Limitations of ozone data assimilation with adjustment of NOx 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], 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

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

Abstract

This study investigates a cross-variable ozone data assimilation (DA) method based on an ensemble Kalman filter (EnKF) that has been used in the companion study to improve ozone forecasts over Beijing and surrounding areas. The main purpose is to delve into the impacts of the cross-variable adjustment of nitrogen oxides (NOx) emissions on the nitrogen dioxide (NO_2) forecasts over this region during the 2008 Beijing Olympic Games. A mixed effect on the NO_2 forecasts was observed through application of the cross-variable assimilation approach in the real-data assimilation (RDA) experiments. The method improved the NO_2 forecasts 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 in night and morning versus significant deterioration in daytime (with 190% increase of the RMSEs), suggesting that the negative DA impacts mainly occurred during daytime. Ideal data assimilation (IDA) experiments with a box model and the same cross-variable assimilation method confirmed the mixed effects found in the RDA experiments. In the same tendency, NOx emission estimation was improved in night and morning even under large biases in the prior emission, while deteriorated in daytime (except for the case of minor errors in the prior emission). The mixed effects observed in the cross-variable DA, 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, highlighted the limitations of the EnKF under strong nonlinear relationships between chemical variables. Under strong nonlinearity between daytime ozone concentrations and NOx emissions uncertainties (with large biases in the a prior emission), the EnKF may come up with inefficient or wrong adjustment to NOx emissions. The present findings reveal that bias correction is essential for the application of the

EnKF in dealing with the DA inconsistency over strong nonlinear system.

1. Introduction

Chemical data assimilation (CDA) integrates models and observations to better represent the chemical state of the atmosphere and 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 that the impact of the initial conditions on the forecast of air pollutants such as ozone decreases with simulation time (Gaubert et al., 2014; Jimenez et al., 2006). To overcome such obstacle, 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 also been 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 provided significant improvement of ozone forecasts (e.g., 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). Elbern 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 (NOx) emissions after data assimilation in most cases even if the model underestimated the NOx 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 (SOx) emissions using an ensemble Kalman filter. The method enhanced the emission rates of SOx when significant over-prediction of SO_2 concentrations subsisted. Such inconsistencies, i.e., the emissions enhanced under the overestimation of concentrations or the emissions reduced under the underestimation of concentrations, reveal some gaps between ozone forecast improvement and precursor emission optimization and call for a comprehensive evaluation of the cross-variable chemical data assimilation techniques.

Tang et al. (2011) employed a high horizontal resolution (9km) model to perform the assimilation of O_3 observations with the ensemble Kalman filter and the adjustment of NOx 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 investigate the state of NO_2 and NOx 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 assimilation experiments and the ideal data assimilation experiments are presented in Sect. 3. Section 4 presents conclusions and discussion.

2. Methodology

(1) Chemical transport model

The chemical transport model used for O_3 simulations was 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 were simulated by the Carbon-Bond Mechanism Z (CBM-Z) which includes 133 reactions for 53 species (Zaveri and Peter, 1999). The dry deposition modeling followed the scheme of Wesely (1999). The vertical eddy diffusivity was parameterized based on a scheme by Byun and Dennis (1995). The O_3 simulations were configured with three nested domains and the horizontal resolutions were 81km, 27km and 9km respectively. The first domain covered East Asia with a 81km resolution and the second domain contained North China with a 27km resolution. The third domain displayed in Fig. 1 covered Beijing and its surrounding areas with 9km resolution. Vertically, the model was set as twenty terrain-following layers, nine of which were within the lowest 2 km of the atmosphere and the height of the first layer near the surface was 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 0.5° × 0.5° 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) Data assimilation algorithm

The assimilation algorithm employed was the ensemble Kalman filter (EnKF) proposed by Evensen (1994). The main feature of this method consists of 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). EnKF can directly calculate the background error covariance from the ensemble forecasts of the highly nonlinear model, which is very suitable for data assimilation in complex high-dimensional models (Carmichael et al., 2008). Its implementation is very simple and does not require an adjoint model which is a very cumbersome task for complex high-dimensional model. 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 dust and ozone forecast skills through emission optimization have been reported (e.g., Constantinescu et al., 2007; Eben et al., 2005; Lin et al., 2008; Tang et al., 2011).

In the present study, the EnKF was employed to assimilate ozone observations for the corrections of NOx 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 was to perturb ozone concentrations, NOx emissions and other key uncertainty sources of ozone modeling, i.e., photolysis rates and vertical diffusion coefficients, as described by the following equations:

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

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

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

where $\boldsymbol{x}$, $\boldsymbol{e}$, and $\boldsymbol{q}$ are ozone concentrations, emissions, and other parameters (NO_2 photolysis rates and vertical diffusion coefficients) respectively, 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 $\boldsymbol{\phi}$. The random samples were extracted from a normal distribution using the method proposed by Evensen (1994). N is the ensemble size. The ensemble size (set as 50) was chosen based on several sensitivity experiments of ozone data assimilation. The experiments were performed with the same model domains and observation network as those employed in this study. The results suggest that an ensemble of 50 members keeps good balance between computational efficiency and assimilation performance of ozone analysis.

In order to avoid filter divergence, the NO_2 photolysis rate and vertical diffusion coefficient were perturbed by Gaussian distributed random noise, and the NOx emissions (to be updated by the EnKF) were perturbed by a time-correlated Gaussian distributed random noise. Estimating the uncertainty of the NOx emissions used for the modeling during the Beijing Olympic Games was a hard task. The INTEX-B Asia inventory (Zhang et al., 2009) was estimated to contain 31% uncertainty in NOx emission estimation. But the base year of this inventory is 2006. Another key factor affecting the emission uncertainty is the temporary air pollution control measures during the Beijing Olympic Games. The control measures were estimated to reduce the NOx emissions by 36% to 47% (Wang et al., 2009; 2010). This would induce large biases into the emission inventory and lead to significant increase of the uncertainties of the emission inventory. Therefore, we estimated the uncertainty of the NOx 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 initial time was 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 were implemented through adding a pseudo smooth random field. The random samples were Gaussian distributed with zero mean. The horizontal and vertical scales of initial error correlations could be effectively controlled using this method. The scales were set as 54 km in the horizontal and 3 model grids in the vertical (approximately 200 m) as in Tang et al. (2011).

Ensemble samples of the emissions, the vertical diffusion coefficients, the photolysis rates and the O_3 concentrations were used to derive ensemble forecasts of ozone. In order to achieve cross-variable adjustment for NOx emissions, an extended state variable was defined as:

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

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

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

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 were perturbed according to the observation errors (Gaussian with mean zero and covariance $\boldsymbol{R}$, including both measurement errors and representativeness errors):

$$\boldsymbol{y}'(\boldsymbol{i}) = \boldsymbol{y} + \Upsilon(\mathbf{i}), \mathrm{i} = 1, 2, \dots, \mathrm{N} \qquad (6)$$

$$\boldsymbol{\Upsilon} \in N(0, \boldsymbol{R}).$$

As suggested by von Loon et al. (2000), the observation errors were 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 could be updated through assimilating the ozone observations:

$$\boldsymbol{U}^{\boldsymbol{a}}(\boldsymbol{i}) = \boldsymbol{U}'(\boldsymbol{i}) + \mathbf{K}(\boldsymbol{y}'(\boldsymbol{i}) - \boldsymbol{H}\boldsymbol{U}'(\boldsymbol{i})), \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{N} \quad (7)$$

$$\boldsymbol{K} = \boldsymbol{P}\boldsymbol{H}^{\boldsymbol{T}}(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^{\boldsymbol{T}} + \boldsymbol{R})^{-\mathbf{1}} \quad (8)$$

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. $\boldsymbol{U}^{\boldsymbol{a}}(\boldsymbol{i})$ is the updated ensemble sample of the extended state variable and was used for the sequential ozone forecast. The updating of the ensemble ensembles of the extended variables was conducted one time every 1 hour (1h), and the updated NOx emissions were then used for the NO_2 forecast of the next hour. The ensemble mean of $U^a(i)$ was taken as the best estimation after assimilating observations and was used as the output analysis state for comparison (e.g. the blue dots in Figures 4 and 5). To reduce the spurious impact caused by the finite ensemble size, localization was performed for analysis and only observations within a localization scale were used to update the N0x emissions at a model grid. The localization scale was set as 45km following the configuration of Tang et al. (2011).

(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., 2010). Figure 1 displays the distributions of these stations and the non-industrial NOx 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 NOx emission rates, and the other 6 (LF, XH, XL, YJ, YuF, YLD) are in the suburban areas with relatively low non-industrial NOx emission rates. The network provides observations of O_3 and NO_2 at the same temporal resolution as the model (i.e., 1h). The measurements of NO_2 and O_3 were observed by online instruments (Model 42C& 42I NO-NO_2-NOx Analyzer and Model 49C&49I O_3 Analyzer from Thermo Scientific). The O_3 observations were assimilated hourly into the model to adjust NOx emissions. The direct comparison between the simulated and observed NO_2 data often suffered from the representativeness errors of the NO_2 measurements. In this study, the stations close to the main roads with heavy traffic were not included in order to reduce the influence of the representativeness errors of the NO_2 measurements. Nevertheless, under certain resolutions (9km for example), the representativeness errors still persisted in NO_2 measurements over urban areas. In order to independently validate the assimilation results, three of the observation stations were withdrawn from the assimilation and were used for the validation. NO_2 observations not used in the assimilation were also used to assess 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 assimilated the surface ozone observations over Beijing and surrounding areas to adjust the NOx emissions over these areas in the NAQPMS. The experiment was based on the study of Tang et al. (2011) in which the assimilation of real O_3 observations with the EnKF was performed to correct NOx emissions. The experiment focused 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 were from a two-week spin-up model run. The initial conditions of ozone, NOx emissions and vertical diffusion parameters were perturbed at 19:00 LT on 8 August 2008 according to the equations (1), (2) and (3) and were used to derive ensemble runs of NAQPMS. After 5h free ensemble runs, the observed ozone data started at 00:00 LT on 9 August to be assimilated hourly into the third model domain (displayed in Fig. 1) of NAQPMS to adjust the NOx emissions. Adjusted factors of the NOx emissions were then used for the NO_2 forecast of the next hour. Both daytime and nighttime observations were assimilated. By considering possible large errors in the modeling of vertical profiles of air pollutants, we only adjust the variables in the first three vertical layers near the surface, which could reduce the influence of the modeling errors of vertical mixing on data assimilation. A free run of NAQPMS without data assimilation (NonDA) was also performed as a reference run to validate the assimilation results of the RDA experiment.

Figure 2 compares the root mean square errors (RMSEs) of the 1 h ensemble mean forecast of NO_2 at the 17 stations in the RDA experiment with the RMSEs in the NonDA experiment. The

RMSE of each site was calculated based on the hourly differences between NO_2 observation and the ensemble mean forecast of NO_2 from 00:00 LT 9 August to 00:00 LT 23 August in 2008. The number of valid observations used for each station is listed in Figure 2. The differences of the RMSEs before and after DA were statistically significant over 11 stations (TJ, BY, YF, IAP, CP, XH, CZ, PEK, QHD, SJZ and TS) at the 95% level of the t-test, while there were no statistically significant differences of the RMSEs before and after DA over 6 stations (XL, YuF, YJ, YLD, LF and BD). The RMSEs of the NO_2 forecasts in the free run of the model were dominated by the biases which accounted for 55~90% of the RMSEs (Bias/RMSE). Biases noticed in simulations performed over urban sites are relatively larger than those over the suburban ones. The free model run overestimated NO_2 concentrations at most of the urban stations, while underestimated that most of the suburban ones. The DA impacts on the NO_2 forecast varied substantially from the suburban to the urban stations. At urban station such as BD, PEK, CZ, QHD, SJZ, and TS, the RMSEs were reduced by 15%~36% after DA, resulting in improvement of NO_2 forecasts in contrast to large increases, ranging from 56~239% of the RMSEs at CP, BY, IAP, YF and TJ. At the suburban sites, the DA showed minor influence on NO_2 forecasts and had no statistically significant impacts on the RMSEs over 5 of the 6 suburban sites. Such minor DA impacts over the suburban sites could be explained firstly, by the fact that emission rates of NOx in the model were very low over suburban regions and the simulation without DA significantly underestimated the NO_2 concentrations. Even with the perturbations on the NOx emission, the ensemble spread was significantly weaker than the errors in the real case, and thereby reduced the DA impacts of the EnKF. On the other hand, in regards to the influences of the air pollutants transport from urban regions, observed negative DA impacts over some urban areas may have induced significant errors into the NO_2 forecasts. The above results suggest the adjustment of the NOx emission by the ozone data assimilation has a mixed effect on the NO_2 forecast (i.e., weak DA impacts over suburban sites, positive DA impacts over some urban sites and negative DA impacts over others). Nevertheless, the assimilation produced significant improvement of ozone forecasts over all these sites, as reported by Tang et al. (2011).

Further investigations were conducted on the variation of such mixed effects of the data assimilation on NO_2 forecasts over both first week (from 00:00 LT 9 August to 00:00 LT 16 August in 2008) and second week (from 00:00 LT 16 August to 00:00 LT 23 August in 2008). As a result, the DA mixed effects were relatively stable during the Beijing Olympic Games. Figures 3 (a-c) display daily variation of the 1h NO_2 forecast RMSEs in DA experiment and NonDA experiment over the urban stations with positive DA impacts (CZ, PEK, QHD, SJZ, and TS), those with negative DA impacts (BY, CP, IAP, TJ and YF) and the suburban stations (LF, XH, YLD, YJ and YuF with weak DA impacts). At the suburban stations, the cross-variable DA also showed very weak impacts on the NO_2 forecast in both the daytime and nighttime. At the urban stations with positive DA impacts, the cross-variable assimilation presented consistent positive DA impacts in daytime, nighttime and morning, with a 23% reduction of RMSEs during daytime and a 21% reduction in night and morning.

At the urban sites with negative DA impacts, the performance of the DA was different between daytime, nighttime and morning hours. Adjusting NOx emissions improves the forecasts of NO_2 concentrations during most of the night and the morning time by reducing 7% of the RMSEs in contrast to the deterioration of the forecast in the daytime with 190% increase of the RMSEs. This finding suggests that the impacts of the cross-variable assimilation on the NO_2 forecast during daytime are opposite to those in night and morning at these urban sites. In clear, negative DA impacts mainly occur in the daytime. As described by Tang et al. (2010b), daytime ozone is strongly nonlinearly related to high NOx 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 daytime ozone and nighttime ozone chemistry, further experiments were carried out in following section to elucidate the impact of the chemistry on the cross-variable assimilation.

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

3.2 Ideal data assimilation experiment

An ideal experiment with a known true state provided a simple way to investigate the potential consequences of some key inspected factors in a highly complex system. In order to investigate the possible cause of observed mixed effects in RDA experiment, this study employed a simplified box model including the main chemical processes of NAQPMS (Xiang et al., 2010). Within conducted ideal data assimilation (IDA) experiments, the true state of ozone concentrations and NOx emissions were assumed to be known. The main purpose is to closely monitor the impacts of ozone chemistry on the cross-variable assimilation method experimented in the RDA. However, this investigation did not take into account complex transport processes and the removal processes were simulated by multiplying the concentrations by removal coefficients. The experiments with the box model were conducted on the IAP station where negative impact on NO_2 forecasts is observed in the RDA experiment. Emission rates and meteorological parameters are from the inputs used by NAQPMS.

Firstly, the IDA experiments focused on the negative DA impacts on the daytime NO_2 forecasts. The a priori emission rates from NAQPMS and their corresponding O_3 concentrations modeled with the box model were assumed to be the true state and were used for validation of the optimized emissions from DA. Ensemble runs of the box model were initialized by the ensemble forecasts of the chemical species of NAQPMS at 19:00 LT on 11 August 2008; NOx emissions were perturbed to provide ensemble samples of emissions during the following ensemble runs of the model. At 12:00 LT on 12 August 2008, the artificial O_3 observation was assimilated into the box model to adjust the NOx emissions. Artificial O_3 observations were generated through adding slight random errors to the true state of O_3 concentrations. To be consistent with the RDA experiment, the random errors for perturbing observations were also assumed to be within 10% of the true value. Three error scenarios for NOx emissions (10%, 30% and 50% underestimations) were assumed and separately applied to simulations of the box model. In order to avoid dealing with complex model errors, the errors in NOx emissions were assumed to be the only error sources of ozone modeling. For each error scenario, cross-variable adjustment of the NOx emissions through assimilating the artificial O_3 observations with the EnKF was conducted. Figures 4(a-c) show the O_3 concentrations and NOx emissions before and after DA, with their ensemble samples before DA at 12:00 August 12, 2008.

Figure 4a presents the results under the first scenario with 10% underestimation of NOx emissions (S1). The analyzed O_3 concentration and NOx emission after DA were close to their true state, suggesting an improvement of the NOx emission estimation from the cross-variable assimilation. Figure 4b shows the results under the second scenario with 30% underestimation of NOx emissions (S2). The DA inefficiently reduced the error in NOx emission, since large errors (about 20%) still persisted in the optimized NOx emission. Ensemble samples of O_3 concentrations shown in Fig.4b were obtained from the ensemble runs of the box model that were derived from the ensemble samples of NOx emissions (also shown in Fig.4b). Obviously, the ensemble forecasts of O_3 concentrations presented high nonlinear responses to the perturbations of NOx emissions. This suggests that the EnKF with Monte Carlo simulations can properly predict the nonlinear evolutions of error statistics of the O_3 modeling. At the analysis step, the ensemble samples of O_3 concentrations and NOx emissions were integrated into the EnKF to calculate the background error covariance in Eq. (5). The linearized relationship between the O_3 concentrations and the NOx emissions is presented in Fig. 4b. Noticeable discrepancies appear 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.

In the third scenario (S3) with NOx emissions underestimated by 50%, enhanced deterioration of the NOx emission estimations was observed (Fig. 4c). The DA closely adjusted the simulated O_3 concentration to the true state, but induced additional bias to previously underestimated NOx emission. Such negative DA impact on NOx emission estimation was similar to the phenomenon observed on the daytime NO_2 forecast over some urban stations in the RDA experiment. From the results in Fig. 4(a-c), the most plausible cause of the negative DA impact on NOx emission estimation is the linearizing analysis of the EnKF in dealing with the cross-variable (O_3 to NOx emission) DA problem of a highly nonlinearly chemical system. With large bias in the a priori estimation of NOx emissions, the cross-variable assimilation may induce enhancement of the bias in NOx 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 over strongly nonlinear chemical system.

In order to consider error scenarios with overestimations of NOx emission, four idealized DA experiments in which NOx emission was assumed to be being overestimated by 10%, 30%, 50% and 100% respectively were performed. The results are shown in Fig. 5(a-d). In the first three experiments with 10%, 30% and 50% overestimations of the a priori NOx emission, the DA worked well and significantly reduced the biases of the emission. In the fourth experiment with the largest bias in the a priori emission estimation, the DA enhanced the bias of the emission estimation in daytime. These mixed DA effects under different biases of the a priori emission estimation are similar to those observed in previous idealized experiments conducted with underestimate scenarios. Both underestimate and overestimate scenarios clearly confirm the mixed effects of the DA.

Note that above IDA experiments do not consider the complex model errors (e.g., errors in boundary layer or transport modeling). In the real case, model errors exist, and the DA scheme needs to properly quantify model uncertainties and deal with the nonlinearity between assimilated observations and adjusted variables simultaneously. Model errors may affect the results of the real DA. Thus, in order to investigate the DA performance of adjusting NOx emissions under the presence of biases on other factors, we assumed that the NO_2 photolysis rate was overestimated by 20% in the idealized box modeling, since the errors of the NO_2 photolysis rates were found to be the top five uncertainty sources of ozone modeling over Beijing and surrounding areas during the Beijing Olympic Games (Tang et al., 2010a).

Firstly, we were blind to the bias of the simulated NO_2 photolysis rate, so that no perturbation was operated on it in the DA experiment. The NOx emission was adjusted in the same way as the above-idealized experiments. Fig. 6a displays the results of the DA experiment under the error scenario of 30% overestimation in the a priori NOx emission. The DA corrected the NOx emission, but led to an underestimation of the emission. This over-correction of NOx emission by the DA could be associated with the bias in simulated NO_2 photolysis rate. Therefore, in the second experiment (Fig. 6b), we considered the uncertainty of the simulated NO_2 photolysis rate and perturbed the NO_2 photolysis rate in the DA. The error scenario was the same as in the first experiment. Under that condition, the DA performed better than that of the first experiment, without over-correction of NOx emission. The results of above experiments suggest that considering the model errors is crucial for the assimilation performance; otherwise the DA leads to over-correction to the state variable. In order to deal with this issue, simulated NO_2 photolysis rates and vertical diffusion coefficients (considered as the key uncertainty sources of the O_3 modeling) were perturbed to account their uncertainties into the real DA experiment. The third DA experiment is quite similar to the second one, but we increased the bias of the a priori NOx emission to 100% overestimation. The results are shown in Fig. 6c. Under large bias in the a priori NOx emission, the DA deteriorated NOx emission estimation. In short, in sight of considering the influence of the model errors, the limitations of the DA method in dealing with the large bias of a highly nonlinear system are still persistent.

To investigate the DA impacts on the NOx emissions in night and morning, variations of O_3 concentrations and NOx emissions before and after DA and their ensemble samples before DA at 8:00 August 13, 2008 (morning time) are shown in Figs. 7(a-c). Similar trends (not shown here) were obtained for other night and morning times. In Figs. 7(a-c), different level errors (10%, 30% and 50% underestimations) in NOx emissions were significantly reduced through the cross-variable assimilation with the EnKF. The ensemble forecasts of morning O_3 concentrations show near-linear responses to the uncertainties (or perturbations) of NOx emissions; the linearization of the EnKF at the analysis step worked properly to correct the biases in NOx emissions. The positive DA impacts on the NOx emission estimation in IDA experiments in night and morning were consistent with the improvement of the NO_2 forecasts after data assimilation in RDA experiment. In comparison with the mixed effects of the DA in daytime, the positive DA impacts in night and morning in both RDA and IDA experiments indicate that the assimilation of O_3 observations with the EnKF might be useful in optimizing NOx emissions and NO_2 forecasts in night and morning. Furthermore, the ensemble forecasts of O_3 concentrations show strong nonlinear responses to the perturbations of NOx emissions during daytime in Figs. 4(a-c) but present near-linear responses in night and morning in Figs. 7(a-c). This suggests the variability of nonlinearity of the chemical system leads to different DA impacts during different periods of the day.

4. Conclusion and discussion

The impacts of cross-variable adjustment of NOx 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 DA experiments with a 3-dimensional chemical transport model and ideal DA experiments with a simplified box chemical model were performed.

The results of the data assimilation experiments revealed mixed effects of the cross-variable assimilation with the EnKF. The DA worked properly in improving the NO_2 forecasts and optimizing the NOx emissions in night and morning when the uncertainties of O_3 concentrations were almost linearized to those of NOx emissions. During daytime, the data assimilation resulted in positive DA impacts on NO_2 forecasts over some urban sites, negative over other urban sites and weak impacts over suburban sites. Through idealized DA experiments, the mixed effects were found to be strongly associated with the difficulty in dealing with highly nonlinear DA problem, especially under large model biases. The results highlight a critical limitation of the EnKF for the chemical DA despite its strong performance for improving ozone forecasts (e.g., Tang et al., 2011).

The results suggest that bias correction is crucial for the application of the EnKF in highly nonlinear chemical DA problems. Alternatively, avoiding the cross-variable DA between two strong-nonlinearly related variables such as NOx emissions and O_3 is also a possible way to overcome this issue. For example, assimilating NO_2 observations directly to optimize NOx emissions might produce better result than assimilating O_3 observations to improve the NO_2 forecasts and NOx emission estimations. Nevertheless, strong nonlinearity issue remains a critical challenge in the chemical DA. In summary, DA approaches that enable dealing with high nonlinearity in both model evolution and analysis step are needed. Particle filters, as nonlinear filter method (e.g., Moral et al., 1996; van Leeuwen, 2009; 2010) might have potential in this field if its limitation for high dimensional system application (Stordal et al., 2011) can be overcome.

Acknowledgements

This study was funded by the CAS Strategic Priority Research Program (Grant No. XDB05030200) and the National Natural Science Foundation (Grant No. 41205091 and No. 41305111).

References

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

Byun, D.W., 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., Daescu, D.: Predicting air quality: Improvements through advanced methods to integrate models and measurements, J. Comput. Phys. 227, 3540–3571, 2008.

Constantinescu, E.M., Sandu, A., Chai, T.F. and Carmichael, G.R.: Ensemble-based chemical data assimilation. II: Covariance localization, Q. J. R. Meteorol. Soc. 133: 1245–1256, 2007.

Eben, K., Jurus, P., Resler, J., Belda, M., Pelikan, E., Kruger, B. C., and Keder, J.: An ensemble Kalman filter for short-term forecasting of tropospheric ozone concentrations, Q. J. R. Meteorol. Soc. (2005), 131, 3313–3322, 2005.

Elbern, H., Strunk, A., Schmidt, H., Talagrand, O.: Emission rate and chemical state estimation by

4-dimensional variational inversion, Atmos. Chem. Phys. 7, 3749–3769, 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., Beekmann, M.: Regional scale ozone data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model, Geosci. Model. Dev. 7, 283–302, 2014.

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

Hanna, S.R., Chang, J.C., 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(21), 3619-3628, 1998.

Hanna, S.R., Lu, Z.G., Frey, H.C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., Hansen, D.A.: Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, Atmos. Environ. 35, 891-903, 2001.

Houtekamer, P.L., Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique, Mon. Wea. Rev. 126, 796–811, 1998.

Houtekamer, P.L., Mitchell, H.L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Wea. Rev. 129, 123–137, 2001.

Jimenez, P., Parra, R., and Baldasano, J. M.: Influence of initial and boundary conditions for ozone modeling in very complex terrains: A case study in the northeastern Iberian Peninsula, Environmental Modelling & Software, 22, 1294-1306, 2007.

Kleczek, M.A., Steeneveld, G., 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, Boundary-Layer Meteorol. 152, 213–243, 2014.

Koohkan, M.R., Bocquet, M., Roustan, Y., Kim, Y., Seigneur, C.: Estimation of volatile organic compound emissions for Europe using data assimilation, Atmos. Chem. Phys. 13, 5887–5905, 2013.

Lin, C., J. Zhu, and Z. Wang: Model bias correction for dust storm forecast using ensemble Kalman filter, J. Geophys. Res., 113, D14306, doi:10.1029/2007JD009498, 2008.

Moore, G.E., Londergan, R.J.: Sampled Monte Carlo uncertainty analysis for photochemical grid models, Atmos. Environ., 35(28), 4863-4876, 2001.

Moral, P.D.: Nonlinear filtering: interacting particle solution, Markov Processes and Related Fields, 2 (4), 555–580, 1996.

Pagowski, M., Liu, Z., Grell, G.A., Hu, M., Lin, H.-C., 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, 2014.

Sandu, A., Chai, T.: Chemical Data Assimilation—An Overview, Atmosphere 2, 426-463, 2011.

Stordal, A.S., Karlsen, H.A., Nævdal, G., Skaug, H.J., Vallès, 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., Gbaguidi, A.: Preliminary application of Monte Carlo uncertainty analysis in ozone simulation, Clim. Environ. Res. (in Chinese) 15 (5), 541-550, 2010a.

Tang, X., Wang, Z.F., Zhu, J., Gbaguidi, A., Wu, Q.Z., Li, J., 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., 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, 2011.

Tang, X., Zhu, J., Wang, Z.F., Wang, M., Gbaguidi, A., Li, J., Shao, M., Tang, G.Q., 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. Wea. Rev. 137, 4089–4114, 2009.

van Leeuwen P.J.: Nonlinear data assimilation in geosciences: an extremely efficient particle filter, Q. J. R. Meteorol. Soc. 136, 1991–1999, doi:10.1002/qj.699, 2010.

van Loon, M., Builtjes, P., Segers, A.J.: Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, Environ. Model. Softw. 15, 603–609, 2000.

Wang, S.X., Zhao, M., Xing, J., Wu, Y., Zhou, Y., Lei, Y., He, K.B., Fu, L.X., Hao, J.M.: Quantifying the Air Pollutants Emission Reduction during the 2008 Olympic Games in Beijing, Environ. Sci. Technol., 44 (7), 2490–2496, 2010.

Wang, Y., Hao, J., McElroy, M. B., Munger, J. W., Ma, H., Chen, D., Nielsen, C. P.: Ozone air quality during the 2008 Beijing Olympics: effectiveness of emission restrictions, Atmos. Chem. Phys., 9, 5237–5251, 2009.

Wang, Z., Maeda, T., Hayashi, M., Hsiao, L.F., Liu, K.Y.: A nested air quality prediction modeling system for urban and regional scales: application for high-ozone episode in Taiwan, Water Air and Soil Pollution 130, 391-396, 2001.

Wang, Z.F., Xie, F.Y., Wang, X.Q., An, J.L., Zhu, J.: Development and application of Nested Air Quality Prediction Modeling System, Chinese Journal of Atmospheric Sciences (in Chinese) 30, 778–790, 2006.

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., Hamill, T.M.: Ensemble data assimilation without perturbed observations. Mon. Wea. Rev. 130, 1913–1924, 2002.

Wu, L., Mallet, V., Bocquet, M., 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., Tang, X.: Application of CBM-Z chemical mechanism during Beijing Olympics, Clim. Environ. Res. (in Chinese) 15 (5), 551-559, 2010.

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., Liu, G.R.: Variability and reduction of atmospheric pollutants in Beijing and its surrounding area during the Beijing 2008 Olympic Games, Chinese Sci. Bull. (in Chinese) 55, 1937–1944, 2010.

Zaveri, R.A., 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., Yao, Z.L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys. 9, 5131-5153, 2009.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., 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.

Figures

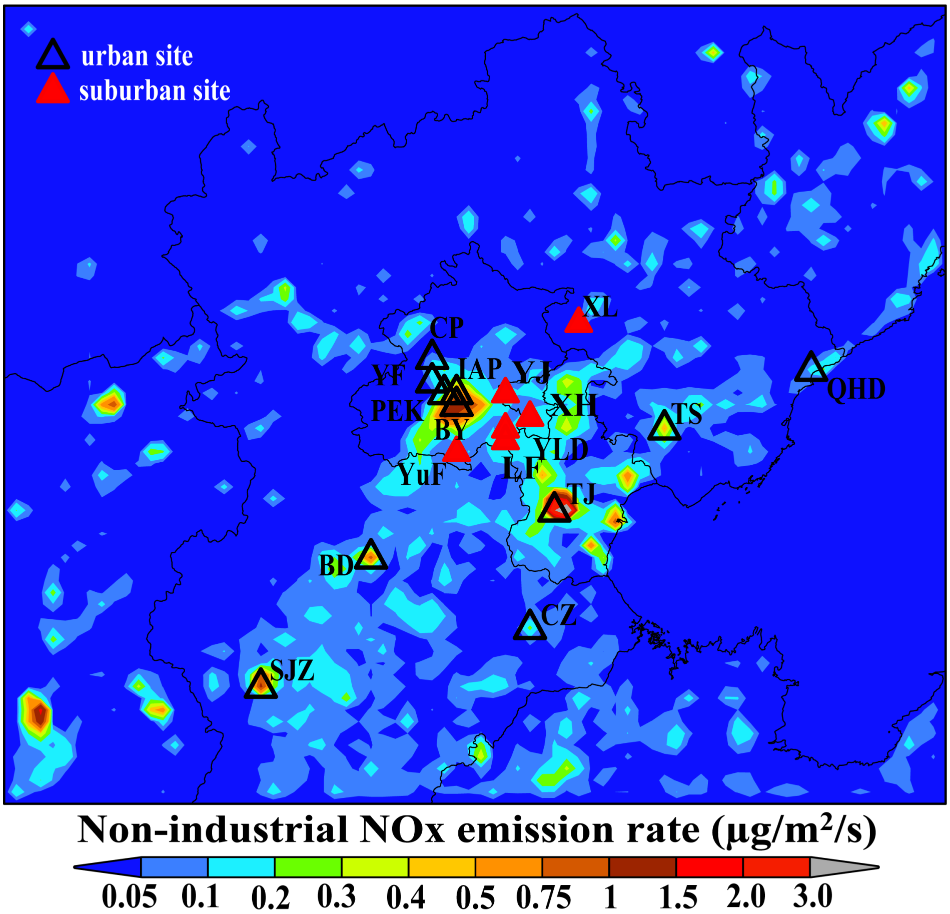


Figure 1 Distribution of the observation stations and non-industrial NOx emission rates in the third model domain (9km resolution) that covers Beijing and its surrounding areas. The non-industrial NOx emission rates (μg/m^2/s) are divided into different bins (<0.05; 0.01-0.1; 0.1-0.2; 0.2-0.3; 0.3-0.4; 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 NOx emission rates are marked by the brown and red colors, and the suburban or rural areas with low non-industrial NOx 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.

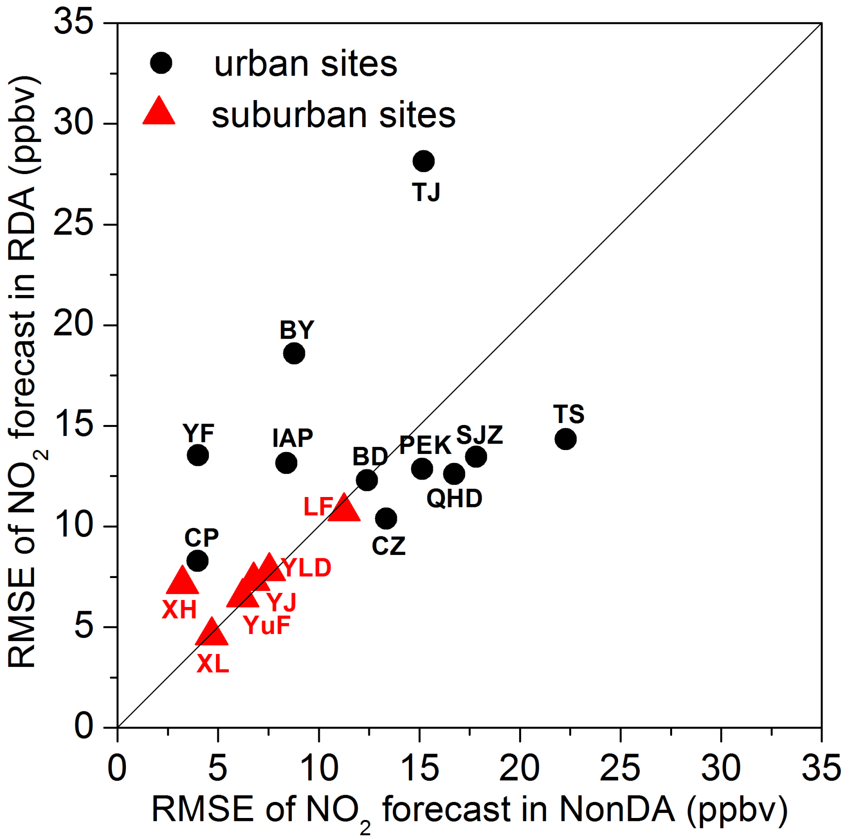
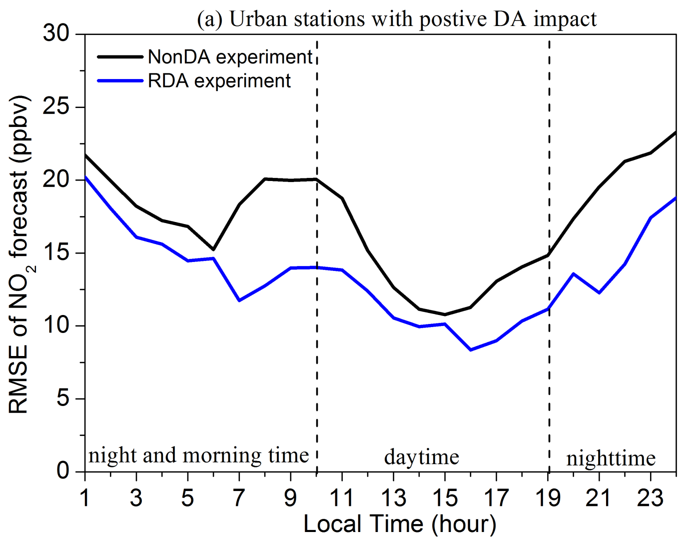
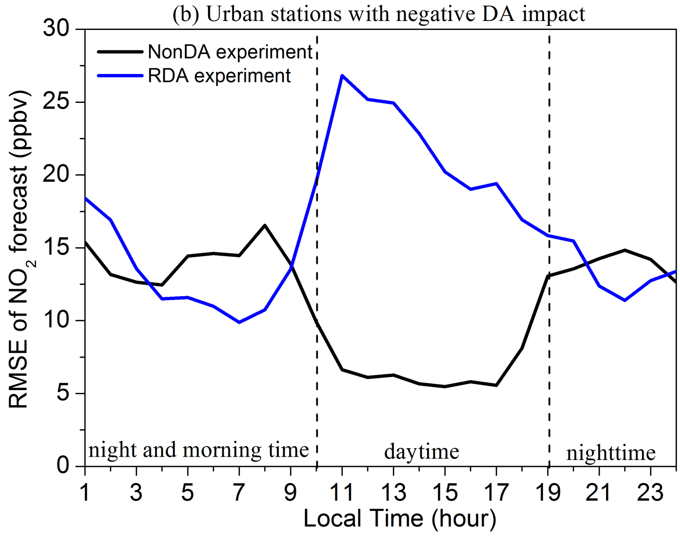


Figure 2 Comparison of the root mean square err[illegible] (RMSEs) (ppbv) of 1h NO_2 forecasts at the 17 stations of Beijing and its surrounding areas durin[illegible]e period of 00:00 L[illegible] August to 00:00 LT 23 August in 2008 in the real data assimilation (RDA) experiments and those [illegible] 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. The number of the valid observations used for the calculation is 336 at QHD, SJZ, TS, IAP, LF, YF and XH, and the numbers are 292, 226, 326, 317, 326, 320, 333, 321, 311, 323 at BD, PEK, BY, CZ, CP, TJ, XL, YJ, YLD and YuF respectively.





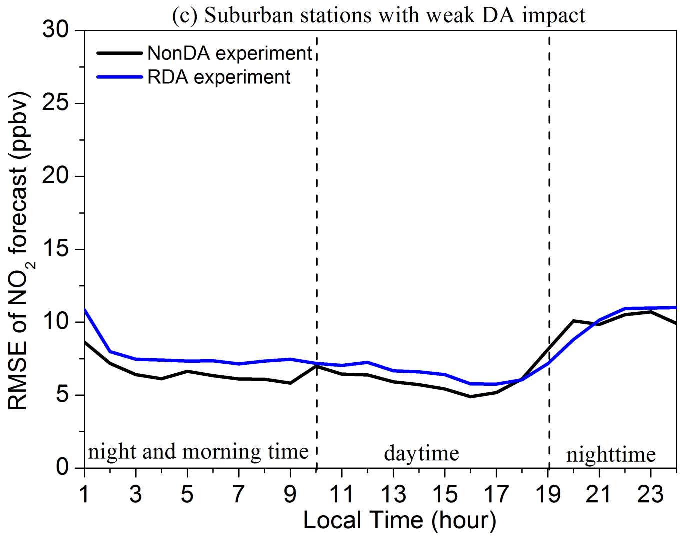
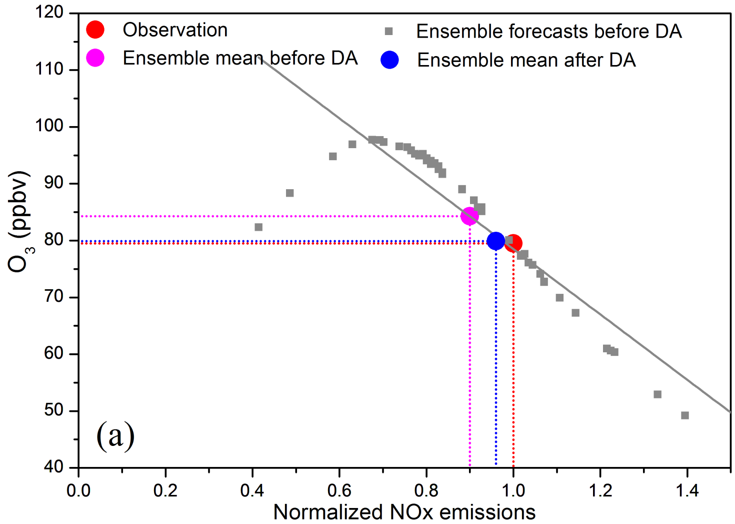


Figure 3 Daily variation of the 1h NO_2 forecast RMSEs (ppbv) in the real data assimilation (RDA) experiments (blue line) and the reference (NonDA) experiment with a free run of the model (black line) over: (a) urban stations (CZ, PEK, QHD, SJZ, and TS) with positive DA impacts; (b) urban sites (BY,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



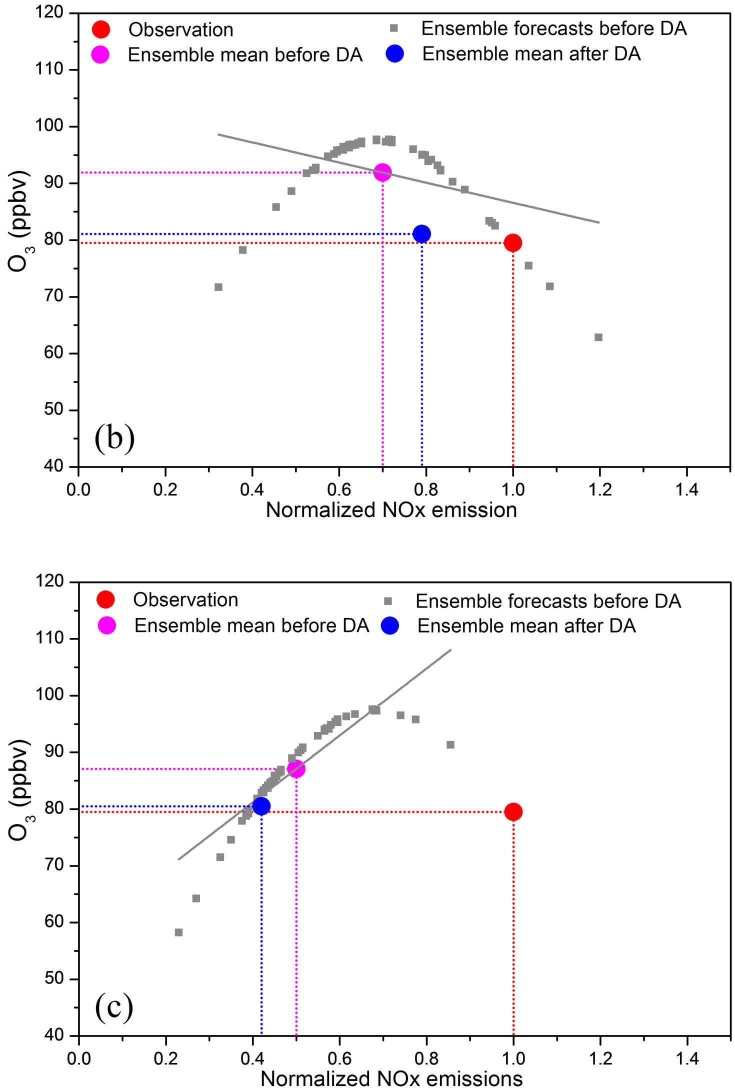
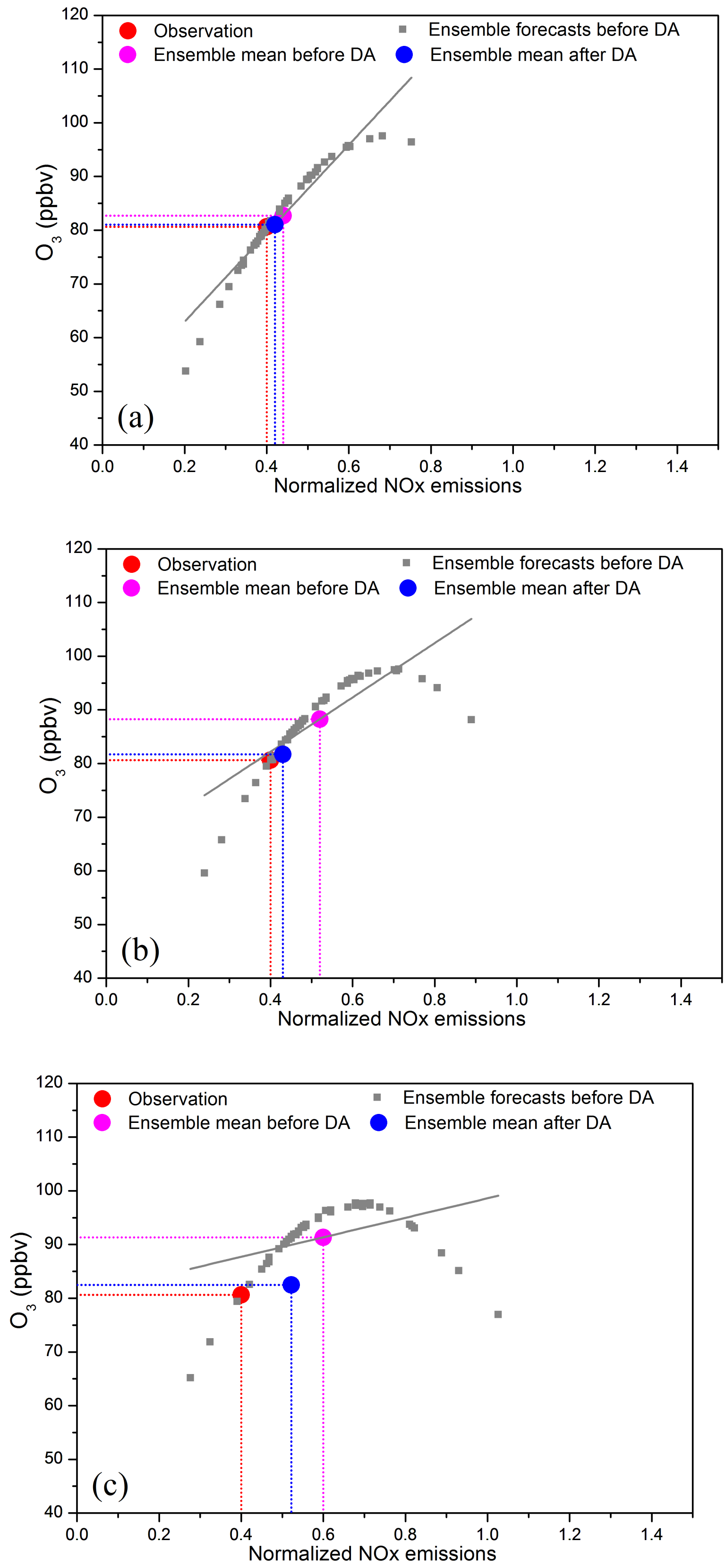


Figure 4 (a-c) O_3 concentrations (ppbv) and NOx emissions (no unit, normalized by the true NOx emission) before and after data assimilation (DA) and their ensemble samples before DA at 12:00 LT on August 12, 2008 in the three ideal ozone data assimilation experiments with the prior NOx emissions underestimated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


Observation
Ensemble mean before DA
Ensemble forecasts before DA
Ensemble mean after DA
O_3 (ppbv)
Normalized NOx emissions
(a)
(b)
(c)

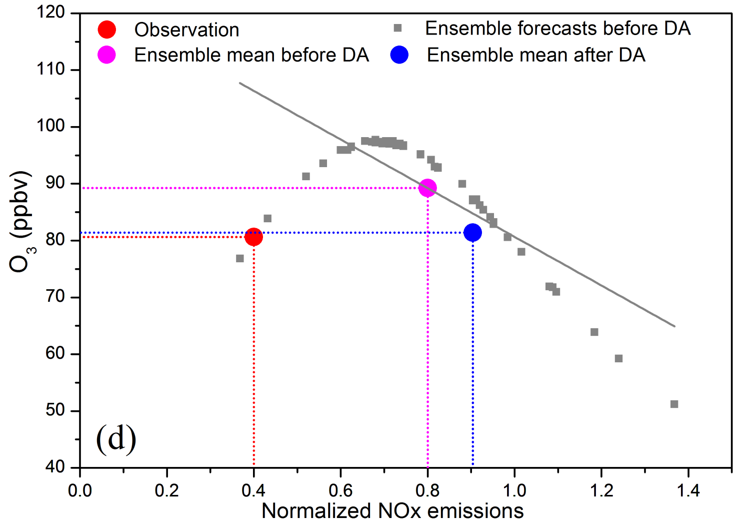
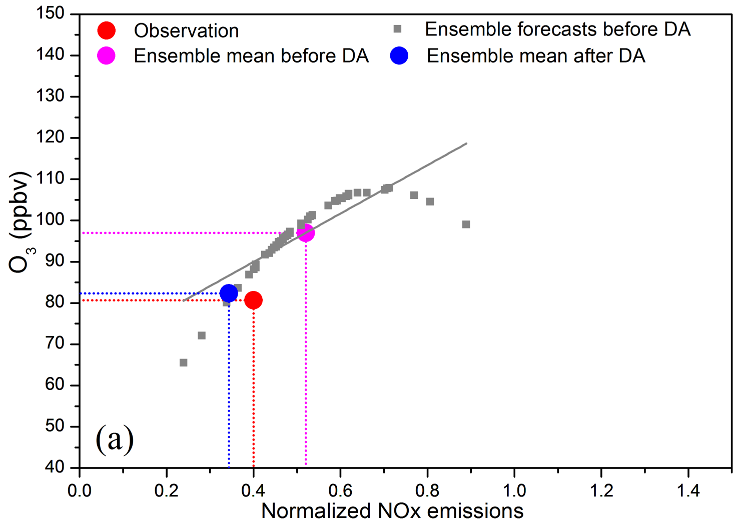
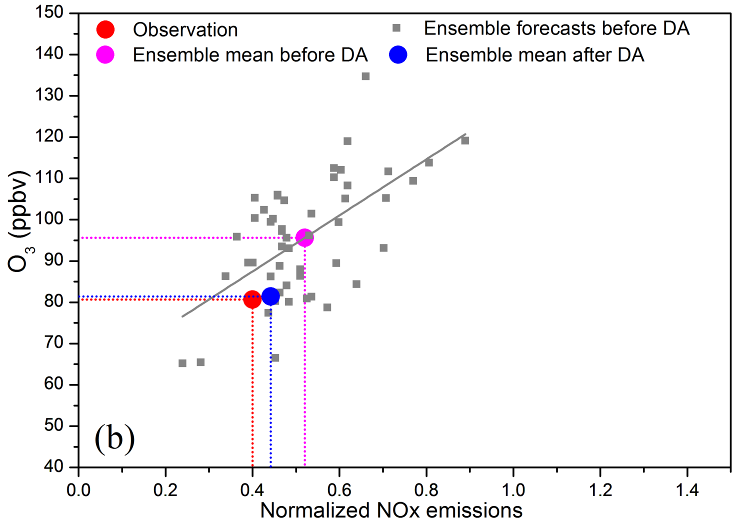


Figure 5 (a-d) O_3 concentrations (ppbv) and NOx emissions (no unit, normalized by the true NOx emission) before and after data assimilation (DA) and their ensemble samples before DA at 12:00 LT on August 12, 2008 in the four idealized DA experiments. (a) DA experiment with 10% overestimation in the a NOx emission estimation; (b) DA experiment with 30% overestimation in the a priori NOx emission estimation; (c) DA experiment with 50% overestimation in the a priori NOx emission; (d) DA experiment with 100% overestimation in the a priori NOx emission. The magenta dot, the gray squares, the gray line, the red dot and the blue dot represent the samples in Fig. 4.





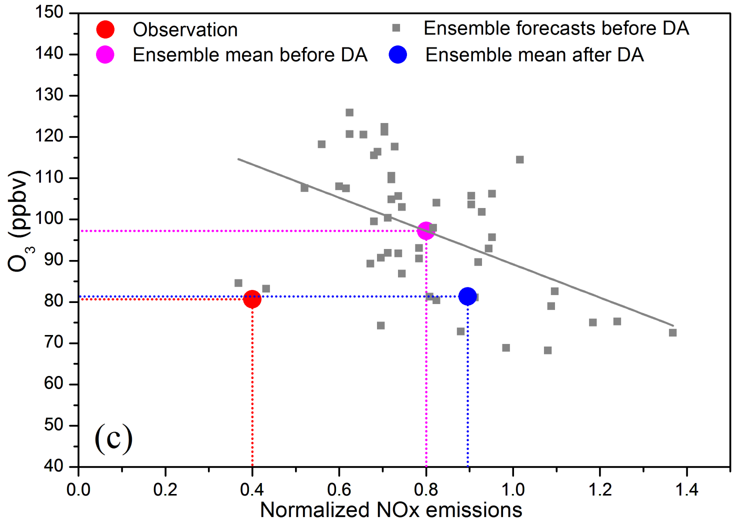


Figure 6 (a-c) O_3 concentrations (ppbv) and NOx emissions (no unit, normalized by the true NOx emission) before and after data assimilation (DA) and their ensemble samples before DA at 12:00 LT on August 12, 2008 in the three ideal DA experiments. The NO_2 photolysis rate is assumed to being overestimated by 20%. (a) The prior NOx emission is overestimated by 30% and adjusted by the DA, but the uncertainty of the NO_2 photolysis rate is missed (without perturbations on the NO_2 photolysis rate) in the DA. (b) The same as the DA experiment in (a), but the uncertainty of the NO_2 photolysis rate is taken into account through perturbing it. (c) The same as the DA experiment in (b), but the bias in the prior NOx emission is increased to 100%. The magenta dot, the gray squares, the gray line, the red dot and the blue dot represent the same as in Fig. 4.

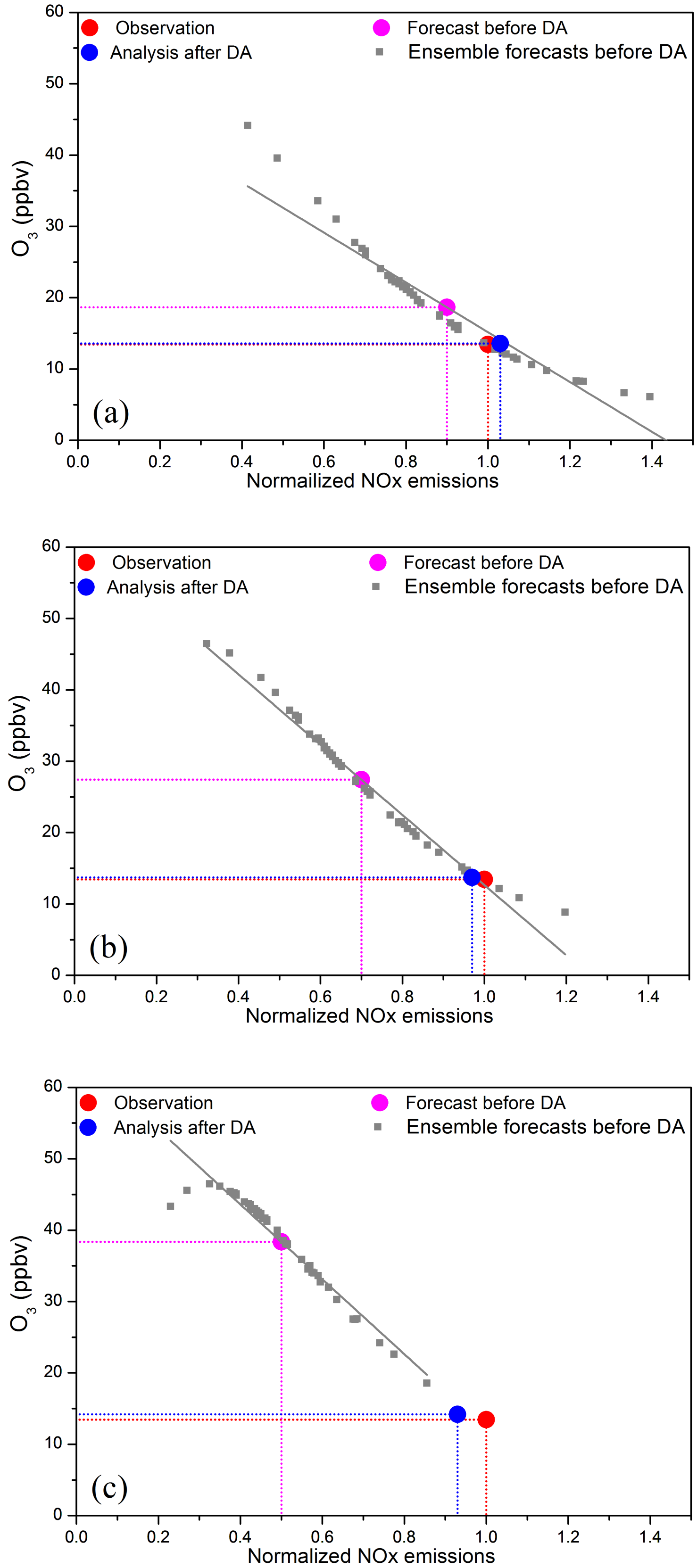


Figure 7 (a-c) O_3 concentrations (ppbv) and NOx emissions (no unit, normalized by the true NOx emission) before and after data assimilation (DA) and their ensemble samples before DA 08:00 LT on August 12, 2008 in the three ideal ozone data assimilation experiments with the prior NOx emissions underestimated 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 Figs. 4.