficient or a wrong adjustment to the NO_x emissions during the daytime. The results of this study provided further explanations for the negative DA impacts observed in previous chemical DA studies and highlighted the limitation of the existing chemical DA methods under high nonlinearity.

1 Introduction

Chemical data assimilation (CDA) that closely integrates models and observations to better represent the chemical state of atmosphere is recognized as a technique for improving the simulations and forecasts of air pollutants such as ozone and aerosols (Carmichael et al., 2008; Sandu et al., 2011; Zhang et al., 2012). The role of CDA in optimizing initial and boundary conditions has been explored in several applications to improve forecasts of ozone and aerosol (Gaubert et al., 2014; Pagowski et al., 2014). Nevertheless, significant challenges persist in CDA.

One of the major challenges in CDA is the divergence of the influences of the initial condition optimization that makes the improvement of air quality forecasts difficult (Carmichael et al., 2008; Gaubert et al., 2014). To overcome such obstacles, emissions with large uncertainties and strong impacts on air quality modeling, identified as the crucial sources of uncertainties and considered to be the key control variables (Beekmann and Derognat, 2003; Hanna et al., 2001), have been integrated into the CDA. The importance of emissions as control variables in the CDA has been also documented recently (Carmichael et al., 2008; Koohkan et al., 2013; Zhang et al., 2012). Accordingly, advanced CDA techniques that enable inverse or cross-variable adjustments of emissions have been established and their applications have brought out notable impacts on improving ozone forecasts (Hanea et al., 2004; Tang et al., 2011).

However, the performances of such advanced CDA on the forecasts of other pollutants related to ozone are rarely reported and have not aroused enough attention. In this field, few studies stand out (Elbern et al., 2007; van Loon et al., 2000). El-

ACPD
15, 35693–35718, 2015

bern et al. (2007) carried out two sets of data assimilation experiments with a four dimensional variational inversion method: (1) assimilation of ozone (O_3) and nitrogen dioxide (NO_2) observations simultaneously, and (2) assimilation of only O_3 observations. Both experiments resulted in reductions of nitrogen oxides (NO_x) emissions after data assimilation in most cases even if the model underestimated the NO_x concentrations before data assimilation. Similar results were reported by van Loon et al. (2000) through the assimilation of O_3 observations and adjustments of sulfur oxides (SO_x) emissions using an ensemble Kalman filter. The method enhanced the emission rates of SO_x when significant over-prediction of SO_2 concentrations existed. Such inconsistencies, i.e., the emissions enhanced under the overestimation of concentrations or the emissions reduced under the underestimation of concentrations, indicates gaps between ozone forecast improvement and precursor emission optimization exist and calls for a comprehensive evaluation of the cross-variable chemical data assimilation techniques.

Tang et al. (2011) employed a high horizontal resolution (9 km) model to perform the assimilation of O_3 observations with the ensemble Kalman filter and the adjustment of NO_x emissions for O_3 forecast improvement over Beijing and its surrounding areas. However, the impact of ozone assimilation on the precursor (NO_2 & volatile organic compounds) uncertainty was not elucidated. This paper, as an extension of Tang et al. (2011), based on the assimilation experiments performed by Tang et al. (2011), attempts to analyze in detail the impacts of the cross-variable ozone data assimilation on NO_2 forecasts over Beijing and surrounding areas during the 2008 Beijing Olympic Games. Both real O_3 data assimilation (with a 3-dimensional chemical transport model) and ideal O_3 data assimilation experiments (with a box model) are performed to delve into the state of NO_2 and NO_x emissions during assimilation processes in order to provide further insights into the scientific potential of the assimilation method.

Section 2 describes the chemical transport model employed, the data assimilation algorithm and the surface observation network. Results from the real data assimila-

ACPD
15, 35693–35718, 2015

tion experiments and the ideal data assimilation experiments are presented in Sect. 3. Section 4 presents the conclusions and discussions of this study.

2 Methodology

2.1 Chemical transport model

The chemical transport model used for O_3 simulations is the Nested Air Quality Prediction Modeling System (NAQPMS) (Wang et al., 2001). Several applications of NAQPMS have been reported for simulating the chemical processes and transports of ozone, modeling the processes of aerosol and acid rain, and providing operational air quality forecasts in megacities such as Beijing and Shanghai (Wang et al., 2006). It contains modules for modeling the processes of emissions, advection, diffusion, dry and wet deposition, gaseous phase, aqueous phase, heterogeneous and aerosol chemical reactions. The gas-chemistry processes are simulated by the Carbon-Bond Mechanism Z (CBM-Z) which includes 133 reactions for 53 species (Zaveri and Peter, 1999). The dry deposition modeling follows the scheme of Wesely (1999). The vertical eddy diffusivity is parameterized based on a scheme by Byun and Dennis (1995). The O_3 simulations are configured with three nested domains and the horizontal resolutions are 81, 27 and 9 km respectively. The first domain covers East Asia with 81 km resolution and the second domain contains North China with 27 km resolution. The third domain displayed in Fig. 1 covers Beijing and its surrounding areas with 9 km resolution. Vertically, the model is set as twenty terrain-following layers, nine of which are within the lowest 2 km of the atmosphere and the height of the first layer near the surface is 50 m. The Fifth-Generation National Center for Atmospheric Research (NCAR)/Penn State Mesoscale Model (MM5; Grell et al., 1994) was employed to provide the hourly meteorological inputs for NAQPMS. The regional emission data of the Intercontinental Chemical Transport Experiment-Phase B (INTEX-B) Asia inventory for 2006 with

ACPD
15, 35693–35718, 2015

$0.5^{\circ} \times 0.5^{\circ}$ resolution (Zhang et al., 2009) and the local high-resolution emission inventory were combined to provide the emission data for NAQPMS (Tang et al., 2011).

2.2 Data assimilation algorithm

The assimilation algorithm employed is the ensemble Kalman filter (EnKF) proposed by Evensen (1994). The main feature of this method is a series of ensemble samples generally produced via ensemble forecasts to calculate the background error covariance of state variables. It serves as an approximate version of the Kalman filter (Kalman, 1960). The simplicity in calculating error covariance from ensemble forecasts fully supports nonlinear evolution of a model that is very suitable for data assimilation in complex high-dimensional models (Carmichael et al., 2008). Its implementation is very simple without needing for an adjoint model which is a very cumbersome task for complex high-dimensional models. It can be used for combined state and parameter estimation (Evensen, 2009). In the field of air pollution, the EnKF has been shown to be an efficient method in optimizing concentrations. Further applications of the EnKF in improving forecast skills of dust and ozone through emission optimization have been reported (e.g., Hanea et al., 2004; Lin et al., 2008; Tang et al., 2011; van Loon et al., 2000).

In the present study, the EnKF is employed to assimilate ozone observations for the corrections of NO_x emissions. The main purpose is to elucidate the performances of that method during the cross-variable assimilation of O_3 observations. The sequential algorithm proposed by Houtekamer and Mitchell (2001), as a variant of EnKF, was adopted for its efficiency in computation. The first step of the implementation is to perturb ozone concentrations, NO_x emissions and other key uncertainty sources of ozone modeling, i.e., photolysis rates and vertical diffusion coefficients, as described

ACPD
15, 35693–35718, 2015

by the following equations:

$$\boldsymbol{x}'(i) = \boldsymbol{x}^{\mathrm{b}} + \boldsymbol{\zeta}(i), \quad i = 1,2,\ldots,N \tag{1}$$

$$\boldsymbol{e}'(i) = \boldsymbol{e}^{\mathrm{b}} + \boldsymbol{\varepsilon}(i), \quad i = 1,2,\ldots,N \tag{2}$$

$$\boldsymbol{q}'(i) = \boldsymbol{q}^{\mathrm{b}} + \boldsymbol{\phi}(i), \quad i = 1,2,\ldots,N \tag{3}$$

where $\boldsymbol{x}$, $\boldsymbol{e}$, and $\boldsymbol{q}$ are ozone concentrations, emissions, and other parameters (NO_2 photolysis rates and vertical diffusion coefficients), and the superscript b represents their background values in the model. The superscript $'$ represents the ensemble samples of these variables after perturbing the background values by random samples of $\boldsymbol{\zeta}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\varphi}$. N is the ensemble size, set as 50 and taken as a choice for a balance between computational efficiency and assimilation performance.

The estimations of the uncertainty in this study are mainly based on the results of Tang et al. (2010a) who provided a detailed uncertainty analysis for the ozone forecasts over Beijing and surrounding areas during the 2008 Beijing Olympic Games. The uncertainty of NO_x emissions is estimated on the basis of the uncertainty estimation of NO_x emissions (31 %) given by the INTEX-B Asia inventory for 2006 with a $0.5^\circ \times 0.5^\circ$ resolution (Zhang et al., 2009). Because the downscaling processes of the emission data from the $0.5^\circ \times 0.5^\circ$ resolution of the inventory to 9 km × 9 km resolution of the model and the changes of emissions over Beijing and surrounding areas during the 2008 Beijing Olympic Games would induce considerable new uncertainties into the emission inventory of model, we estimated the uncertainty of the NO_x emissions to be 60 % of the first guess emission rates, about twice the uncertainty in the INTEX-B Asia inventory. The uncertainties of vertical diffusion coefficients in ozone modeling have been estimated by Beekmann and Derognat (2003), Hanna et al. (1998) and Moore et al. (2001), ranging from 25 to 50 %. We estimated the uncertainty of vertical diffusion coefficients to be 35 % of the first guess values which are close to the average estimation of the above three estimations. Also with reference to the studies of Hanna et al. (1998) and Moore et al. (2001), the uncertainty of the modeled photolysis rates was estimated to be 30 %. The uncertainty of the modeled O_3 concentrations at the

ACPD
15, 35693–35718, 2015

initial time is estimated to be 50 % after comparing the modeled O_3 concentrations with observations. Based on the method suggested by Evensen (1994), the perturbations of the variables in three dimensions are implemented through adding a pseudo smooth random field. The horizontal and vertical scales of initial error correlations can be well controlled using this method. The scales are set as 54 km in the horizontal and 3 model grids in the vertical (approximately 200 m) as Tang et al. (2011).

Ensemble samples of the emissions, the vertical diffusion coefficients, the photolysis rates and the O_3 concentrations provide various estimations for the four variables (different from their initial estimations) and are used to derive ensemble forecast of ozone. In order to achieve cross-variable adjustment for NO_x emissions, an extended state variable is defined as:

$$\boldsymbol{U}'(i) = \begin{vmatrix} \boldsymbol{x}'(i) \\ \boldsymbol{e}'(i) \end{vmatrix}, \quad i = 1, 2, \ldots, N, \tag{4}$$

where $\boldsymbol{x}'(i)$ and $\boldsymbol{e}'(i)$ represent the ozone concentrations and the emissions after perturbations as in Eq. (1). Through the ensemble forecast $\boldsymbol{x}'(i)$ is strongly dependent on $\boldsymbol{e}'(i)$, which makes it convenient for estimating the correlation between $\boldsymbol{x}$ and $\boldsymbol{e}$ and for cross-variable adjustment of NO_x emissions. The background error covariance of the extended variable can be directly calculated from the ensemble forecast results during the simulation period:

$$\boldsymbol{P} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\boldsymbol{U}'(i) - \overline{\boldsymbol{U}'}\right) \left(\boldsymbol{U}'(i) - \overline{\boldsymbol{U}'}\right)^T, \tag{5}$$

where $\overline{\boldsymbol{U}'}$ is the mean of the ensemble samples of the extended state variable and N is the ensemble size.

This algorithm treats the observations as random variables and perturbs them to prevent filter divergence of the EnKF (Houtekamer and Mitchell, 1998). When ozone observations are available, they are perturbed according to the observation errors (Gaus-

ACPD
15, 35693–35718, 2015

sian with mean zero and covariance R, including both measurement errors and representativeness errors):

$$\boldsymbol{y}'(i) = \boldsymbol{y} + \boldsymbol{\Upsilon}(i), \quad i = 1, 2, \ldots, N \tag{6}$$
$$\boldsymbol{\Upsilon} \in N(0, \boldsymbol{R}).$$

As suggested by von Loon et al. (2000), the observation errors are assumed to be within 10 % of the original observation value and uncorrelated in time and space. It is worth noting that some other variants of the EnKF (e.g., the ensemble square root filter (EnSRF) proposed by Whitaker and Hamill, 2002) do not need the perturbations on observations but can also provide accurate analyses.

Then the ensemble samples of the extended variables from the ensemble forecasts can be updated through assimilating the ozone observations:

$$\boldsymbol{U}^a(i) = \boldsymbol{U}(i) + \boldsymbol{K}(\boldsymbol{y}'(i) - \boldsymbol{H}\boldsymbol{U}(i)), \quad i = 1, 2, \ldots, N \tag{7}$$
$$\boldsymbol{K} = \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T + \boldsymbol{R})^{-1}, \tag{8}$$

where $\boldsymbol{H}$ represents a linear operator mapping the extended state variable from model space to observational space, and $\boldsymbol{K}$ is the Kalman weight calculated based on the background error covariance and the observation error covariance. $\boldsymbol{U}^a(i)$ is the updated ensemble sample of the extended state variable and used for the sequential ozone forecast. The updating of the ensemble ensembles of the extended variables is conducted one time every 1 hour (1 h), and the updated NO_x emissions are then used for the NO_2 forecast of the next hour.

2.3 Surface observation network

We employed a regional surface air quality network over Beijing and its surrounding areas during the 2008 Beijing Olympic Games including 17 stations established by the Beijing Environment Monitoring Center and Chinese Academy of Science (Xin et al.,

ACPD
15, 35693–35718, 2015

2010). Figure 1 displays the distributions of these stations and the non-industrial NO_x emission rates of the observation regions in the third model domain. As can be seen, 11 urban stations (CP, PEK, BY, IAP, YF, BD, CZ, QHD, SJZ, TS, TJ) are located in the urban areas with high non-industrial NO_x emission rates, and the other 6 (LF, XH, XL, YJ, YuF, YLD) are in the suburban areas with relatively low non-industrial NO_x emission rates. The network provides observations of O_3 and NO_2 at the same temporal resolution as the model, i.e., 1 h. In order to independently validate the assimilation results, three of the observation stations are withdrawn from the assimilation and used for the validation. NO_2 observations not used in the assimilation are also used for assessing the impacts of the cross-variable assimilation on the NO_2 forecasts.

3 Results

3.1 Real data assimilation experiment

The real data assimilation (RDA) experiment assimilates the surface ozone observations over Beijing and surrounding areas to adjust the NO_x emissions over these areas in the NAQPMS. The experiment is based on the study of Tang et al. (2011) in which the assimilation of real O_3 observations with the EnKF is performed to correct NO_x emissions. The experiment focuses on a two-week period from 00:00 LT 9 August to 00:00 LT 23 August in 2008. The initial conditions of the simulation are from a two-week spin-up model run. The ozone initial conditions, NO_x emissions and vertical diffusion parameters are perturbed at 19:00 LT on 8 August 2008 according to the Eqs. (1), (2) and (3) and used to derive ensemble runs of NAQPMS. After 5 h free ensemble runs and at 00:00 LT on 9 August, the observed ozone data starts to be hourly assimilated into the third model domain (displayed in Fig. 1) of NAQPMS to adjust the NO_x emissions. Adjusted factors of the NO_x emissions are then used for the NO_2 forecast of the next hour. Both daytime and nighttime observations are assimilated. Considering possible large errors in the modeling of vertical profiles of air pollutants, we only adjust

ACPD
15, 35693–35718, 2015

the variables in the first three vertical layers near the surface, which can reduce the influence of the modeling errors of vertical mixing on data assimilation. A free run of NAQPMS without data assimilation (NonDA), as a reference run, is also performed for the validation of the assimilation results of the RDA experiment.

Figure 2 compares the root mean square errors (RMSEs) of the 1 h NO_2 forecast at the 17 stations in the RDA experiment with the RMSEs in the NonDA experiment. The impacts of the cross-variable adjustment of the NO_x emissions through ozone data assimilation on the NO_2 forecast vary substantially from suburban to urban stations. Over suburban sites, the assimilation shows minor influence on NO_2 forecasts, and the RMSEs in the RDA experiment are close to those in the NonDA experiment. This is likely due to the low emission rates of NO_x over suburban areas, so that the predicted NO_2 concentrations could not undergo significant changes through the data assimilation process. Over urban sites, the NO_2 predictions are significantly affected by the data assimilation. Over the stations of BD, PEK, CZ, QHD, SJZ, and TS, the RMSEs of NO_2 forecast are reduced by 15–36 % after data assimilation, resulting in improvement of NO_2 forecasts in contrast to large increase within 56–239 % of the RMSEs over the stations of CP, BY, IAP, YF and TJ. This result suggests a mixed effect on the NO_2 forecast from the adjustment of the NO_x emission by the ozone data assimilation, i.e., weak DA impacts over suburban sites, positive DA impacts over some urban sites and negative DA impacts over the other urban sites. Nevertheless, the assimilation produced significant improvement of ozone forecasts over all these sites, as reported by Tang et al. (2011).

Further investigations are conducted on the mixed effects of the data assimilation (weak DA impacts over suburban sites, positive DA impacts over some urban sites and negative DA impacts over the other urban sites) on the NO_2 forecast during various periods. Figure 3 a–c displays the daily variation of the 1 h NO_2 forecast RMSEs in the RDA experiment and the NonDA experiment over the urban stations with positive DA impacts (CZ, PEK, QHD, SJZ, and TS), the urban sites with negative DA impacts (BY, CP, IAP, TJ and YF) and the suburban stations with weak DA impacts (LF, XH, YLD,

ACPD
15, 35693–35718, 2015

YJ and YuF) respectively. Over the suburban stations, the cross-variable assimilation also brings out very weak impacts on the NO_2 forecast during both the day and the night. Over the urban stations with positive DA impacts, the cross-variable assimilation presents consistent positive DA impacts during both the day, the night and the morning, with 23 % reductions of RMSEs during the daytime and 21 % reductions during the night and the morning.

Over the urban sites with negative DA impacts, the performance of the data assimilation varies from the day to the night and the morning. Adjusting NO_x emissions improves the forecasts of NO_2 concentrations during most of the night and the morning time and reduces the RMSEs by 7 % in contrast to the deterioration of the forecast in daytime with 190 % increase of the RMSEs. This result suggests that the impacts of the cross-variable assimilation on the NO_2 forecast during daytime are opposite to those during the night and the morning at these urban sites. It also highlights that the negative DA impacts mainly occurs during the daytime. As described by Tang et al. (2010b), daytime ozone is strongly nonlinearly related to high NO_x emissions over urban areas (in particular over central Beijing), whereas nighttime ozone is mainly controlled by the titration reaction of O_3-NO with weak nonlinearity. Due to the obvious discrepancy between the daytime ozone chemistry and the nighttime ozone chemistry, further experiments are carried out in Sect. 3.2 to elucidate the impacts of ozone chemistry on the cross-variable assimilation.

Another phenomenon observed in Fig. 3a and b is that the forecast errors of NO_2 forecasts with the free run of model during the night and the morning are much higher than those during the daytime. This may be related to the large uncertainties in the modeling of the nighttime boundary layer over urban regions (Kleczek et al., 2014). Although the modeling of vertical diffusion is taken as a key uncertainty source in our data assimilation, its uncertainty is not constrained by the data assimilation. Therefore, high errors still exist in the nighttime NO_2 forecasts after data assimilation, as shown in Fig. 3a and b.

ACPD
15, 35693–35718, 2015

3.2 Ideal data assimilation experiment

To perform an in-depth analysis on the mixed effects observed in the RDA experiment, ideal data assimilation (IDA) experiments with known true state of ozone concentrations and NO_x emissions are carried out. To closely monitor the impacts of ozone chemistry on the cross-variable assimilation method used in the RDA experiment, a simplified box model including the main chemical processes of NAQPMS is employed (Xiang et al., 2010). It does not take into account the complex transport processes and simulates the removal processes through multiplying the concentrations by removal coefficients. The experiments with the box model focus on the IAP station where a negative impact on NO_2 forecasts is observed in the RDA experiment. Emission rates and meteorological parameters are from the inputs used by the NAQPMS. The prior emission rates from the NAQPMS and their corresponding O_3 concentrations modeled with the box model are assumed to be the true state of emissions and concentrations and would be used for the validation of the optimized emissions from data assimilation.

At the first step, the IDA experiments focus on the negative DA impacts on the daytime NO_2 forecasts. Artificial O_3 observations are generated through adding small random errors to the true state of O_3 concentrations. To be consistent with the RDA experiment, the random errors for perturbing observations are also assumed to be within 10 % of the true value. Three error scenarios for the NO_x emissions, i.e., with 10, 30 and 50 % underestimations respectively, are assumed and separately used for three simulations of the box model. In order to avoid dealing with complex model errors, the errors in the NO_x emissions are assumed to be the only error sources of ozone modeling. For each error scenario, cross-variable adjustment of the NO_x emissions through assimilating the artificial O_3 observations with the EnKF is conducted. Figure 4a–c shows the O_3 concentrations and NO_x emissions before and after DA and also their ensemble samples before DA at 12:00 12 August 2008.

Figure 4a presents the results under the first scenario with 10 % underestimation of NO_x emissions (S1), and the analyzed O_3 concentration and NO_x emission after

ACPD
15, 35693–35718, 2015

DA are close to their true state, suggesting an improvement of the NO_x emission estimation from the cross-variable assimilation. Figure 4b shows the results under the second scenario with 30 % underestimation of NO_x emissions (S2). The DA also reduces the error in the NO_x emission but this is not done efficiently, since large errors (about 20 %) still persist in the optimized NO_x emission. The ensemble samples of O_3 concentrations shown in Fig. 4b are obtained from the ensemble runs of the box model that are derived by the ensemble samples of the NO_x emissions (also shown in Fig. 4b). We can see that the ensemble forecasts of O_3 concentrations present high nonlinear responses to the perturbations of the NO_x emissions. This suggests that the EnKF with Monte Carlo simulations can well forecast the nonlinear evolutions of error statistics of the O_3 modeling. However, at the analysis step, the ensemble samples of the O_3 concentrations and the NO_x emissions are combined by the EnKF to produce linear correlations between them during the calculation of the background error covariance in Eq. (5), which is a linearization process. The linearized relationship between the O_3 concentrations and the NO_x emissions is presented in Fig. 4b. We can see noticeable discrepancies between the nonlinear relationship denoted by the ensemble samples and the linearized relationship at the analysis step. This significantly weakens the performance of the EnKF in the cross-variable adjustment.

Under the third scenario (S3) with the NO_x emissions underestimated by 50 %, enhanced deterioration of the NO_x emission estimations is observed in Fig. 4c. The DA adjusts the simulated O_3 concentration close to the true state, but induces additional bias to the previously underestimated NO_x emission. Such negative DA impact on the NO_x emission estimation is similar to the phenomenon observed for the daytime NO_2 forecast over some urban stations in the RDA experiment. With the true state of O_3 concentrations and NO_x emissions as known in the IDA experiment, we can find that the most plausible cause of such negative DA impact on the NO_x emissions is the linearization process at the analysis step of the EnKF for the strong nonlinear relationship between the uncertainty of O_3 concentrations and that of NO_x emissions. With a large bias in the prior estimation of the NO_x emissions, the cross-variable assimilation may

ACPD
15, 35693–35718, 2015

induce an increase of the bias in the NO_x emissions. The results of the three IDA experiments, i.e., positive DA impact under the first and second scenarios and negative impact under the third scenario, confirm the mixed effects of the cross-variable assimilations observed in the RDA experiments, and suggest a strong link between the mixed effects and the linearization process at the analysis step of the EnKF for the strong nonlinear chemical relationships.

To investigate the DA impacts on the NO_x emissions at night and morning hours, the O_3 concentrations and NO_x emissions before and after DA and their ensemble samples before DA at 08:00 13 August 2008 (representing the morning time) are shown in Fig. 5a–c. Similar figures are obtained for other night and morning times and not shown here. As can be seen in Fig. 5a–c, different level errors (10, 30 and 50 % underestimations) in the NO_x emissions are significantly reduced through the cross-variable assimilation with the EnKF. The ensemble forecasts of the morning O_3 concentrations show near-linear responses to the uncertainties (or perturbations) of the NO_x emissions, which makes the linearization of the EnKF at the analysis step work properly to correct the biases in the NO_x emissions.

The positive DA impacts on the NO_x emission estimation in the IDA experiments during the night and the morning are consistent with the improvement of the NO_2 forecasts after data assimilation in the RDA experiment during the night and the morning. Compared with the mixed effects of the DA during the daytime, the positive DA impacts during the night and the morning in both the RDA and IDA experiments indicate that the assimilation of O_3 observations with the EnKF might be useful in optimizing NO_x emissions and NO_2 forecasts during the night and the morning. Furthermore, the ensemble forecasts of O_3 concentrations show strong nonlinear responses to the perturbations of the NO_x emissions during the daytime in Fig. 4a–c but present near-linear responses during the night and the morning in Fig. 5a–c, which suggests a fast variability of the relationships between the uncertainties of O_3 concentrations and those of NO_x emissions and leads to the different DA impacts during different periods of the day. It should be pointed out that the IDA experiments do not consider the complex model errors

ACPD
15, 35693–35718, 2015

(e.g., errors in boundary layer modeling). In the real case, model errors exist and may be larger than the errors in the NO_x emissions. In such situations, the cross-variable adjustment of the NO_x emission would be more complicated and challenging, because the DA scheme needs to quantify the uncertainties of both the NO_x emissions and other key model processes in order to avoid over-adjustment of the NO_x emissions to compensate the missing model errors except for dealing with the nonlinear relationships between the assimilated observations and the adjusted variables.

4 Conclusion and discussion

The impacts of the cross-variable adjustment of NO_x emissions on NO_2 forecasts were investigated through assimilating O_3 observations with a variant of the EnKF (proposed by Houtekamer and Mitchell, 2001) over Beijing and surrounding areas during the 2008 Beijing Olympic Games. Both real data assimilation experiments with a 3-dimensional chemical transport model and ideal data assimilation experiments with a simplified box chemical model were performed.

The results of the data assimilation experiments highlighted a mixed effect of the cross-variable assimilation with the EnKF in which the NO_x emissions are adjusted through assimilating O_3 observations. The cross-variable assimilation worked properly in improving the NO_2 forecasts and optimizing NO_x emissions during the night and the morning when the uncertainties of O_3 concentrations are almost linearly related to those of NO_x emissions. During daytime, the data assimilation resulted in positive DA impacts on NO_2 forecasts over some urban sites, negative DA impacts over the other urban sites and weak DA impacts over suburban sites. These mixed effects are found to be strongly associated with the rapid variations of the relationships between ozone uncertainties and NO_x emission uncertainties and the negative impacts of the linearization analysis of the EnKF under strong chemical nonlinearities during daytime. This revealed some critical limitations of the EnKF in its application to the cross-variable chemical data assimilation despite its representation of the fully nonlinear evolution of

ACPD
15, 35693–35718, 2015

the model and strong performance for improving ozone forecasts (e.g., van Loon et al., 2000; Tang et al., 2011). Assimilation approaches that enable dealing with high nonlinear problems in both model evolution and analysis step are needed. Particle filters as a nonlinear filter method (e.g., Moral et al., 1996; van Leeuwen, 2009, 2010) might have potential in this field if its limitation in application for high dimensional system (Stordal et al., 2011) can be overcome.

For the application of the EnKF in chemical data assimilation, due to the needs of linearization at analysis step, the assimilation should avoid the linearization of a strong nonlinear relationship such as the relations between O_3 concentrations and NO_x emissions. Alternatively, using the observations of air pollutants that are linearly or weakly nonlinearly related to the optimized emissions would reduce the negative impacts of the linearization analysis of the EnKF assimilation. As for optimizing the NO_x emissions discussed in this study, the assimilation of NO_2 observations with the EnKF may avoid linearizing a strong nonlinear relationship and might be useful in improving NO_2 forecasts and NO_x emission estimations. Nevertheless, how to deal with the strong nonlinear problem is a challenge in chemical data assimilation, and further explorations on this subject are needed. For example, the nonlinear relationships between concentrations and emissions might change with model resolution and affect the performance of the inverse estimation of emissions at different model resolutions. Except for inversely estimating emissions, state estimation of air pollutants through assimilation of observations that are not directly related to the estimated variable, e.g., state estimation of volatile organic compounds through assimilation of ozone observations, may also need to deal with the chemical nonlinear problem.

Acknowledgements. This study was funded by the CAS Strategic Priority Research Program (Grant No. XDB05030200) and the National Natural Science Foundation (Grant No. 41205091).

ACPD
15, 35693–35718, 2015

References

Beekmann, M. and Derognat, C.: Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements fromthe Atmospheric Pollution Over the Paris Area (ESQUIF) campaign, J. Geophys. Res., 108, 8559, doi:10.1029/2003JD003391, 2003.

Byun, D. W. and Dennis, R.: Design artifacts in Eulerian air quality models: evaluation of the effects of layer thickness and vertical profile correction on surface ozone concentrations, Atmos. Environ., 29, 105–126, 1995.

Carmichael, G., Chai, T., Sandu, A., Constantinescu, E., and Daescu, D.: Predicting air quality: improvements through advanced methods to integrate models and measurements, J. Comput. Phys., 227, 3540–3571, 2008.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, doi:10.5194/acp-7-3749-2007, 2007.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, J. Geophys. Res., 99, 10143–10162, 1994.

Gaubert, B., Coman, A., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., and Beekmann, M.: Regional scale ozone data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model, Geosci. Model Dev., 7, 283–302, doi:10.5194/gmd-7-283-2014, 2014.

Grell, G. A., Dudhia, J., and Stauffer, D. R.: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5), NCAR Technical Note NCAR/TN-398+STR, 117 pp., 1994.

Hanea, R. G., Velders, G., and Heemink, A.: Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model, J. Geophys. Res., 109, D10302, doi:10.1029/2003JD004283, 2004.

Hanna, S. R., Chang, J. C., and Fernau, M. E.: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, Atmos. Environ., 32, 3619–3628, 1998.

Hanna, S. R., Lu, Z. G., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., and Hansen, D. A.: Uncertainties in predicted ozone concentrations due to input uncertainties

ACPD
15, 35693–35718, 2015

for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, Atmos. Environ., 35, 891–903, 2001.
Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., 126, 796–811, 1998.
Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, 2001.
Kleczek, M. A., Steeneveld, G., and Holtslag, A. A. M.: Evaluation of the Weather Research and Forecasting mesoscale model for GABLS3: impact of boundary-layer schemes, boundary conditions and spin-up, Bound.-Lay. Meteorol., 152, 213–243, 2014.
Koohkan, M. R., Bocquet, M., Roustan, Y., Kim, Y., and Seigneur, C.: Estimation of volatile organic compound emissions for Europe using data assimilation, Atmos. Chem. Phys., 13, 5887–5905, doi:10.5194/acp-13-5887-2013, 2013.
Moore, G. E. and Londergan, R. J.: Sampled Monte Carlo uncertainty analysis for photochemical grid models, Atmos. Environ., 35, 4863–4876, 2001.
Moral, P. D.: Nonlinear filtering: interacting particle solution, Markov Processes and Related Fields, 2, 555–580, 1996.
Pagowski, M., Liu, Z., Grell, G. A., Hu, M., Lin, H.-C., and Schwartz, C. S.: Implementation of aerosol assimilation in Gridpoint Statistical Interpolation (v. 3.2) and WRF-Chem (v. 3.4.1), Geosci. Model Dev., 7, 1621–1627, doi:10.5194/gmd-7-1621-2014, 2014.
Sandu, A. and Chai, T.: Chemical Data Assimilation – An Overview, Atmosphere, 2, 426–463, 2011.
Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., and Vallés, B.: Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter, Comput. Geosci., 15, 293–305, doi:10.1007/s10596-010-9207-1, 2011.
Tang, X., Wang, Z. F., Zhu, J., Wu, Q. Z., and Gbaguidi, A.: Preliminary application of Monte Carlo uncertainty analysis in ozone simulation, Clim. Environ. Res., 15, 541–550, 2010a (in Chinese).
Tang, X., Wang, Z. F., Zhu, J., Gbaguidi, A., Wu, Q. Z., Li, J., and Zhu, T.: Sensitivity of ozone to precursor emissions in urban Beijing with a Monte Carlo scheme, Atmos. Environ., 44, 3833–3842, 2010b.
Tang, X., Zhu, J., Wang, Z. F., and Gbaguidi, A.: 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, doi:10.5194/acp-11-12901-2011, 2011.

ACPD
15, 35693–35718, 2015

Tang, X., Zhu, J., Wang, Z. F., Wang, M., Gbaguidi, A., Li, J., Shao, M., Tang, G. Q., and Ji, D. S.: Inversion of CO emissions over Beijing and its surrounding areas with ensemble Kalman filter, Atmos. Environ., 81, 676–686, 2013.
van Leeuwen, P. J.: Particle filtering in geophysical systems, Mon. Weather Rev., 137, 4089–4114, 2009.
van Leeuwen, P. J.: Nonlinear data assimilation in geosciences: an extremely efficient particle filter, Q. J. Roy. Meteor. Soc., 136, 1991–1999, doi:10.1002/qj.699, 2010.
van Loon, M., Builtjes, P., and Segers, A. J.: Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, Environ. Modell. Softw., 15, 603–609, 2000.
Wang, Z., Maeda, T., Hayashi, M., Hsiao, L. F., and Liu, K. Y.: A nested air quality prediction modeling system for urban and regional scales: application for high-ozone episode in Taiwan, Water Air Soil Pollut., 130, 391–396, 2001.
Wang, Z. F., Xie, F. Y., Wang, X. Q., An, J. L., and Zhu, J.: Development and application of Nested Air Quality Prediction Modeling System, Chinese J. Atmos. Sci., 30, 778–790, 2006 (in Chinese).
Wesely, M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmos. Environ., 23, 1293–1304, 1999.
Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, 2002.
Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, J. Geophys. Res., 113, D20310, doi:10.1029/2008JD009991, 2008.
Xiang, W. L., An, J. L., Wang, Z. F., Wu, Q. Z., and Tang, X.: Application of CBM-Z chemical mechanism during Beijing Olympics, Clim. Environ. Res., 15, 551–559, 2010 (in Chinese).
Xin, J. Y., Wang, Y. S., Tang, G. Q., Wang, L. L., Sun, Y., Wang, Y. H., Hu, B., Song, T., Ji, D. S., Wang, W. F., Li, L., and Liu, G. R.: Variability and reduction of atmospheric pollutants in Beijing and its surrounding area during the Beijing 2008 Olympic Games, Chinese Sci. Bull., 55, 1937–1944, 2010 (in Chinese).
Zaveri, R. A. and Peters, L. K.: A new lumped structure photochemical mechanism for large-scale applications, J. Geophys. Res., 104, 30387–30415, 1999.
Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian

ACPD
15, 35693–35718, 2015

emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys., 9, 5131–5153, doi:10.5194/acp-9-5131-2009, 2009.
Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects, Atmos. Environ., 60, 656–676, 2012.

ACPD
15, 35693–35718, 2015

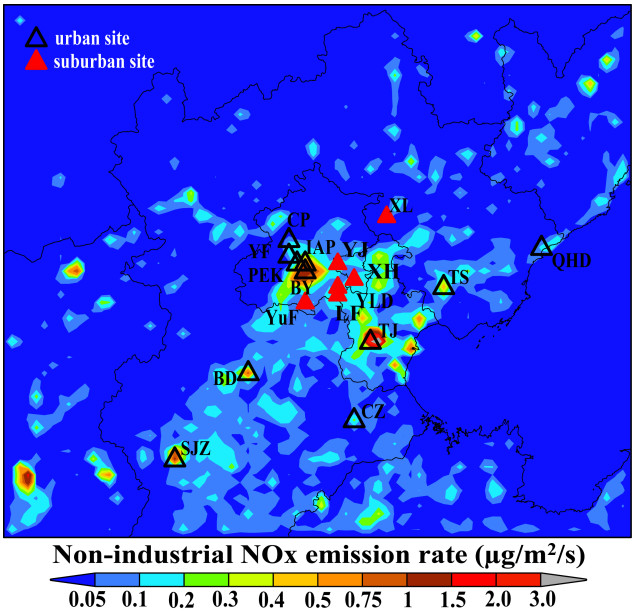

Figure 1. Distribution of the observation stations and non-industrial NO_x emission rates in the third model domain (9 km resolution) that covers Beijing and its surrounding areas. The non-industrial NO_x emission rates ($\mu g\,m^{-2}\,s^{-1}$) are divided into different bins (< 0.05; 0.01–0.1; 0.1–0.2; 0.2–0.3; 0.3–0.4; 0.4–0.5; 0.5–0.75; 0.75–1.0; 1.0–1.5; 1.5–2.0; 2.0–3.0) and represented by different shaded colors. The urban areas with high non-industrial NO_x emission rates are marked by the brown and red colors, and the suburban or rural areas with low non-industrial NO_x emission rates are marked by the green or blue colors. The 11 urban sites are denoted by the black triangles, and the 6 suburban stations are represented by the red triangles. The abbreviations of the station names are displayed close to the marks.

ACPD
15, 35693–35718, 2015

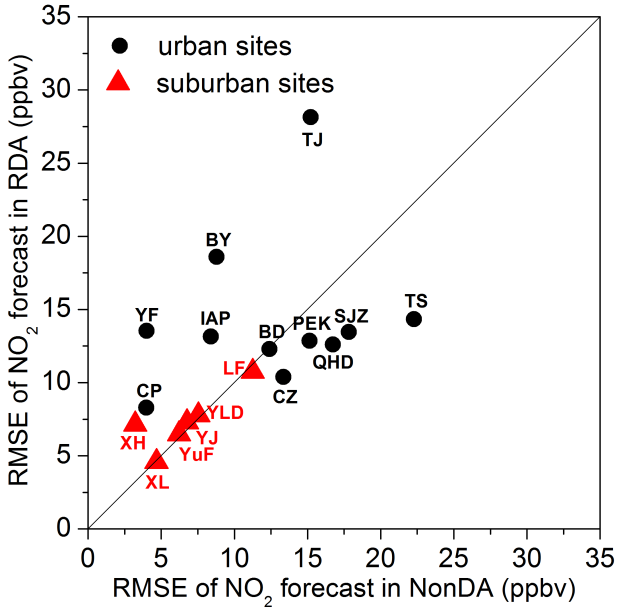

Figure 2. Comparison of the root mean square errors (RMSEs) (ppbv) of 1 h NO_2 forecasts at the 17 stations of Beijing and its surrounding areas during the period of 00:00 LT 9 August to 00:00 LT 23 August in 2008 in the real data assimilation (RDA) experiments and those in the reference (NonDA) experiment with a free run of the model. The comparisons at urban sites are denoted by the dots and those over suburban stations are represented by the triangles. The abbreviations of the station names are displayed close to the marks.

ACPD
15, 35693–35718, 2015

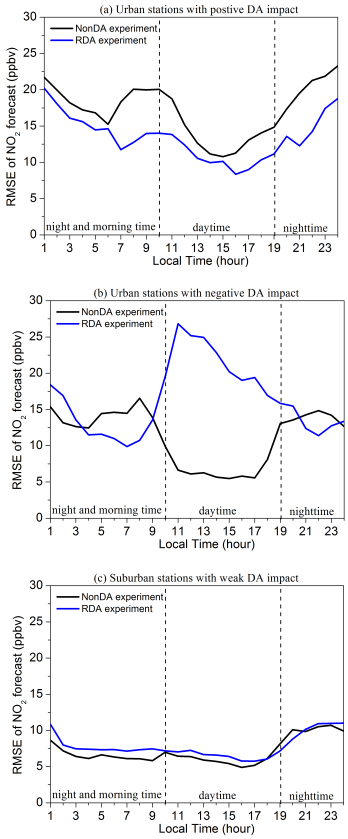

Figure 3. Daily variation of the 1 h NO_2 forecast RMSEs (ppbv) in the real data assimilation (RDA) experiments (blue line) and the reference (NonDA) experiment with a free run of the model (black line) over: **(a)** urban stations (CZ, PEK, QHD, SJZ, and TS) with positive DA impacts; **(b)** urban sites (BY, CP, IAP, TJ and YF) with negative DA impacts; **(c)** suburban stations (LF, XH, YLD, YJ and YuF) with weak DA impacts.

ACPD
15, 35693–35718, 2015

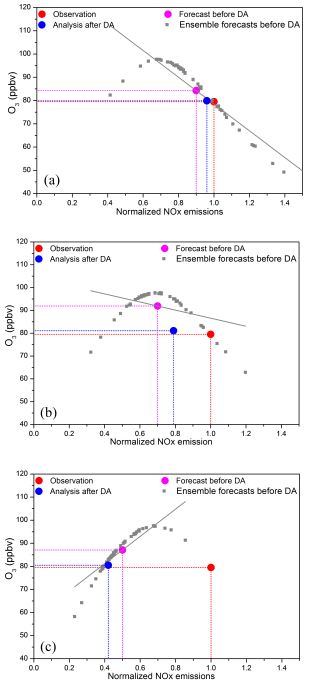

Figure 4. (a–c) O_3 concentrations (ppbv) and NO_x emissions (no unit, normalized by the true NO_x emission) before and after data assimilation (DA) and their ensemble samples before DA at 12:00 LT on 12 August 2008 in the three ideal ozone data assimilation experiments with the prior NO_x emissions underestimated by 10 % **(a)**, 30 % **(b)** and 50 % **(c)** respectively. The magenta dot represents the O_3 concentrations and NO_x emissions before DA, and the gray squares denote the ensemble forecast O_3 concentrations corresponding to the perturbations of the NO_x emissions. The gray line represents a linear relationship calculated from the ensemble samples of O_3 concentrations and NO_x emissions. The red dot represents the true state of NO_x emission and the observed O_3 concentration. The analyzed O_3 concentration and NO_x emission are denoted by the blue dot.

ACPD
15, 35693–35718, 2015

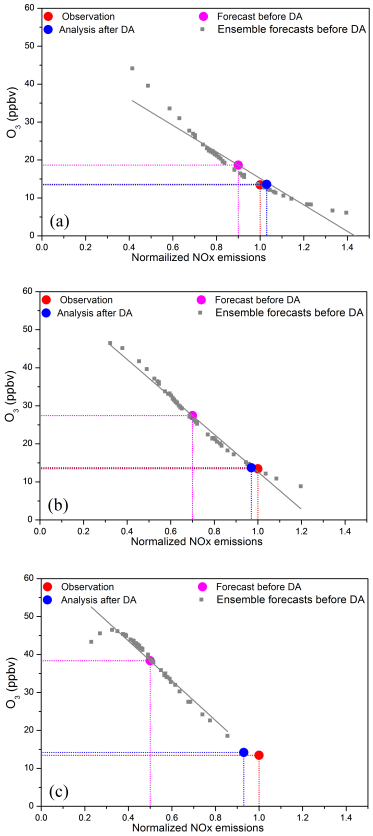

Figure 5. (a–c) O_3 concentrations (ppbv) and NO_x emissions (no unit, normalized by the true NO_x emission) before and after data assimilation (DA) and their ensemble samples before DA 08:00 LT on 12 August 2008 in the three ideal ozone data assimilation experiments with the prior NO_x emissions underestimated by 10 % **(a)**, 30 % **(b)** and 50 % **(c)** respectively. The magenta dot, the gray squares, the gray line, the red dot and the blue dot represent the same information as Fig. 4.

ACPD
15, 35693–35718, 2015