



Background error covariance with balance constraints for aerosol species and applications in variational data assimilation

Zengliang Zang¹, Zilong Hao¹, Yi Li¹, Xiaobin Pan¹, Wei You¹, Zhjin Li², and Dan Chen³

¹College of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China

²Joint Institute For Regional Earth System Science and Engineering, University of California, Los Angeles, California 90095, USA

³National Center for Atmospheric Research, Boulder, Colorado 80305, USA

Correspondence to: Zengliang Zang (zzlqxy@163.com)

Received: 18 September 2015 – Published in Geosci. Model Dev. Discuss.: 16 November 2015

Revised: 9 June 2016 – Accepted: 11 July 2016 – Published: 10 August 2016

Abstract. Balance constraints are important for background error covariance (BEC) in data assimilation to spread information between different variables and produce balance analysis fields. Using statistical regression, we develop a balance constraint for the BEC of aerosol variables and apply it to a three-dimensional variational data assimilation system in the WRF/Chem model; 1-month forecasts from the WRF/Chem model are employed for BEC statistics. The cross-correlations between the different species are generally high. The largest correlation occurs between elemental carbon and organic carbon with as large as 0.9. After using the balance constraints, the correlations between the unbalanced variables reduce to less than 0.2. A set of data assimilation and forecasting experiments is performed. In these experiments, surface PM_{2.5} concentrations and speciated concentrations along aircraft flight tracks are assimilated. The analysis increments with the balance constraints show spatial distributions more complex than those without the balance constraints, which is a consequence of the spreading of observation information across variables due to the balance constraints. The forecast skills with the balance constraints show substantial and durable improvements from the 2nd hour to the 16th hour compared with the forecast skills without the balance constraints. The results suggest that the developed balance constraints are important for the aerosol assimilation and forecasting.

1 Introduction

Aerosol data assimilation in chemical transport models has received an increasing amount of attention in recent years as a basic methodology for improving aerosol analysis and forecasting. In a data assimilation system, the background error covariance (BEC) plays a crucial role in the success of an assimilation process. The BEC and the observation error determine analysis increments from the assimilation process (Derber and Bouttier, 1999; Chen et al., 2013).

However, accurate estimation of the BEC remains difficult due to a lack of information about the true atmospheric states and also due to computational requirement arising from the large dimension of the BEC (typically $10^7 \times 10^7$). For a variational data assimilation system, a few methods have been developed to estimate and simplify the expression of the BEC, such as the analysis of innovations, the NMC (National Meteorological Center) and the ensemble-based (Monte Carlo) methods (Constantinescu et al., 2007; Singh et al., 2011). The NMC method is extensively used in operational atmospheric and meteorology–chemistry data assimilation systems. It assumes that the forecast errors are approximated by differences between pairs of forecasts that are valid at the same time (Parrish and Derber, 1992). Pagowski et al. (2010) estimated the BEC of PM_{2.5} (particles having an aerodynamic diameter less than 2.5 μm) by calculating the differences between the forecasts of 24 and 48 h, and used the estimated BEC in a Grid-point Statistical Interpolation (GSI) system (Wu et al., 2002). Benedetti and Fisher (2007) estimated the BEC of the sum of the mixing ratios of all aerosol

species for an operational analysis and forecast systems at ECMWF (The European Centre for Medium-Range Weather Forecasts). The BEC with multiple species and size bins of aerosols have been calculated and employed in data assimilation. Liu et al. (2011) estimated the BEC with 14 aerosol species in the Goddard Chemistry Aerosol Radiation and Transport scheme of the Weather Research and Forecasting/Chemistry (WRF/Chem) model and applied it to the GSI system. Schwartz et al. (2012) increased the number of the species to 15 based on the study of Liu et al. (2011). Li et al. (2013) estimated the BEC for five species derived from the Model for Simulation Aerosol Interactions and Chemistry (MOSAIC) scheme.

One important role that the BEC plays in meteorological data assimilation is to spread information between different variables to produce balanced analysis fields, which employ balance constraints to convert original variables into new independent variables. Balance constraints have been employed in atmospheric and oceanic data assimilation, such as geostrophic balance or temperature–salinity balance (Bannister, 2008a, b). To incorporate balance constraints, the model variables are usually transformed to balanced and unbalanced parts. The unbalanced parts as control variables are can be assumed independent, and the balanced parts are constrained by balance constraints (Derber and Bouttier, 1999). Instead of using an empirical function as a balance constraint, balance constraints are also derived using regression techniques (Ricci and Weaver, 2005). Although distinct empirical relations between some variables (such as temperature and humidity) may not exist, the regression equation can also be estimated as balance constraints (Chen et al., 2013).

In current aerosol variational data assimilation with multiple variables, balance constraints are not yet incorporated in the BEC. The state variables are assumed to be independent variables without cross-correlation. However, the aerosol species are frequently highly correlated due to their common emission sources and diffusion processes. For example, the correlations in terms of the R^2 between elemental carbon and black carbon exceed 0.6 in many locations across Asia and the South Pacific in both urban and suburban locations (Salako et al., 2012), and the correlations between different size bins, such as $PM_{2.5}$ and $PM_{10-2.5}$ (the diameter of particles being between 2.5 and 10 μm), are also generally significant (Sun et al., 2003; Geller et al., 2004). Thus, the cross-correlations between different species or size bins are necessary to produce balanced analysis fields. Cross-correlations spread the observation information from one variable to other variables, which can produce more balanced initial fields. For the data assimilation of the ensemble Kalman filter method, the BEC with balance constraints is assured (Pagowski and Grell, 2012; Schwartz et al., 2014), although the balance may break down because of localization.

Recently, several studies have suggested that the BEC with balanced cross-correlation should be introduced into

aerosol variational data assimilation (Kahnert, 2008; Liu et al., 2011; Li et al., 2013; Saide et al., 2013). Kahnert (2008) exhibited cross-correlations of the 17 aerosol variables from the Multiple-scale Atmospheric Transport and Chemistry (MATCH) model. He found that the statistical cross-correlations between aerosol components are primarily influenced by the interrelations between emissions and by interrelations due to chemical reactions to a much lesser degree. Saide et al. (2012, 2013) incorporated the capacity to add cross-correlations between aerosol size bins in GSI for assimilating observations of aerosol optical depth (AOD) data. The cross-correlations between the two connecting size bins for each species were considered using recursive filters, whereas the cross-correlation is not considered for the other size bins that are not connecting.

In this paper, we explore incorporating cross-correlations between different species in BEC using balance constraints. The balance constraints are established using statistical regression. We apply the BEC with the balance constraints to a data assimilation and forecasting system with the MOSAIC scheme in WRF/Chem. The MOSAIC scheme includes a large number of variables with eight species, and flexibility of eight or four size bins. The scheme of four size bins is used in our studies. The four bins are located between 0.039–0.1, 0.1–1.0, 1.0–2.5, and 2.5–10 μm , and the total mass of the first three bins are $PM_{2.5}$. A 3DVAR system for the MOSAIC (4-bin) scheme has been developed by Li et al. (2013). For comparisons, we employ this 3DVAR system with the same model configurations as employed by Li et al. (2013). The data assimilation and forecasting experiments are performed with a focus on assessing the impact of cross-correlations of the BEC on analyses and forecasts.

The paper is organized as follows: Sect. 2 describes the 3DVAR system and the formulation of the BEC. Section 3 describes the WRF/Chem configuration and estimates the correlations among the emissions. The statistical characteristics of the BEC, including the regression coefficient of the cross-correlation, are discussed in Sect. 4. Using the BEC, experiments of assimilating surface $PM_{2.5}$ observations and aircraft observations are discussed in Sect. 5. Shortcomings, conclusions and future perspectives are presented in Sect. 6.

2 Data assimilation system and BEC

In this section, we present a formulation of the BEC with cross-correlation between different species using a regression technique. Then, the cost function with the new BEC is derived and the calculating factorization of the BEC is described.

The control variables of the data assimilation are obtained from the MOSAIC (4-bin) aerosol scheme in the WRF/Chem model (Zaveri et al., 2008). The MOSAIC scheme includes eight aerosol species, that is, elemental carbon or black carbon (EC / BC), organic carbon (OC), nitrate (NO_3^-), sulfate

(SO_4), chloride (Cl), sodium (Na), ammonium (NH_4), and other inorganic mass (OIN). Each species is separated into four bins with different sizes: 0.039–0.1, 0.1–1.0, 1.0–2.5 and 2.5–10 μm . The scheme involves 32 aerosol variables with eight species and four size bins. These variables cannot be directly introduced as control variables in an assimilation system with regard to computational efficiency. The number of variables must be decreased prior to assimilation. Li et al. (2013) lumped these variables into five species as control variables in the 3DVAR system. The five species consist of EC, OC, NO_3 , SO_4 and OTR (other species). Here, OTR is the sum of Cl, Na, NH_4 and OIN. Note that the data assimilation system aims to assimilate the observation of $\text{PM}_{2.5}$; only the first three of four size bins are utilized to lump as one control variable for each species.

For a 3DVAR system, the cost function (J), which measures the distance of the state vector to the background and observations, can be written as follows:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y} - \mathbf{Hx})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{Hx}). \quad (1)$$

Here, \mathbf{x} is the vector of the state variables, including EC, OC, NO_3 , SO_4 and OTR; \mathbf{x}^b is the background vector of these five species, which are generated by the MOSAIC scheme; \mathbf{y} is the observation vector; \mathbf{H} is the observation operator that maps the model space to the observation space and is assumed to be linear here; \mathbf{R} is the observation error covariance associated with \mathbf{y} ; and \mathbf{B} is the background error covariance associated with \mathbf{x}^b . Equation (1) is usually written in the incremental form

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{H}\delta\mathbf{x} - \mathbf{d}), \quad (2)$$

where $\delta\mathbf{x}(\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b)$ is the incremental state variable. The observation innovation vector is known as $\mathbf{d} = \mathbf{y} - \mathbf{Hx}^b$. The minimization solution is the analysis increment $\delta\mathbf{x}$, and the final analysis is $\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}$. This analysis is statistically optimal as a minimum error variance estimate (e.g., Jazwinski, 1970; Cohn, 1997).

In Eq. (1) or Eq. (2), \mathbf{x}^b is a $(N \times m)$ – vector, where N is the number of model grid points, and m is the number of state variables. \mathbf{B} is a symmetric matrix with a dimension of $(N \times m)^2$. For a high-resolution model, the number of vector \mathbf{x}^b is on the order of 10^7 . Therefore, the number of elements in \mathbf{B} is approximately 10^{14} . With this dimension, \mathbf{B} cannot be explicitly manipulated. To pursue simplifications of \mathbf{B} , we employ the following factorization:

$$\mathbf{B} = \mathbf{DCD}^T, \quad (3)$$

where \mathbf{D} and \mathbf{C} are the standard deviation matrix and the correlation matrix, respectively. \mathbf{D} and \mathbf{C} can be described and separately prescribed after the factorization. \mathbf{D} is a diagonal

matrix, whose elements include the standard deviation of all state variables in the three-dimensional grids. To reduce the computational cost, we use the average value of standard deviations that are at the same level. Thus, the standard deviation is simplified with vertical levels. \mathbf{C} is a symmetric matrix, having the form

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{\text{EC}} & \mathbf{C}_{\text{EC}}^{\text{OC}} & \mathbf{C}_{\text{EC}}^{\text{NO}_3} & \mathbf{C}_{\text{EC}}^{\text{SO}_4} & \mathbf{C}_{\text{EC}}^{\text{OTR}} \\ \mathbf{C}_{\text{OC}}^{\text{EC}} & \mathbf{C}_{\text{OC}} & \mathbf{C}_{\text{OC}}^{\text{NO}_3} & \mathbf{C}_{\text{OC}}^{\text{SO}_4} & \mathbf{C}_{\text{OC}}^{\text{OTR}} \\ \mathbf{C}_{\text{NO}_3}^{\text{EC}} & \mathbf{C}_{\text{NO}_3}^{\text{OC}} & \mathbf{C}_{\text{NO}_3} & \mathbf{C}_{\text{NO}_3}^{\text{SO}_4} & \mathbf{C}_{\text{NO}_3}^{\text{OTR}} \\ \mathbf{C}_{\text{SO}_4}^{\text{EC}} & \mathbf{C}_{\text{SO}_4}^{\text{OC}} & \mathbf{C}_{\text{SO}_4}^{\text{NO}_3} & \mathbf{C}_{\text{SO}_4} & \mathbf{C}_{\text{SO}_4}^{\text{OTR}} \\ \mathbf{C}_{\text{OTR}}^{\text{EC}} & \mathbf{C}_{\text{OTR}}^{\text{OC}} & \mathbf{C}_{\text{OTR}}^{\text{NO}_3} & \mathbf{C}_{\text{OTR}}^{\text{SO}_4} & \mathbf{C}_{\text{OTR}} \end{bmatrix}, \quad (4)$$

where \mathbf{C}_{EC} , \mathbf{C}_{OC} , \mathbf{C}_{NO_3} , \mathbf{C}_{SO_4} and \mathbf{C}_{OTR} at diagonal locations are the background error auto-correlation matrices that are associated with each species. They represent the correlation among pairs of grid points for one species. Other sub-matrices represent the correlations between different species, known as cross-correlations. For example, $\mathbf{C}_{\text{OC}}^{\text{EC}}$ represents the cross-correlations between EC and OC, and $\mathbf{C}_{\text{OC}}^{\text{EC}} = (\mathbf{C}_{\text{EC}}^{\text{OC}})^T$. In Li et al. (2013), these cross-correlations were disregarded; that is, the five species are considered independently and \mathbf{C} is thus a block diagonal matrix.

In this study, the cross-correlations between different species are considered by introducing control variable transforms (Derber and Bouttier, 1999; Barker et al., 2004; Huang et al., 2009). We divide the model aerosol variables into balanced components ($\delta\mathbf{x}_b$) and unbalanced components ($\delta\mathbf{x}_u$):

$$\delta\mathbf{x} = \delta\mathbf{x}_b + \delta\mathbf{x}_u. \quad (5)$$

Note the EC does not need to be divided. There is not unbalanced component for EC that is similar to the variable of vorticity in the data assimilation of ECMWF (Derber and Bouttier, 1999), or the variable of stream function in the data assimilation of MM5 (Barker et al., 2004). The transformation from unbalanced variables ($\delta\mathbf{x}_u$) to full variables ($\delta\mathbf{x}$) by the balance operator \mathbf{K} is given by

$$\delta\mathbf{x} = \mathbf{K}\delta\mathbf{x}_u. \quad (6)$$

Eq. (6) can be written as

$$\begin{bmatrix} \delta\text{EC} \\ \delta\text{OC} \\ \delta\text{NO}_3 \\ \delta\text{SO}_4 \\ \delta\text{OTR} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 & 0 \\ \rho_{21} & \mathbf{I} & 0 & 0 & 0 \\ \rho_{31} & \rho_{32} & \mathbf{I} & 0 & 0 \\ \rho_{41} & \rho_{42} & \rho_{43} & \mathbf{I} & 0 \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\text{EC}_u \\ \delta\text{OC}_u \\ \delta\text{NO}_3_u \\ \delta\text{SO}_4_u \\ \delta\text{OTR}_u \end{bmatrix}, \quad (7)$$

where ρ_{ij} is the submatrix of \mathbf{K} , which represents the statistical regression coefficients between the variables i and j (Chen et al., 2013). Note that ρ_{ij} is a diagonal matrix with the dimension of model grid points. Each model grid point has a regression coefficient. For convenience, we assumed that the elements of ρ_{ij} is a constant value for all grid points, which are denoted as ρ_{ij} and are calculated by linear regression with all grid points. For example, ρ_{21} can be obtained

from the regression equation of OC and EC as

$$\delta\text{OC} = \rho_{21}\delta\text{EC} + \boldsymbol{\epsilon}, \quad (8)$$

where $\boldsymbol{\epsilon}$ is the residual. δEC and δOC can be estimated from the forecast differences of 24 h forecasts and 48 h forecasts, similar to the statistics of the BEC. Equation (8) contains the slope but no intercept. The intercept is nearly zero because δEC and δOC represent forecast differences that can be considered to be zero mean values. After obtaining ρ_{21} , the balanced part (e.g., the value of the regression prediction) of δOC can be obtained by

$$\delta\text{OC}_b = \hat{\delta}\text{OC} = \rho_{21}\delta\text{EC}. \quad (9)$$

Where $\hat{\delta}\text{OC}$ represents the predicted value of Eq. (8), which is equal to the balanced part (δOC_b). Remove the δOC_b from the full variables to obtain the unbalanced part (δOC_u), that is, $\boldsymbol{\epsilon}$ in Eq. (8). Thus, the calculation of δOC_u can be written as

$$\delta\text{OC}_u = \delta\text{OC} - \rho_{21}\delta\text{EC}. \quad (10)$$

Here, δOC_u and δEC are employed as predictors in the next regression equation to obtain δNO_{3b} . Then, we can obtain the unbalanced parts of the remaining variables, which are defined as follows:

$$\delta\text{NO}_{3u} = \delta\text{NO}_3 - (\rho_{31}\delta\text{EC} + \rho_{32}\delta\text{OC}_u), \quad (11)$$

$$\delta\text{SO}_{4u} = \delta\text{SO}_4 - (\rho_{41}\delta\text{EC} + \rho_{42}\delta\text{OC}_u + \rho_{43}\delta\text{NO}_{3u}), \quad (12)$$

$$\begin{aligned} \delta\text{OTR}_u = \delta\text{OTR} - (\rho_{51}\delta\text{EC} + \rho_{52}\delta\text{OC}_u \\ + \rho_{53}\delta\text{NO}_{3u} + \rho_{54}\delta\text{SO}_{4u}). \end{aligned} \quad (13)$$

The coefficient of determination (R^2) can be employed to measure the fit of these regressions. It can be expressed as

$$R^2 = \frac{\text{SSR}}{\text{SST}}, \quad (14)$$

where SSR and SST are the regression sum of squares and the sum of squares for total, respectively.

These unbalanced parts can be considered to be independent because they are residual and random. \mathbf{B}_u denotes the unbalanced variables of the BEC and can be factorized as

$$\mathbf{B}_u = \mathbf{D}_u \mathbf{C}_u \mathbf{D}_u^T, \quad (15)$$

where \mathbf{D}_u and \mathbf{C}_u are the standard deviation matrix and the correlation matrix, respectively. \mathbf{C}_u should be a block diagonal without cross-correlations as follows:

$$\mathbf{C}_u = \begin{bmatrix} \mathbf{C}_{\text{EC}} & & & & \\ & \mathbf{C}_{\text{OC}_u} & & & \\ & & \mathbf{C}_{\text{NO}_{3u}} & & \\ & & & \mathbf{C}_{\text{SO}_{4u}} & \\ & & & & \mathbf{C}_{\text{OTR}_u} \end{bmatrix}. \quad (16)$$

According the definition of the BEC,

$$\mathbf{B} = \langle (\delta\mathbf{x})(\delta\mathbf{x}^T) \rangle, \quad (17)$$

and \mathbf{B}_u can be written as

$$\mathbf{B}_u = \langle (\delta\mathbf{x}_u)(\delta\mathbf{x}_u^T) \rangle. \quad (18)$$

Using Eqs. (6), (17) and (18), the relationship between \mathbf{B} and \mathbf{B}_u is

$$\mathbf{B} = \mathbf{K}\mathbf{B}_u\mathbf{K}^T. \quad (19)$$

$\mathbf{B}^{\frac{1}{2}}$ and $\mathbf{B}_u^{\frac{1}{2}}$ are defined as the square root of \mathbf{B} and the square root of \mathbf{B}_u , respectively. Their transformation is

$$\mathbf{B}^{\frac{1}{2}} = \mathbf{K}\mathbf{B}_u^{\frac{1}{2}}. \quad (20)$$

Using Eq. (15), Eq. (20) can be written as follows:

$$\mathbf{B}^{\frac{1}{2}} = \mathbf{K}\mathbf{D}_u\mathbf{C}_u^{\frac{1}{2}}. \quad (21)$$

Generally, a transformed cost function of Eq. (2) is expressed as a function of a preconditioned state variable:

$$\begin{aligned} J(\delta\mathbf{z}) = & \frac{1}{2}\delta\mathbf{z}^T\delta\mathbf{z} \\ & + \frac{1}{2}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\delta\mathbf{z} - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\delta\mathbf{z} - \mathbf{d}). \end{aligned} \quad (22)$$

Here, $\delta\mathbf{z} = \mathbf{B}^{-\frac{1}{2}}\delta\mathbf{x}$. Using Eq. (21), Eq. (22) can be written as

$$\begin{aligned} J(\delta\mathbf{z}) = & \frac{1}{2}\delta\mathbf{z}^T\delta\mathbf{z} + \frac{1}{2}(\mathbf{H}\mathbf{K}\mathbf{D}_u\mathbf{C}_u^{\frac{1}{2}}\delta\mathbf{z} - \mathbf{d})^T \\ & \mathbf{R}^{-1}(\mathbf{H}\mathbf{K}\mathbf{D}_u\mathbf{C}_u^{\frac{1}{2}}\delta\mathbf{z} - \mathbf{d}). \end{aligned} \quad (23)$$

Equation (23) is the last form of the cost function with the cross-correlation of \mathbf{B} .

According to Li et al. (2013), the correlation matrix of the unbalanced parts (\mathbf{C}_u) is factorized as

$$\mathbf{C}_u = \mathbf{C}_{ux} \otimes \mathbf{C}_{uy} \otimes \mathbf{C}_{uz}. \quad (24)$$

Here, \otimes denotes the Kronecker product, and \mathbf{C}_{ux} , \mathbf{C}_{uy} and \mathbf{C}_{uz} represent the correlation matrices between grid points in the x direction, the y direction and the z direction, respectively, with the sizes $n_x \times n_x$, $n_y \times n_y$ and $n_z \times n_z$, respectively. Here, n_x , n_y and n_z represent the numbers of grid points in the x direction, y direction and z direction, respectively. This factorization can decrease the size of the dimension of \mathbf{C}_u . Another desirable property of Eq. (24) is

$$\mathbf{C}_u^{\frac{1}{2}} = \mathbf{C}_{ux}^{\frac{1}{2}} \otimes \mathbf{C}_{uy}^{\frac{1}{2}} \otimes \mathbf{C}_{uz}^{\frac{1}{2}}. \quad (25)$$

\mathbf{C}_{ux} and \mathbf{C}_{uy} are expressed by Gaussian functions, and \mathbf{C}_{uz} is directly computed from the proxy data. They will be discussed in Sect. 4.2.

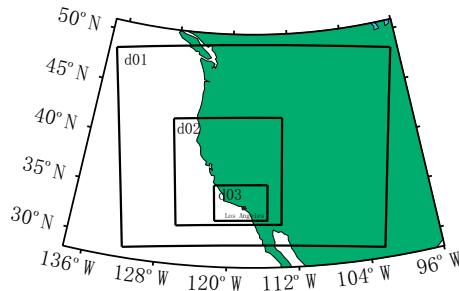


Figure 1. Geographical display of the three-nested model domains. The innermost domain covers the Los Angeles basin; the black point denotes the location of Los Angeles.

3 WRF/Chem configuration and cross-correlations of emission species

In this section, we describe the configuration of WRF/Chem, whose forecasting products will be employed in the following BEC statistics and data assimilation experiments. In addition, the cross-correlations of emission species from the WRF/Chem emission data are investigated to understand the cross-correlation between different species of the BEC.

3.1 WRF/Chem configuration

WRF/Chem (V3.5.1) is employed in our study. This is a fully coupled online model with a regional meteorological model that is coupled to aerosol and chemistry models (Grell et al., 2005). The model domain with three spatial domains is shown in Fig. 1. The horizontal grid spacing for these three domains are 36 km (80×60 points), 12 km (97×97 points) and 4 km (144×96 points), respectively. The outer domain spans southern California and the innermost domain encompasses Los Angeles. All domains have 31 vertical levels with the top at 50 hPa. The vertical grid is stretched to place the highest resolution in the lower troposphere. The discussion of the BEC and the emissions presented in this paper will be confined to the innermost domain. The initial meteorology conditions for WRF/Chem are prepared using the North American Regional Reanalysis (NARR) (Mesinger et al., 2006). The meteorology boundary conditions and sea surface temperatures are updated at each initialization. For the forecast running, the initial meteorological conditions are obtained from the NARR data. The initial aerosol conditions are obtained from the former forecast. The emissions are derived from the National Emission Inventory 2005 (NEI'05) for both aerosols and trace gases (Guenther et al., 2006). For more details, the reader is referred to Li et al. (2013).

3.2 Cross-correlations of emission species

The emission source is necessary for running the WRF/Chem model. It is an important factor for the distribution of the aerosol forecasts. The analysis of the correlations among the

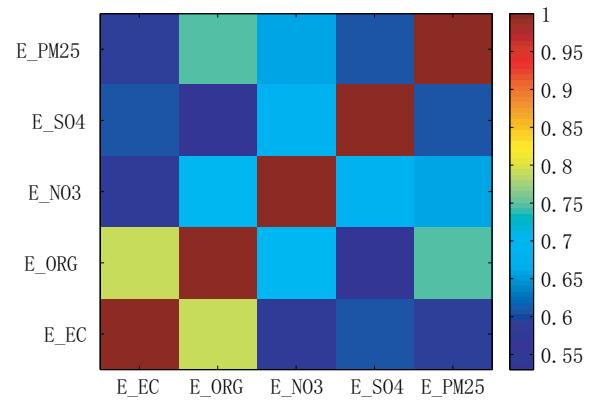


Figure 2. Cross-correlations between emission species of E_EC, E_ORG, E_NO3, E_SO4 and E_PM25. The emission species data are derived from the NEI'05 emissions set for the innermost domain of the WRF/Chem model.

emission species can help us to understand the BEC statistics. The emission species is derived from the emission file that is produced by the NEI'05 data for each model domain. Only the emission data for the innermost domain is used to calculate the correlation among the emission species. The emission file contains 37 variables, including gas species and aerosol species. An aerosol species also comprises a nuclei mode and accumulation model species (Peckam et al., 2013). From these aerosol emission species, five lumped aerosol species are calculated, which is consistent with the variables in the data assimilation. These five lumped species are E_EC (sum of the nuclei mode and the accumulation mode of elemental carbon PM_{2.5}), E_ORG (sum of the nuclei mode and the accumulation mode of organic PM_{2.5}), E_NO3 (sum of the nuclei mode and the accumulation mode of nitrate PM_{2.5}), E_SO4 (sum of the nuclei mode and the accumulation mode of sulfate PM_{2.5}) and E_PM25 (sum of the nuclei mode and the accumulation mode of unspeciated primary PM_{2.5}).

Figure 2 shows the cross-correlations of the five lumped aerosol emission species. All cross-correlations exceed 0.5. This result reveals that the emission species are correlated, which may be attributed to the common emission sources and diffusion processes that are controlled by the same atmospheric circulation. The most significant cross-correlation is between E_EC and E_ORG with a value of approximately 0.8. This high correlation demonstrates that the emission distributions of these two species are very similar. Their emissions are primary in urban and suburban areas with small emissions in rural areas and along roadways (not shown). As shown in Fig. 2, the lowest cross-correlation is between E_ORG and E_SO4; the latter emissions are primary in the urban and suburban areas with few emissions in rural areas and roadways (not shown).

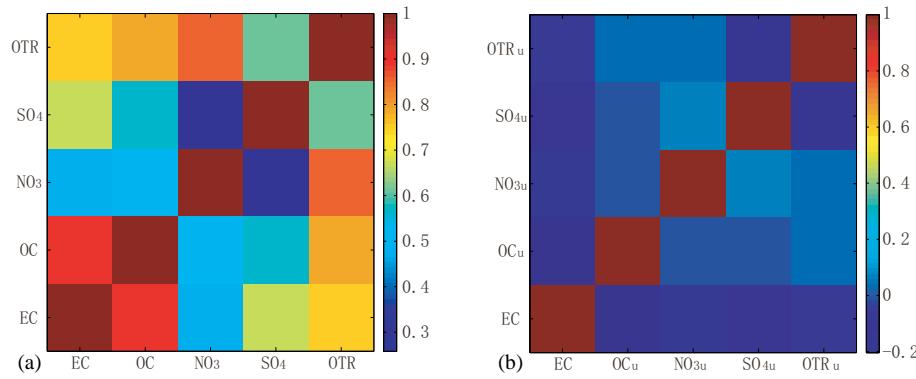


Figure 3. Cross-correlations between the five variables of the BEC. These variables are (a) full variables and (b) unbalanced variables of EC, OC, NO₃, SO₄ and OTR.

4 Balance constraints and BEC statistics

With the configuration of the WRF/Chem model described in Sect. 3.1, forecasts for 1 month (from 00:00 UTC on 15 May to 00:00 UTC on 14 June 2010) were performed for the balance constraints and the BEC statistics. Forecast differences between 24 h forecasts and 48 h forecasts are available at 00:00 UTC; 30 forecast differences are employed as inputs in the NMC method. For this method, 30 forecast differences are sufficient; however, a longer time series may be more beneficial for the BEC statistics (Parrish and Derber, 1992).

4.1 Balance regression statistics

Using the 30 forecast differences between 24 and 48 h forecasts, we can obtain δ EC, δ OC, δ NO₃, δ SO₄ and δ OTR. The size of these variables is $(N \times 30)$, where N is the number of model grid points. We put these data into Eqs. (6)–(13) to calculate the regression coefficients of ρ_{ij} and the unbalanced parts of the variables. Note the process of calculation should be step by step, since the latter equation will use the unbalanced parts of former equations. Table 1 shows the regression coefficients, whose column and row are consistent with $\rho_{i,j}$ in Eq. (7). The last column in Table 1 is the coefficient of determination (R^2) of the regression equations. For the regression equation of OC, the regression coefficient is 0.90 and the coefficient of determination of Eq. (7) is 0.86, which indicates that EC and OC are highly correlated and their mass concentration scales are approximate. Their correlation is similar to the correlation of the stream function and velocity potential; thus, we set them as the first and second variables in the regression statistics. For the regression equation of NO₃, the regression coefficients of EC and OC_u are 4.01 and 3.76, respectively, because the mass concentration scale of NO₃ exceeds the mass concentration scales of EC and OC_u. The coefficient of determination is only 0.32, which indicates that the correlations between NO₃ and EC and between NO₃ and OC_u are weak. This result reveals that the forecast errors of NO₃ differ from the forecast errors of

EC and OC_u. A possible reason is that NO₃ is the secondary particle that is primarily derived from the transformation of NO_x, but EC and OC_u are derived from direct emissions. Similar to NO₃, SO₄ is also primarily derived from the transformation of SO₂ and the coefficient of determination for SO₄ is also low. For the last variable OTR, the maximum coefficient of determination is 0.96 because OTR includes some different compositions that are correlated with the first four variables.

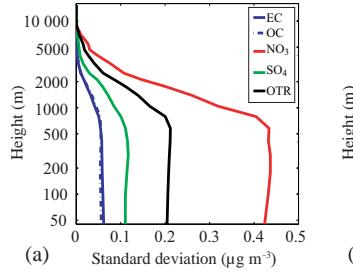
Figure 3 shows the cross-correlations of the five full variables and the unbalanced variables. In Fig. 3a, the cross-correlations of the full variables exceed 0.3 and most of them exceed 0.5. In Fig. 3b, however, the cross-correlations of the unbalanced variables are less than 0.2. Some of the cross-correlations are close to zero, which indicates that these unbalanced variables are approximatively independent and can be employed as control variables in the data assimilation system.

4.2 BEC statistics

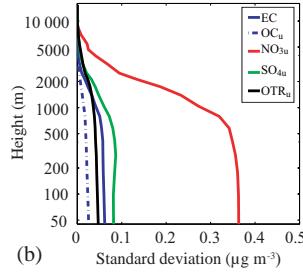
Using the original full variables and the unbalanced variables obtained by the regression equations, the BEC statistics are obtained. Figure 4 shows the vertical profiles of the standard deviations of the original \mathbf{D} and the unbalanced \mathbf{D}_u . In Fig. 4a, the original standard deviation of NO₃ is the largest value, whereas the smallest value is OC, whose profile is close to the profile of EC. All profiles show a significant decrease at approximately 800 m because the aerosol particulates are usually limited under the boundary level. In Fig. 4b, all standard deviations decrease in different degrees, with the exception of EC, which remains as the control variable in the unbalanced BEC statistics. Note that the standard deviation of OTR_u decreases by approximately 80 % compared with NO_{3u}, which decreases by approximately 10 %. This result is attributed to the small coefficient of determination for the regression of NO₃ (in Table 1), which indicates that a small portion of NO₃ can be predicted by the regression and a large

Table 1. Regression coefficients of balance operator \mathbf{K} and the coefficient of determination (regression coefficients correspond to ρ_{ij} in Eq. 7).

Species	Regression coefficient (ρ)					Coefficient of determination (R^2)
EC	1					/
OC	0.90	1				0.86
NO_3	4.01	3.76	1			0.32
SO_4	1.35	-0.21	-3.15	1		0.48
OTR	2.93	2.35	0.28	0.60	1	0.96



(a) Standard deviation ($\mu\text{g m}^{-3}$)

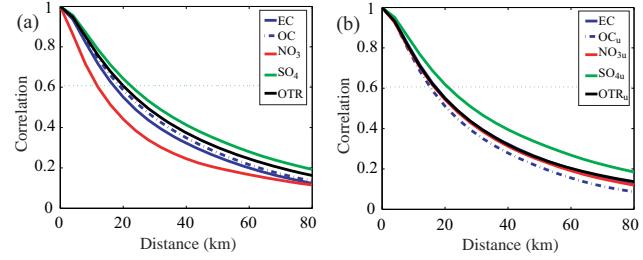


(b) Standard deviation ($\mu\text{g m}^{-3}$)

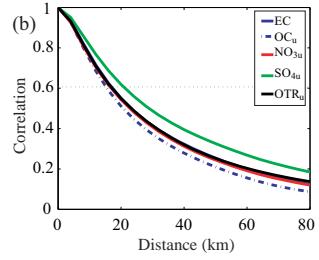
Figure 4. Vertical profiles of the standard deviation of the variables. (a) Full variables and (b) unbalanced variables.

portion is an unbalanced component. In contrast with NO_3 , a small portion of OTR is the unbalanced component.

For the correlation matrix of \mathbf{C} and \mathbf{C}_u , they are factorized as three independent one-dimensional correlation matrices in Eq. (24). The horizontal correlation \mathbf{C}_x or \mathbf{C}_y is approximately expressed by a Gaussian function. The correlation between two points r_1 and r_2 can be written as $e^{-\frac{(r_2-r_1)^2}{2L_s^2}}$, where L_s is the horizontal correlation scale and is a constant value for \mathbf{C}_x and \mathbf{C}_y , which are considered to be isotropic (Li et al., 2013). This scale can be estimated by the curve of the horizontal correlations with distances. Figure 5 shows the curves of the horizontal correlations for the five control variables. For the full variables (Fig. 5a), the sharpest decrease in the curves is observed for NO_3 and the slowest decrease in the curves is observed for SO_4 . We assume that the decline curve is according to the Gaussian function. Then the intersection of the decline curve and the line of $e^{-\frac{1}{2}}$ (≈ 0.61) can be approximately as the value of horizontal correlation scale. The horizontal correlation scales of EC, OC, NO_3 , SO_4 and OTR are 25, 27, 20, 30 and 28 km, respectively. For the unbalanced variables (Fig. 5b), their curves are closer than the curves of the full variables. The correlation scales of EC, OC_u , NO_3u , SO_4u and OTR_u are 25, 23, 24, 28 and 25 km, respectively. These results suggest that the unbalanced variables are expressed by some common factors, such as EC, OC_u and NO_3u , in the regression equations of Eqs. (10)–(13), which produces consistent horizontal correlation scales.



(a)



(b)

Figure 5. Same as Fig. 4, with the exception of the horizontal auto-correlation curves of the variables. The horizontal thin line is the reference line of $e^{-\frac{1}{2}}$ (≈ 0.61) for determining the horizontal correlation scales.

For the vertical correlation between \mathbf{C}_z and \mathbf{C}_{uz} , they are directly estimated using the forecasting differences in the data assimilation system, but not estimated from an approximately alternative function. Because it is only an $n_z \times n_z$ matrix. Figure 6 shows the vertical correlation matrices \mathbf{C}_z and \mathbf{C}_{uz} for the full variables (left column) and the unbalanced variables (right column), respectively. A common feature of both the full variables and the unbalanced variables is the significant correlation between the levels of the boundary layer height, which is consistent with the profile of the standard deviation in Fig. 4. Some weak adjustments to the correlations between the full and unbalanced variables are made. For example, the correlation of NO_3u is stronger than the correlation of NO_3 between the boundary layers. Compared with the analysis of horizontal correlation scale, the vertical correlation scale of NO_3u is larger than the vertical correlation scale of NO_3 . Conversely, the vertical correlation scale of OTR_u is smaller than the vertical correlation scale of OTR. These results demonstrate that the vertical correlations for the unbalanced variables are more consistent than the vertical correlations of the full variables, which is similar to the adjustments to the horizontal correlation scale. Note that the differences in vertical correlation are slight, compared with those of horizontal correlation. The main reason is that the vertical correlations are generally affected by the atmospheric boundary layer height. Thus, all vertical correlation decreases rapidly for the levels above the boundary layer height.

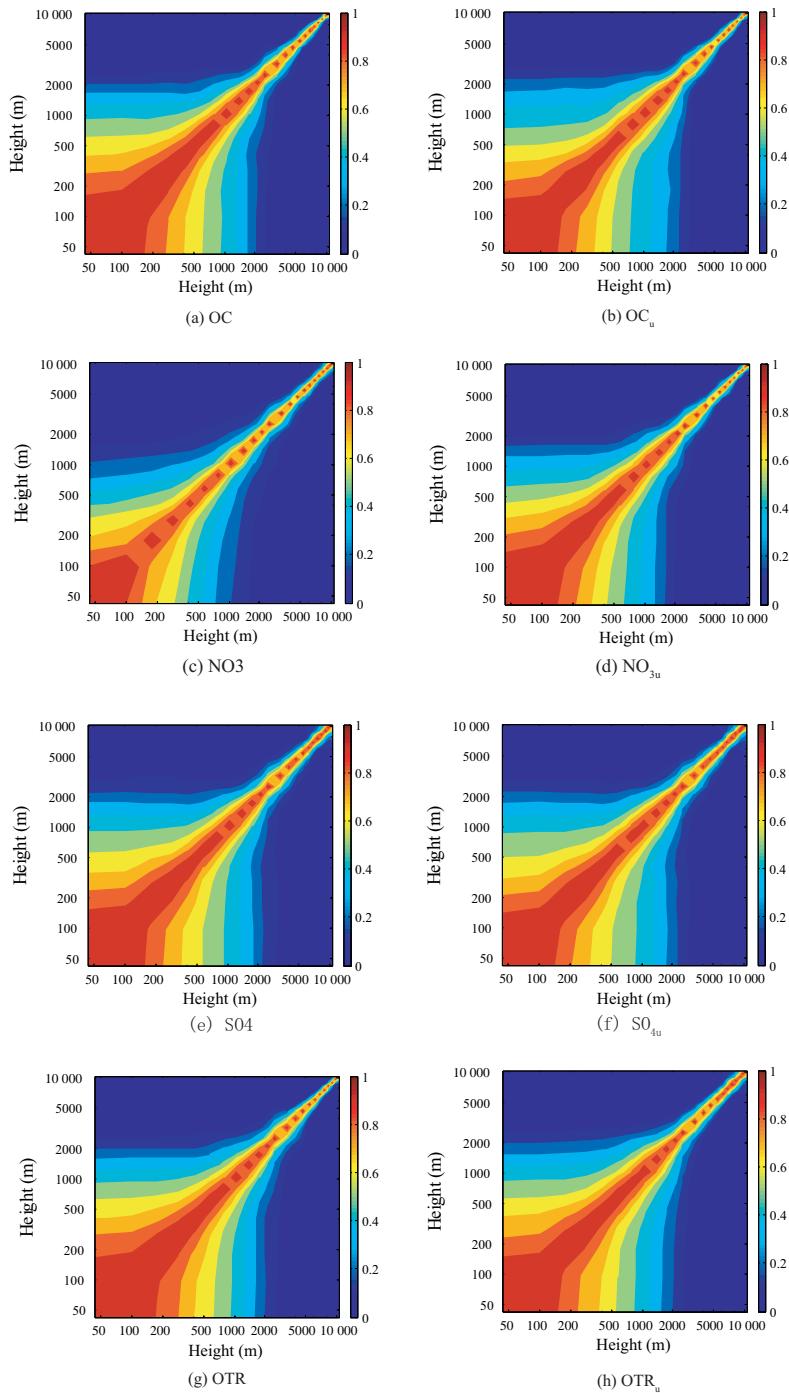


Figure 6. Vertical correlations of the five variables of the BEC. The left column represents the full variables, and the right column represents the unbalanced variables.

5 Application to data assimilation and prediction

To exhibit the effect of the balance constraint of the BEC, the data assimilation experiments and 24 h forecasts for nine cases are run using WRF/Chem model. The surface $\text{PM}_{2.5}$ and aircraft-speciated observations are assimilated using dif-

ferent BEC, and the evaluations are presented for the data assimilation and subsequent forecasts. Three basic statistical measures including mean bias (BIAS), root mean square error (RMSE) and correlation coefficient (CORR) are utilized for the evaluations.

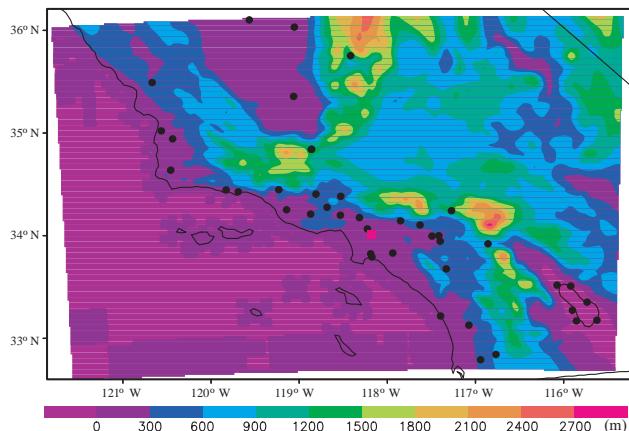


Figure 7. The topography of the innermost domain and the locations of surface monitoring stations (black dots). The red square is the location of Los Angeles

5.1 Observation data and experiment scheme

Two types of observation data are employed in our experiments. The first type of observation data consists of hourly surface $\text{PM}_{2.5}$ concentrations from the California Air Resources Board (ARB). There are 42 surface $\text{PM}_{2.5}$ monitoring sites that existed in the innermost domain of the WRF/Chem model (Fig. 7). The second type of observation data is the speciated concentration along the aircraft flight track. The aircraft observations were investigated from the California Research at the Nexus of Air Quality and Climate Change (CalNex) field campaign in 2010. Nine flights data around Los Angeles from 15 May to 14 June 2010 are selected as the cases of data assimilation. Table 2 shows the start time and end time of each flight. The species of the aircraft observations include OC, NO_3 , SO_4 and NH_4 . Note that NH_4 is not a control variable; thus, the aircraft observation of NH_4 is disregarded in the data assimilation. Because the particle size of the aircraft observations is less than $1.0 \mu\text{m}$, some adjustments to the flight observations are made according to the ratios between the concentration under $2.5 \mu\text{m}$ and the concentration under $1.0 \mu\text{m}$ for each species using model products. With the ratios multiplied by the aircraft observed concentrations, the speciated concentrations under $2.5 \mu\text{m}$ can be obtained.

The initial time of data assimilation cases are designed according to the period of flights, shown in Table 2. The time window of assimilation for the flight data is $\pm 1.5 \text{ h}$, though some flight times do not completely cover the time windows. Figure 8 shows the aircraft tracks during the time window of data assimilation. It is obvious that the aircraft data on 21, 25 May and 14 June are relative few as the tracks are almost outside of the study domain. For the surface data, it is only the observations at the initial time that assimilated. For each case, three parallel experiments are performed. The first experiment is the control experiment without aerosol data as-

similation, which is frequently known as a free run and denoted as Control. The second experiment is a data assimilation experiment that assimilates surface $\text{PM}_{2.5}$ and aircraft observations using the full variables without balance constraints; it is denoted as DA-full. The third experiment is also a data assimilation experiment that also assimilates surface $\text{PM}_{2.5}$ and aircraft observations, but employs the unbalanced variables as control variables conducted by the balanced constraint; it is denoted as DA-balance. The backgrounds for DA-full and DA-balance are the forecasting results from the previous runs without DA. These previous forecasting results have been obtained when we run the model for the BEC statistics. The observation error is half of the standard deviation of the original background variable, and a vertical profile of observation errors is applied with the average profile of standard deviation of the background variable. In each experiment, a 24 h forecasting is run using the WRF/Chem model with the same configuration described in Sect. 3.1, and the case on 3 June 2010 is presented in detail as an example.

5.2 Increments of data assimilation

Figure 9 shows the horizontal increments of EC, OC, NO_3 , SO_4 and OTR at the first model level for the DA-full (left column) and DA-balance experiments (right column) of the case on 3 June 2010. In the DA-full experiment, the increment of EC and OTR (Fig. 9a and i) are similar. They are obtained from the surface $\text{PM}_{2.5}$ observations because no direct aircraft observations correspond to these two variables. In the DA-balance experiment, significant adjustments are made to the increments of EC (Fig. 9b) under the action of the balance constraints. The observations of OC greatly affect the increments of EC for the high cross-correlation between EC and OC. Thus, the increments of EC are similar with the increments of OC. Similarly, significant adjustments are made to the increment of OTR (Fig. 9j), though there are no species observation of OTR. There are also some slight adjustments for the increments of OC, NO_3 and SO_4 for the crossing spread among species.

Figure 10 shows the vertical increments along 35.0°N for the DA-full and DA-balance experiments. Similar to Fig. 9, the increments of EC and OTR (Fig. 10a and i) spread upward from the surface in the DA-full experiment, which are obtained from the surface $\text{PM}_{2.5}$ observation. In the DA-balance, the increments of EC and OTR (Fig. 10b and j) exhibit observation information from the aircraft height at approximately 500 m, and the value of the increments show significant increases. The distributions of the increments for these five variables in the DA-balance (Fig. 10, right column) generally tend to coincide compared with the distributions of the increments in the DA-full (Fig. 10, left column). The results of the DA-balance are reasonable due to their influence on each other across the balance constraints.

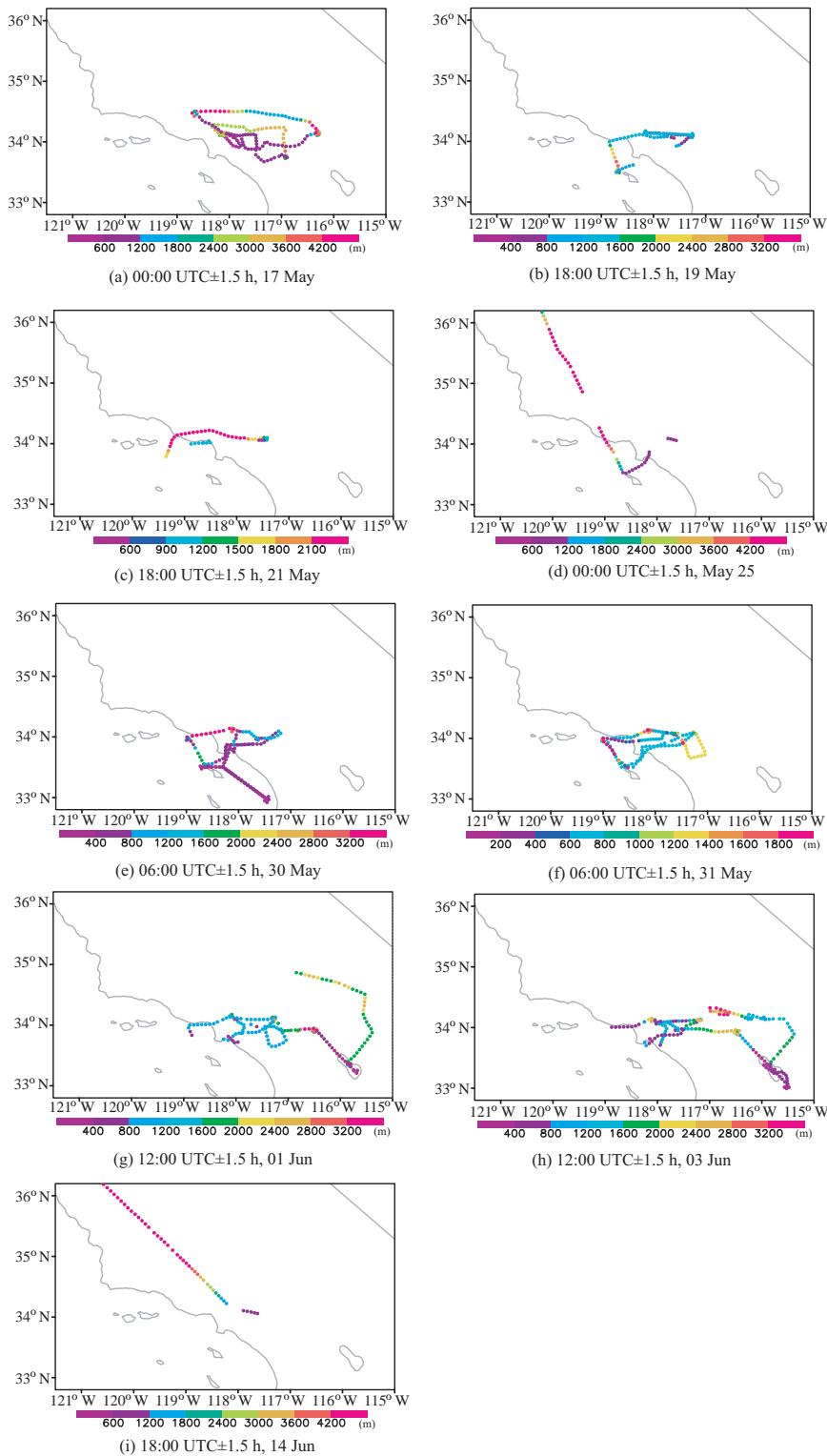


Figure 8. Aircraft flight tracks during the time window of data assimilation for nine cases. The color of the track indicates the aircraft height.

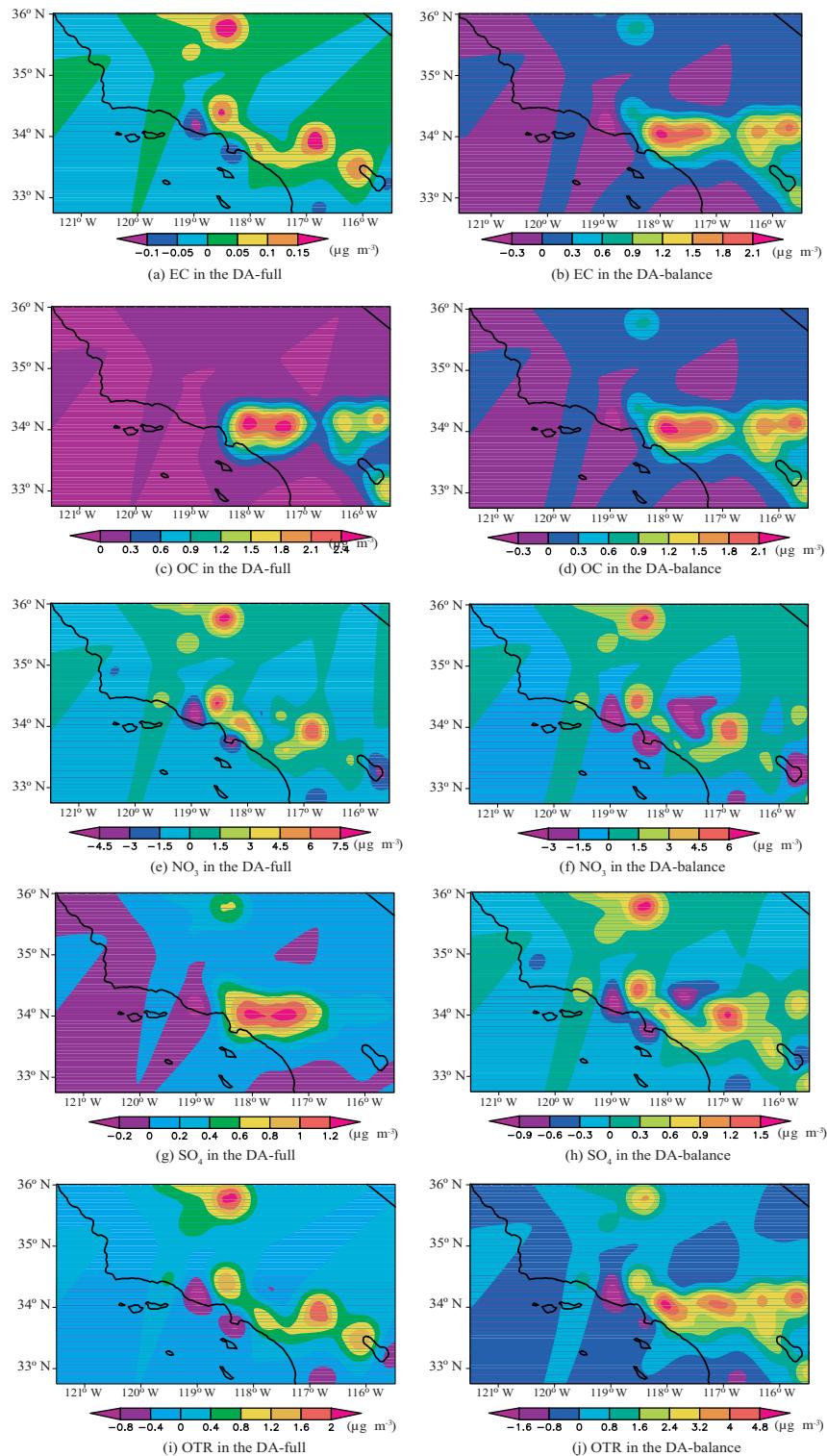


Figure 9. Surface distributions of increments of the five variables of EC, OC, NO_3 , SO_4 and OTR at 12:00 UTC on 3 June 2010. The left column and right column are from DA-full and DA-balance, respectively.

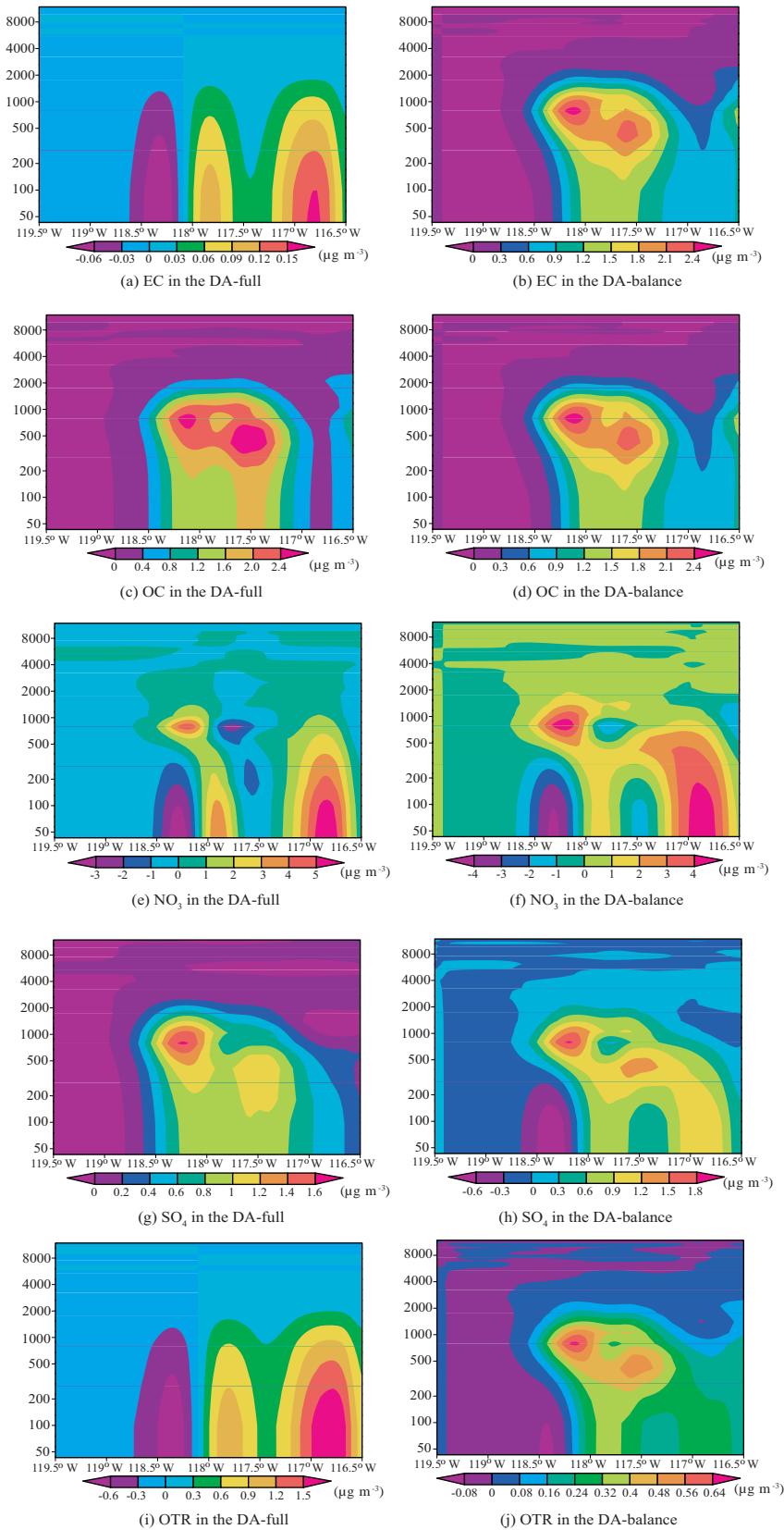


Figure 10. Same as Fig. 9, with the exception of the vertical sections along 35° N.

Table 2. The periods of flight during CalNex 2010 and the initial time of assimilation.

Number of cases	Start time of flight	End time of flight	Initial time of assimilation
1	18:00 UTC, 16 May	01:42 UTC, 17 May	00:00 UTC, 17 May
2	17:28 UTC, 19 May	00:10 UTC, 20 May	18:00 UTC, 19 May
3	17:28 UTC, 21 May	00:10 UTC, 21 May	18:00 UTC, 21 May
4	23:08 UTC, 24 May	05:23 UTC, 25 May	00:00 UTC, 25 May
5	01:59 UTC, 30 May	07:45 UTC, 30 May	06:00 UTC, 30 May
6	05:00 UTC, 31 May	10:54 UTC, 31 May	06:00 UTC, 31 May
7	07:59 UTC, 2 June	14:09 UTC, 2 June	12:00 UTC, 2 June
8	07:59 UTC, 3 June	14:041 UTC, 3 June	12:00 UTC, 3 June
9	17:56 UTC, 14 June	23:35 UTC, 14 June	18:00 UTC, 14 June

5.3 Evaluation of data assimilation and forecasts

Figure 11 shows the scatter plots of the initial model fields vs. the surface observation for all nine cases. In Fig. 11a, the simulated concentrations of the Control experiment display a significant underestimation with a BIAS of $-3.66 \mu\text{g m}^{-3}$. The mean concentration of Control is $10.90 \mu\text{g m}^{-3}$, about 25.1 % lower than observed mean concentrations ($14.56 \mu\text{g m}^{-3}$). In the DA-full and DA-balance experiments, there are remarkable increases for the simulated concentrations, and the BIASSs reduce to as small as -1.21 and $-0.94 \mu\text{g m}^{-3}$. The RMSE is $9.53 \mu\text{g m}^{-3}$ in the Control experiment. The RMSE reduces to 4.82 and $4.48 \mu\text{g m}^{-3}$ in the DA-full and DA-balance experiment, respectively. There are also significant improvements for the CORR in the DA-full and DA-balance experiments, compared with the Control experiment. Furthermore, these three statistical measures of the DA-balance experiments show some slight improvement, compared with that of the DA-full experiments. The result demonstrates that more observation information spread by balance constraints can improve assimilation performance.

To evaluate the effects of the data assimilation, the CORR, RMSE and BIAS during the forecast time are calculated for each case, and their averaged results are shown in Fig. 12. The CORRs of the DA-balance and DA-full experiments are very close (Fig. 12a). But, the difference increases after the 1st hour with a higher CORR in the DA-balance experiment. The CORR of the DA-balance experiment is substantially higher than that of the DA-full experiment from the 2nd hour to the 16th hour. Similar improvements for the RMSE and the BIAS of the DA-balance experiment are observed in Fig. 12b and c, respectively. The improvement for the BIAS in the DA-balance experiment is the most significant among these three statistical measures. The peak value of the improvement for the BIAS (Fig. 12c) is at the 4th hour, and the improvement is distinct until the end of forecasts. These improvements indicate that the balance constraint is positive for the subsequent forecasts, which derives from the balanced initial distribution among species.

6 Summary and discussion

We examined the BEC in a 3DVAR system, which uses five control variables (EC, OC, NO_3 , SO_4 and OTR) that are derived from the MOSAIC aerosol scheme in the WRF/Chem model. Based on the NMC method, differences within a month-long period between 24 and 48 h forecasts that are valid at the same time were employed in the estimation and analyses of the BEC. The background errors of these five control variables are highly correlated. Especially between EC and OC, their correlation is as large as 0.9.

A set of balance constraints was developed using a regression technique and incorporated in the BEC to account for the large cross-correlations. We employ the balance constraint to separate the original full variables into balanced and unbalanced parts. The regression technique is used to express the balanced parts by the unbalanced parts. These unbalanced parts can be assumed independent. Then, the unbalanced parts are employed as control variables in the BEC statics. Accordingly, the standard deviations of these unbalanced variables are less than the standard deviations of the original variables. The horizontal correlation scales of unbalanced variables are closer than that of full variables on the effect of the balance constraints, and the vertical correlations of unbalanced variables show a similar trend.

To evaluate the impact of the balance constraints on the analyses and forecasts, three groups of experiments, including a control experiment without data assimilation and two data assimilation experiments with and without balance constraints (DA-full and DA-balance), were performed. In the data assimilation experiments, the observations of surface $\text{PM}_{2.5}$ concentration and aircraft-specified concentration of OC, NO_3 and SO_4 were assimilated. The observations of these three variables can spread to the two remaining variables in the increments of the DA-balance, which results in a more complex distribution. The evaluations of CORR, RMSE and BIAS for the initial analysis fields show more improvement in the DA-balance experiments, compared with the DA-full experiments, although some of these improvements are slight. An important reason is that the surface $\text{PM}_{2.5}$ observations are independent from the aircraft obser-

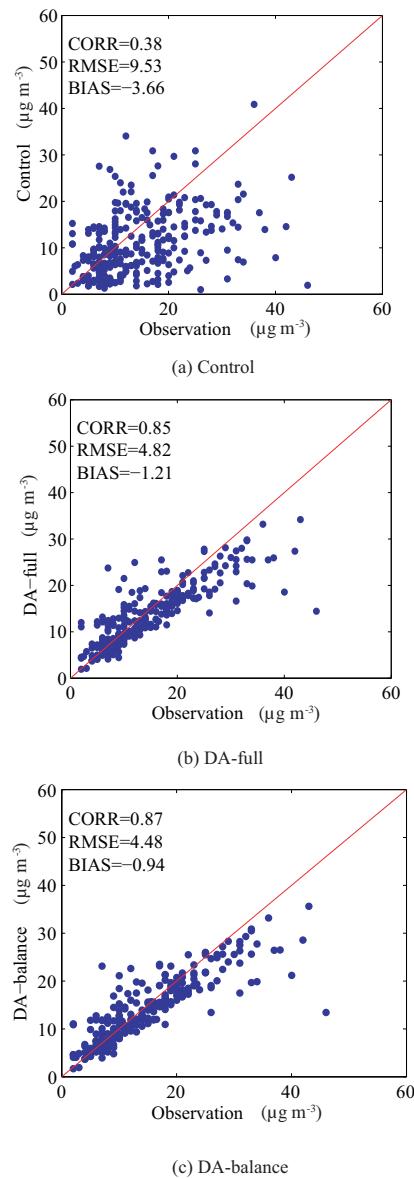


Figure 11. Scatter plots of observed concentrations of $\text{PM}_{2.5}$ vs. simulated $\text{PM}_{2.5}$ concentrations of the experiments of (a) Control, (b) DA-full, and (c) DA-balance for all nine cases.

vations. If we evaluate the analysis fields by the species observation of aircraft, there may be more significant improvements in the DA-balance experiments.

While the improvements increase after the first forecasting hour in the DA-balance experiments, compared with forecasts of the DA-full experiments, the improvements persist to the end of forecasts, and are substantial from the 2nd hour to the 16th hour (Fig. 12). These results suggested that the balance constraints can serve an important role for continually improving the skill of sequent forecasts. Note that some aircraft data are relatively few, and some flight tracks are not around Los Angeles in some cases (Fig. 8). If there are more

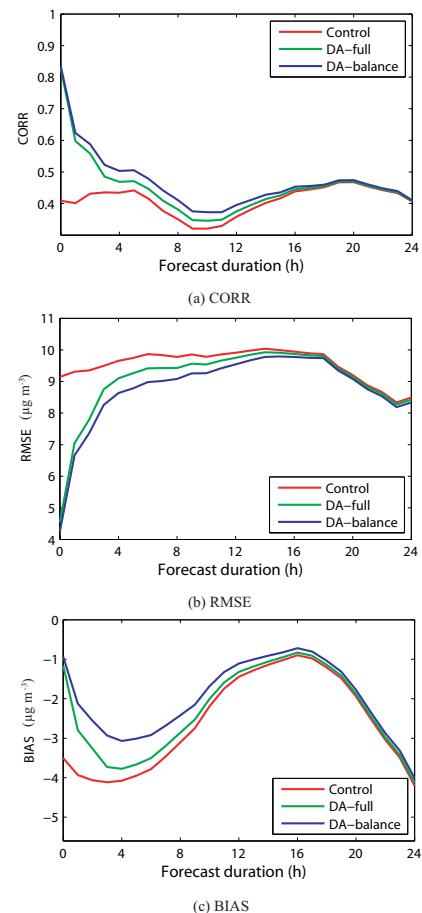


Figure 12. The averaged (a) correlations, (b) root mean square errors (RMSE in $\mu\text{g m}^{-3}$) and (c) mean bias (BIAS in $\mu\text{g m}^{-3}$) of the $\text{PM}_{2.5}$ concentration forecasts against observations as a function of forecast duration.

aircraft observations, the improvements of the DA-balance experiments should be more significant and durable.

The developed method for incorporating balance constraints in aerosol data assimilation can be employed in other areas or other applications for different aerosol models. For the aerosol variables in different models, some cross-correlations between different species or size bins should exist because their common emissions and diffusion processes are controlled by the same atmospheric circulation. Although these cross-correlations may be stronger than the cross-correlations of atmospheric or oceanic model variables, theoretic balance constraints, such as geostrophic balance or temperature–salinity balance, do not exist. We expected to discover a universal balance constraint that can describe the physical or chemical balanced relationship of aerosol variables, and utilize it in the data assimilation system. In addition, we expected to expand the balance constraint to include gaseous pollutants, such as nitrite (NO_2), sulfur dioxide (SO_2) and (carbon monoxide) CO . These gaseous pollu-

tants are correlated with some aerosol species, such as NO_3 , SO_4 and EC, which can improve the data assimilation analysis fields of aerosols by assimilating these gaseous observations. The assimilation of aerosol observations may improve the analysis fields of gaseous pollutants.

7 Code availability

This data assimilation system is established by ourself. The code of this system can be obtained on request from the first author (zzlqxxy@163.com).

Acknowledgements. This research was supported by the National Natural Science Foundation of China (41275128). We gratefully thank the California Air Resources Board (<http://www.arb.ca.gov/homepage.htm>) and NOAA Earth System Research Laboratory Chemical Sciences Division (<http://esrl.noaa.gov/csd/groups/csd7/measurements/2010calnex/>), for providing the download of surface and aircraft aerosol observations.

Edited by: A. Sandu

Reviewed by: two anonymous referees

References

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, *Q. J. Roy. Meteor. Soc.*, 134, 1951–1970, 2008a.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, *Q. J. Roy. Meteor. Soc.*, 134, 1971–1996, 2008b.

Barker, D. M., Huang, W., Guo, Y. R., and Xiao, Q. N.: A Three-Dimensional (3DVAR) data assimilation system for use with MM5: implementation and initial results, *Mon. Weather Rev.*, 132, 897–914, 2004.

Benedetti, A. and Fisher, M.: Background error statistics for aerosols, *Q. J. Roy. Meteor. Soc.*, 133, 391–405, 2007.

Chen, Y., Rizvi, S., Huang, X., Min, J., and Zhang, X.: Balance characteristics of multivariate background error covariances and their impact on analyses and forecasts in tropical and Arctic regions, *Meteorol. Atmos. Phys.*, 121, 79–98, 2013.

Cohn, S. E.: Estimation theory for data assimilation problems: Basic conceptual framework and some open questions, *J. Meteorol. Soc. Jpn.*, 75, 257–288, 1997.

Constantinescu, E. M., Chai, T., Sandu, A., and Carmichael, G. R.: Autoregressive models of background errors for chemical data assimilation, *J. Geophys. Res.*, 112, D12309, doi:10.1029/2006JD008103, 2007.

Derber, J. and Bouttier, F.: A reformulation of the background error covariance in the ECMWF global data assimilation system, *Tellus*, 51, 195–221, 1999.

Geller, M. D., Fine, P. M., and Sioutas, C.: The relationship between real-time and time-integrated coarse (2.5–10 m), intermodal (1–2.5 m), and fine (< 2.5 m) particulate matter in the Los Angeles basin, *J. Air Waste Manage.*, 54, 1029–1039, 2004.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled “online” chemistry within the WRF model, *Atmos. Environ.*, 39, 6957–6976, doi:10.1016/j.atmosenv.2005.04.027, 2005.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, 6, 3181–3210, doi:10.5194/acp-6-3181-2006, 2006.

Huang, X. Y., Xiao, Q., Barker, D. M., Zhang, X., Michalakes, J., Huang, W., and Henderson, T.: Four-dimensional variational data assimilation for WRF: formulation and preliminary results, *Mon. Weather Rev.*, 137, 299–314, 2009.

Jazwinski, A. H.: *Stochastic processes and filtering theory*, Academic Press, New York, 376 pp., 1970.

Kahnert, M.: Variational data analysis of aerosol species in a regional CTM: background error covariance constraint and aerosol optical observation operators, *Tellus B*, 60, 753–770, 2008.

Li, Z., Zang, Z., Li, Q. B., Chao, Y., Chen, D., Ye, Z., Liu, Y., and Liou, K. N.: A three-dimensional variational data assimilation system for multiple aerosol species with WRF/Chem and an application to $\text{PM}_{2.5}$ prediction, *Atmos. Chem. Phys.*, 13, 4265–4278, doi:10.5194/acp-13-4265-2013, 2013.

Liu, Z., Liu, Q., Lin, H. C., Schwartz, C. S., Lee, Y. H., and Wang, T.: Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia, *J. Geophys. Res.*, 116, D23206, doi:10.1029/2011JD016159, 2011.

Mesinger, F., DiMego, G., Kalnay, E., Shafran, P., Ebisuzaki, W., Jovic, D., Woollen, J., Mitchell, K., Rogers, E., Ek, M., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American Regional Reanalysis, *B. Am. Meteorol. Soc.*, 87, 343–360, 2006.

Peckham, S. E., Grell, G. A., McKeen, S. A., and Ahmadov, R.: WRF/Chem Version 3.5 User’s Guide, Colorado, NOAA Earth System Research Laboratory, 2013.

Pagowski, M. and Grell, G. A.: Experiments with the assimilation of fine aerosols using an ensemble Kalman filter, *J. Geophys. Res.*, 117, D21302, doi:10.1029/2012JD018333, 2012.

Pagowski, M., Grell, G. A., McKeen, S. A., Peckham, S. E., and Devenyi, D.: Three-dimensional variational data assimilation of ozone and fine particulate matter observations: Some results using the Weather Research and Forecasting–Chemistry model and Grid-point Statistical Interpolation, *Q. J. Roy. Meteor. Soc.*, 136, 2013–2024, doi:10.1002/qj.700, 2010.

Parrish, D. F. and Derber, J. C.: The national meteorological center spectral statistical interpolation analysis, *Mon. Weather Rev.*, 120, 1747–1763, 1992.

Ricci, S. and Weaver, A. T.: Incorporating State-Dependent Temperature–Salinity Constraints in the Background Error Covariance of Variational Ocean Data Assimilation, *Mon. Weather Rev.*, 133, 317–338, 2005.

Saide, P. E., Carmichael, G. R., Spak, S. N., Minnis, P., and Ayers, J. K.: Improving aerosol distributions below clouds by assimilating

satellite-retrieved cloud droplet number, *P. Natl. Acad. Sci. USA*, 109, 11939–11943, doi:10.1073/pnas.1205877109, 2012.

Saide, P. E., Carmichael, G. R., Liu, Z., Schwartz, C. S., Lin, H. C., da Silva, A. M., and Hyer, E.: Aerosol optical depth assimilation for a size-resolved sectional model: impacts of observationally constrained, multi-wavelength and fine mode retrievals on regional scale analyses and forecasts, *Atmos. Chem. Phys.*, 13, 10425–10444, doi:10.5194/acp-13-10425-2013, 2013.

Salako, G. O., Hopke, P. K., Cohen, D. D., Begum, B. A., Biswas, S. K., Pandit, G. G., Chung, Y. S., Rahman, S. A., Hamzah, M. S., Davy, P., Markwitz, A., Shagjamba, D., Lodoysamba, S., Wimolwattanapun, W., and Bunprapob, S.: Exploring the Variation between EC and BC in a Variety of Locations, *Aerosol Air Qual. Res.*, 12, 1–7, 2012.

Schwartz, C. S., Liu, Z., Lin, H.-C., and McKeen, S. A.: Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth, *J. Geophys. Res.*, 117, D13202, doi:10.1029/2011JD017383, 2012.

Schwartz, C. S., Liu, Z., Lin, H.-C., and Cetola, J. D.: Assimilating aerosol observations with a “hybrid” variational-ensemble data assimilation system, *J. Geophys. Res.-Atmos.*, 119, 4043–4069, doi:10.1002/2013JD020937, 2014.

Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., and Jones, D.: Construction of non-diagonal background error covariance matrices for global chemical data assimilation, *Geosci. Model Dev.*, 4, 299–316, doi:10.5194/gmd-4-299-2011, 2011.

Sun C.-H., Lin Y.-C., and Wang C.-S.: 7. Relationships among Particle Fractions of Urban and Non-urban Aerosols, *Aerosol Air Qual. Res.*, 3, 7–15, 2003.

Wu, W. S., Purser, R. J., and Parrish, D. F.: Three-dimensional variational analysis with spatially inhomogeneous covariances, *Mon. Weather Rev.*, 130, 2905–2916, 2002.

Zaveri, R. A., Easter, R. C., Fast, J. D., and Peters, L. K.: Model for Simulating Aerosol Interactions and Chemistry(MOSAIC), *J. Geophys. Res.*, 113, D13204, doi:10.1029/2007JD008782, 2008.