



Inverse modeling of black carbon emissions over China using ensemble data assimilation

P. Wang, H. Wang, Y. Q. Wang, X. Y. Zhang, S. L. Gong, M. Xue, C. H. Zhou, H. L. Liu, X. Q. An, T. Niu, and Y. L. Cheng

Institute of Atmospheric Composition, Key Laboratory of Atmospheric Chemistry (LAC) of China Meteorological Administration (CMA), Chinese Academy of Meteorological Sciences (CAMS), Beijing, 100081, China

Correspondence to: P. Wang (pwang@cma.scma.cn)

Received: 24 March 2015 – Published in Atmos. Chem. Phys. Discuss.: 4 August 2015

Revised: 23 December 2015 – Accepted: 6 January 2016 – Published: 27 January 2016

Abstract. Emissions inventories of black carbon (BC), which are traditionally constructed using a *bottom-up* approach based on activity data and emissions factors, are considered to contain a large level of uncertainty. In this paper, an ensemble optimal interpolation (EnOI) data assimilation technique is used to investigate the possibility of optimally recovering the spatially resolved emissions bias of BC. An inverse modeling system for emissions is established for an atmospheric chemistry aerosol model and two key problems related to ensemble data assimilation in the *top-down* emissions estimation are discussed: (1) how to obtain reasonable ensembles of prior emissions and (2) establishing a scheme to localize the background-error matrix. An experiment involving 1-year-long simulation cycle with EnOI inversion of BC emissions is performed for 2008. The bias of the BC emissions intensity in China at each grid point is corrected by this inverse system. The inverse emission over China in January is 240.1 Gg, and annual emission is about 2539.3 Gg, which is about 1.8 times of bottom-up emission inventory. The results show that, even though only monthly mean BC measurements are employed to inverse the emissions, the accuracy of the daily model simulation improves. Using top-down emissions, the average root mean square error of simulated daily BC is decreased by nearly 30 %. These results are valuable and promising for a better understanding of aerosol emissions and distributions, as well as aerosol forecasting.

1 Introduction

Black carbon (BC) refers to light-absorbing carbon aerosols produced by all kinds of incomplete combustion processes (Bond and Bergstrom, 2006). It is an important component of atmospheric particulate matter, affecting weather, climate and air quality, and therefore attracts much attention among the scientific community. Its absorptive nature, which directly causes reductions in incoming shortwave solar radiation, is a key contributor to climate forcing by aerosols (Liousse et al., 1993; Menon et al., 2002; Hansen et al., 2005; Ramanathan and Carmichael, 2008). BC aerosols have been shown to act as cloud condensation nuclei when they become hydrophilic, affecting cloud microphysical properties and rainfall processes (Lary et al., 1997; Bond et al., 2013). The lifetime of BC is about 3–10 days, and it can be transported far from its source to affect remote and pristine areas (Hansen et al., 1988; Hara et al., 2008). Its light-absorbing properties reduce atmospheric visibility (Wolff, 1981). Qiu and Yang (2000) showed that BC contributes to the considerable degradation in optical depths and visibility noted in northern China. Furthermore, from the human health perspective, these particles, which are generally sub-micron in size, contribute greatly to the threat of pulmonary diseases, as they can penetrate into the lungs while also carrying a variety of toxic elements with them. Therefore, an accurate picture of the distribution and variation of BC is crucial to our understanding of climate change and pollutant dynamics, and ultimately helps us to develop better policies to tackle associated environmental problems.

However, there is considerable uncertainty involved in the estimation of the distribution of BC and its contribution to

BC emissions inventory, which is traditionally constructed from the *bottom up* approach based on activity data and emissions factors, is considered to have large uncertainty (Cao et al., 2006). The overall uncertainty of BC over all of Asia is estimated at about 360 % (Streets et al., 2003a, b). Emissions of BC are difficult to determine under the best of circumstances, largely because of the uncertainty in quantifying the fraction of total particulate matter that is elemental carbon of less than approximately 1 μm in diameter. This fraction is highly sensitive to the fuel type and combustion conditions (Wehner et al., 1999), necessitating a detailed treatment of emission factors by fuel, sector and the degree of emissions control. Such problems are particularly compounded in China, where no statistics are available on the types of combustor and particulate controls, neither in terms of the prevalence nor the performance of each type of device.

Inverse modeling is a powerful approach to observation-based inferences about atmospheric model inputs (e.g., emissions). Hakami et al. (2005) developed a 4DVar inverse modeling method for the recovery of BC emissions, and sizable improvements were found at sub-regional levels. However, the domain-wide emissions inventory did not change significantly because measurements at only four observation sites were used to inverse and assimilate. Employing a multiple linear regression model, Fu et al. (2012) derived a *top-down* emissions estimate of annual BC and organic carbon (OC). They discovered that, when emissions of BC increased from 1.92 to 3.05 Tg yr^{-1} , the average model simulation of annual mean BC increased from 1.1 to 1.9 $\mu\text{g m}^{-3}$, which showed better agreement with the observed amount (2.5 $\mu\text{g m}^{-3}$).

The ensemble Kalman filter (EnKF), introduced by Evensen (1994), a technology based on ensemble forecasting and Kalman filter theory, has been successfully employed in atmospheric chemistry analyses, such as dust storm and aerosol data assimilation (Lin et al., 2008; Sekiyama et al., 2010; Tang et al., 2013). EnKF has some advantages over 4DVar insofar as it does not require the reconstruction of an adjoint model, which is technically difficult and cumbersome for the complex chemical transport model. However, the algorithm is highly sensitive to ensemble size (Mitchell and Houtekamer, 2002), and therefore tends to be computationally demanding and has limited use in large-scale and on-line atmospheric chemical transport models. Moreover, the EnKF method assumes that the probability density functions (PDFs) of the initial conditions, emissions and observations are Gaussian in their distributions. When there is large bias, problems such as filter divergence will lead to analysis failure.

In this paper, an ensemble optimal interpolation (EnOI) data assimilation method is used to investigate the possibility of optimally correcting the spatially resolved emissions bias of BC. The background-error covariances are estimated using the ensemble, but the model only needs a single forecast, allowing for the use of a larger ensemble than EnKF. The preliminary results for the inverse emissions for BC are

presented in this paper. The details of the methodology are described in Sect. 2, followed by a description of the model and the observations used in Sect. 3. The inverse modeling results are presented in Sect. 4, and a summary and discussion is provided in Sect. 5.

2 Methodology

2.1 Inverse theory and formulation

Air quality models can be generally written as

$$\mathbf{C}(x, t) = \mathbf{M}\mathbf{C}(x, t-1) + \mathbf{Q}(x, t) + q(x, t), \quad (1)$$

where \mathbf{C} is the vector of pollutant concentrations, x is the spatial location, \mathbf{Q} represents emissions, \mathbf{M} is the model time forward operator (could be nonlinear) and q is the model error, a random variable.

The observations are assumed to be also available at time t :

$$y(t) = \mathbf{H}[\mathbf{C}(x, t)] + r, \quad (2)$$

where r represents observation errors. The operator \mathbf{H} in Eq. (2) is the projection from the whole model domain to observation locations.

Without observations, we can carry out a simulation with a given initial concentration, emission inventory, and ignore model error to solve Eq. (1) and to obtain a numerical solution $\mathbf{C}^b(x, t)$. The estimation problem discussed here is that, given observations (Eq. 2), can we determine a better estimation of $\mathbf{C}(x, t)$ and $\mathbf{Q}(x, t)$ than $\mathbf{C}^b(x, t)$ and $\mathbf{Q}^b(x, t)$, where the superscript b represents the background or the first guess. Generally, this problem can be approached from the perspective of how to find the PDFs of the emissions $\mathbf{Q}(x, t)$ and associated model (1) solution $\mathbf{C}(x, t)$, given observations $y(t)$.

Zhu and Wang (2006) introduced the formulation of this estimation theory given the PDFs of the initial condition, emissions and observations are Gaussian distributed. The inverse method seeks an optimal estimate of the emission that is consistent with both the observation and prior constraints of source by minimizing the following cost function:

$$J(\mathbf{C}, \mathbf{Q}) = \iint \begin{bmatrix} \mathbf{C}(x) - \mathbf{C}^b(x) \\ \mathbf{Q}(x) - \mathbf{Q}^b(x) \end{bmatrix}^T \begin{pmatrix} \mathbf{W}_C(x, x') & \mathbf{W}_{CQ}(x, x') \\ \mathbf{W}_{CQ}(x, x') & \mathbf{W}_Q(x, x') \end{pmatrix} \begin{bmatrix} \mathbf{C}(x') - \mathbf{C}^b(x') \\ \mathbf{Q}(x') - \mathbf{Q}^b(x') \end{bmatrix} dx dx' + [\mathbf{y} - \mathbf{H}\mathbf{C}^b]^T \mathbf{R} [\mathbf{y} - \mathbf{H}\mathbf{C}^b], \quad (3)$$

where \mathbf{W}_C is the error covariance matrix of concentration between different grids, \mathbf{W}_{CQ} is the error covariance matrix of concentration and emission and \mathbf{R} is the error covariance matrix of observation.

Based on the Kalman filter, we have the following analysis equations to minimizing the cost function:

$$\begin{pmatrix} C^a \\ Q^a \end{pmatrix} = \begin{pmatrix} C^b \\ Q^b \end{pmatrix} + \mathbf{K}(y - \mathbf{H}C^b), \quad (4)$$

$$\mathbf{K} = \begin{pmatrix} \mathbf{W}_C(x, x') & \mathbf{W}_{CQ}^T(x, x') \\ \mathbf{W}_{CQ}(x, x') & \mathbf{W}_C(x, x') \end{pmatrix} \mathbf{H}^T \left(\mathbf{H} \begin{pmatrix} \mathbf{W}_C(x, x') & \mathbf{W}_{CQ}^T(x, x') \\ \mathbf{W}_{CQ}(x, x') & \mathbf{W}_C(x, x') \end{pmatrix} \mathbf{H}^T + \mathbf{R} \right)^{-1}.$$

2.2 Ensemble approach

To solve Eq. (4), we can use the ensemble method. First, we generate the N ensemble of emission, and then integrate model (1) N times from $t - 1$ to t . The resulting N model outputs C_i^b and N emission ensembles Q_i^b form a joint state emissions ensemble,

$$\mathbf{A} = \begin{pmatrix} C_1^b & C_2^b & \dots & C_N^b \\ Q_1^b & Q_2^b & \dots & Q_N^b \end{pmatrix}. \quad (5)$$

This joint ensemble can be used to construct the joint error covariance

$$\mathbf{P} = \frac{1}{N-1} \mathbf{A}' \mathbf{A}'^T, \quad (6)$$

where \mathbf{A}' denotes the ensemble anomaly referenced to the ensemble mean.

The analysis includes updating each C_i^a and Q_i^a in ensemble form as follows:

$$\begin{pmatrix} C_i^a \\ Q_i^a \end{pmatrix} = \begin{pmatrix} C_i^b \\ Q_i^b \end{pmatrix} + \mathbf{K}(y_i - \mathbf{H}C_i^b) \quad (7)$$

$$\mathbf{K} = \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R}),$$

where y_i , $j = 1, \dots, N$ are the ensembles of perturbed observation samples.

Then, the estimated emission is

$$\hat{Q}^a = \frac{1}{N} \sum_{i=1}^N Q_i^a. \quad (8)$$

2.3 Ensemble optimal interpolation

The EnOI analysis is computed by solving an equation similar to Eq. (4), written as

$$\begin{pmatrix} C^a \\ Q^a \end{pmatrix} = \begin{pmatrix} C^b \\ Q^b \end{pmatrix} + \mathbf{K}(y - \mathbf{H}C^b), \quad (9)$$

where the analysis is computed in the stationary ensemble of model states and emissions sampled during a long-term integration. Once the ensemble is created, only one single model state and emission are updated. The method greatly reduces the computation cost and provides a sub-optimal solution compared to the EnKF approach.

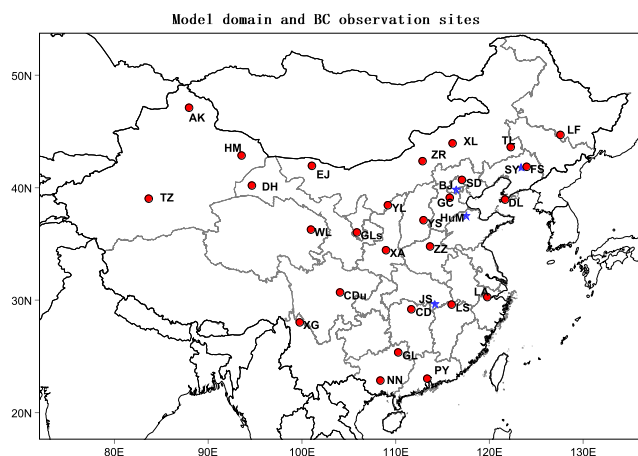


Figure 1. The GRAPES–CUACE model domain and BC observation sites used in this study. Observations considered for assimilation are shown as red points, while observations considered for verification are shown as blue stars. Information about stations is shown in Table 1.

3 Model description and BC measurements

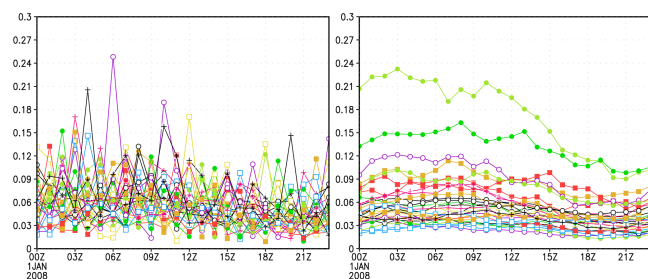
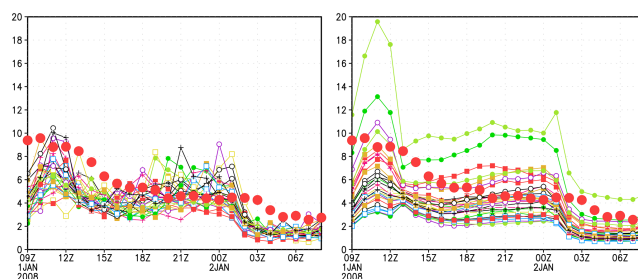
The model used in this study is an online-coupled chemical weather forecasting system, Global/Regional Assimilation and Prediction System–Chinese Unified Atmospheric Chemistry Environment model (GRAPES–CUACE), which consists of two components: (1) GRAPES is a mesoscale meteorological model developed by the China Meteorological Administration (CMA). It produces meteorological fields (winds, turbulence, precipitation, etc.) to drive the CUACE chemistry model. (2) CUACE includes emissions, transport, dry and wet depositions, and removal both in and below clouds of both gases and aerosols (Gong and Zhang, 2008; Wang et al., 2009; Zhou et al., 2012). These two parts are coupled online. Using the national official basic information of emission sources published in 2005 based on the bottom-up inventory developed by Cao et al. (2006), the emission subsystem (EMIS) provides hourly gridded offline emission intensities of 32 species, including seven categories of aerosol species (sea salt, sand/dust, BC, OC, sulfates, nitrates and ammonium salts) in the aerosol modules over the model domain. The basic gridded BC emissions inventory is based on energy consumption and activity information for various emission sectors: industry, residential, transport, power generation, agriculture, biomass burning and others.

The model domain for this study is approximately (70–140°, 17–57° E), as shown in Fig. 1. The computational domain consists of $140 \times 80 \times 32$ grid cells, with a horizontal resolution of 0.5° , and 31 vertical layers from the ground to 200 hPa.

The near-real-time (NRT) data used in this work are the surface daily and hourly BC concentrations collected from over 30 CMA Atmosphere Watch Network (CAWNET) sta-

Table 1. Observation site information.

No.	Observation sites	Long	Lat	Alt	Description
1	AKeDaLa (AK)	87.97	47.12	562	background site
2	TaZhong (TZ)	83.67	39	1099.3	rural site
3	HaMi (HM)	93.52	42.82	737.2	rural site
4	EJiNaQi (EJ)	101.07	41.95	940.5	urban site
5	DunHuang (DH)	94.68	40.15	1140	rural site
6	WaLiGan (WL)	100.92	36.28	3816	background site
7	ZhuRiHe (ZR)	112.9	42.4	1151.9	rural site
8	YuLin (YL)	109.2	38.43	1105	urban site
9	YuShe (YS)	112.98	37.07	1041.4	urban site
10	LongFengShan (LF)	127.6	44.73	330.5	rural site
11	XiLinHaoTe (XL)	116.12	43.95	1003	rural site
12	TongLiao (TL)	122.27	43.6	178.7	urban site
13	FuShun (FS)	123.95	41.88	163	urban site
14	GuCheng (GC)	115.8	39.13	11	urban site
15	DaLian (DL)	121.63	38.9	91.5	urban site
16	ChengDu (CDu)	104.04	30.65	553	urban site
17	ZhuZhang (XG)	99.73	28.02	3580	background site
18	ZhengZhou (ZZ)	113.68	34.78	110	urban site
19	XiAn (XA)	108.97	34.43	410	urban site
20	GuiLin (GL)	110.3	25.32	164.4	rural site
21	LinAN (LA)	119.73	30.3	138.6	rural site
22	LuShan (LS)	115.99	29.57	1165	rural site
23	NanNing (NN)	108.35	22.82	172	urban site
24	PanYu (PY)	113.35	23	131	urban site
25	GaoLanShan (GLs)	105.85	36	2161.5	rural site
26	ChangDe (CD)	111.71	29.17	565	rural site
27	ShangDianZi (SDz)	117.12	40.65	293.3	background site
28	ShenYang(SY)	123.41	41.76	110	urban site
29	Beijing (BJ)	116.47	39.8	31.3	urban site
30	HuiMin (HuM)	117.53	37.48	11.7	urban site
31	JinSha (JS)	114.2	29.63	330.5	rural site

**Figure 2.** Emission ensembles at grid (116.5° N, 40° E): sample strategy with no time correlation (left); sample strategy with time correlation $\alpha = 0.9$ (right) (units: $\text{mg m}^{-2} \text{s}^{-1}$).**Figure 3.** The ensembles of BC model simulation at grid 116.5° N, 40° E. The left one uses sample strategy with no time correlation. The right one uses sample strategy with time correlation $\alpha = 0.9$. The red points is the daily variation of BC observations (units: $\mu\text{g m}^{-3}$).

tions. The locations of these sites are shown in Fig. 1. Observations considered for assimilation are shown as red circles, and observations considered for verification are shown as blue stars. Information on the stations is presented in Table 1. The rural stations are typically located some 100 km away from local pollutant sources or nearby major cities, and

at moderate height above the area's local elevation. At the urban stations, the sampling heights are 50–100 m higher than the ground level. This enables the production of samples that are representative of the region, rather than the immediate locality.

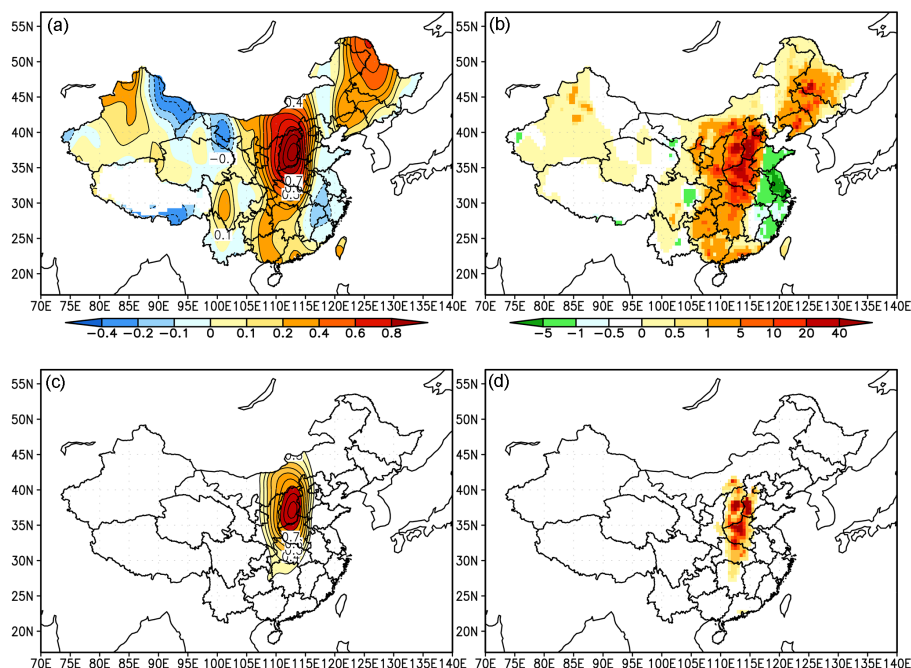


Figure 4. BC emissions inverted by observations from site Zhengzhou (ZZ): (a) back-ground error correlation between site ZZ and other grids without localization; (b) BC emissions inverted by observations from site ZZ without localization; (c) background-error correlation between site ZZ and other grids with localization; (d) BC emissions inverted by observations from site ZZ with localization.

The BC observations are obtained using Aethalometer (Model AE-31, Magee Scientific, Berkeley, California, USA) instruments, which measure optically absorbing filterable aerosol material at a 5 min time interval (Zhang et al., 2008) at seven wavelengths of 370, 470, 520, 590, 660, 880 and 950 nm. The BC concentrations used in this study are derived from the optical absorption at 880 nm.

4 Inverse modeling of BC emissions over China

4.1 Experiment design

We conduct 1-year-long simulation using bottom-up emission for 2008. NCEP $1^\circ \times 1^\circ$ reanalysis data are employed for the model's initial and 6 h meteorological lateral boundary input field and forecast time is every 24 h throughout every month to ensure the model's fair ability to simulation of meteorological field (Wang et al., 2015). The monthly mean of simulated BC is put into the equation as C , and the y also uses the monthly mean of the BC observations. To obtain the ensemble for calculating the K , we make perturbations of the emissions (30 members), put them in the model, and run the model for 24 h using monthly mean meteorological field. Subsequently, we obtain 24×30 ensembles. The reason we use monthly mean data is to reduce the effects of other factors of model error.

4.2 Ensemble strategy

A good ensemble system should satisfy at least two conditions: (1) the ensemble mean should be close to the truth; (2) the ensemble spread should be a reasonable representation of the root mean square error (RMSE) between the ensemble and the truth. Because the atmospheric components are usually log-normal distributed, we produced N ensembles of the emissions according to the formula below:

$$Q'_i(x, t) = Q(x, t) \exp \left(\rho_i(x) \times \log(\sigma) - \frac{\log(\sigma)^2}{2} \right), \quad (10)$$

where $i = 1, 2, \dots, N$, x is the spatial location, $Q(x, t)$ represents bottom-up emission, $Q'_i(x, t)$ is emission ensembles, $\rho_i(x)$ is two-dimensional pseudorandom fields, and σ is standard deviation of the emission. Here, we use the procedure proposed by Evensen (1994) to generate the pseudorandom fields $\rho_i(x)$, which is able to compute them with the mean equating to zero, variance equating to one, and a specified covariance that determines the smoothness of the fields. Using the two-dimensional fast Fourier transform (FFT), we obtain two-dimensional perturbations of the emissions, and then use singular vector decomposition (SVD) to obtain the optimal ensembles. In this paper, we compare two schemes: scheme A, which only considers the spatial correlation as Eq. (10), and scheme B, which considers both the spatial and temporal correlation:

$$\rho_i(x, t) = \alpha \times \rho_i(x, t) + \sqrt{1 - \alpha^2} \times \rho_i(x, t), \quad (11)$$

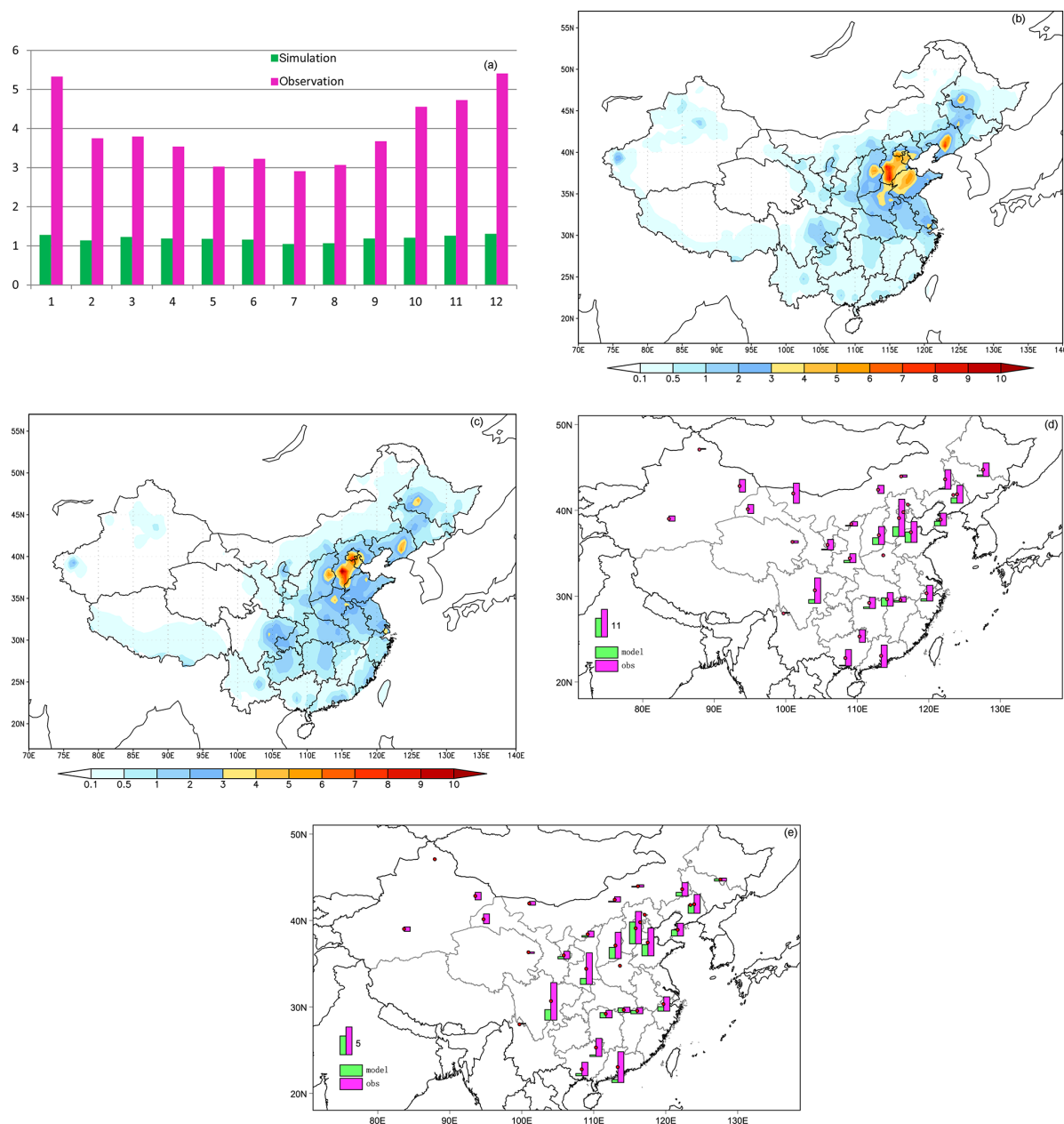


Figure 5. BC monthly mean concentrations (units: $\mu\text{g m}^{-3}$). (a) Seasonality of BC concentration; (b) model simulation with bottom-up emission for January; (c) model simulation with bottom-up emission for July; (d) comparison between model simulation and observation at BC observation sites for January; (e) comparison between model simulation and observation at BC observation sites for July; green bars show the model-simulated BC using the bottom-up emissions inventory; pink bars show the observed surface BC concentrations.

where α is time correlation coefficient. To test whether the sample strategies are good, we compare the ensemble of the emissions and BC concentrations. Figure 2 shows the ensemble of the emissions for the two schemes. Although the two schemes produce the perturbations of the emissions with the same variance, the ensemble of the emissions by scheme B has a much larger spread than scheme A. Figure 3 shows the ensemble of BC concentration simulations computed by the corresponding emissions, with the red dots representing

the observations. It is clear that the ensemble simulation of scheme B can include the observation more reasonably.

4.3 Localization in the inverse modeling

The inverse modeling of the BC emissions in this study uses the ensemble member to calculate the background-error covariance matrix between the concentrations and emissions. The accuracy of the matrix is a very important factor for the

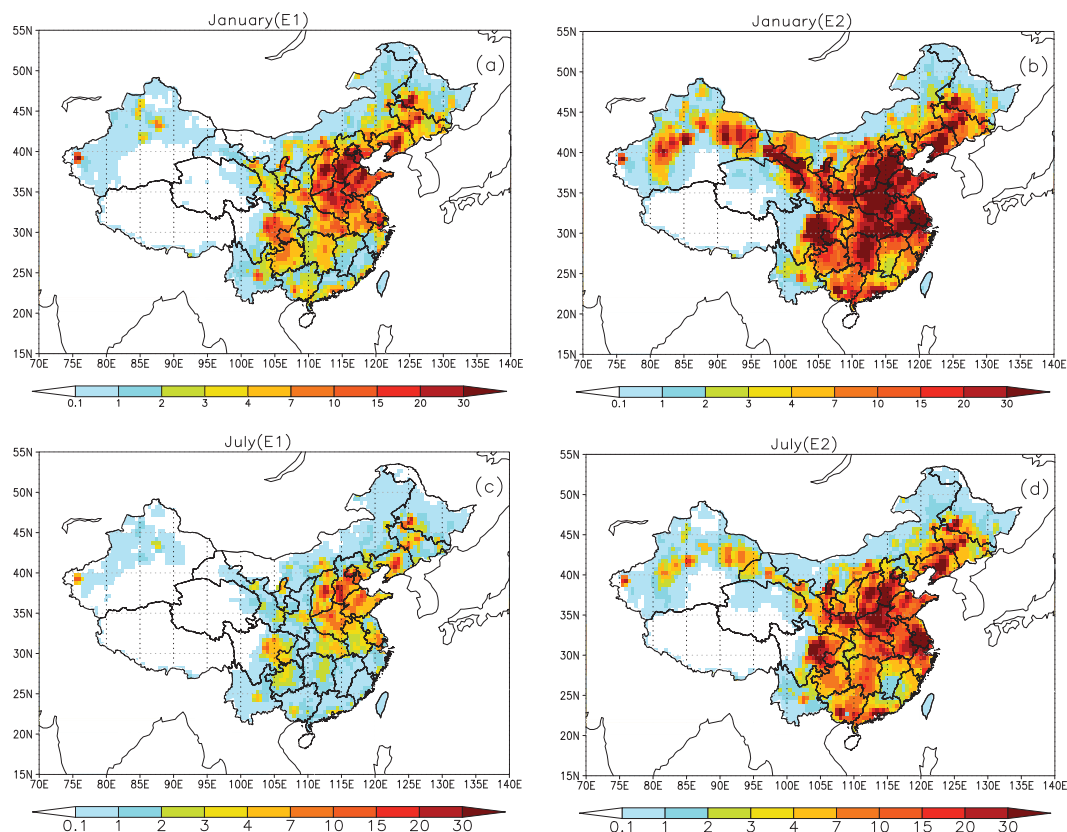


Figure 6. BC emissions from China inverted by 27 observation sites (units: $\mu\text{g m}^{-2} \text{s}^{-1}$): (a) bottom-up emissions E1 for January; (b) inverse emissions E2 for January; (c) bottom-up emissions E1 for July; (d) inverse emissions E2 for July.

inverse result. Since we cannot obtain as many ensembles as model dimensions, we should employ appropriate techniques that eliminate the effects of sample error and that localize the impact of an observation to a subset of the model state variables. Figure 4a shows the correlation of the background-error between the sites concentrations and emissions. The figure shows that the site concentration is highly correlated with the emissions near the site, which is expected. However, there are some emissions from remote regions that are also correlated with the site concentration, i.e., spurious correlations. Localization is an essential tool for an ensemble-based assimilation to adequately span the model sub-space.

Usually, a typical implementation of localization involves the multiplication of the ensemble-based covariance by a correlation function, so the gain matrix is re-expressed as Eq. (12). We use the distance-dependent covariance localization scheme,

$$\mathbf{K} = (\mathbf{C} \circ \mathbf{P})\mathbf{H}^T [\mathbf{H}(\mathbf{C} \circ \mathbf{P})\mathbf{H}^T + \mathbf{R}]^{-1}, \quad (12)$$

where \mathbf{C} is a correlation matrix and the operation denoted by the open circle is an element-by-element matrix multiplication (also called a Schur product).

Here, we use an elliptic function with e -folding scales of L_a (major axis) and L_b (minor axis). When the distance–

dependent covariance localization scheme is used, these covariances are artificially reduced to near zero with e -folding length scales of L_a and L_b . The localization scales are set by calculating the shape of the covariance matrix.

Figure 4c shows the correlation of the background error between the site concentrations and emissions. Using this localization scheme, the spurious correlations far away from the site concentration are reduced, but the pattern of correlation around the site remains. Therefore, the inversion provides a more accurate estimation of emissions.

4.4 Results of BC emissions bias correction

An 1-year-long simulation is conducted, driven by the current Chinese bottom-up emissions inventories for BC, and the results are compared with surface-observed BC concentrations (Fig. 5). The comparison of model results to observations evaluates the bottom-up inventories. Figure 5a compares the seasonal variation of observed and simulated surface BC. The simulated concentrations are significantly lower than the observed throughout the year. This indicates a region-wide underestimation in monthly and annual bottom-up emission inventory. The average simulated annual mean BC concentration is $1.19 \mu\text{g m}^{-3}$, much lower than the ob-

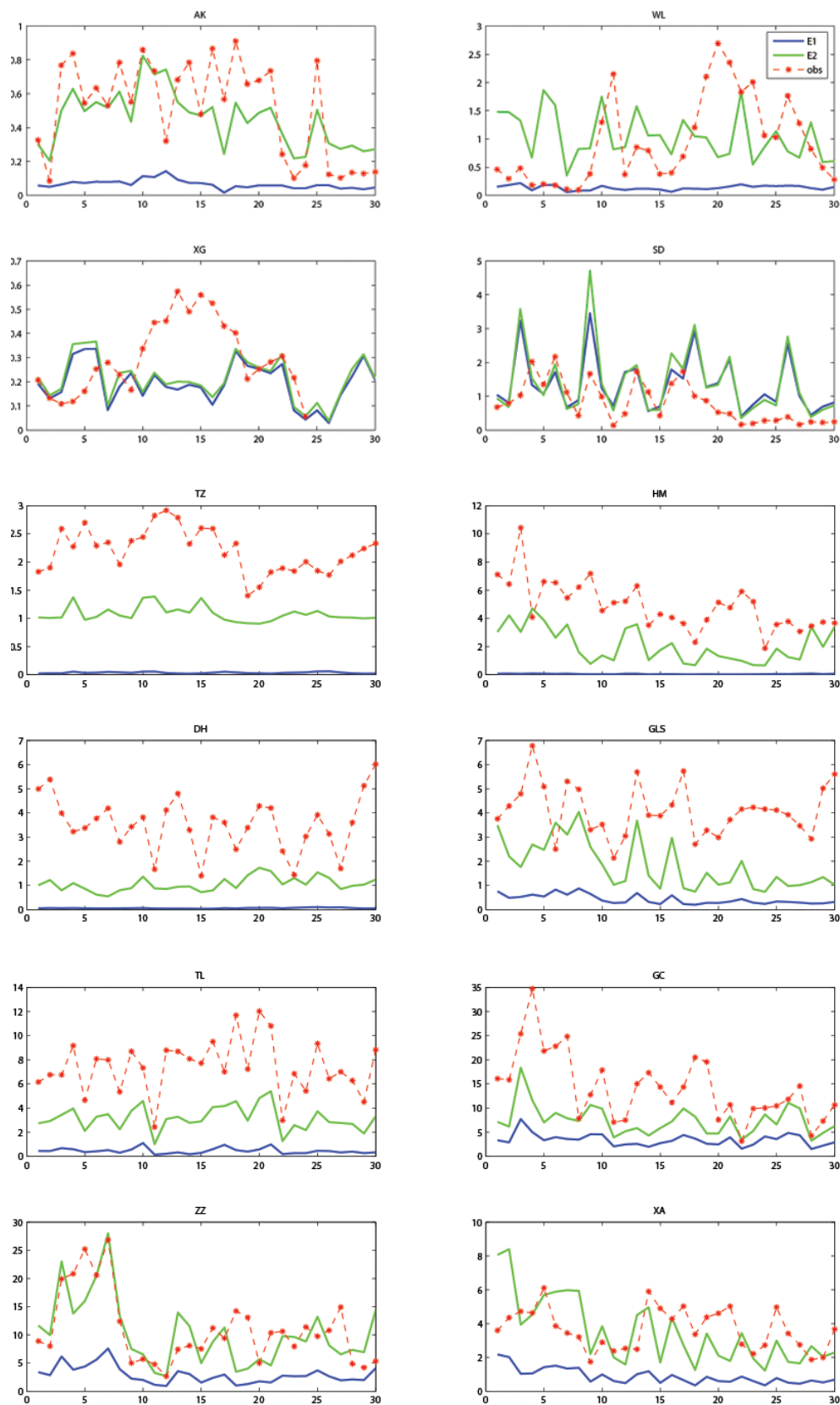


Figure 7. BC daily concentrations in January 2008 (units: $\mu\text{g m}^{-3}$). The red dotted line shows the observation; the blue line is the model simulation driving by prior emissions (E1); and the green line is the model simulation driven by inverse emissions (E2).

Table 2. Estimates of provincial BC emission (Gg) by bottom-up and inverse method.

Province	January			July		
	E1	E2	E2/E1	E1	E2	E2/E1
Hebei	12.1	22.3	1.8	9.5	14.2	1.5
Shanxi	11.0	17.8	1.6	7.8	11.8	1.5
ShangDong	10.3	13.6	1.3	8.2	11.5	1.4
Henan	9.9	19.3	1.9	6.9	10.7	1.5
Sichuan	6.4	16.4	2.6	4.5	13.6	3.0
Liaoning	6.3	13.1	2.1	4.6	10.2	2.2
Nei Mongol	6.2	11.9	1.9	4.3	7.5	1.7
Jiangsu	5.5	13.2	2.4	4.3	8.4	2.0
Hubei	5.3	11.3	2.1	4.7	7.1	1.5
Anhui	4.8	8.2	1.7	4.4	7.5	1.7
Heilongjiang	4.5	6.5	1.4	3.8	4.5	1.2
Hunan	4.1	10.8	2.6	3.7	5.4	1.4
Guizhou	4.1	5.0	1.2	3.4	4.9	1.4
Yunnan	3.5	4.0	1.1	2.9	3.2	1.1
Guangxi	3.2	6.4	2.0	2.7	5.6	2.1
Shannxi	3.1	7.7	2.5	2.2	5.7	2.6
Jilin	3.9	7.8	2.0	3.1	5.3	1.7
Zhejiang	1.8	4.7	2.6	1.6	3.2	2.0
Fujian	1.4	2.3	1.6	1.0	1.9	1.9
Jiangxi	1.9	3.7	1.9	1.9	3.2	1.7
Guangdong	2.7	5.0	1.9	2.2	4.4	2.0
hainan	0.2	0.2	1.0	0.1	0.1	1.0
Xizang	0.3	0.3	1.0	0.2	0.2	1.0
Gansu	3.1	8.1	2.6	2.8	5.8	2.1
Qinhai	0.5	0.9	1.8	0.5	0.8	1.7
Ningxia	1.1	2.8	2.5	0.7	1.6	2.2
Xinjiang	3.2	9.4	2.9	2.4	6.1	2.6
Chongqin	1.8	2.9	1.6	1.3	2.4	1.9
Beijing	2.8	2.6	0.9	1.9	1.5	0.8
Tianjin	1.1	1.3	1.2	0.8	1.0	1.2
Shanghai	0.3	0.7	2.3	0.2	0.4	1.7
China	126.4	240.1	1.9	98.5	169.7	1.7

served $3.92 \mu\text{g m}^{-3}$. BC observations are higher in Winter than summer, suggesting strong emission-associated heating. Figure 5b–e compared the spatial distribution of monthly mean observed and simulated BC concentrations for January and July. The model shows higher BC concentrations in January than in July, which is similar to the observation. The simulated BC concentration had much higher values in the east than in the west with the highest concentration over northern China, corresponding to the strong emission there. However, the model simulations not only underestimate BC concentration at urban sites but also significantly underestimate at rural sites such as TaZhong, Hami, Dunhong, Gaolanshan in northwest China where the bottom-up emissions have very few emissions both for January and July. Figure 7 shows the daily BC concentrations variation in January at background, rural and urban sites. At background sites AkeDala and WLG, the model still produces a relatively low concentration indicating underestimation of BC

emissions in northwest China. Little bias in simulations at the background site ZhuZhang shows that emissions around that region in Yunnan province are relatively accurate. The simulated BC concentration at the background site SDz is a little higher than the observation. That is because SDz is located in northwest China, which is a densely populated and industrialized area, and the emission rate around the SDz is relatively high. Besides these four background sites, the model performance of daily BC concentration based on bottom-up inventory at rural and urban sites is very poor, largely because of the underestimation in emission. This suggests that the bottom-up emission is very low and misrepresented in space and time.

We use the 27 BC monthly observations to inverse the emissions. Figure 6 shows the bottom-up or prior emissions (E1) and inverse emissions (E2) for January and July. There are significant increases in the E2 emissions over most regions of China, including eastern China, central China,

Table 3. Model simulations and surface observations of monthly mean BC concentrations at assimilation sites and verification sites (units: $\mu\text{g m}^{-3}$) and the relative error percentage (i.e., $(\text{model} - \text{obs})/\text{obs} \times 100\%$).

Site	Model (E1)	Model (E2)	Observation	Observation SD	Relative error percentage (E1)	Relative error percentage (E2)
AK	0.07	0.44	0.51	0.29	86.9	13.0
TZ	0.04	1.08	2.20	0.38	98.4	51.1
HM	0.06	2.21	4.90	1.74	98.9	54.9
EJ	0.05	2.67	7.84	2.85	99.4	66.0
DH	0.06	1.02	3.55	1.14	98.4	71.4
WL	0.13	1.03	0.94	0.76	85.7	9.3
ZR	0.14	1.00	3.37	1.01	95.7	70.2
YL	0.31	0.89	1.88	2.98	83.6	52.9
YS	2.70	5.56	6.94	2.75	61.1	19.9
LF	0.58	2.23	5.16	2.49	88.8	56.8
XL	0.14	0.37	0.93	0.52	84.7	59.8
TL	0.47	2.97	7.42	2.24	93.6	59.9
FS	2.00	4.82	7.06	1.73	71.7	31.6
GC	3.79	7.60	14.24	7.02	73.4	46.7
DL	1.74	4.13	4.85	2.44	64.1	14.8
CDu	1.45	7.14	9.71	8.87	85.0	26.5
XG	0.20	0.21	0.30	3.79	34.8	29.2
ZZ	3.28	10.68	10.89	6.27	69.9	2.0
XA	1.02	3.57	3.66	1.20	72.1	2.5
GLs	0.26	1.92	4.99	6.20	94.8	61.5
LA	1.00	4.19	6.19	5.57	83.8	32.3
LS	0.73	2.69	2.08	1.35	65.0	29.4
NN	0.55	2.26	6.36	4.35	91.3	64.5
PY	0.55	2.15	8.69	6.21	93.6	75.2
GL	0.46	1.82	4.11	1.08	88.8	55.8
CD	0.71	3.02	4.38	3.69	83.7	31.0
SDz	1.39	1.38	0.81	0.61	71.8	70.5
BJ	3.45	8.68	11.96	5.26	71.2	27.4
HuM	4.27	6.41	8.06	7.30	47.0	20.4
JS	3.20	4.47	5.35	3.74	40.1	16.4
SY	0.89	3.14	3.05	2.29	70.7	2.8
Assi_sites mean	0.88	2.93	4.96	2.95	82.2	42.9
Veri_sites mean	2.95	5.67	7.10	4.65	57.2	16.8
All_sites mean	1.15	3.28	5.24	3.17	79.0	39.5

Sichuan basin and western China. Only the regions around Beijing and the northeast part of Hebei province present a bit of a decrease in E2, because model simulation at the assimilated site SDz is lower than the observation. The results show that the basic emissions produced by the bottom-up method are underestimated and have been corrected by EnOI in most regions of China in January, not only in eastern China and central China where the rural population density and economic level are high, but also in northwestern China, which has a lower rural population densities and lower economic level.

Table 2 compares our inverse provincial and national emission of BC with bottom-up inventories. The five largest contributions by province in E1 for January are from Hebei,

Shangdong, Henan, Shanxi and Sichuan. In E2 by the inversion, the five largest BC emission province are same with E1, but emissions are greatly enhanced in many provinces. We find emission from half of the provinces in China to be enhanced by a factor of over 2 for January. The resulting estimate emission over China in January and July is 240.1 and 169.5 Gg, respectively. Figure 9 shows the seasonality of BC emission in China. The emissions in every month were enhanced after the inversion. The annual emission of bottom-up inventory is about 1449.6 Gg, and inverse inventory is about 2539.3 Gg.

While inverse modeling can provide a simplified solution, the processes contributing to the model bias and error go beyond emission. Therefore, emission inversion is likely to

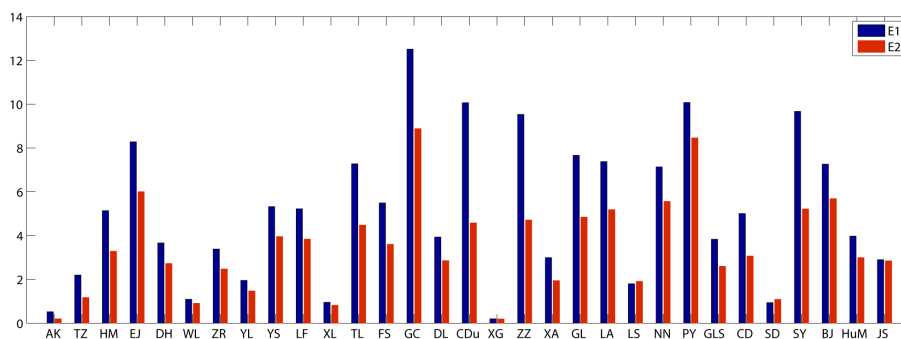


Figure 8. The root mean square error (RMSE) between the daily model simulation and observation. The blue bars show the RMSE between the daily model simulation driven by prior emissions (E1) and observations, and the red bars show the RMSE between the daily model simulation driven by inverse emissions (E2) and observations (units: $\mu\text{g m}^{-3}$).

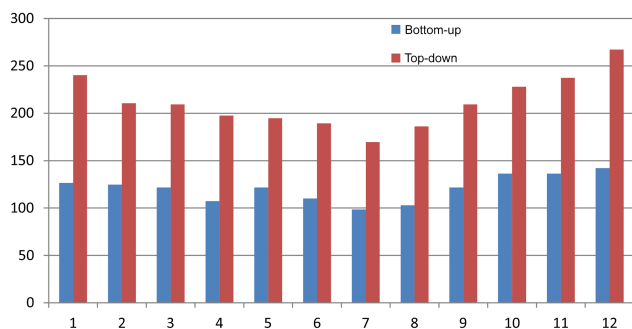


Figure 9. Seasonality of BC emission in China (units: Gg).

lump up uncertainties from other processes into emission. Although we had used monthly mean concentration to eliminate the effect, it is still important to evaluate the inverse emission estimation and uncertainty of the result. Table 3 lists the monthly mean BC concentrations at the observation sites. It shows that there are large errors between the monthly BC concentration simulations driven by the bottom-up emissions (E1) and observations. Most of them feature a negative bias. After EnOI inversion, the model simulation for most observation sites in China is much closer to the observation, even for the verification sites. The error percentage decreases from 78.97 to 39.54 %, which is an almost 50 % decrease.

Figure 7 shows the variations in daily BC concentrations in January 2008. The red line shows the observations, the blue line is the model simulation driven by prior emissions (E1), and the green line is the model simulation driven by inverse emissions (E2). Even though we only employ the monthly mean BC measurements to inverse the emissions, the accuracy of the daily model simulation is also improved. Take the Zhengzhou site (ZZ) for example, driven by inverse emission inventory, the daily simulation concentration and its variation are more consistent with observation. Observations in the ZZ site exhibit a peak during 2–8 January the simulation with E1 does not present, whereas the simulation with E2 does. We calculate the RMSE between the daily model simulation and

observation (Fig. 8). The blue bar is the RMSE between the daily model simulation driven by prior emissions (E1) and the observation, and the red bar is the RMSE between the daily model simulation driven by the inverse emissions (E2) and the observation. As we can see, almost all of the RMSEs decrease, with the average RMSE dropping from 5.08 to 3.47, which is a decrease of about 31.56 %. Because there are large region-wide underestimations in the bottom-up emission, not only in the densely populated and industrialized areas such as northern China, the Yangtze River Delta and the Sichuan basin, but also in northwest China, which has a lower population densities and lower economic level, the model performance of daily BC concentration is very poor. The simulation at rural and urban sites are significantly lower than the observations. With inversion by EnOI, the emission negative bias has been corrected, the simulated concentrations increase and improve. However, there are still large difference between the daily observations and simulation, because there are some other sources of uncertainty such as meteorology and other factors of model error. We have used monthly mean data in the inversion process to reduce these effects, but when it comes to hourly and daily simulation, these effects should be considered reasonably, which is part of the future work we have planned.

From the figures and tables above, it is apparent that the biases of the BC emission intensities in China at each grid point are corrected by the EnOI inversion system. Where the bias is large, the RMSE decreases significantly. However, if there is small bias, such as in the ZhuZhang site, the correction is tiny, and the RMSE changes very slightly. We also find that, near Beijing, emissions around the SDz site are overestimated because the observation in that site is lower than the simulation. However, after the inversion, the BC concentration at the verification site Beijing still increases, the reason is the emission around Beijing is underestimated, and has been corrected. The accuracy of the SDz site changes negligibly. We believe that the reason is that the horizontal grid resolution of 50 km is not high enough to distinguish between two very close observation sites with different variation. This

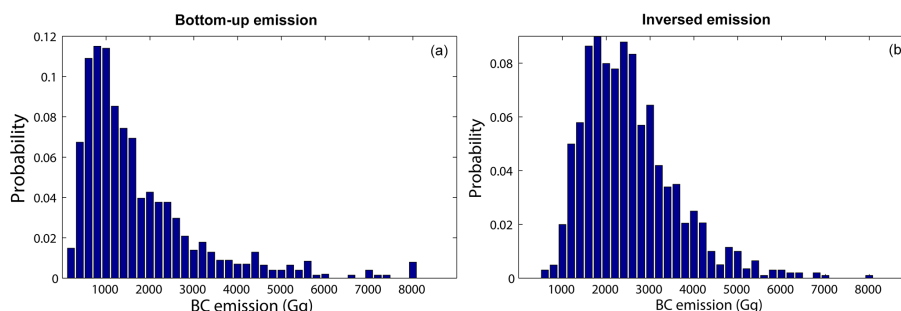


Figure 10. Uncertainty analysis for annual Chinese BC bottom-up and inverse emission inventory.

problem could be solved if we use a model and emissions inventory with higher resolution.

We also conduct the Monte Carlo simulation to quantify the uncertainty of the total bottom-up emission and the inverse emission inventory in China. The lognormal distribution is assumed, and the standard deviation is calculated by combining the root mean square error between observation and simulation with standard deviation of the inventory. Monte Carlo simulations with randomly selected values within the PDFs are repeatedly implemented 10 000 times. The uncertainty in Chinese BC bottom-up emission and inversed emission inventory at 95 % are obtained, as shown in Fig. 10. The mean value, 2.5th percentile value, and 97.5th percentile value are 1570, 321, and 5138 Gg (bottom-up) and 2650, 1114, 5471 Gg (inverse emission), respectively. Therefore, the uncertainty of these two emission inventory are about $[-80, 227 \%$], and $[-58, 102 \%$], correspondingly. Using the ensemble inversion modeling, the uncertainty of BC emission inventory decreased significantly. We also compare our estimation with results from previous study.

Streets et al. (2003a) estimated the 1.05 Tg BC emission in China for the year 2000 with $\pm 360 \%$ uncertainty measured as 95 % confidence intervals. Zhang et al. (2009) estimation of the China BC emission is 1.61 Tg. Qin and Xie (2012) estimated the 1.57 Tg BC emission in China for the year 2005 with $[-51, 148 \%$] uncertainty. Our estimations are nearly 50–100 % higher than these bottom-up inventories. One reason is there were very few emissions for northwest China and northeast China in all of these emission inventory. They are so similarly low in these regions probably because these bottom-up inventories are based on the same statistical data source. Based on the top-down regression method, Fu et al. (2012) estimated that the annual BC emission is 3.05 ± 0.78 Tg, which is higher than our estimation. One possible reason is that their estimation was produced by applying domain-wide scaling factors and may be biased high in central China, which had been pointed out in their paper.

5 Summary and discussion

An inverse modeling system is developed for BC emissions in an online-coupled chemical weather forecasting system, GRAPES-CUACE, using the inexpensive EnOI methodology. The emissions sampling strategy is discussed and improved. With its time correlation strategy, the ensemble forecast can have a larger spread to include the observations. The effect of localization in the analysis is also studied. With reasonable localization, the effects of sample error and spurious correlations are reduced.

BC aerosols in China are simulated and compared to surface measurements, with the goal of deriving top-down BC emissions bias estimates. We conduct 1-year-long simulation for 2008, driven by the current Chinese bottom-up BC emissions inventories. Comparison of the model results to observations at background and rural sites evaluates the bottom-up inventories. The simulated average monthly mean BC concentration in January for all rural and background sites is $1.152 \mu\text{g m}^{-3}$, which is 78 % lower than the observed amount ($5.238 \mu\text{g m}^{-3}$), indicating the BC source is underestimated on a national scale, likely due to uncertainties in emissions from small-scale industrial activity, domestic combustion, and transportation (Qin and Xie, 2012). There is a missing source in western China, likely associated with the use of biofuels or other low-quality fuels for heating. This indicates that the bottom-up BC emissions are too low and spatiotemporally misrepresented. The monthly mean eliminates the meteorological factors and other uncertainties; therefore, the underestimation comes mainly from the emissions, which explains the large biases in the simulated concentrations. The monthly mean of the simulated BC is put into Eq. (9) as the background forecast, and the monthly mean of the BC observations is considered as the observation. With the inversion of EnOI, the bias of the BC emissions intensity in China at each grid point is corrected by the inverse system. We find emission from provinces in China to be enhanced by a factor of 1.8. The emission over China in January by our estimation is 240.1 Gg, and annual BC emission is about 2539.3 Gg.

Applying top-down emissions estimates, the simulated average annual mean concentration at rural and background

sites for January is improved to $3.282 \mu\text{g m}^{-3}$. Despite only employing the monthly mean BC measurements to inverse the emissions, the accuracy of the daily model simulation also improves. The average RMSE drops from 5.08 to 3.47 (a decrease of approximately 31.56 %), showing that where the bias is large, the RMSE decreases significantly. It is also found that the performance of EnOI may fail under two circumstances: where the bias is small, such as in the ZhuZhang (XG) site; and where the horizontal grid resolution is not high enough to distinguish between two very close observation sites with different variations. This problem could be solved if we use a model and emissions inventory with a higher resolution. Our study finds that EnOI is a useful and computation-free method to correct the bias of aerosol emissions. These results are valuable and promising for a better understanding of aerosol emissions and distributions, as well as aerosol forecasting. We plan to explore the use of other data, such as satellite-derived aerosol optical depth observations, to evaluate our results. We suggest that EnOI may provide a practical and cost-effective alternative to the EnKF for correction of the aerosol emissions bias where computational cost is a limiting factor. In future work, we intend to employ EnKF and ensemble Kalman smoother to optimally recover temporally resolved (e.g., daily timescale) emissions inventories of BC and other aerosols.

Acknowledgements. This work is supported by the National Basic Research Program (973) (2011CB403404), the National Natural Scientific Foundation of China (41305117), Basic Research Fund of CAMS (2013Y005) and the National Natural Scientific Foundation of China (41205081).

Edited by: J. Brandt

References

- Bond, T. C. and Bergstrom, R. W.: Light absorption by carbonaceous particles: An investigative review, *Aerosol Sci. Technol.*, 40, 27–67, doi:10.1080/02786820500421521, 2006.
- Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, *J. Geophys. Res.*, 118, 5380–5552, 2013.
- Cao, G., Zhang, X., and Zheng, F.: Inventory of black carbon and organic carbon emissions from China, *Atmos. Environ.*, 40, 6516–6527, doi:10.1016/j.atmosenv.2006.05.070, 2006.
- 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.
- Fu, T.-M., Cao, J. J., Zhang, X. Y., Lee, S. C., Zhang, Q., Han, Y. M., Qu, W. J., Han, Z., Zhang, R., Wang, Y. X., Chen, D., and Henze, D. K.: Carbonaceous aerosols in China: top-down constraints on primary sources and estimation of secondary contribution, *Atmos. Chem. Phys.*, 12, 2725–2746, doi:10.5194/acp-12-2725-2012, 2012.
- Gong, S. L. and Zhang, X. Y.: CUACE/Dust – an integrated system of observation and modeling systems for operational dust forecasting in Asia, *Atmos. Chem. Phys.*, 8, 2333–2340, doi:10.5194/acp-8-2333-2008, 2008.
- Hakami, A., Henze, D. K., Seinfeld, J. H., Chai, T., Tang, Y., Carmichael, G. R., and Sandu, A.: Adjoint inverse modeling of black carbon during the Asian Pacific Regional Aerosol Characterization Experiment, *J. Geophys. Res.*, 110, D14301, doi:10.1029/2004JD005671, 2005.
- Hansen, A. D. A., Bodhaine, B. A., Dutton, E. G., and Schnell, R. C.: Aerosol black carbon measurements at the South Pole: initial results, 1986–1987, *Geophys. Res. Lett.*, 15, 1193–1196, 1988.
- Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M., and Zhang, S.: Efficacy of climate forcings, *J. Geophys. Res.*, 110, D18104, doi:10.1029/2005JD005776, 2005.
- Hara, K., Osada, K., Yabuki, M., Hayashi, M., Yamanouchi, T., Shiobara, M., and Wada, M.: Measurement of black carbon at Syowa station, Antarctica: seasonal variation, transport processes and pathways, *Atmos. Chem. Phys. Discuss.*, 8, 9883–9929, doi:10.5194/acpd-8-9883-2008, 2008.
- Lary, D. J., Lee, A. M., Toumi, R., Newchurch, M. J., Pirre, M., and Renard, J. B.: Carbon aerosols and atmospheric photochemistry, *J. Geophys. Res.*, 102, 3671–3682, 1997.
- Lin, C., Wang, Z., and Zhu, J.: An Ensemble Kalman Filter for severe dust storm data assimilation over China, *Atmos. Chem. Phys.*, 8, 2975–2983, doi:10.5194/acp-8-2975-2008, 2008.
- Lioussé, C., Cachier, H., and Jennings, S. G.: Optical and thermal measurements of black carbon aerosol content in different environments: Variation of the specific attenuation cross-section, *sigma*, *Atmos. Environ.*, 27A, 1203–1211, 1993.
- Menon, S., Hansen, J., Nazarenko, L., and Luo, Y.: Climate effects of black carbon aerosols in China and India, *Science*, 297, 2250–2253, 2002.
- Mitchell, H. L. and Houtekamer, P. L.: Ensemble size balance and model-error representation in an ensemble Kalman filter, *Mon. Weather Rev.* 130, 2791–2808, 2002.
- Qin, Y. and Xie, S. D.: Spatial and temporal variation of anthropogenic black carbon emissions in China for the period 1980–2009, *Atmos. Chem. Phys.*, 12, 4825–4841, doi:10.5194/acp-12-4825-2012, 2012.
- Qiu, J. and Yang, L.: Variation characteristics of atmospheric aerosol optical depths and visibility in North China during 1980–1994, *Atmos. Environ.*, 34, 603–609, 2000.
- Ramanathan, V. and Carmichael, G.: Global and regional climate changes due to black carbon, *Nat. Geosci.*, 1, 221–227, 2008.

- Sekiyama, T. T., Tanaka, T. Y., Shimizu, A., and Miyoshi, T.: Data assimilation of CALIPSO aerosol observations, *Atmos. Chem. Phys.*, 10, 39–49, doi:10.5194/acp-10-39-2010, 2010.
- Streets, D. G., Bond, T. C., Carmichael, G. R., Fernandes, S. D., Fu, Q., He, D., Klimont, Z., Nelson, S. M., Tsai, N. Y., Wang, M. Q., Woo, J.-H., and Yarber, K. F.: An inventory of gaseous and primary aerosol emissions in Asia in the year 2000, *J. Geophys. Res.*, 108, 8809, doi:10.1029/2002JD003093, 2003a.
- Streets, D. G., Yarber, K. F., Woo, J.-H., and Carmichael, G. R.: Biomass burning in Asia: Annual and seasonal estimates and atmospheric emissions, *Global Biogeochem. Cy.*, 17, 1099, doi:10.1029/2003GB002040, 2003b.
- Tang, X., Zhu, J., Wang, Z. F., Wang, M., Gbaguidi, A., Li, J., Shao, M., Tang, G. Q., and Ji, D. S.: Inversion of CO emissions over Beijing and its surrounding areas with ensemble Kalman filter, *Atmos. Environ.*, 81, 676–686, 2013.
- Wang, H., Gong, S. L., Zhang, H. L., Chen, Y., Shen, X. S., Chen, D. H., Xue, J. S., Shen, Y. F., Wu, X. J., and Jin, Z. Y.: A new-generation sand and dust storm forecasting system GRAPES_CUACE/Dust: Model development, verification and numerical simulation, *Chin. Sci. Bull.*, 55, 635–649, doi:10.1007/s11434-009-0481-z, 2009.
- Wang, H., Xue, M., Zhang, X. Y., Liu, H. L., Zhou, C. H., Tan, S. C., Che, H. Z., Chen, B., and Li, T.: Mesoscale modeling study of the interactions between aerosols and PBL meteorology during a haze episode in Jing-Jin-Ji (China) and its nearby surrounding region – Part 1: Aerosol distributions and meteorological features, *Atmos. Chem. Phys.*, 15, 3257–3275, doi:10.5194/acp-15-3257-2015, 2015.
- Wehner, B., Bond, T. C., Birmili, W., Heintzenberg, J., Wiedensohler, A., and Charlson, R. J.: Climate-relevant particulate emission characteristics of a coal fired heating plant, *Environ. Sci. Technol.*, 33, 3881–3886, 1999.
- Wolff, G. T.: Particulate elemental carbon in the atmosphere, *JAPCA J. Air Waste Ma.*, 31, 935–939, 1981.
- Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, *Atmos. Chem. Phys.*, 9, 5131–5153, doi:10.5194/acp-9-5131-2009, 2009.
- Zhang, X. Y., Wang, Y. Q., Zhang, X. C., Guo, W., Gong, S. L., Zhao, P., and Jin, J. L.: Carbonaceous aerosol composition over various regions of China during 2006, *J. Geophys. Res.*, 113, D14111, doi:10.1029/2007JD009525, 2008.
- Zhou, C. H., Gong, S. L., Zhang, X. Y., Liu, H. L., Xue, M., Cao, G. L., An, X. Q., Chen, H. Z., Zhang, Y. M., and Niu, T.: Towards the improvements of simulating the chemical and optical properties of Chinese aerosols using an on-line coupled model – CUACE/Aero, *Tellus B*, 64, 18965, doi:10.3402/tellusb.v64i0.18965, 2012.
- Zhu, J. and Wang, P.: Ensemble Kalman smoother and Ensemble Kalman filter approaches to the joint air quality state and emission estimation problem, *Chin. J. Atmos. Sci.*, 30, 871–882, 2006.