

# Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions

X. Tang, J. Zhu, Z. F. Wang, and A. Gbaguidi

LAPC and ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Received: 18 February 2011 – Published in Atmos. Chem. Phys. Discuss.: 8 March 2011

Revised: 31 October 2011 – Accepted: 9 December 2011 – Published: 20 December 2011

**Abstract.** In order to improve the surface ozone forecast over Beijing and surrounding regions, data assimilation method integrated into a high-resolution regional air quality model and a regional air quality monitoring network are employed. Several advanced data assimilation strategies based on ensemble Kalman filter are designed to adjust O<sub>3</sub> initial conditions, NO<sub>x</sub> initial conditions and emissions, VOCs initial conditions and emissions separately or jointly through assimilating ozone observations. As a result, adjusting precursor initial conditions demonstrates potential improvement of the 1-h ozone forecast almost as great as shown by adjusting precursor emissions. Nevertheless, either adjusting precursor initial conditions or emissions show deficiency in improving the short-term ozone forecast at suburban areas. Adjusting ozone initial values brings significant improvement to the 1-h ozone forecast, and its limitations lie in the difficulty in improving the 1-h forecast at some urban site. A simultaneous adjustment of the above five variables is found to be able to reduce these limitations and display an overall better performance in improving both the 1-h and 24-h ozone forecast over these areas. The root mean square errors of 1-h ozone forecast at urban sites and suburban sites decrease by 51 % and 58 % respectively compared with those in free run. Through these experiments, we found that assimilating local ozone observations is determinant for ozone forecast over the observational area, while assimilating remote ozone observations could reduce the uncertainty in regional transport ozone.

## 1 Introduction

As one of the typical city clusters in China, Beijing and its surrounding areas are facing serious challenges in surface ozone pollutions within urbanization and motorization processes (Chan and Yao, 2008; Shao et al., 2006). Exposure to high ozone concentrations leads to heavy damages on both human health and plant life (Anderson et al., 1996; Chameides et al., 1999). Providing ozone forecast is undoubtedly of great importance, not only for the public, but also for the policy makers. Forecasting and early warning of ozone pollution, performed during the Beijing Olympic Games to ensure a health environment for athletes and attendees, constituted an important issue for the Campaign of Air Quality Research in Beijing and the Surrounding Region (CAREBEIJING-2008, Wang et al., 2011). However, ozone forecast has not been integrated in the current operational air quality forecast over these areas, and little studies focused on this issue.

In previous studies on ozone forecast over Beijing, An et al. (2010) and Yu et al. (2011) developed statistical forecast models to forecast ozone concentrations based on several statistical techniques including multiple linear regressions, principal component analysis and neural network methods, while Tang et al. (2010a) and Zhang et al. (2010) employed ensemble forecast methods based on chemical transport model (CTM) to forecast ozone. A main drawback with the statistical forecast model is the difficulty in describing non-local influences such as emission changes, transport processes and complex chemical reactions (Flemming et al., 2001). The ensemble forecasting methods with CTM contain the influences from the complex chemical and dynamical processes and do not have the conceptional limitations with statistical



Correspondence to: J. Zhu  
(jzhu@mail.iap.ac.cn)

forecast method. Its limitations lie in that large uncertainty in CTM is still a great challenge so that forecast with ensemble mean brings limited improvement of forecast skill (Mallet et al., 2009; von Loon et al., 2007). Linear combination of each ensemble member based on past observations and past forecasts can produce better forecast performances (Mallet et al., 2009; Zhang et al., 2010), but the effectiveness of the combining weights is limited to the locations and variables with observations, and the errors in observation are normally not taken into account (Mallet et al., 2010).

In this paper, advanced data assimilation method as an alternative approach is employed to improve the ozone forecast of CTM over Beijing and its surrounding areas. This is the first attempt to employ data assimilation method to overcome difficulties related to regional ozone forecast over these areas. This method consists of integration of observational information into a numerical model with the purpose of obtaining the estimate of the model state that minimizes the error variance. The attractive features of data assimilation method reside not only in its ability to improve the estimations of the physical properties that are not observed directly, but also in its adequate consideration of the observation errors. Several applications of data assimilation method in ozone modeling brought out relevant findings for ozone forecast improvement in many locations (Chai et al., 2006; Elbern et al., 2007; van Loon et al., 2000; Hanea et al., 2004; Constantinescu et al., 2007a). However, adoption of suitable data assimilation method with well-fitting data strategy for ozone forecast improvement over Beijing and its surrounding areas has not yet been addressed in previous publications.

Compared with the previous ozone data assimilation studies focusing on the ozone forecast over Europe or North America, ozone data assimilation over Beijing and surrounding regions has its particularities. At first, as a typical city cluster within urbanization and motorization processes, Beijing and surrounding regions is undergoing rapid and complex variations of air pollutant emissions. The number of motor vehicles in Beijing rocketed from about 3.5 million in 2008 to around 5 million in 2010, while several emission reduction measures such as traffic restriction are conducted. These rapid changes can hardly be described by the existing emission inventory updated once every 2–3 years, inducing large uncertainty into ozone modeling. What is more, complex regional transport processes of air pollutants over these areas (Streets et al., 2007; Wu et al., 2011) complicate this problem and become a barrier to improve the ozone forecast. Another difficulty for improving ozone forecast over these areas is the lack of routine observations of ozone and precursors. Therefore, how to deal with the above problems and overcome such difficulties is a key ingredient for improving the ozone forecast over these areas.

The objective of this study is to investigate the possibility to overcome these difficulties with data assimilation method. Several data assimilation strategies are designed to adjust ozone initial conditions, precursor initial conditions

and emission rates separately or jointly through assimilating ozone observations, exploring possible solutions to reducing the uncertainty in precursor emission and initial conditions when precursor observations are scarce. A regional air quality observation network covering Beijing, Tianjin and Hebei Province is employed to provide ozone observations for data assimilation, and the roles of the monitoring stations of this network in data assimilation are evaluated. Ensemble Kalman filter (EnKF) is employed as the data assimilation method for its strong attractive features in application for complex models (Carmichael et al., 2008; Evensen, 2009). It supports fully nonlinear evolution of the error statistics through the highly nonlinear model and is convenient to dealing with model error. Furthermore, its implementation is very simple and suitable for parallel computation without needing for tangent linear or adjoint model.

Section 2 describes the adopted data assimilation method, regional air quality model, regional air quality observation network and the designed experiments. Results and discussions are presented in Sect. 4 and conclusions are given in Sect. 5.

## 2 Methodology

### 2.1 Data assimilation with EnKF

EnKF, proposed by Evensen (1994), is an approximate version or extension of Kalman filter (Kalman, 1960). It uses ensembles of random samples to obtain error statistics of model state variable or parameter. Error statistics can be propagated with linear or nonlinear dynamic model through simply implementing ensemble simulations of the dynamic model. Several variants of EnKF are available for application in large geophysical system (Anderson, 2001; Houtekamer and Mitchell, 2001; Keppenne, 2000; Sakov and Oke, 2008). We adopt the sequential algorithm proposed by Houtekamer and Mitchell (2001) to implement EnKF for its efficiency in computation. The implementation process and setup are detailed below.

#### 2.1.1 Definition of state vector

In CTM, the state vector  $\mathbf{x}$  of model system evolves from time  $k-1$  to time  $k$  can be represented in discrete form:

$$\mathbf{x}_k^f = \mathbf{M}_{k-1}(\mathbf{x}_{k-1}^b, \boldsymbol{\theta}_{k-1}^b) \quad (1)$$

where the superscripts  $f$  and  $b$  denote forecast and background (or first guess) respectively,  $\mathbf{M}_{k-1}$  denotes the model dynamic operator.  $\boldsymbol{\theta}$  represents model inputs such as meteorological inputs, chemical reaction parameters and emissions. The state vector  $\mathbf{x}$  defined in this study contains  $\text{O}_3$  initial conditions, VOCs and  $\text{NO}_x$  initial conditions. Note that the state vector can be extended to including more other variables.

### 2.1.2 Initial perturbation of state vector

A key step of EnKF is generating an initial set of samples of state vector to provide initial conditions for Monte Carlo ensemble simulations. The initial ensemble samples are obtained through perturbing the background value of state vector:

$$\mathbf{x}'_0(i) = \mathbf{x}^b_0 + \delta\mathbf{x}_0(i), \quad i = 1, 2, \dots, N \quad (2)$$

where  $\delta\mathbf{x}_0$  represents the random perturbation samples added to the background value at the initial time, and  $i$  denotes its index in ensemble.  $\mathbf{x}'_0$  is the obtained initial ensemble sample of state vector. The ensemble size  $N$  is set as 50, which can keep a good balance between computational efficiency and assimilation performance and has been proved to be credible for application in ozone data assimilation by previous publications (Carmichael et al., 2008; Constantinescu et al., 2007b).

Ideally, initial perturbation should reflect the statistic characterization of the error in background values (Evensen, 2003). For application in CTM, the importance of initial perturbation seems not as great as that in application for meteorological and ocean model. Wu et al. (2008) present credible results of ozone data assimilation without initial perturbation in EnKF. In this study, we employ the method suggested by Evensen (1994) to generate a pseudo smooth random perturbation field in three dimensions. This method is convenient for setting the amplitude, horizontal and vertical scales of perturbations. The perturbation ranges of initial conditions of O<sub>3</sub>, NO<sub>x</sub> and VOCs are assumed to be 50 % of the background values. The spatial correlation scale of initial perturbation fields is set as 54 km in horizontal and 3 model grids (about 200 m) in vertical after several sensitivity tests. The correlation scale in horizontal or vertical is chosen independently from several scales with the purpose of minimizing the root mean square error (RMSE) of the analyzed ozone. In addition, during initial perturbation, the variables in state vector are assumed to be independent with each other considering the difficulty in obtaining their correlations directly.

### 2.1.3 Perturbation of input parameter

The forecast error of state vector comes not only from the initial error of state vector, but also from error in other sources such as model parameter or numerical technique. We define the latter as model error. A critical issue for data assimilation is how to deal with model errors. Neglecting model errors in EnKF may lead to filter divergence which is characterized by too small ensemble spread and disregard of observation during analysis (Mitchell and Houtekamer, 2000). A simple method to compensate the missed model errors is inflating background error covariance (Constantinescu et al., 2007c). However, lacking physical basis and inducing spurious linear increase of background error at the area far away from

observation sites make this method limited. In this study, we adopt an alternative approach to deal with model errors.

First, we assume model errors are mainly from uncertain model parameters ( $\theta$  in Eq. 1). In order to represent their uncertainty in the form of ensemble, Gaussian distributed random noise is generated and added to the first guess value ( $\theta^b$ ) of model parameters at each integration step as below:

$$\boldsymbol{\theta}'_{k-1}(i) = \boldsymbol{\theta}^b_{k-1} + \delta\boldsymbol{\theta}_{k-1}(i), \quad i = 1, 2, \dots, N \quad (3)$$

where  $\delta\boldsymbol{\theta}$  is the random perturbation sample from Gaussian distribution, and  $i$  denotes its index in ensemble. Precursor emissions, photolysis rates and vertical diffusion coefficients are assumed to be dominant sources of model error and are perturbed in Eq. (3). Referring to the uncertainty analysis result of Tang et al. (2010b), the perturbation magnitude of NO<sub>x</sub> emissions is restricted within 60 % of the first guess emission rates and that of VOCs emissions within 80 %. The perturbations of NO<sub>2</sub> photolysis rate and vertical diffusion coefficient are restricted within 30–35 % of the first guess values respectively.

Here, we split the above parameters into two categories: the parameters needed to be updated by EnKF (denoted by  $\eta$ ); the parameters without needing update (denoted by  $\lambda$ ). For  $\eta$ , we adopt a time-correlated noise to simulate the temporal evolution of their errors:

$$\boldsymbol{\eta}'_k(i) = \boldsymbol{\eta}^b_k + \delta\boldsymbol{\eta}_k(i)$$

$$\delta\boldsymbol{\eta}_k(i) = \alpha\delta\boldsymbol{\eta}_{k-1}(i) + \sqrt{1-\alpha^2}\sigma w_{k-1}(i), \quad i = 1, 2, \dots, N \quad (4)$$

where  $\boldsymbol{\eta}^b_k$  is the first guess of the parameter and  $\delta\boldsymbol{\eta}_{k+1}(i)$  denotes the random perturbation sample.  $w_{k-1}(i)$  is random sample drawn from Gaussian distribution with mean zero and standard deviation one,  $\sigma$  denotes the standard deviation of the error in parameter.  $\alpha$  represents the smooth coefficient that is dependent on time-decorrelation scale ( $\tau$ ):

$$\alpha = \exp(-1/\tau) \quad (5)$$

We use 24 h as the first guess value of the time-decorrelation scale, similar as the decorrelation length employed by Segers (2002). Note that this assumption may not be true; other options might improve the performance of EnKF. The advantages of using a colored noise process for the adjusted parameter lie in that it can avoid rapid fluctuations of perturbations (e.g., the fluctuation of perturbation from -30 % to 30 % within 2 h). Correlations between the red noise of the parameter and the ensemble of state variable will be developed, making it possible to update the parameter consistently with observations of state variable (Evensen, 2003). For  $\lambda$ , their errors are simulated as white noise process without temporal correlations:

$$\boldsymbol{\lambda}'_k(i) = \boldsymbol{\lambda}^b_k + \delta\boldsymbol{\lambda}_k(i) \quad (6)$$

$\delta\boldsymbol{\lambda}_k(i)$  is random sample drawn from Gaussian distribution with mean zero and standard deviation set as the perturbation

magnitude of parameters. It should be noted that in the actual simulations only the parameters perturbed with red noise are modified by EnKF before being used by the simulation.

#### 2.1.4 Update of state vector and input parameter

After obtaining ensemble of initial conditions of state vector and ensemble of parameters, ensemble runs of original CTM are conducted to propagate these errors in model. Each initial ensemble sample of state vector from Eq. (2) and ensemble sample of parameter from Eqs. (4) and (6) serve as inputs in each ensemble run, and the Eq. (1) is transformed as:

$$\mathbf{x}_k^f(i) = \mathbf{M}_{k-1}(\mathbf{x}'_{k-1}(i), \boldsymbol{\eta}'_{k-1}(i), \boldsymbol{\lambda}'_{k-1}(i)), \quad i = 1, 2, \dots, N \quad (7)$$

Note that  $\boldsymbol{\theta}$  is replaced by  $\boldsymbol{\eta}$  and  $\boldsymbol{\lambda}$  here. Then, a key step of EnKF is to update the forecast state vector  $\mathbf{x}^f$  and the parameter  $\boldsymbol{\eta}$  in Eq. (7) through assimilating observation data.

To facilitate the description, we define an extended state vector including parameter  $\boldsymbol{\eta}$ :

$$\mathbf{b}_k^f(i) = \begin{vmatrix} \mathbf{x}_k^f(i) \\ \boldsymbol{\eta}_k^f(i) \end{vmatrix}, \quad i = 1, 2, \dots, N \quad (8)$$

The forecast error covariance  $\mathbf{P}^f$  of the extended state vector is estimated based on the forecast ensemble from Eq. (9):

$$\mathbf{P}_k^f = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{b}_k^f(i) - \overline{\mathbf{b}}_k^f)(\mathbf{b}_k^f(i) - \overline{\mathbf{b}}_k^f)^T \quad (9)$$

where the overline denotes the ensemble mean. The observation error is assumed to be Gaussian with mean zero and covariance  $\mathbf{R}$ . An ensemble of observation samples is generated accordingly:

$$\begin{aligned} \mathbf{y}_k'(i) &= \mathbf{y}_k + \boldsymbol{\gamma}_k(k), \quad i = 1, 2, \dots, N \\ \boldsymbol{\gamma}_k &\in N(0, \mathbf{R}_k) \end{aligned} \quad (10)$$

where  $\mathbf{y}_k$  is the original observation value and  $\boldsymbol{\gamma}_k$  is the random perturbation sample from Gaussian distribution. The observation errors, including both representation error and measurement error, are assumed to be uncorrelated in time and space and are set as 10 % of the original observation value with reference to von Loon et al. (2000).

Based on the error statistics of the forecast and observational state vector, the extended state vector is updated according to the following formulations:

$$\begin{aligned} \mathbf{b}_k^a(i) &= \mathbf{b}_k^f(i) + \mathbf{K}_k(\mathbf{y}_k' - \mathbf{H}\mathbf{b}_k^f(i)), \quad i = 1, 2, \dots, N \\ \mathbf{K}_k &= \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \end{aligned} \quad (11)$$

$\mathbf{H}$  is the observation operator that maps the extended state vector from model space to the observation space.  $\mathbf{K}$  represents the Kalman gain dependent on forecast error covariance and observation error covariance.  $\mathbf{b}_k^a$  is the updated extended

state vector (analysis). In order to avoid storing and inverting very large matrices during analysis, ozone observations at different sites are assimilated into model in a sequential way with observations assimilated site by site. The updated extended state vector from assimilating observation of one site is used as the background for assimilating observation at next sites. The sequential way has been reported to be better than the way with observations of all sites assimilated simultaneously as long as the observation errors of different sites are uncorrelated (Houtekamer and Mitchell, 2001).

A major limitation of EnKF is the use of finite ensemble size which can induce spurious correlation between two independent variables leading to underestimation of analysis error covariance and spurious increment of state vector (Evensen, 2009). In order to reduce the spurious influence caused by the finite ensemble size, we employ a local analysis scheme in which only observations within a cutoff radius (localization scale) of the analysis grid are used to update the state vector. Its advantage lies in its convenience in eliminating the weak influence from remote observation, which is difficult to be accurately estimated. After several sensitivity tests with various localization scales (81 km, 72 km, 63 km, 54 km, 45 km, 36 km), the cutoff radius for updating ozone initial conditions is set as 54 km and that for updating precursor initial conditions and emissions 45 km. The current localization scale brings better forecast skill than the other tested scales. It is worth noting that many factors such as ensemble size, dynamic system and lifetime of chemical species would influence the choice of localization scale.

### 3 Air quality model and observation network

#### 3.1 Regional air quality model

The CTM used with EnKF is the Nested Air Quality Prediction Modeling System (NAQPMS) developed by the Institute of Atmospheric Physics of Chinese Academy of Sciences (Wang et al., 1997, 2006). Several applications of NAQPMS have been reported for simulating chemical process and transport of ozone (Li et al., 2009; Tang et al., 2010c), modeling process of aerosol and acid rain (Li et al., 2011; Wang et al., 2002) and providing operational air quality forecast in mega cities such as Beijing and Shanghai (Wang et al., 2006).

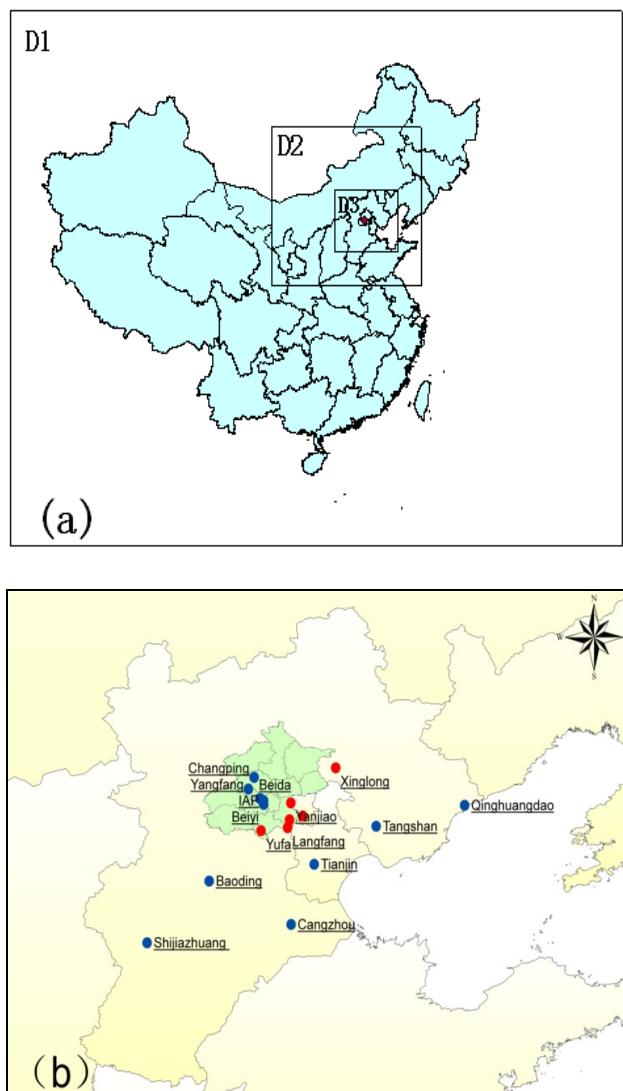
As a multi-scale air quality model, NAQPMS can provide forecast of both primary and secondary pollutant from regional to urban scale. It includes modules of emissions, diffusion, advection/convection, deposition and gas/aqueous chemistry. The process of gas phase chemistry is modeled with the Carbon-Bond Mechanism Z (CBM-Z) proposed by Zaveri and Peters (1999). A revised version of RADM aqueous-phase chemistry (Wang et al., 2002) is served to simulate the aqueous-phase chemistry. The dry deposition modeling follows the scheme of Wesely (1989).

In this study, the model is configured with three nested model domains (displayed in Fig. 1a) in order to reduce the influence of boundary conditions on our focus areas. The coarse domain (D1 and D2) is to provide boundary conditions for its nested domain with one-way nesting technique. The boundary conditions of the largest domain are extracted from a global transport model CHASER (Sudo et al., 2002). 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 50 m. The hourly meteorological driver of chemical transport model is provided by the Fifth-Generation NCAR/Penn State Mesoscale Model (MM5) (Grell, 1994) which is configured with the same horizontal grid structure as the chemical transport model. The initial and boundary conditions for MM5 run are from NCAR/NCEP  $1^\circ \times 1^\circ$  reanalysis data.

The gridded emission data of the three domains is prepared via the Sparse Matrix Operator Kernel Emissions (SMOKE) model (Houyoux et al., 2000) to combine and process the following emission inventories. The INTEX-B Asia inventory for 2006 with a  $0.5^\circ \times 0.5^\circ$  resolution (Zhang et al., 2009) serves as the regional emission inventory for all model domains. Then the power emissions of Beijing and its surrounding provinces (Tianjin, Hebei, Shanxi, Inner Mongolia and Shandong) in this inventory are updated by the power plant emission dataset with exact longitude and latitude of point sources (Hao et al., 2007). Other emissions of Beijing in INTEX-B emission are also updated with the industrial boiler, domestic, industrial process emissions, and the mobile emissions derived from Mobile 6 of SMOKE based on the traffic flow from the annual report of Beijing Traffic Development Research Center in 2006 (Wu et al., 2010). In order to reflect the emission control measures conducted over Beijing and its surrounding areas during the Beijing Olympic Games, we removed the emission sources corresponding to the control measures from the above updated emission inventory according to the “29th Olympic Games Beijing air quality protection measures”. More details for the emission control measures during the Beijing 2008 Olympic Games are provided in Wang et al. (2009).

### 3.2 Regional air quality observation network

A regional air quality observation network covering Beijing, Tianjin and Hebei Province is used to provide surface ozone observational data for data assimilation. This network was established in 2008 as part of the air quality protection project for the Beijing 2008 Olympic Games. It provided hourly observed data of  $O_3$ ,  $NO_x$ ,  $SO_2$ ,  $CO$ ,  $PM_{2.5}$  in near real-time during the Beijing 2008 Olympic Games. The aims of this network are to monitor the variations of the regional air quality and assess the influence of regional pollutant transport on Beijing's air quality. It has been applied to evaluate model performance (Zou et al., 2010), improve ozone forecast with an ensemble forecast system (Tang et



**Fig. 1.** (a) Model domains. The first domain covers East Asia with  $81\text{ km} \times 81\text{ km}$  resolution; the second domain consists of North China with  $27\text{ km} \times 27\text{ km}$  resolution; the third domain includes Beijing and its surrounding areas with  $9\text{ km} \times 9\text{ km}$  resolution. (b) Monitoring stations. The six suburban stations are marked as red dots and the eleven urban stations are represented by blue dots.

al., 2010a), and assess the air quality of Beijing and its surrounding areas during the Beijing 2008 Olympic Games (Xin et al., 2010).

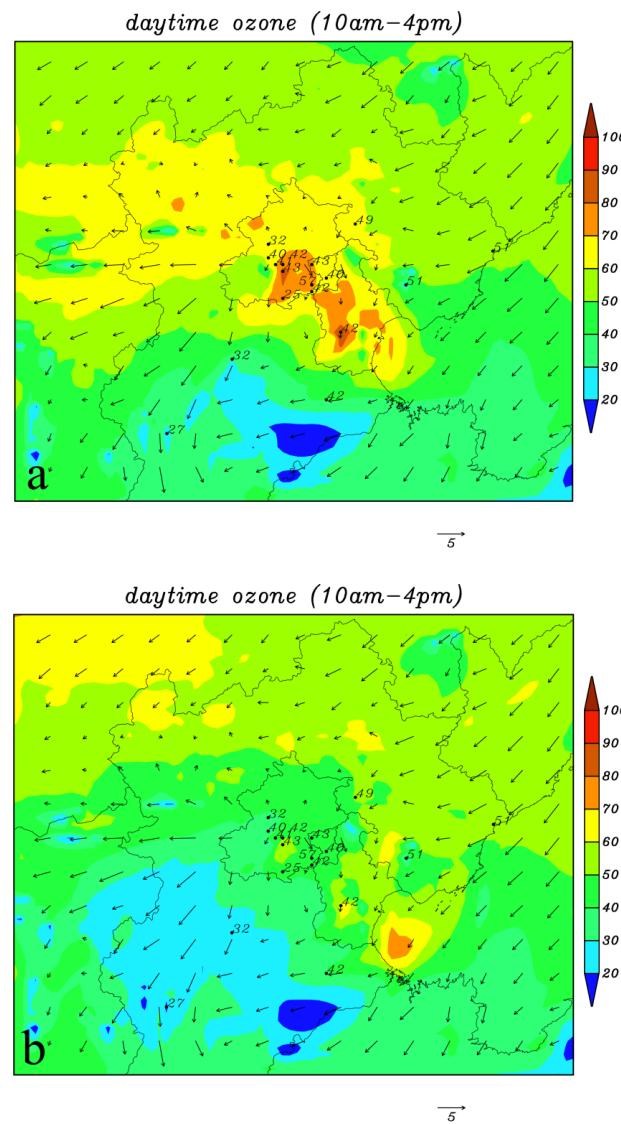
In this study, this network is employed for data assimilation to investigate its potentials in improving ozone forecast. The distribution of the 17 monitoring stations of this network is displayed in Fig. 1b. It contains five urban sites at Beijing (Changping, Beida, Beiyi, IAP, Yangfang), six suburban sites close to Beijing (Langfang, Xianghe, Xinglong, Yanjiao, Yufa, Yongledian), and six urban sites at the surrounding cities (Baoding, Cangzhou, Qinghuangdao,

Shijiazhuang, Tangshan, Tianjin). The urban sites are located at the central urban area with high precursor emission rates, while the suburban sites are located away from the central urban area with a relatively low precursor emission rates.

### 3.3 Experiment design

In order to identify the advantages and disadvantages of adjusting initial conditions and emissions through assimilating ozone observations, we design a series of experiments as shown in Table 1. The focus is on using EnKF method to adjust ozone initial conditions, VOCs initial conditions, NO<sub>x</sub> initial conditions, VOCs emission rates and NO<sub>x</sub> emission rates separately or jointly. Firstly, a free run of model is conducted as the reference experiment EXP0 for further comparison with the data assimilation experiments. Ozone initial conditions are adjusted in EXP1, while initial conditions of VOCs and NO<sub>x</sub> are adjusted in EXP2 and EXP4 respectively. EXP3 and EXP5 are implemented to correct emissions of VOCs and NO<sub>x</sub> respectively. In EXP6, a combined experiment is designed to simultaneously adjust the initial conditions and emissions. In order to verify the effects of data assimilation on the non-observational areas, ozone observations at two urban sites (IAP and Yangfang) and one suburban site (Langfang) are withdrawn from assimilation in the six data assimilation experiments and used as independent data for validation. Moreover, two cross validation data assimilation experiments, EXP6u and EXP6s, are designed to further investigate the effects of data assimilation on ozone forecast over the non-observational areas.

The analyses focus on a two-week simulating period from 00:00 LT 9 August to 00:00 LT 23 August in 2008. The meteorological inputs of air quality modeling each day are obtained from a 1.5-day MM5 run, with the first 12-hour as spin-up time and the remaining one day to provide the meteorological inputs. A free run of MM5 without nudging is conducted with considering of the lack of suitable meteorological observations in the fine-resolution nested domains and the short forecast period of each run. Overall, the predicted wind, temperature and relative humidity agree well with the observations except for the underestimation of surface temperature over urban areas and inconsistencies of modeling wind during 13–17 August (not shown here). For air quality modeling, we firstly conducted two-week spin-up simulation with clean initial conditions in order to minimize the influence of artifacts from initial conditions. Then we perturbed the initial conditions of state vector at 19:00 LT on 8 August 2008 in the way as Eq. (2) and started implementation of ensemble simulation. The first five hours of ensemble simulation are also set as spin-up time, which would be helpful for recovering the balance of model after perturbation and obtaining a flow-dependent background error covariance. At 00:00 LT on 9 August 2008 the observed ozone data started to be assimilated hour by hour. The assimilation ended at 00:00 LT on 23 August 2008. In order to reduce the compu-



**Fig. 2.** Simulated daytime ozone concentrations (ppbv) on August 14 in the third domain obtained from (a) free run and (b) EXP1. The observed daytime ozone values (numbers near dots) at monitoring stations are also shown.

tational cost, data assimilation is only performed in the third model domain. To investigate the quick response of ozone forecast to assimilation, we mainly assess the performance of 1-h ozone forecast after assimilating observational data. Furthermore, the effects of data assimilation on 24-h forecast are also discussed in Sect. 4.6.

## 4 Results and discussions

### 4.1 Adjustment of ozone initial conditions

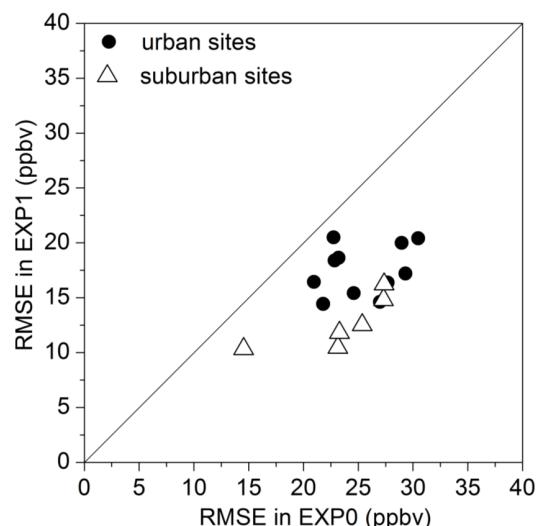
As the most frequently employed control variable in ozone data assimilation studies (e.g., Eben et al., 2005; Elbern et al.,

**Table 1.** Experiments designed for evaluating the performance of different data assimilation strategies.

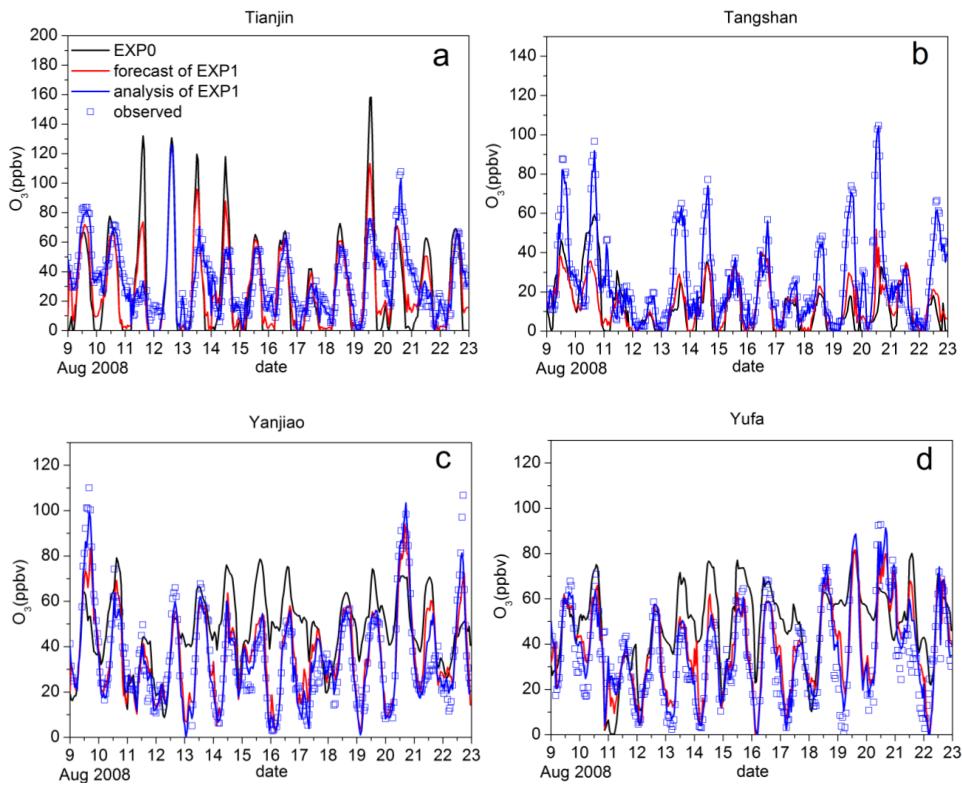
Variable adjusted in EnKF							
Experiment	O <sub>3</sub> initial conditions	VOCs initial conditions	VOCs emission rates	NO <sub>x</sub> initial conditions	NO <sub>x</sub> emission rates	Observations assimilated	Observations for forecast evaluation
EXP0						None	17 sites
EXP1	✓					9 urban sites 5 suburban sites	17 sites
EXP2		✓				the same sites as EXP1	17 sites
EXP3			✓			the same sites as EXP1	17 sites
EXP4				✓		the same sites as EXP1	17 sites
EXP5					✓	the same sites as EXP1	17 sites
EXP6	✓	✓	✓	✓	✓	the same sites as EXP1	17 sites
EXP6u	✓	✓	✓	✓	✓	11 urban sites	17 sites
EXP6s	✓	✓	✓	✓	✓	6 suburban sites	17 sites

2007; Wu et al., 2008), ozone initial conditions are adjusted separately in EXP1 through assimilating ozone observations. A comparison of the observed daytime ozone on 14 August 2008 against the 1-h forecast daytime ozone concentrations after adjustment in EXP1 is presented in Fig. 2b, while a corresponding comparison in the reference experiment EXP0 is shown in Fig. 2a. Obvious overestimation by reference simulation is observed in the urban areas of Beijing, Tianjin and their trans-boundary suburban areas. Adjusting ozone initial conditions in EXP1 significantly reduces this overestimation especially at the urban and suburban areas of Beijing with abundant monitoring stations. Additionally, at the downwind areas of Beijing without monitoring stations, the forecasted ozone concentrations after assimilation also show significant decrease. The present results indicate that adjusting ozone initial conditions with the current regional monitoring network can bring significant improvement of ozone forecast over Beijing and its surrounding areas. The results demonstrate that ozone concentrations over Beijing and its surrounding areas are regional concern. Thus, improving ozone forecast at a regional scale is crucial in improving ozone forecast over Beijing and its surrounding areas.

In order to identify the effects of adjusting ozone initial conditions at individual sites, the RMSEs of 1-h ozone forecast at each site in EXP1 is compared with that in EXP0, as shown in Fig. 3. At most of the stations, the RMSE in EXP1 is much lower than that in EXP0. The RMSE averaged over urban sites is reduced by 31 % after adjusting ozone initial conditions and that over suburban sites reduced by 46 %. The large decreases of RMSE highlight the importance of adjust-

**Fig. 3.** Comparison of RMSE of 1-h ozone forecast at urban sites (dots) and suburban sites (triangles) in free run against those in EXP1.

ing ozone initial values for short-term ozone forecast over Beijing and its surrounding areas. The fact that the impact of correcting ozone initial conditions at suburban sites is greater than that obtained from urban sites, might be due to the different ozone formation mechanisms and their related different roles of uncertainty sources in ozone forecast in urban and suburban areas.



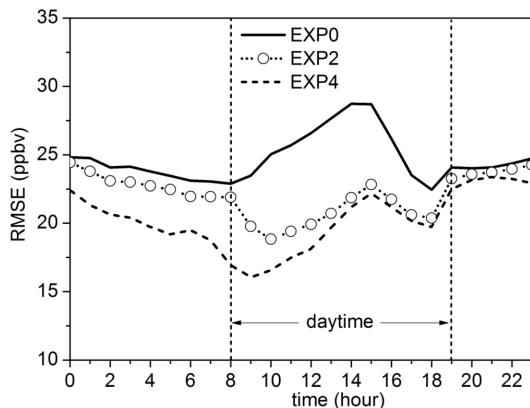
**Fig. 4.** Time series of the hourly ozone concentrations obtained from observation (rectangle), free run (black line), analysis (blue line) and 1-h forecast in EXP1 (red line) at two urban sites (Tianjin and Tangshan) and at two suburban sites (Yanjiao and Yufa).

In order to further understand this discrepancy, a comparison of the hourly ozone concentrations from the 1-h forecast in EXP1 against those from the analysis in EXP1, simulation in EXP0 and observation at two urban sites (Tianjin and Tangshan) and two suburban sites (Yanjiao and Yufa) is presented in Fig. 4. At the two suburban sites, assimilating local ozone observations in EXP1 strongly adjust the analysis and greatly improve the forecast with the analyzed values and forecasted values quite close to the observations. At the two urban sites, the simulation in EXP0 overestimates the daytime ozone at Tianjin and underestimates the daytime ozone at Tangshan during most of the time. Assimilating local observations can correct the analysis as well as showed at the two suburban sites. However, the 1-h forecast shows a large difference of performances between the two urban sites and the two suburban sites. The 1-h forecast at the two urban sites rapidly relaxes toward the reference simulation especially at Tangshan. This result suggests that ozone forecast error at urban areas can increase quickly within 1-h model integration even with good ozone initial conditions. The most plausible reasons of this phenomenon might be: (1) freshly emitted precursors at urban areas and their uncertainties in model emission inventory and initial values would affect the short term ozone forecast immediately; (2) the influence of ozone observations of urban site during analysis may limit to a rela-

tively small area due to the short life cycle of ozone at urban areas. Further attention should be paid to reducing the uncertainties of other factors (e.g., initial conditions and emissions of ozone precursors) for improving ozone forecast at urban areas. In addition, it is interesting to remark on Fig. 4 that at the two urban sites, EnKF show a deficiency in correcting ozone initial values during the nighttime on several days (e.g., 11–12 August at Tianjin; 13 August at Tangshan). This is probably caused by the almost zero values of the simulated nighttime ozone which can lead to a very small background error covariance, disregarding the observation in EnKF.

#### 4.2 Adjustment of precursor initial conditions

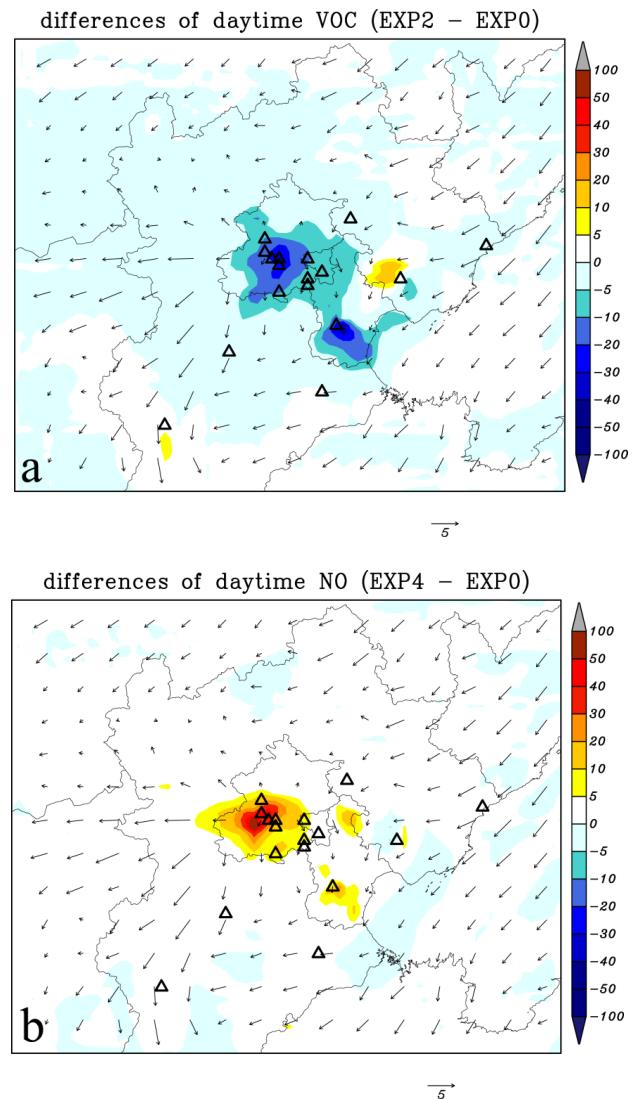
Uncertainty in ozone forecast not only originates from the uncertainty of its own initial conditions, but also from the uncertainty in the initial values of those precursors. One of the difficulties in improving ozone forecast over Beijing and its surrounding areas lies in the fact that present observations of precursors are scarce and not sufficient for constraining the uncertainty of the precursor initial values. Thus, we explore the possibility to improve ozone forecast with assimilating ozone observations to adjust the precursor initial values. EXP2 and EXP4 are designed to adjust VOCs initial conditions and  $NO_x$  initial conditions respectively.



**Fig. 5.** Time series of hourly RMSE of 1-h ozone forecast in free run (solid line), EXP2 (dash line with circles) and EXP4 (dash line).

In Fig. 5, the hourly RMSEs of 1-h ozone forecast in EXP2 and EXP4 are compared with those in EXP0. Obviously, adjustment of VOCs or NO<sub>x</sub> initial values brings significant improvement of ozone forecast during the daytime, and the daily-averaged RMSEs of ozone forecast decrease by 10% in EXP2 and 18% in EXP4. Such similar relatively significant improvements of daytime ozone forecast from adjusting NO<sub>x</sub> initial values and VOCs initial values suggest that both NO<sub>x</sub> and VOCs initial conditions have an important role in the forecast daytime ozone levels. This might relate to the rapid photochemical reactions between ozone, NO<sub>x</sub> and VOCs during daytime. It also indicates that precursor initial conditions may serve as good control variables for ozone data assimilation. Moreover, a performance of nighttime ozone forecast from adjusting NO<sub>x</sub> initial values, better than that obtained from adjusting VOCs initial values, is observed. Such performance might relate to the important role of titration reaction between ozone and NO in nighttime ozone chemistry.

In order to investigate the response of precursor concentrations to the cross-species data assimilation, the departures of forecasted daytime VOCs in EXP2 and daytime NO in EXP4 from those in the reference experiment on 14 August 2008 are shown in Fig. 6a and 6b respectively. Obvious responses of the forecasted precursor concentrations are obtained in urban areas of Beijing, Tianjin and Tangshan where the precursor emission rates are higher and ozone observations available. A noteworthy phenomenon is that the responses of daytime VOCs concentrations in EXP 2 are almost in contrast to those of daytime NO<sub>x</sub> concentrations in EXP4. For instance, adjustment of VOCs initial values in EXP2 leads to increase of the forecasted daytime VOCs concentrations in the urban areas of Beijing, while adjustment of NO<sub>x</sub> initial values in EXP4 induces decrease of the forecasted NO concentrations. The discrepancy between these responses is probably related to the nonlinear relationship between ozone and its precursors. At the areas with a “VOCs-limited” ozone formation

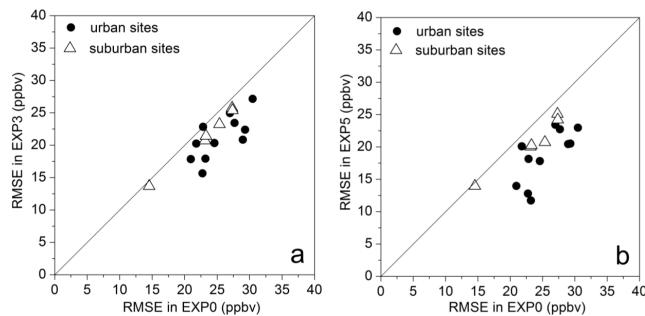


**Fig. 6.** (a) Departure of daytime VOCs concentrations (ppbv) in EXP2 from those in free run during 10 a.m.–4 p.m. on 14 August 2008 (LT); (b) Departure of daytime NO concentrations (ppbv) in EXP4 from those in free run during 10 a.m.–4 p.m. on 14 August 2008 (LT).

scheme such as urban areas of Beijing and Tianjin (Tang et al., 2010c; Shao et al., 2006), the statistical correlations in the background error covariance are positive for O<sub>3</sub>-VOCs correlation and negative for O<sub>3</sub>-NO correlation.

#### 4.3 Adjustment of precursor emission rates

Precursor emissions have been pointed out as the important uncertainty sources of ozone forecast by many previous studies (Carmichael et al., 2008; Constantinescu et al., 2007b; Hanna et al., 1998). For ozone forecast over Beijing and its surrounding areas, Tang et al. (2010b) also reported the importance of the uncertainty in precursor emissions during

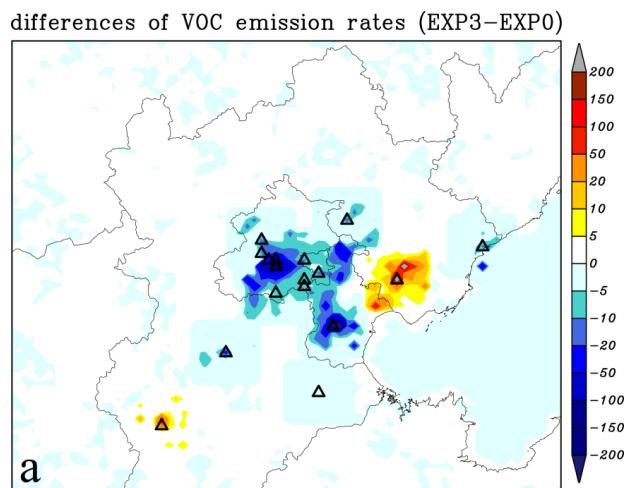


**Fig. 7.** Comparison of (a) RMSE of 1-h ozone forecast at urban sites (dots) and suburban sites (triangles) in free run against those in EXP3 and (b) against those in EXP5.

Beijing 2008 Olympic Games. In this study, we attempt to adjust precursor emissions through assimilating ozone observations in order to reduce the uncertainty in ozone forecast. EXP3 and EXP5 are conducted to adjust VOCs and NO<sub>x</sub> emissions respectively. The correction factors of emission rate are restricted in the range between 0.2 and 5 in order to exclude unreasonable correction values. The updated emission rates at each analysis step are used for the model to forecast ozone during next hour.

Figure 7a presents a comparison of the RMSE of 1-h ozone forecast at each site in EXP3 against that in the reference experiment. A similar comparison between EXP5 and the reference experiment is made in Fig. 7b. At the urban sites, the site-averaged RMSE is reduced by 16 % in EXP3 and by 27 % in EXP5. On the other hand, a relatively minor influence from adjusting precursor emissions is observed for ozone forecast at the suburban sites, and the site-averaged RMSE is reduced by 8 % in EXP3 and by 12 % in EXP5. The present results suggest that adjusting precursor emissions through assimilating ozone observations can serve as an efficient way to improve ozone forecast at urban areas over Beijing and its surrounding areas. The discrepancy between the performances of adjusting precursor emissions over urban areas and those over suburban areas might due to the different impacts of emission uncertainty on the ozone forecast over various regions. As reported by Tang et al. (2010b), precursor emissions are the most important uncertainty sources for short-term ozone forecast over urban areas beside ozone initial conditions, in contrast to the short-term ozone forecast over suburban areas where the precursor emissions show minor role.

Figure 8a displays the change of daytime VOCs emission rates after adjusting VOCs emissions in EXP3 on 14 August 2008, while the change of daytime NO<sub>x</sub> emission rates after adjusting NO<sub>x</sub> emissions in EXP5 is presented in Fig. 8b. The horizontal distribution of the daytime VOCs emission discrepancy between EXP3 and EXP0 is showed to be similar to that of the daytime VOCs initial value discrepancy between EXP2 and EXP0 (Fig. 6a). The distribution of the day-



**Fig. 8.** (a) Departure of VOCs emission rates ( $\text{ton km}^{-2} \text{yr}^{-1}$ ) in EXP3 from those in free run during 10 a.m.–4 p.m. on 14 August 2008 (LT); (b) Departure of NO<sub>x</sub> emission rates ( $\text{ton km}^{-2} \text{yr}^{-1}$ ) in EXP5 from those in free run during 10 a.m.–4 p.m. on 14 August 2008 (LT).

time NO<sub>x</sub> emission discrepancy between EXP5 and EXP0 is also quite similar to that of the daytime NO initial value discrepancy between EXP4 and EXP0 (Fig. 6b). The most obvious changes of the precursor emission rates are observed at the urban areas of Beijing, Tianjin and Tangshan. This suggests that assimilating ozone observations can effectively adjust precursor emission rates. An explanation for the similar responses of precursor initial conditions and emission rates after adjustment is that the characterization of the correlation between ozone and its precursor initial values is similar to that of the correlation between ozone and its precursor emission rates. We also examined the temporal variations of the emission correction factors during the simulation time (not shown here). The factors vary greatly over time, but a relatively strong signal with diurnal cycle can be identified.

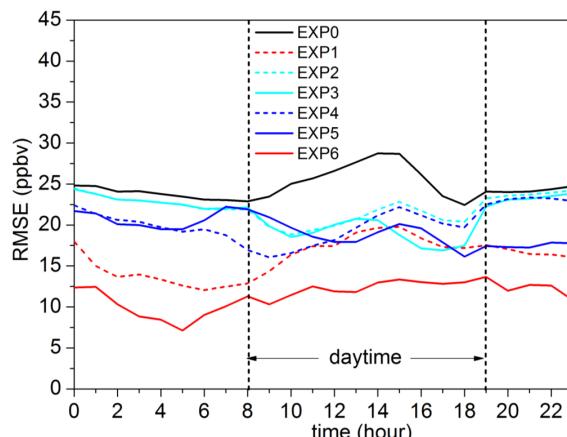
For example, adjusting NO<sub>x</sub> emissions in urban Beijing frequently result in increase of NO<sub>x</sub> emission rates during daytime and decrease of NO<sub>x</sub> emission rates during nighttime. This result indicates that it is possible to extract a systematic correction factor as a bias correction factor for ozone simulation. Nevertheless, such a revision of precursor emissions may not be interpreted as optimization of precursor emissions, since all these adjustments are based on the purpose of minimizing the error variance of the estimated ozone instead of the error variance of the adjusted factors. Further investigations are therefore necessary in order to assess the impacts of these inversed adjustments on the forecast of precursors such as NO<sub>2</sub> and VOCs species.

#### 4.4 Simultaneous adjustment of initial conditions and emissions

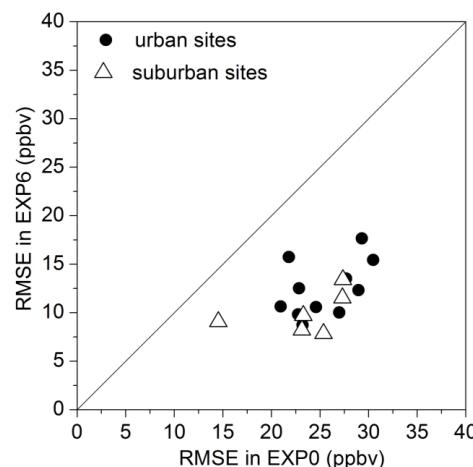
In the above data assimilation experiments, ozone initial values, VOCs initial values, NO<sub>x</sub> initial values, VOCs emission rates and NO<sub>x</sub> emission rates are adjusted separately through assimilating ozone observations. All these assimilation strategies exhibit good performances in improving ozone forecast over Beijing and its surrounding areas. However, strategies are limited. As evoked above, adjusting ozone initial values shows poor efficiency in improving ozone forecast at some urban sites, while adjusting precursor initial values or emissions shows a deficiency in improving the short-term ozone forecast at suburban areas. Therefore, we propose a comprehensive data assimilation strategy in EXP6 with simultaneous adjustment of the above five variables in order to explore the possibility of overcoming the limitations of the previous strategies.

The daily variations of the RMSEs of 1-h ozone forecast in EXP6 and in the previous six experiments are shown in Fig. 9. The adjusting precursor initial conditions demonstrates a potential for improving the short-term ozone forecast almost as great as shown by adjusting precursor emissions. The improvements of ozone forecast from separately adjusting precursor initial conditions or emissions are mainly exhibited during daytime especial for adjusting VOC initial conditions and emissions. Adjusting ozone initial conditions is efficient in reducing the RMSEs of both daytime and nighttime ozone forecast, and the averaged RMSE is 35 % lower than that of the reference experiment. The best performances of ozone forecast are obtained in EXP6 with simultaneously adjusting the five variables, and the averaged RMSE of ozone forecast is 54 % lower than that of the reference experiment. Both daytime and nighttime ozone forecasts are significantly improved.

In order to investigate the impact of the simultaneous adjustment strategy on ozone forecast at individual sites, we compare the RMSE of 1-h ozone forecast at each site in EXP6 with that in reference experiment in Fig. 10. Significant decreases of the RMSEs at all sites can be claimed after employing the simultaneous adjustment strategy. Compared



**Fig. 9.** Daily variation of the RMSEs of 1-h ozone forecast in EXP0 (free run), EXP1 (adjusting ozone initial conditions only), EXP2 (adjusting VOCs initial conditions only), EXP3 (adjusting VOCs emission rates only), EXP4 (adjusting NO<sub>x</sub> initial conditions only), EXP5 (adjusting VOCs emission rates only) and EXP6 (adjusting ozone initial conditions and precursor NO<sub>x</sub>, VOCs initial conditions and emission rates jointly).



**Fig. 10.** Comparison of RMSE of 1-h ozone forecast at urban sites (dots) and suburban sites (triangles) in free run against those in EXP6.

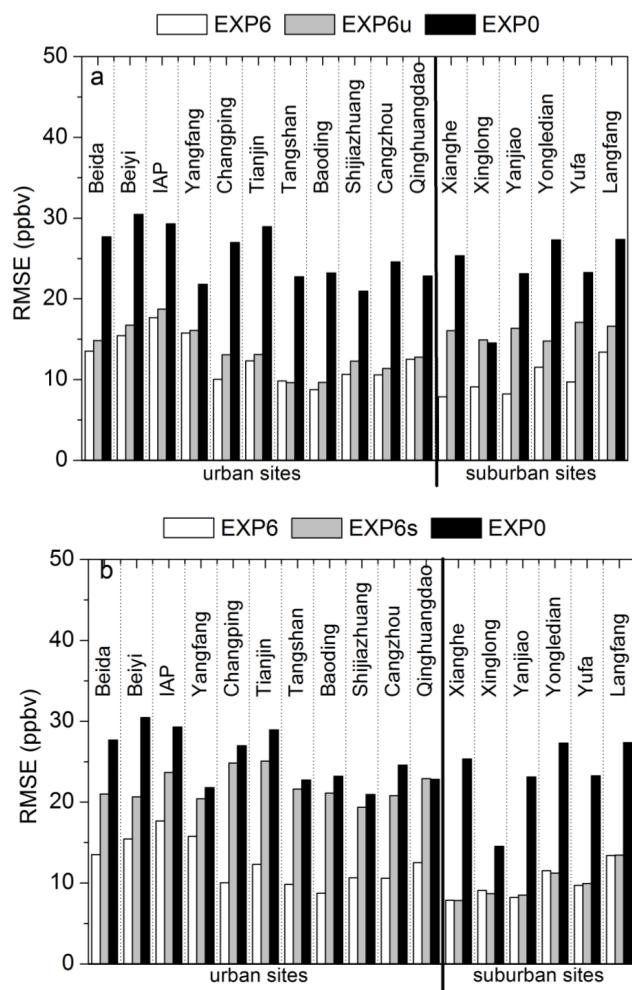
with the strategy consisting of separately adjusting ozone initial values, the simultaneous adjustment brings further improvement of ozone forecast at urban sites, and the reduction rate of RMSE of ozone forecast at urban sites is improved from 31 % in EXP1 to 51 % in EXP6. Moreover, the inefficiency of separately adjusting precursor emissions in improving ozone forecast at suburban sites disappears in the simultaneous adjustment experiment, and the site-averaged RMSE at suburban sites is 58 % lower than that of the reference experiment.

In short, the simultaneous adjustment data assimilation strategy is founded to be able to reduce the limitations of the previous strategies and produce overall better performances in improving the 1-h ozone forecast over Beijing and its surrounding areas. EnKF is a powerful tool for improving ozone forecast. It can simultaneously adjust ozone initial conditions, precursor initial conditions and emissions in an efficient way, despite the complex relationships between ozone and its precursors. The present result also implies that further improvement of ozone forecast may be achieved through adjusting more uncertainty factors simultaneously. In fact, there is still 11.6 parts per billion by volume (ppbv) of daily-averaged RMSE for 1-h ozone forecast in EXP6. The residual errors may originate from other uncertainty sources such as meteorological simulation and deposition parameters.

#### 4.5 Cross validation data assimilation experiment

In the previous six data assimilation experiments, ozone observations at two urban sites and one suburban site are withdrawn from assimilation and used as independent data for validation. In order to further investigate the effects of data assimilation on ozone forecast over the non-observation areas, we design two cross validation data assimilation experiments. The 17 monitoring stations are split into two subsets, 11 urban sites and 6 suburban sites. The experiment EXP6u assimilates ozone observations at the 11 urban sites with the simultaneous adjustment strategy of EXP6, and the remaining 6 suburban sites are withheld for validation. In the other experiment EXP6s, the 6 suburban sites are used for assimilation with the same data assimilation strategy and the 11 urban sites served as validation sites.

In Fig. 11a, the RMSE of 1-h ozone forecast at each of the 17 sites in EXP6u is compared with those in reference experiment and in EXP6. The RMSEs at both assimilation and independent sites in EXP6u are lower than those in the reference experiment except for the RMSE at Xinglong. This result suggests that the current data assimilation strategy with EnKF can improve the ozone forecast not only over the observation areas but also over the non-observation areas. On the other hand, the reduction rate of RMSE at independent site in EXP6u is lower than that at the same site in EXP6. This finding highlights the importance of assimilating local observations in improving ozone forecast over the observation area. Another interesting phenomenon in Fig. 11a is that the values of RMSEs at several assimilation sites (e.g., Beida and Changping) in EXP6u are slightly higher than those in EXP6. This might relate to the role of the 6 independent sites in reducing the uncertainty of regional-transport ozone. These sites are located at the suburban areas between three megacities (Beijing, Tianjin and Tangshan), and assimilating their ozone observations in EXP6 may not only improve ozone initial values over these areas, but also further increase the performance of ozone forecast at their downwind areas. This result implies that assimilating remote ozone observa-



**Fig. 11.** (a) A comparison of the RMSEs of 1-h ozone forecast at the 17 sites in EXP6u (with ozone observations at eleven urban sites assimilated) against those in reference experiment EXP0 and those in EXP6 (with ozone observations at all sites except for three independent sites assimilated). (b) A comparison of the RMSEs of 1-h ozone forecast at the 17 sites in EXP6s (with ozone observations at six suburban sites assimilated) against those in reference experiment and those in EXP6.

tions could be useful in reducing the uncertainty in regional transport ozone.

In Fig. 11b, the RMSE of 1-h ozone forecast at each of the 17 sites in EXP6s is compared against that in reference experiment and that in EXP6. At the six assimilation sites of EXP6s, data assimilation exhibits almost the same performances as in EXP6. At the 11 validation sites of EXP6s, the RMSEs are significantly reduced by data assimilation in EXP6s except for the RMSE at Qinghuangdao. This validates the effectiveness of data assimilation on ozone forecast at the areas without ozone observation. Additionally, the reduction rates of RMSE at Tangshan, Shijiazhuang, and Qinghuangdao are relatively lower than those at Beida,

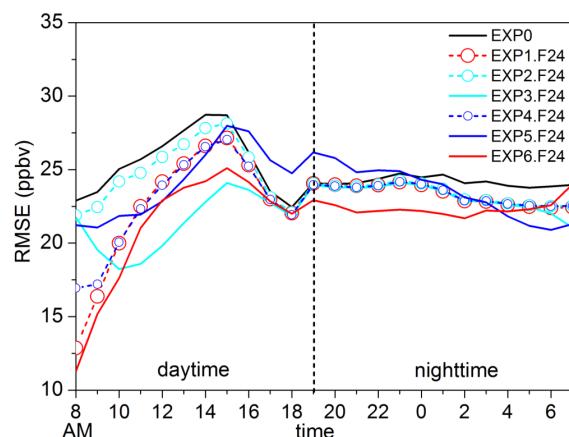
IAP and Tianjin. This deficiency of data assimilation is probably due to the long distance of these urban sites from the six assimilation stations. Based on the cross validation experiments, we can conclude that assimilating local ozone observations is determinant in ozone forecast over the observational area, although the current data assimilation strategy with EnKF is able to improve the ozone forecast over both the observational areas and the non-observational areas.

#### 4.6 Impacts of data assimilation on 24-h ozone forecast

In the previous data assimilation experiments, we evaluate the impacts of different data assimilation strategies on 1-h ozone forecast. In order to further investigate the effects of data assimilation on much longer ozone forecast cycle, we also conduct a 24-h ozone forecast experiment for each of the six data assimilation strategies. The 24-h free forecast (without data assimilation) is implemented once each day during 9–22 August. It starts every time at 8 a.m. and ends at 7 a.m. the next day. The initial conditions and the correction factors of emission are obtained from the last hour output of the assimilation window in which ozone observations are assimilated with the same setting as the 1-h forecast experiment. Here, we use EXP1.F24 to represent the 24-h forecast experiment and employ the same data assimilation strategy as the experiment EXP1. Similarly, EXP2.F24, EXP3.F24, EXP4.F24, EXP5.F24 and EXP6.F24 denote the 24-h forecast experiments with the corresponding strategies of EXP2, EXP3, EXP4, EXP5 and EXP6 respectively.

In Fig. 12, a comparison between the hourly RMSEs of ozone forecast in the six 24-h forecast experiments and those in the reference experiment EXP0 is made. Among all the 24-h forecast experiments, the simultaneous adjustment strategy brings overall the best forecast skill with the RMSEs reduced by 17 % during daytime and 8 % during nighttime. It validates the high efficiency of this simultaneous adjustment strategy in reducing the forecast error in a relatively long-range forecast cycle. Nevertheless, the forecast skill in the 24-h forecast experiment with the simultaneous adjustment strategy is much lower than that in the 1-h forecast experiment. This result suggests that an hour-by-hour data assimilation can provide a stronger constraint for model and restrict the growth of ozone forecast error.

The other five strategies have shown less obvious influence on improving 24-h ozone forecast. Adjustment of VOCs emissions is efficient in improving the daytime ozone forecast and the RMSE during daytime is lower than the reference simulation by 15 %. Especially during the high ozone period, it provides a better forecast performance than the simultaneous adjustment. Adjusting NO<sub>x</sub> initial conditions exhibits similar impact as adjusting O<sub>3</sub> initial values on 24-h ozone forecast and the daily RMSEs decrease by 7 % and 8 % respectively. The smallest influence is shown by adjusting VOCs initial values with only 3 % decrease of the daily RMSEs.



**Fig. 12.** Daily variation of the RMSEs of ozone forecast in the six 24-h forecast experiments (EXP1.F24, EXP2.F24, EXP3.F24, EXP4.F24, EXP5.F24, EXP6.F24) and those in free run EXP0.

What is noteworthy is that adjusting NO<sub>x</sub> emissions presents relatively complex influence on 24-h ozone forecast. It has significant effect in reducing the ozone forecast error during the first 8 h (covering the high ozone period). However, a negative effect on ozone forecast is induced by this strategy during the period from 3 p.m. to 11 p.m., with the RMSEs higher than those in the reference simulation. This phenomenon is probably related to the use of emission correction factors obtained from the last-hour (7am) analysis of the assimilation window for the 24-h forecast. The effectiveness of the correction factor is limited in similar error contexts as the corresponding last-hour analysis where the factor is obtained. The correction factors in the present experiment are obtained at 7 a.m. when the reference simulation generally overestimates the urban ozone concentrations. Their employment during the evening and nighttime, however, is quite possible to exacerbate further the underestimation of ozone concentrations. This phenomenon can be observed at Tianjin site obviously (seen Fig. 4).

An important message from the 24-h forecast experiments is that there are still great challenges in improving long-range ozone forecast with the emission correction factor based on a temporary error statistic. Ozone forecast errors can be induced by different sources, many of which have complex temporal variation. For instance, small initial errors in weather forecast can grow rapidly with model integration in a strongly nonlinear way, leading to significant forecast uncertainties especially in long-range ozone forecast. In this study, lack of considering temporal variation of error sources largely accounts for the relatively low efficiency of the simultaneous adjustment strategy in improving the 24-h ozone forecast compared to the 1-h ozone forecast experiment. Further investigating in the error sources and their temporal variation will be needed in future studies of ozone data assimilation for long-term forecast.

## 5 Conclusions

Aiming at improving the ozone forecast over Beijing and surrounding regions, this study explores several advanced data assimilation strategies designed to adjust ozone initial conditions, precursor initial conditions and precursor emission rates separately or jointly through assimilating ozone observations. Ensemble Kalman filter integrated with a high-resolution regional air quality model and a regional air quality monitoring network is employed.

The results suggest that EnKF is a powerful tool for improving ozone forecast over Beijing and surrounding regions. It can efficiently adjust ozone initial conditions, precursor initial conditions and emissions through assimilating ozone observations, despite the complex relationships between ozone and its precursors, and brings improvement to the short-term ozone forecast. This result implies that the inversed or cross-species adjustment through assimilating ozone observations is a possible solution for the large uncertainty in precursor initial conditions and emissions. Among all the data assimilation strategies, the simultaneous adjustment of initial conditions and emission rates is found to be the most efficient data assimilation strategy for both 1-h and 24-h ozone forecast which can reduce the limitations in the separate adjustment of initial conditions or emission rates. Assimilating local ozone observations of the regional air quality monitoring network is decisive in the performance of ozone forecast over the observation area, while assimilating remote ozone observations is also useful in reducing the uncertainty in regional transport ozone.

This study also highlights several limitations of the current data assimilation strategies that should be stated. First, there still subsist more than 10 ppbv of RMSE for 1-h ozone forecast even with ozone initial conditions, precursor initial conditions and precursor emission rates simultaneous adjustment. It implies that some other important errors such as the errors in atmospheric chemistry mechanism, deposition parameter and especially the errors in meteorological forecasting (diffusion, wind, cloudiness, humidity, temperature etc.) may have unneigligible influence on ozone forecast. Further investigation in improving ozone forecast should be therefore performed to extend the control vector of EnKF and include more sensitive variable. A particularly beneficial approach would be the coupling of meteorological and chemical data assimilation which can not only reduce the possibilities of initial or emission error compensating for meteorological error but also improve the estimation of flow-dependent background error covariance. Moreover, model errors are taken into account and the errors in precursor emission, photolysis rate and vertical diffusion coefficient are incorporated in this study. In order to better represent model error, more error sources such as the above mentioned errors and the possible errors from Gaussian assumption and the linearity of the ozone-precursor relationship in EnKF should be incorporated into the model error module. Secondly, how to

use the emission correction factor obtained from a temporary error statistic for long-range ozone forecast is still a great challenge due to the complex temporal variation of the error sources of ozone forecast. Further investigating in the error sources and their temporal variation is needed in future studies of ozone data assimilation for long-term forecast. Last but not the least, precursor initial conditions and emissions are adjusted based on the purpose to minimize the error variance of the estimated ozone instead of the error variance of the adjusted factors. Therefore, these inversed or cross-species adjustments may not lead to an improvement of precursor forecast, and further investigations are needed to assess their impacts on the forecast of precursors such as NO<sub>2</sub> and VOCs species, providing multi-species view on the current data assimilation strategies.

**Acknowledgements.** We thank the reviewers for their valuable comments contributed to significantly improving this manuscript. We thank Yuesi Wang, Tong Zhu, Min Shao and Min Hu for observational data support. This study was funded by Beijing Municipal Science & Technology Commission (No. D09040903670904) and National Natural Science Foundation (Grant No. 41075064).

Edited by: D. Parrish

## References

- An, J. L., Wang, Y. S., and Zhu, B.: Principal component and multiple regress analysis predicting ozone concentrations: case study in summer in Beijing, *Acta Scientiae Circumstantiae* (in Chinese), 30, 1286–1294, 2010.
- Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, 129, 2884–2903, 2001.
- Anderson, H. R., Ponce de Leon, A., Bland, J. M., Bower, J. S., and Strachan, D. P.: Air pollution and daily mortality in London: 1987–92, *British Medical Journal*, 312, 665–669, 1996.
- Carmichael, G. R., Sandu, A., Chai, T. F., Daescu, D. N., Constantinescu, E. M., and Tang, Y. H.: Predicting air quality: improvements through advanced methods to integrate models and measurements, *J. Comput. Phys.*, 227, 3540–3571, 2008.
- Chai, T. F., Carmichael, G. R., Sandu, A., Tang, Y. H., and Daescu, D. N.: Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements, *J. Geophys. Res.*, 111, D02301, doi:10.1029/2005JD005883, 2006.
- Chameides, W. L., Li, X. S., Tang, X. Y., Zhou, X. J., Chao, L., Kiang, C.S., John, J. St., Saylor, R. D., Liu, S. C., Lam, K. S., Wang, T., and Giorgi, F.: Is ozone pollution affecting crop yields in China?, *Geophys. Res. Lett.*, 26, 867–870, 1999.
- Chan, C. K. and Yao, X.: Air pollution in mega cities in China, *Atmos. Environ.*, 42, 1–42, 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, 2007a.
- Constantinescu, E. M., Sandu, A., Chai, T. F., and Carmichael, G. R.: Assessment of ensemble-based chemical data assimilation in an idealized setting, *Atmos. Environ.*, 41, 18–36, 2007b.

- Constantinescu, E. M., Sandu, A., Chai, T. F., and Carmichael, G. R.: Ensemble-based chemical data assimilation. I: General approach, *Q. J. Roy. Meteorol. Soc.*, 133, 1229–1243, 2007c.
- 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. Roy. Meteorol. Soc.*, 131, 3313–3322, 2005.
- 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.
- Evensen, G.: The ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- Evensen, G.: The Ensemble Kalman Filter for combined state and parameter estimation, *IEEE Control Systems Magazine*, 29, 83–104, 2009.
- Feng, Y., Zhang, W. F., Sun., D. Z., and Zhang, L. Q.: Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification, *Atmos. Environ.*, 45, 1979–1985, 2011.
- Flemming, J., Reimer, E., and Stern, R.: Long term evaluation of the ozone forecast by an Eulerian model, *Phys. Chem., Earth*, 26, 775–779, 2001.
- 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. J. M., 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.
- Hao, J. M., Wang, L. T., Shen, M. J., Li, L., and Hu, J. N.: Air quality impacts of power plant emissions in Beijing, *Environ. Pollut.*, 147, 401–408, 2007.
- Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, 129, 123–137, 2001.
- Houyoux, M. R., Vukovich, J. M., Coats, C. J., Wheeler, N. J. M., and Kasibhatla, P. S.: Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SM-RAQ) project, *J. Geophys. Res.*, 105, 9079–9090, 2000.
- Kalman, R. E.: A new approach to linear filtering and prediction problems, *Transactions of the ASME – Journal of Basic Engineering*, 82, 35–45, 1960.
- Keppenne, C. L.: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter, *Mon. Weather Rev.*, 128, 1971–1981, 2000.
- Li, J., Wang, Z. F., Akimoto, H., Tang, J., and Uno, I.: Modeling of the impacts of China's anthropogenic pollutants on the surface ozone summer maximum on the northern Tibetan Plateau, *Geophys. Res. Lett.*, 36, L24802, doi:10.1029/2009GL041123, 2009.
- Li, J., Wang, Z., Wang, X., Yamaji, K., Takigawa, M., Kanaya, Y., Pochanart, P., Liu, Y., Irie, H., Hu, B., Tanimoto, H., and Akimoto, H.: Impacts of aerosols on summertime tropospheric photolysis frequencies and photochemistry over Central Eastern China, *Atmos. Environ.*, doi:10.1016/j.atmosenv.2011.01.016, in press, 2011.
- Mallet, V.: Ensemble forecast of analyses: Coupling data assimilation and sequential aggregation, *J. Geophys. Res.*, 115, D24303, doi:10.1029/2010JD014259, 2010.
- Mallet, V., Stoltz, G., and Mauricette, B.: Ozone ensemble forecast with machine learning algorithms, *J. Geophys. Res.*, 114, D05307, doi:10.1029/2008JD009978, 2009.
- Mitchell, H. L. and Houtekamer, P. L.: An adaptive ensemble Kalman filter, *Mon. Weather Rev.*, 128, 416–433, 2000.
- Sakov, P. and Oke, P. R.: Implications of the form of the ensemble transformation in the ensemble square root filters, *Mon. Weather Rev.*, 136, 1042–1053, 2008.
- Segers, A.: Data assimilation in atmospheric chemistry models using Kalman filtering, Ph.D. thesis, Delft Univ., Delft, The Netherlands, 2002.
- Shao, M., Tang, X. Y., Zhang, Y. H., and Li, W. J.: City clusters in China: air and surface water pollution, *Front. Ecol. Environ.*, 4, 353–361, 2006.
- Streets, D. G., Fu, J. H. S., Jang, C. J., Hao, J. M., He, K. B., Tang, X. Y., Zhang, Y. H., Wang, Z. F., Li, Z. P., Zhang, Q., Wang, L. T., Wang, B. Y., and Yu, C.: Air quality during the 2008 Beijing Olympic Games, *Atmos. Environ.*, 41, 480–492, 2007.
- Sudo, K. M., Takahashi, J., Kurokawa, and Akimoto, H.: CHASER: A global chemical model of the troposphere 1. Model description, *J. Geophys. Res.*, 107, 4339, doi:10.1029/2001JD001113, 2002.
- Tang, X., Wang, Z. F., Zhu, J., Wu, Q. Z., and Gbaguidi, A.: Ensemble-based surface ozone forecast over Beijing, *Climate and Environmental Research* (in Chinese), 15, 677–684, 2010a.
- Tang, X., Wang, Z. F., Zhu, J., Wu, Q. Z., and Gbaguidi, A.: Preliminary application of Monte Carlo uncertainty analysis in ozone simulation, *Climate and Environmental Research* (in Chinese), 15, 541–550, 2010b.
- 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, 2010c.
- van Loon, M., Builtjes, P. J. H., and Segers, A. J.: Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, *Environ. Model. Softw.*, 15, 603–609, 2000.
- van Loon, R., Vautard, R., Schaap, M., Bergstrom, R., Bessagnet, B., Brandt, J., Builtjes, P. J. H., Christensen, J. H., Cuvelier, C., Graff, A., Jonson, J. E., Kroli, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., White, L., and Wind, P.: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, *Atmos. Environ.*, 41, 2083–2097, 2007.
- Wang, M., Zhu, T., Zhang, J. P., Zhang, Q. H., Lin, W. W., Li, Y., and Wang, Z. F.: Using a mobile laboratory to characterize the distribution and transport of sulfur dioxide in and around Beijing, *Atmos. Chem. Phys.*, 11, 11631–11645, doi:10.5194/acp-11-11631-2011, 2011.
- Wang, Y., Hao, J., McElroy, M. B., Munger, J. W., Ma, H., Chen, D., and Nielsen, C. P.: Ozone air quality during the 2008 Beijing Olympics: effectiveness of emission restrictions, *Atmos. Chem.*

- Phys., 9, 5237–5251, doi:10.5194/acp-9-5237-2009, 2009.
- Wang, Z., Akimoto, H., and Uno, I.: Neutralization of soil aerosol and its impact on the distribution of acid rain over East Asia: Observations and model results, *J. Geophys. Res.*, 107, doi:10.1029/2001JD001040, 2002.
- Wang, Z. F., Huang, M. Y., He, D., Xu, H. Y., and Zhou, L.: Studies on transport of acid substance in China and East Asia part I: 3-D Eulerian transport model for pollutants, *Chin. J. Atmos. Sci.*, 21, 367–375, 1997.
- 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 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, 1989.
- 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.
- Wu, Q. Z., Wang, Z. F., Gbaguidi, A., Gao, C., Li, L. N., and Wang, W.: A numerical study of contributions to air pollution in Beijing during CAREBeijing-2006, *Atmos. Chem. Phys.*, 11, 5997–6011, doi:10.5194/acp-11-5997-2011, 2011.
- 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.* (in Chinese), 55, 1937–1944, 2010.
- 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., Huo, H., Kanbari, A., Klimont, Z., Park, I., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L., and Yao, Z.: Asian emissions in 2006 for the NASA INTEX-B mission, *Atmos. Chem. Phys.*, 9, 5131–5153, doi:10.5194/acp-9-5131-2009, 2009.
- Zhang, W., Wang, Z. F., An, J. L., Yang, T., and Tang, X.: Update the ensemble air quality modeling system with BP model during Beijing Olympics (in Chinese), *Clim. Environ. Res.*, 15, 595–601, 2010.
- Zou, Y. F., Wu, Q., Wang, Z. F., and Zhu, J.: Update of industrial pollution source in Hebei and its effects on regional air quality modeling during 2008 Beijing Olympics Games (in Chinese), *Clim. Environ. Res.*, 15, 624–635, 2010.