DEVELOPMENT OF A REGIONAL DATA ASSIMILATION SYSTEM AND ITS APPLICATION IN TWO DISTINCT AREAS OF CHINA FOR ESTIMATING CO SURFACE FLUX

LU, L. J.[1,2*] – WANG, X. B.[3]

[1]College of Resource and Environment, Anhui Science and Technology University
Bengbu 233000, China

[2]School of Environment Science and Spatial Information, China University of Mining and
Technology, Xuzhou 221116, China

[3]State Key Laboratory of Safety and Health for Metal Mines, Maanshan 243000, China

*Corresponding author
e-mail: lijiang_lu@163.comnstitution

Abstract. Traditional methods are time-consuming and labor-intensive for CO flux estimations, there are generally significant uncertainties concerning the results. Data assimilation method has been adopted in recent years for flux optimization in many studies and has proven to be an effective way to improve the accuracy of the CO flux. In this study, a regional data assimilation system, i.e., TracersTracker, was developed based on the POD4DVAR (Proper Orthogonal Decomposition Four-dimensional Variational) method, the CMAQ (Community Multi-scale Air Quality) model and the WRF (Weather Research and Forecasting) model. The system was then applied in two distinct representative areas, i.e., Shangdianzi and Waliguan, for estimating the surface anthropogenic CO flux in 2016. Results show that the CO posteriors in the two study areas were generally higher than the CO priories, but the variations of the posteriors in Shangdianzi and Waliguan are quite different. The overall increase rates of the posteriors are 29.1% and 61.2% in Shangdianzi and Waliguan, respectively. Posteriors optimized by the TracersTracker system significantly improve the accuracy and the correlation of CO simulations in both study areas, which proves that the TracersTracker system is an effective tool for improving the accuracy of the CO emission flux.
Keywords: *carbon monoxide, flux inversion, CMAQ model, proper orthogonal decomposition, four-dimensional variational assimilation*

Introduction

CO is one of the six major atmospheric pollutants, these pollutants are generally generated from human productive and living activities, e.g., cement production, coal combustion. During the past 40 years of urbanization and industrialization, a large amount of CO was emitted in China, which seriously damaged the living environment and caused some social issues (Gurjar et al., 2010; Kamimura et al., 2017; Tao et al., 2019). According to the statistics of the Ministry of Environmental Protection, over 75% of cities in China cannot met the CO concentration standard established by the Ministry of Environmental Protection (MEPC, 2017). Therefore, it is very important to determine the spatial distribution and variation characteristics of anthropogenic CO emissions on the earth surface, which can help us understand the mechanism of CO migration and transformation, and provide effective way for environmental governance. In order to quantify the anthropogenic CO emission flux, many statistical methods have been carried out to obtain surface CO emissions, which is called the "bottom-up"

method (Law et al., 2015). Over the last few decades, researchers have established several atmospheric pollutant emission inventory datasets at different scales in China, from province to national levels (Kilmont, 2002; Ma, 2007; Ohara, 2007; Shi et al., 2014). However, these inventories generallyhave low spatial-temporal resolutions, which increases the difficulty for forecasting and diagnosing air quality (Zheng et al., 2009). Besides, "bottom-up" CO fluxes generally are time-consuming and labor-intensive (Li et al., 2007; Lu et al., 2019), and the obtained CO fluxes generally have significant uncertainties, which needs to be verified carefully before using in realistic applications.

In recent years, the "top-down" method has been used in many researches for obtain high-accuracy fluxes. The data assimilation is one "top-down" method in which fluxes are optimized by minimizing the difference between the observations and the simulations, this method has been proved to be an effective way for improving the accuracy of flux estimations (Saide et al., 2009; Tian et al., 2014). The four-dimensional variational data assimilation (4DVAR) is one of the most widely used data assimilation methods, it can realize the effective integration of the transport model, observational sampling and emission–concentration relationships (Lu et al., 2015; Park et al., 2016). Besides, it can make full use of effective information in the surface observations in the assimilation procedure (Zhang et al., 2008; Peng et al., 2015). However, the main hamper for using the 4DVAR method is the large dimension (generally with order of 10^7–10^8) in flux estimation, which requires large amount of computing resources that it is hard to satisfy. Besides, the adjoint model must be integrated in the use of 4DVAR (Gou and Sandu, 2011; Park et al., 2016), which requires large efforts for development and maintenance because of the high nonlinearity of the chemical transport model (Tian et al., 2008; Gou and Sandu, 2011; Lu et al., 2015).

In order to solve these problems, many dimension-reduction methods have been introduced into the 4DVAR, among whom the POD4DVAR method is a good way to reduce the dimension. A data assimilation system coupled with POD method can largely reduce the dimension in the assimilation procedure that was used frequently in many realistic applications (Tian and Xie, 2009; Tian et al., 2011; Tian and Feng, 2015; Zhang et al., 2017). Some researchers also tried to integrate other mathematical algorithms, e.g., Ensemble Kalman Filter, into the assimilation system for the flux inversion, which also can reduce computing resources and improve inversion accuracy (Lu et al., 2015; Kim et al., 2018). However, current CO fluxes mainly come from statistical investigation, wich generally have a low spatial resolution.

In order to improve the spatial resolution of the CO flux, an efficient inversion algorithm, POD4DVar, was developed by coupling the POD and the 4DVAR in this paper, then massive computing problem was solved in a low dimension. Afterwards, a regional high-resolution assimilation inversion system, TracersTracker, was conducted by coupling the POD4DVAR algorithm and the CMAQ model. The TracersTracker system can effectively absorb the information from observations for eliminating "errors" in the prior fluxes, and then improve significantly the accuracy of the posterior simulations. In order to dianose the effectiveness of the system, two regions with different geographical and climatic conditions were chosen for experiments.

Details of materials and methods are given in section 2, results of the background concentrations, sensitivity experiments and flux inversions of CO are shown in section 3, conclusions are shown in section 5.

Materials and Methods

The TracersTracker system is constructed by integrating the WRF model, the CMAQ model and the POD4DVAR method. The WRF model and the CMAQ model are used for providing meteorological fields and atmospheric transport simulations respectively. The POD method has proved to be an effective method for the dimension reduction, in which problems can be solved in a low-dimension space, the POD4DVAR is a combination of POD and the traditional 4DVAR.

Generally, 4DVAR can been described by minimizing the difference between the observations and the simulations as follows:

$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_{t=0}^{n} \left(y_i - H_i \left[M_{t_0 \to t_k}(x) \right] \right) T R_i^{-1} \left(y_i - H_i \left[M_{t_0 \to t_k}(x) \right] \right) \quad \text{(Eq.1)}$$

where the M means the transport model, the superscript T indicates a transpose, x_b is the background filed, k is the index of the observation time, H is the observation operator, R_i is the observational error covariance and B is the background error covariance.

We suppose there are S time steps in each assimilation time window: at the first time step, N first-guesses $x_{0,n} (n = 1,...,N)$ should be prepared by sampling from the historical forecasts or other existing priories. Secondly, transport model, e.g., CMAQ in this paper, can been integrated with the N first-guesses until the last time step of the assimilation time window and N state variable series $x_{k,n} (k = 1,...,S)$ can be obtained. With the N state variable series, the sample ensemble $X_n = (x_{0,n},...,x_{s,n})^T$ can be conducted and matrix $A = (X_1, X_2,...X_N)$ by X_N. From the ensemble perturbation, the PODs can be generated based on the matrix, and the optimization can be linearly represented by the PODs.

With the WRF model, the CMAQ model and the POD4DVAR method, TracersTracker system is developed for optimization of CO fluxes. The detail of the process for using the TracersTracker system is as follows (*Fig. 1*):

(1) Generating the first-guess perturbation ensemble. Start from a first-guess extracted from the priori and get N first-guesses from perturbations generated by Monte Carlo method.

(2) Getting N state variable series. Run the CMAQ model from the N first-guesses generated in last step over the whole assimilation time window and N state variable series can be obtained. With proper perturbations, the space of the N state variable series can capture the spatiotemporal evolution of the model state.

(3) Proper orthogonal decomposition. Obtain the base vectors Pn (PODs) from the space of the N state variable series by the proper orthogonal decomposition technique. In this way, the original N-dimensional problem can be solved in N dimensions, which greatly reduce the computing resources.

(4) Rerunning the CMAQ model. Rerun the CMAQ model from the Pn vectors and from the previous step over the assimilation time window to obtain n simulations of the state series.

(5) Generating a new emission. The optimal emission can be approximately expressed by the POD-base vectors generated from step (3). The weights of the base vectors can be determined by the performance of each POD.
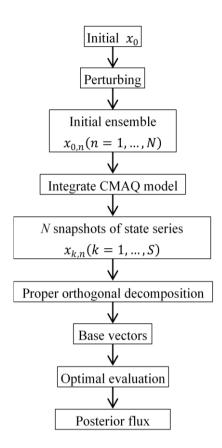
Figure 1. *Schematic diagram of the TracersTracker data assimilation system*

A carbon assimilation system involves many parameters, e.g., lag window, boundary filed, chemical mechanism, measurements, these parameters affect largely in the assimilation process and eventually the inversion results. The data assimilation system should be tested for these parameters before being applied to realistic applications. Ignoring influences of these parameters may decrease the accuracy of the inversion results or lead to incorrect results. In this section, nine sensitivity experiments were conducted for system parameters in the process of flux inversion, the detailed settings of these experiments are as follows:

Reference experiment (RE): two nesting mode (*Fig. 2*), d02 is used for providing boundary field for inner area d01, CO concentration observations are assimilated for flux optimization of the inner area. In this experiment, chemical mechanism is CB05, the number of perturbation samples is 126, lag window is not considered and the spatial resolution is 3 km. This experiment is the main experiment in this paper, results of other experiments are compared with that in this experiment.

Experiment 1 (E1): perturbing sample experiment. The number of perturbation samples is set to 60, the settings of other parameters are the same as that in RE. This experiment mainly diagnoses the influence of the perturbation samples in the TracersTracker inversion system.

Experiment 2 (E2): perturbing sample experiment. The number of perturbation samples is set to 200, the settings of other parameters are the same as that in RE, the purpose of this experiment is the same as that of E1.

Experiment 3 (E3): boundary field experiment. No nesting mode is adopted in this experiment; the settings of other parameters are the same as that in RE. This experiment mainly diagnoses the influence of nested mode in the TracersTracker inversion system.
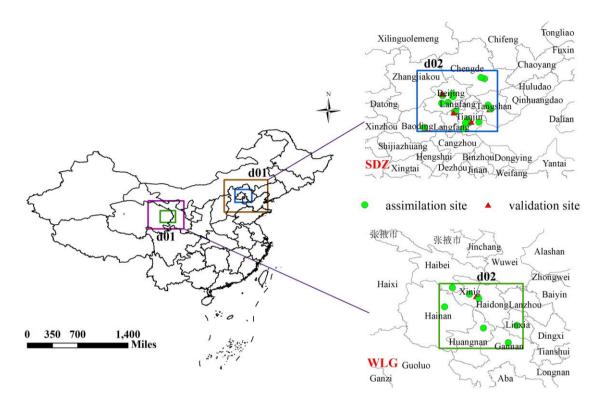

Figure 2. The two study areas. d02 is our inner area and d01 is the boundary layer area for providing boundary fields for d02. The green dots are the CO monitoring sites used for flux inversions and the red dots are used for validation.ext of the introduction

Experiment 4 (E4): boundary field experiment. Three nesting mode is adopted in this experiment; the settings of other parameters are the same as that in RE. The purpose of this experiment is the same as that of E3.

Experiment 5 (E5): lag window experiment. The lag window is set to 2 days in this experiment, the settings of other parameters are the same as that in RE. This experiment mainly diagnoses the influence of lag window in the TracersTracker inversion system.

Experiment 6 (E6): lag window experiment. The lag window is set to 2 weeks, the settings of other parameters are the same as that in RE. The purpose of this experiment is the same as that of E5.

Experiment 7 (E7): chemical mechanism experiment. The chemical mechanism in the model was changed into CB06 in this experiment, the settings of other parameters are the same as that in RE. The main purpose of this experiment is to test the influence of different chemical mechanisms on assimilation inversion results in the TracersTracker inversion system.

Experiment 8 (E8): spatial resolution experiment. The spatial resolution is set to 27 km in this experiment, the settings of other parameters are the same as that in RE. This experiment mainly tests the influence of spatial resolution on assimilation inversion results in the TracersTracker inversion system.

After the sensitivity analysis for the TracersTracker system, the system was then applied to two distinct areas in China, i.e., Shangdianzi (hereinafter referred to as SDZ) and Waliguan (hereinafter referred to as WLG), for improving the resolution of the CO flux inversion in 2016. China has a vast land area that significant differences of CO emission exist in different regions with unbalanced development. The two distinct areas, i.e., SDZ and WLG (*Fig. 2*), to some extent, can represent two kinds of regions with different characteristics of CO emissions. SDZ, located on the North China Plain, covers the region of Jing-Jin-Tang, is one of the highest urbanized regions in china. The total population is over 60 million in this area, emitting a large amount of CO through people's daily activities. There is one Global Atmosphere Watch (GAW) station, i.e., SDZ, in this area, which lies northeast of Beijing city, with low population and industrial density. WLG, unlike SDZ, locate in a region with a underdeveloped industrial foundation, covering parts of Qinghai and Gansu provinces. In this area, variations of CO concentration are significantly different from that of SDZ regions in the whole year. The purpose of choosing these two areas is to investigate the characteristics of emission mechanisms and uncertainties for CO inversions in different areas with the support of the TracersTracker system.

In this paper, the MIX inventory of 2010 released by the MEIC team was used as the a priori in the simulation process, this inventory has information of some major anthropogenic pollutants, e.g., CO_2, CO, PM10, PM2.5, SO_2, NH_3, BC and NMVOC. The inventory files have a resolution of 0.25 degree and include emissions from five sectors, i.e., power, agriculture, residential, industry and transport.

Hourly CO concentration measurements of 2016 in the two study areas were collected for flux inversion and accuracy validation. These CO observations can be downloaded from the air quality monitoring network of The Ministry of Environment Protection of the People's Republic (http://106.37.208.233:20035/). There are 22 and 9 motoring sites in SDZ and WLG respectively, the monitoring sites used in our study are shown in *Fig. 1*, details of these monitoring site are as listed in *Table 1*. In general, even if no pollutants are released, there is still residual amounts of pollutants in the atmosphere, called the background or baseline, often referred to be the measure in pure air masses that are not perturbed by pollutant emissions. In different regions, the background concentration is obviously different, which is affected by many factors, e.g., the local topography, the climate and the CO emissions intensity, the real-time concentrations are the results of long-time migration, transformation and integration. Therefore, the background concentration of CO is not caused by the current emissions and the current CO concentration measurements generally include two part, i.e., the background part and the current emission part, the background part should be subtracted from the observations before assimilation for the current flux inversion. Before the real assimilation process in the TracersTracker system, we evaluated the CO background concentration for the 31 monitoring sites mentioned above. In order to determine the background concentration of CO, the raw data were filtered to remove the regional pollution amount. We defined a threshold for removing obviously local pollution. Pollutant concentrations above the threshold within ±6 hours were removed. Then, we calculated the mean and standard deviation (σ) for every continuous unpolluted period and removed data larger than 3σ above the mean. We repeated this step until the concentration data set was stable.

Table 1. *CO observation and validation sites in SDZ and WLG*

Name	Longitude	Latitude	Location	Usage
BDCPZ	116.238	40.227	SDZ	ASSIMILATION
BJDL	116.237	40.300	SDZ	VALIDATION
BJGC	116.200	39.917	SDZ	ASSIMILATION
BJHRZ	116.647	40.308	SDZ	ASSIMILATION
BJNZG	116.471	39.947	SDZ	ASSIMILATION
BJSYXC	116.673	40.187	SDZ	ASSIMILATION
CDTL	117.978	40.924	SDZ	ASSIMILATION
CDWHZX	117.831	40.981	SDZ	ASSIMILATION
LFJLZX	116.729	39.564	SDZ	ASSIMILATION
LFKFQ	116.785	39.583	SDZ	ASSIMILATION
LFYCGS	116.698	39.525	SDZ	VALIDATION
TSSEZ	118.179	39.657	SDZ	VALIDATION
TSTCGS	118.233	39.678	SDZ	ASSIMILATION
TSWZJ	118.122	39.834	SDZ	ASSIMILATION
TJBCKJY	117.276	39.233	SDZ	ASSIMILATION
TJDLZX	117.316	39.088	SDZ	ASSIMILATION
TJWLJGY	117.432	39.135	SDZ	VALIDATION
TJTFGYY	117.186	39.097	SDZ	ASSIMILATION
TJTPW	117.103	38.861	SDZ	ASSIMILATION
TJSTC	117.744	39.130	SDZ	ASSIMILATION
BDJCZ	115.521	38.870	SDZ	VALIDATION
BDJDZX	115.471	38.914	SDZ	ASSIMILATION
GNJCZ	102.919	34.994	WLG	ASSIMILATION
HBXHZ	100.910	36.964	WLG	ASSIMILATION
HNQBQZ	100.632	36.281	WLG	ASSIMILATION
HNLWZ	102.029	35.516	WLG	ASSIMILATION
LXHBJ	103.251	35.611	WLG	VALIDATION
LXZWDX	103.223	35.609	WLG	ASSIMILATION
XNDWSC	101.522	36.726	WLG	ASSIMILATION
XNJCZ	101.756	36.649	WLG	VALIDATION
XNSLYY	101.857	36.584	WLG	ASSIMILATION

Results and Discussion

CO background concentrations

Fig. 3 is background CO concentrations at different monitoring sites in SDZ and WLG. The economic and geographical conditions of Shangdianzi region and Waliguan region are different, the CO concentration level and the varation characteristics are also different. Overall, the CO concentration in Waliguan is significantly lower than that in Shangdianzi (*Table 2*). The average annual CO concentration in Waliguan is 1.09 mg/m^3, that is 1.43 mg/m^3 in Shangdianzi area, about 1.3 times of that in Waliguan.

From January to December, the CO concentration in Shangdianzi is higher than that in Waliguan, but the difference between December and January is the largest. However, in terms of seasonal varation, the trends are similar in the two regions, the CO concentration is generally low in summer and autumn, and significantly increased in winter and spring. The background concentrations of CO are highly correlated with the observed concentration; its variation trend is basically consistent with the observed CO concentration. The average annual background concentration in Shangdianzi is 1.08 mg/m^3, and that is 0.69 mg/m^3 in Waliguan, which is significantly lower than that in Shangdianzi. In summer and autumn, the background concentration in Shangdianzi is

1.53 times of that in Waliguan. In winter and spring, the background concentration in Shangdianzi is 1.55 times of that in Waliguan. Besides, it can be seen that in the same region, the spatial variation of CO background concentration is also very unbalanced. In Shangdianzi, the background concentration of CO in Tangshan, Langfang and Tianjin is much higher than that in other areas, where are the main areas of CO emissions. In Waliguan, Xining and Linxia are the main areas of CO emissions, background concentrations of where are much higher than that in other areas. These areas are mainly industrial and human settlements with active human activities and large CO emissions.
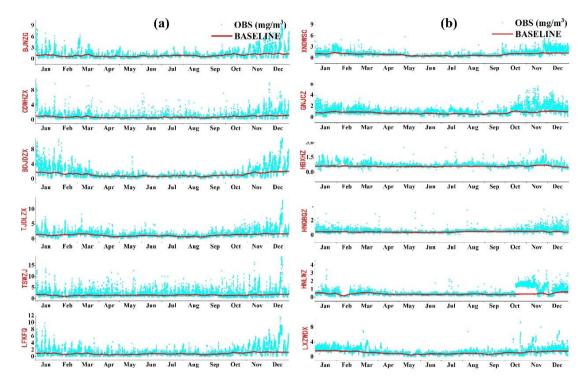

Figure 3. CO background at monitoring sites. (a) and (b) are the CO background in SDZ and WLG respectively

Table 2. CO background concentrations in Shangdianzi and Waliguan. obs and bg are CO monitoring and background concentration respectively

	SDZ (mg/m³)		WLG (mg/m³)	
	obs	bg	obs	bg
Jan	1.98	1.27	1.47	0.87
Feb	1.27	1.05	1.20	0.82
Mar	1.48	1.10	1.04	0.78
Apr	1.02	0.84	0.82	0.63
May	0.88	0.83	0.76	0.54
Jun	1.01	0.92	0.73	0.53
Jul	1.07	0.95	0.76	0.54
Aug	1.02	0.88	0.81	0.59
Sep	1.10	0.95	0.80	0.57
Oct	1.35	1.17	1.13	0.77
Nov	2.03	1.43	1.63	0.85
Dec	2.90	1.54	2.01	0.90

Although there are large differences of CO concentration between the two study areas, CO concentration can both reach to a high level in some heavy pollution periods, e.g., January and December. This illustrates that even in areas with low CO emission intensity, serious pollution events can also appear because of adverse meteorological and geographical conditions. In addition, no matter in Shangdianzi or Waliguan, the CO concentration is higher in spring and winter than in the other two seasons. The main reason is that Shangdianzi and Waliguan are located in the north of China. In spring and winter, more fossil fuel is consumed by living activities, e.g., residents heating, because of low temperatur, and the CO concentration can be pushed up significantly in a short time due to these activities. However, the difference of CO concentrations between the four seasons in Waliguan is significantly smaller than that in Shangdianzi, which shows that the activity intensity of heating in spring and winter in this area is limited to a few areas and has limited impact on the whole area. Unlike the Waliguan, the CO concentration in Shangdianzi changes more violently, and there are more periods with serious pollutions.

Sensitivity experiments

Reference experiment

This experiment is the reference for other experiments in sensitivity test to diagnose the variations of assimilation results when system parameters change. Experiment results (*Table 3*) show that the priori-based CO concentration simulations are significantly lower than that in RE, the RMSE of prior-based CO simulations is 1.30 mg/m^3, it is reduced to 0.94 mg/m^3 in the posterior-based simulations in RE, the simulated accuracy improved obviously.

Table 3. Results of sensitivity experiments for CO

	priori	RE	E1	E2	E3	E4	E5	E6	E7	E8
Jan		40%	4%	44%	20%	41%	39%	23%	42%	47%
Feb		23%	2%	22%	15%	20%	22%	20%	25%	19%
Mar		31%	81%	28%	13%	32%	23%	26%	29%	25%
Apr		29%	58%	26%	20%	25%	28%	10%	29%	20%
May		26%	31%	23%	26%	24%	25%	33%	29%	18%
Jun		26%	64%	22%	25%	28%	26%	17%	27%	18%
ul		40%	43%	36%	27%	39%	36%	38%	39%	31%
Aug		29%	20%	26%	12%	26%	28%	7%	28%	24%
Sep		30%	14%	28%	17%	25%	31%	20%	29%	25%
Oct		28%	25%	26%	17%	29%	24%	41%	24%	26%
Nov		24%	78%	29%	23%	22%	23%	28%	30%	31%
Dec		23%	10%	25%	23%	21%	25%	38%	28%	28%
R	0.63	0.76	0.42	0.77	0.54	0.75	0.77	0.75	0.76	0.69
RMSE	1.30	0.94	1.47	0.86	1.35	0.97	0.87	0.99	0.92	1.05

Besides, the correlation between simulations and observations increased from 0.60 based on the a priori to 0.76 based on the posteriors. From the perspective of the whole year, the annual average flux increased from 1.14 mole/s to 1.47 mole/s with an increase rate of 29%. From the perspective of monthly variations, the CO posteriors of 12 months in 2016 increased by 40%, 23%, 31%, 29%, 26%, 26%, 40%, 29%, 30%, 28%, 24% and 23%, respectively, compared with the a prior, January and July have the

largest growth rates. In terms of seasons, the growth rate of summer and autumn is significantly higher than that of spring and winter, the underestimation of summer and autumn in the a prior is much greater than that of spring and winter. With the parameters described in this experiment, the TracersTracker system can assimilate CO observations effectively, and then obtain the optimized CO flux which can reduce simulation errors and significantly improve the simulations accuracy.

Perturbing sample

The number of perturbing samples greatly affect the accuracy of the posterior fluxes. In theory, the more perturbing samples, the more spatial information the samples can represent, but the more computing resources it needs. Insufficient perturbing samples cannot fully reflect the temporal and spatial variation characteristics of the CO flux and the CO concentrations, which can lead to large errors in the assimilation inversion results. In the sensitivity experiment, 126 perturbing samples are used in the reference experiment, 60 and 200 perturbing samples are used in Experiment 2 and Experiment 3, respectively to verify the influence of perturbing samples on the assimilation inversion results. Experiment results show that there is no significant difference between 126 and 200 perturbing samples in CO simulation errors. Their correlation coefficients with 126 and 200 perturbing samples are 0.76 and 0.77, respectively, and their RMSE are also very close, i.e., 0.94 and 0.86, respectively. This shows that when the number of perturbing samples increases to a certain extent, the accuracy of assimilation inversion cannot be improved obviously. In contrast, when 60 perturbing samples are used for assimilation inversion, there is large errors in the inversion results. As can be seen in *Table 1*, the correlation coefficient of CO simulated concentration is significantly reduced to 0.42, its RMSE increased from 0.94 mg/m^3 to 1.47 mg/m^3, the simulation accuracy also decreased significantly. Flux inversion results show that the posterior with 200 perturbing samples is close to that in the reference experiment, the annual average of the posterior is 1.46 mole/s. Compared with the a prior, the 200-samples-based posterior has growth ratios of 44%, 22%, 28%, 26%, 23%, 22%, 36%, 26%, 28%, 26%, 29% and 25% in 12 months, respectively, which was basically consistent with the trend of the reference experimental inversion results (*Table 3*). In comparison, although there is no significant difference in the total amount of flux inversion results based on 60 perturbing samples, there is obvious fluctuations in monthly variations. The 12-month growth ratios are 4%, 2%, 81%, 58%, 31%, 64%, 43%, 20%, 14%, 25%, 78% and 10%, respectively, the inversion fluxes in January, February, November and December are all abnormal, which shows that there is large uncertainties in the inversion results with 60 perturbing samples (*Table 3*).

Boundary layer

Generally, boundary layers directly impact the simulated surface concentration and flux. Ignoring boundary layers in the process of assimilation can lead to additional errors in the final inversion results. In the reference experiment, two nesting domains mode is set to eliminate errors around the boundary layer, as shown in *Fig. 2*, d01 is our study area and d02 is the boundary layer area providing the boundary field for d01. In Experiment 3, no nesting mode is adopted, the dynamic evolution setting of the boundary field is removed, and the default boundary field inside the CMAQ model is adopted. According to the comparison between the posterior-based simulations and the observations (*Table 3*), the errors of CO simulations without boundary field is

significant, RMSE increases from 0.94 mg/m^3 of the a prior to 1.35 mg/m^3 of the Experiment 3, and the correlation decreases from 0.76 to 0.54. After the boundary field is reset in Experiment 4, RMSE decreased to 0.97 mg/m^3, the correlation also increased to 0.75, which was slightly lower than that in the reference experiment, the simulation accuracy is significantly improved. This shows that the boundary field plays an important role in assimilation inversion system, if the boundary field is not set or the setting is not suitable, the inversion result could have large uncertainties. Suitable boundary field can provide accurate exchange data for the study area, thus significantly improving the accuracy of CO flux inversions. Without suitable boundary fields, flux inversions have great fluctuations compared to that in the reference experiment. The average flux retrieved from the Experiment 3 is 1.31 mole/s which is significantly lower than that in the reference experiment, and the monthly variations in the Experiment 3 is quite different with that in the reference experiment. In an obvious contrast, the average and monthly flux in the Experiment 4 have no significant difference with that in the reference experiment after sufficient boundary information been provided. In this paper, we use two-nested mode in all inversion processes as described in the reference experiment.

Lag window

Due to different physical and chemical characteristics of different pollutants, lag windows of them are also obviously different. In the process of assimilation inversion, the lag window should be set reasonably for a specified substance. The life time of CO is short of about two months and the study areas in this paper are small, the lag window should also be set short. In the reference experiment, the lag window is set to 0, i.e., no consideration for lag window. In contrast, the lag windows in Experiment 5 and Experiment 6 are set to 2 days and 2 weeks respectively to obtain the influence of lag window on the assimilation inversion results. The simulation results show that the errors and correlations in Experiment 5 and Experiment 6 are quite close. Compared to the reference experiment, the simulation correlation in Experiment 5 is slightly improved (R changes from 0.76 to 0.77), in experiment 6 it is slightly reduced (R changes from 0.76 to 0.75), the overall difference is not obvious. Meanwhile, the RMSEs in Experiment 5 and Experiment 6 did not change significantly, which are 0.87 mg/m^3 and 0.99 mg/m^3, respectively. The annual average flux in Experiment 5 and Experiment 6 are 1.46 mole/s and 1.43 mole/s, which only have minor differences with that in the reference experiment of 1.47 mole/s (*Table 3*).

Chemical reactions

In the CMAQ model, there are two chemical mechanisms to describe the chemical reaction mechanism of simulated substances, namely CB05 and CB06. These two chemical mechanisms have some differences in simulating the chemical reaction of trace gases, which will ultimately affect the flux inversion results. In the reference experiment, CB05 was adopted to describe the chemical reactions of CO, it is replaced to CB06 in Experiment 7 to diagnose the sensitivity of different chemical mechanisms in the inversion process. As shown in *Table 1*, simulations errors by CB05 and CB06 are quite close, and RMSE is 0.92 mg/m^3, correlation is 0.76, both are basically same as that in the reference experiment. Besides, the annual average of the CB06-based posterior is 1.48 mole/s that it almost has no difference with that in the reference experiment (*Table 3*).

Spatial resolution

Inversion results of large-scale, e.g., global and continent scale, usually have coarse resolutions in degree, which cannot guarantee the requirement for high-resolution flux at regional scale. The spatial resolution of assimilation inversion system in the reference experiment 3 km, which has greatly improved the resolution in the posterior. In Experiment 8, the resolution of the assimilation inversion system is set to be 27 km to analysis detail affections of the spatial resolution on the results of flux inversions. Experiment results show that simulations errors and RMSEs (from 0.94 mg/m^3 to 1.05 mg/m^3) become larger, correlations R become worse (from 0.76 to 0.69) when the spatial resolution becomes coarser (*Table 3*). Besides, the flux inversions with 27 km have more uncertainties compared to that in the reference experiment, and the simulation accuracy of the inversion flux will be affected by the coarse spatial resolution. In summary, coarse spatial resolution will bring uncertainties to the results of flux inversion.

CO flux inversions

The posterior emissions

9 and 22 groups of CO concentration observations are collected in WLG and SDZ respectively for CO flux inversions. Based on the background concentrations of CO, the "polluted observations" are extracted for the 31 monitoring sites which are actually assimilated in the inversion process, the CO flux inversions are as shown in *Table 3* and *Fig. 4*.
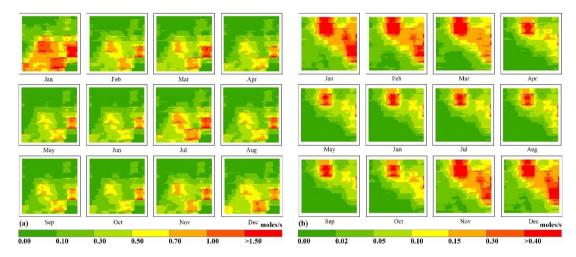

Figure 4. Increment of CO posterior compared with the MIX priori in SDZ (a) and WLG (b)

In SDZ, the CO posterior is higher than the a prior in most periods, flux grow rates in 12 months are 40.2%, 23.2%, 30.7%, 28.6%, 26.2%, 26%, 40.5%, 28.8%, 29.6%, 27.7%, 24.1% and 23.4%, respectively (*Table 4*), most of them are close except January and July, whose monthly increase are higher than that in other months. The average flux of the posterior is 1.47 mole/s, which is about 1.3 times of that in the a priori. The minimum growth of the posterior appears in February and December, the maximum growth is in January and July, CO posterior has significant variation after the optimization of the TracersTracker system.

Table 4. *Increments and proportions of the CO posteriors in SDZ and WLG*

	SDZ (mole/s)			WLG (mole/s)		
	Max	**Ave**	**Percent(%)**	**Max**	**Ave**	**Percent(%)**
Jan	7.39	0.66	40.2	1.28	0.19	68.9
Feb	2.58	0.33	23.2	0.92	0.16	60.7
Mar	4.81	0.37	30.7	1.17	0.14	62.5
Apr	3.30	0.30	28.6	0.80	0.08	58.5
May	3.02	0.27	26.2	0.71	0.06	58
Jun	2.67	0.26	26.0	1.06	0.06	59
Jul	5.06	0.40	40.5	1.14	0.06	61.2
Aug	3.24	0.28	28.8	1.08	0.06	59.4
Sep	2.06	0.26	29.6	0.79	0.06	62.6
Oct	2.74	0.25	27.7	0.59	0.08	57.4
Nov	2.44	0.29	24.1	1.10	0.17	66
Dec	2.00	0.32	23.4	0.96	0.17	60.2
Annual	3.44	0.33	29.1	0.97	0.10	61.2

Compared to SDZ, the CO posterior flux in WLG is also higher than the a prior in most periods, but the growth proportion is much larger than that in SDZ. The posterior flux increased by 68.9%, 60.7%, 62.5%, 58.5%, 58%, 59%, 61.2%, 59.4%, 62.6%, 57.4%, 66% and 60.2%, respectively in 12 months of the whole year (*Table 4*). The overall increase rate is about 4 times of that in SDZ, although the overall increase rate is high, the absolute flux increment (0.10 mole/s) in WLG is much smaller than that in SDZ (0.33 mole/s), which is mainly related to the low CO priori in WLG. The annual average flux of the posterior in WLG is 0.09 mole/s, the minimum is 0.05 mole/s occurring in May and September, the maximum flux in the posterior is 0.15 mole/s appearing in December and January.

According to the spatial distribution of the posterior fluxes, the net increase and increase ration in the posteriors have significant difference. The net increase of the posterior fluxes is mainly concentrated in urban or densely populated areas. In SDZ, the area with the most flux growth is in Tangshan, followed by Tianjin and Beijing, and Chengde has the least flux growth. In WLG, the area with obvious CO flux growth is mainly located near Xining, followed by the urban area and surrounding area of Linxia, and other areas, e.g., Haibei, Hainan and Haidong, have little flux growths. Due to the productive and living activities are concentrated in the urban and nearby areas, the prior flux of CO in these areas is much larger than that in other areas, thus the net increase of CO in these areas is large with the same growth rate (*Fig. 4*). In contrast, the main areas with high growth rates are not entirely located in the urban areas, the spatial distribution of growth rate has no obvious tendency except for some periods, e.g., June, August in WLG. Whether in SDZ or WLG, the proportion of posterior incremental growth is scattered in time and space.

Comparison of observations and simulations

In order to evaluate the accuracy of the posteriors obtained by the TracersTracker system, comparisons between the observations and the posteriori-based simulations were conducted in the validation sites shown in *Fig. 1*. Some indices for the accuracy evaluation are used in this section, i.e., max absolute error (E_{max}), mean error (E_{mean}),

root mean squared error (E_{rmse}), and correlation coefficient (R), indices are determined by the following equations.

$$E_{max} = \max(e_1, e_2, ..., e_n)$$ (Eq.2)

$$E_{mean} = (e_1, e_2, ..., e_n)/n$$ (Eq.3)

$$E_{rmse} = \sqrt{(e_1^2 + e_2^2, ..., + e_n^2)/n}$$ (Eq.4)

$$R = \frac{\sum_{i=1}^{n}(o_i - \bar{o})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}\sqrt{\sum_{i=1}^{n}(s_i - \bar{s})^2}}$$ (Eq.5)

where *n* is the number of comparison couples, *e* is the difference between the CO posterior-based simulation and the CO concentration observation, *o* and *s* mean the observations and simulations, respectively.

The CO concentration in January is significantly higher than that in July, this tendency can be grasped well in the posteriors and the a priori. However, there are large errors between the CO simulations in the a priori and ground-based observations, and these errors are largely eliminated in the posterior-based simulations (*Fig. 5* and *Fig. 6*).
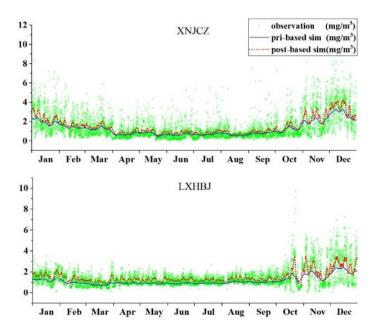

Figure 5. *Comparison of posterior-simulations and observations in WLG*

In SDZ, the correlation coefficients between the observed and simulated CO concentrations based on the a prior are 0.58, 0.66, 0.65, 0.61 and 0.52 in the five validation sites, i.e., BJDL, BDJCZ, LFYCGS, TJWLJGQ, and TSSEZ, respectively (*Table 5*). Priori-based simulations have a poor correlation with the CO observed

concentrations, the main cause of this is the underestimation in the a priori. In an obvious contrast, errors between the posterior-based CO simulations and observations are significantly reduced and the simulation accuracy has been improved greatly (*Fig. 5* and *Fig. 6*). Besides, the correlation coefficients of validation sites, i.e., BJDL, BDJCZ, LFYCGS, TJWLJGQ and TSSEZ have been improved to 0.78, 0.78, 0.76, 0.76 and 0.70, respectively.
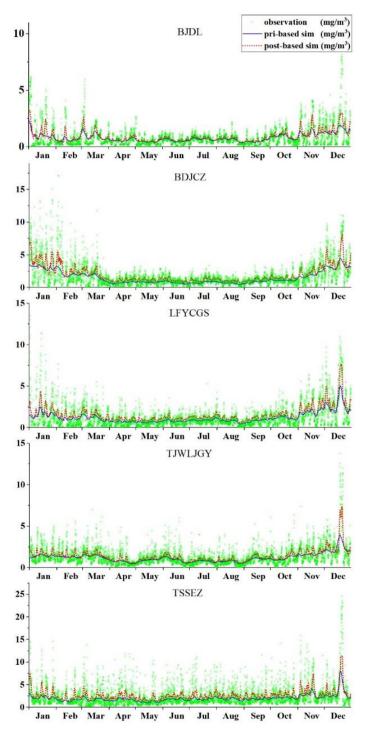

Figure 6. *Comparison of posterior-simulations and observations in SDZ*

Table 5. *Errors of CO posterior-simulations and observations in SDZ and WLG*

	SDZ (mg/m^3)					WLG (mg/m^3)	
	BJDL	BDJCZ	LFYCGS	TJKG	TSSEZ	XNJCZ	LXHBJ
priori-based:							
E_{max}	7.61	13.9	12.0	9.87	18.4	7.22	8.26
E_{mean}	0.49	0.87	0.71	0.60	1.20	0.65	0.54
E_{rmse}	0.62	1.11	0.82	0.74	1.57	0.72	0.60
R	0.58	0.66	0.65	0.61	0.52	0.69	0.64
posterior-based:							
E_{max}	6.42	12.7	10.6	6.94	13.2	6.43	7.19
E_{mean}	0.43	0.76	0.75	0.54	1.21	0.62	0.48
E_{rmse}	0.49	0.90	0.68	0.58	1.19	0.63	0.52
R	0.78	0.78	0.76	0.76	0.70	0.73	0.74

In WLG, the correlation coefficients between the observed and simulated CO concentrations based on the prior flux is 0.69 and 0.64 at XNJCZ and LXHBJ, respectively (*Table 5*). The correlations are slightly higher than that based on the a priori in SDZ, but the priori-based simulations still have great errors. With the posteriori optimized by the TracersTracker system, the correlation between CO simulations and observations increased to 0.73 and 0.74 at XNJCZ and LXHBJ, respectively, and the CO simulations and observations are more consistent in the variation trend in the whole year. In addition, errors between the CO simulations and observations are also reduced obviously, the average error is reduced from 0.65 mg/m³ to 0.62 mg/m³ at XNJCZ and 0.54 mg/m³ to 0.48 mg/m³ at LXHBJ, respectively. At the same time, the maximum errors also show a stable downward trend in both SDZ and WLG. In SDZ, the maximum error of posterior-based simulations decreased by 1.19 mg/m³, 1.2 mg/m³, 1.4 mg/m³, 2.93 mg/m³ and 5.2 mg/m³ at BJDL, BDJCZ, LFYCGS, TJWLJGQ and TSSEZ, respectively, the average decrease ration of the five monitoring sites is over 20%. In the WLG, the maximum error of posterior-based simulations decreased by 0.79 mg/m³ and 1.07 mg/m³ at XNJCZ and LXHBJ respectively with an average decrease of 12% (*Table 5*).

All above performance of the posterior illustrate that the TracersTracker system can greatly optimize the CO flux and significantly improve the simulation accuracy. It also shows that observation information is very important in the assimilation system, enough observations is the effective guarantee for high-precious flux inversions. Insufficient observed information could bring a lot of uncertainties, thus increases the simulations errors of the posteriors. In addition, there are both similarities and differences in simulation errors and tendencies in the a priori and posteriors no matter in SDZ and WLG. CO simulation errors in both areas tend to decrease in all validation sites, and errors range obviously shrinks in the posterior-based simulations. Although the trend of error distribution is similar in all monitoring sites, the shrink trend of error change is not in the same degree. Some of them have large contractions, e.g., TSSEZ, some have small contractions, e.g., BJDL, some have no obvious error contractions, e.g., XNJCZ, which is mainly related to the location of each validation site and the magnitude of CO observation concentration itself.

Seasonal and regional variation of emissions

From the perspective of seasonal net increase of the posteriors, the net flux increase in winter is the largest among the four seasons in SDZ, reaching 0.43 mole/s, there is no significant difference in the other three seasons, which are about 0.3 mole/s, 0.31 mole/s and 0.27 mole/s in spring, summer and autumn respectively, only about two thirds of that in winter. Similar to SDZ, net flux increase in winter is the largest in WLG, reaching 0.17 mole/s, while spring and summer have the smallest net increase of 0.06 mole/s, only one third of that in winter (*Table 6*). From the perspective of annual average net increase, it is much smaller in WLG than that in SDZ.

Table 6. *Seasonal variation and increment proportions of the CO posteriors in SDZ and WLG*

Time	Max	Avg	Percent(%)
SDZ:			
Spring	3.69	0.31	28.5
Summer	3.49	0.31	31.8
Autumn	2.37	0.27	27.2
Winter	3.86	0.43	28.9
WLG:			
Spring	0.15	0.09	60.6
Summer	0.11	0.06	59.4
Autumn	0.17	0.10	59.8
Winter	0.17	0.17	65

From the perspective of seasonal net increase ration of the posteriors, there is no significant difference in the incremental ratio of the four seasons in SDZ. The growth ratio of the four seasons is basically the same, about 30%, summer has the highest increment (31.8%) which is slightly higher than that in other three seasons. In WLG, the proportions of growth in all seasons is much higher than that in SDZ, and the tendency of growth is also significantly different from that in SDZ (*Table 6*). The growth rate of WLG in winter is the largest, reaching 65%, slightly higher than that in other seasons.

Overall, the posteriors in winter have the highest growth rate in both areas. During this period, anthropogenic living activities, e.g., residential heating, push up the CO concentration in the atmosphere. With the development of economic, the demand for living activities in winter increases rapidly, thus makes the CO flux growth rate in winter higher than that in other seasons.

Fig. 7 is the scatter diagram of the seasonal error distribution in SDZ, in which the horizontal axis is the concentration of CO observations, the vertical axis is the posterior-based CO simulation, the black points are the prior-based simulations, and the red points are the posterior-based simulations. On the whole in SDZ, CO concentration in winter is the highest with a range of 0 mg/m^3 to 24 mg/m^3, which is much higher than that in other seasons and indicating that CO concentration changes frequently and greatly in winter, the main cause for this phenomenon is the activities of residents heating. For the CO simulations, the results have obvious different situations in different seasons whether in the a priori or the posteriors. Besides, the simulation accuracy of the a priori and the posteriors are better in autumn and winter than that in spring and summer, correlations of the priori-based simulations are 0.56 and 0.59 in spring and summer, 0.62 and 0.61 in autumn and winter, respectively. After the flux

optimization, the trend of the posterior-based simulations is more obvious. The simulation precision is higher in autumn and winter, and the correlation coefficients are 0.75 and 0.83, respectively.
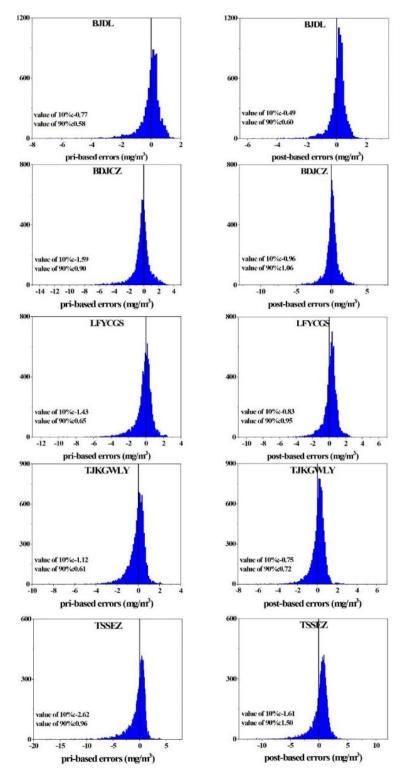

Figure 7. *Simulated errors of validation sites in SDZ. The pri-based and post-based errors are simulations errors based on the a priori and the posterior, respectively*

Fig. 8 is the scatter diagram of the seasonal error distribution in WLG. Similar to that in SDZ, CO concentrations in winter is significantly higher than that in other seasons in WLG, but there are obvious differences of CO simulations between the two study areas. The correlation of priori-based simulations is very poor, especially in autumn (R is only 0.56), which is significantly lower than that in other seasons. After flux optimization, the correlation between the CO observations and the posterior-based simulations has been improved greatly especially in winter (R is 0.79) and spring (R is 0.76).
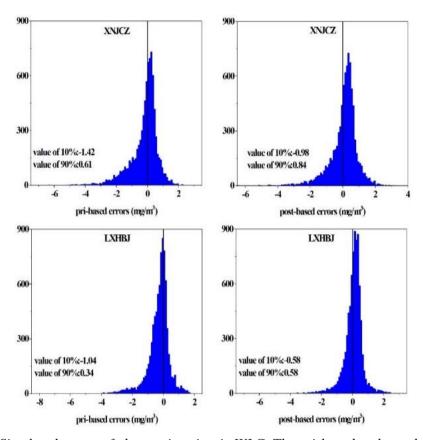

Figure 8. Simulated errors of observation sites in WLG. The pri-based and post-based errors are simulations errors based on the a priori and the posterior, respectively

In winter of northern China, the low temperature will increase the consumption of fossil fuels, e.g., coal and oil, thus the CO concentration in the atmosphere increase sharply in a short period. In addition, adverse weather conditions will exacerbate the increase speed, and eventually led to heavy air pollution events. In general, it will be more difficult to simulate the CO concentration with frequent variations in the atmospheric model. However, from the results shown above, CO concentration in winter with large fluctuations have better simulations than other period, which illustrate that the posterior can effectively optimize the a priori, errors are eliminated significantly in the posterior. For the period with small fluctuation of CO concentration, the accuracy of the posterior simulation is improved, but the accuracy in spring, summer and autumn are not as good as that in winter (*Fig. 9* and *Fig. 10*).
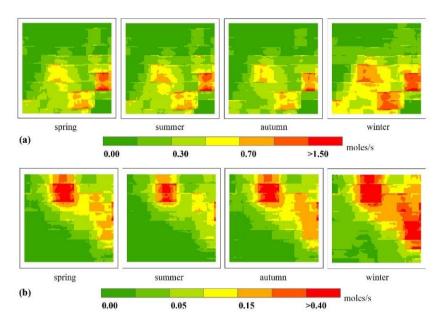
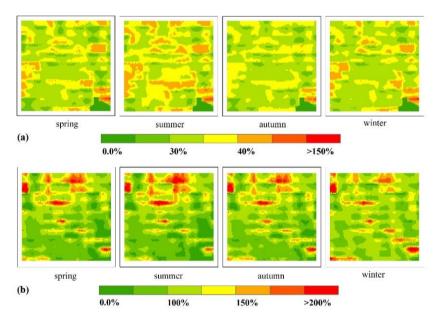*Figure 9.* *Seasonal increment of the CO posteriors in SDZ (a) and WLG (b)*

Figure 10. *Seasonal increment proportions of the CO posteriors in SDZ (a) and WLG (b)*

Conclusions

In this paper, a regional assimilation system, i.e., TracersTracker, was conducted for CO flux inversion by coupling the POD4DVAR method and the CMAQ model. The system was then applied to two distinct areas in china to diagnose the variations of CO flux inversion with different conditions. The TracersTracker system can effectively assimilate the hourly CO observations in the two study areas, the posteriors can significantly improve the accuracy of CO simulations. This study also suggested that the MIX inventory has underestimations of CO emission by 29.1% in SDZ and 61.2% in WLG. Due to the active chemical characteristics of CO, the flux inversion for CO can

only be carried out in a small area, how to obtain global and continental CO fluxes should be considered in next step. In addition, in order to improve the efficiency of the assimilation inversion system, other mathematical methods with high efficiencies should be introduced into the assimilation system to obtain high-resolution CO fluxes.

REFERENCES

[1]   Gou, T., Sandu, A. (2011): Continuous versus discrete advection adjoints in chemical data assimilation with CMAQ. – Atmospheric Environment 45(28): 4868-4881.

[2]   Gurjar, B. R., Jain, A., Sharma, A., Agarwal, A., Gupta, P., Nagpure, A. S., Lelieveld, J. (2010): Human health risks in megacities due to air pollution. – Atmospheric Environment 44(36): 4606-4613.

[3]   Kamimura, A., Armenta, B., Nourian, M., Assasnik, N., Nourian, K., Chernenko, A. (2017): Perceived Environmental Pollution and Its Impact on Health in China, Japan, and South Korea. – J Prev Med Public Health 50(3): 188-194.

[4]   Kilmont, Z. S. (2002): Anthropogenic emissions of non-methane volatile organic compounds in China. – Atmospheric Environment 36: 1309-1322.

[5]   Kim, H., Kim, H. M., Kim, J., Cho, C. (2018): Effect of Data Assimilation Parameters on The Optimized Surface CO_2 Flux in Asia. – Asia-Pacific Journal of Atmospheric Sciences 54(1): 1-17.

[6]   Law, K. J. H., Stuart, A. M., Zygalakis, K. C. (2015): Data Assimilation: A Mathematical Introduction. – Revista Brasileira De Meteorologia 26(3): 433-442.

[7]   Li, P., Chai, T. F., Carmichael, G. R., Tang, Y. H., Streets, D., Woo, J.-H., Friedli, H. R., Radke, L. F. (2007): Top-down estimate of mercury emissions in China using four-dimensional variational data assimilation. – Atmospheric Environment 41(13): 2804-2819.

[8]   Lu, S., Lin, H. X., Heemink, A. W., Fu, G., Segers, A. J. (2015): Estimation of Volcanic Ash Emissions Using Trajectory-Based 4D-Var Data Assimilation. – Monthly Weather Review 144: 575-589.

[9]   Lu, L., Chen, B., Guo, L., Zhang, H., Li, Y. (2019): A regional data assimilation system for estimating CO surface flux from atmospheric mixing ratio observations—a case study of Xuzhou, China. – Environmental Science and Pollution Research 26(9): 8748-8757.

[10]   Ma, Y. T. (2007): The Compilation of Vehicle Emission Inventory in Pearl River Delta Region and its Uncertainty Analysis. – Beijing, Peking University.

[11]   MEPC (2017): Report on the State of the Environment in China 2016. – Ministry of Ecology and Environment, China.

[12]   Ohara, T., Akimoto, H., Kurokawa, J., Horii, N., Yamaji, K., Yan, X., Hayasaka, T. (2007): An Asian emission inventory of anthropogenic emission sources for the period 1980-2020. – Atmos. Chem. Phys. 7: 4419-4444.

[13]   Park, S., Kim, D., Lee, S., Lee, H. W. (2016): Variational data assimilation for the optimized ozone initial state and the short-time forecasting. – Atmospheric Chemistry and Physics 16(5): 3631-3649.

[14]   Peng, Z., Zhang, M., Kou, X., Tian, X., Ma, X. (2015): A regional carbon data assimilation system and its preliminary evaluation in East Asia. – Atmos. Chem. Phys. 15: 1087-1104.

[15]   Saide, P., Osses, A., Gallardo, L., Osses, M. (2009): Adjoint inverse modeling of a CO emission inventory at the city scale: Santiago de Chile's case. – Atmospheric Chemistry and Physics Discussions 9(2): 6325-6361.

[16] Shi, Y., Xia, Y. F., Lu, B. H., Liu, N., Zhang, L., Li, S. J., Li, W. (2014): Emission inventory and trends of NO_x for China, 2000-2020. – Journal of Zhejiang University 15(6): 454-464.

[17] Tian, X., Xie, Z., Dai, A. (2008): An ensemble-based explicit four-dimensional variational assimilation method. – Journal of Geophysical Research 113(D21).

[18] Tian, X., Xie, Z. H. (2009): An explicit four-dimensional variational data assimilation method based on the proper orthogonal decomposition: Theoretics and evaluation. – Science in China Series D: Earth Sciences 52(2): 279-286.

[19] Tian, X., Xie, Z., Sun, Q. (2011): A POD-based ensemble four-dimensional variational assimilation method. – Tellus A: Dynamic Meteorology and Oceanography 63(4): 805-816.

[20] Tian, X., Xie, Z., Liu, Y., Cai, Z., Fu, Y., Zhang, H., Feng, L. (2014): A joint data assimilation system (Tan-Tracker) to simultaneously estimate surface CO_2 fluxes and 3-D atmospheric CO_2 concentrations from observations. – Atmospheric Chemistry and Physics 14(23): 13281-13293.

[21] Tian, X., Feng, X. (2015): A non-linear least squares enhanced POD-4DVar algorithm for data assimilation. – Tellus A: Dynamic Meteorology and Oceanography 67: 25340.

[22] Xue, T., Liu, J., Zhang, Q., Geng, G., Zheng, Y., Tong, D., Liu, Z., Guan, D., Bo, Y., Zhu, T., He, K., Hao, J. (2019): Rapid improvement of PM2.5 pollution and associated health benefits in China during 2013–2017. – Science China Earth Science 62: 1847-1856.

[23] Zhang, L., Constantinescu, E. M., Sandu, A., Tang, Y., Chai, T., Carmichael, G. R., Byun, D., Olaguer, E. (2008): An adjoint sensitivity analysis and 4D-Var data assimilation study of Texas air quality. – Atmospheric Environment 42(23): 5787-5804.

[24] Zhang, B., Tian, X. J., Zhang, L. F., Sun, J. H. (2017): Handling non-linearity in radar data assimilation using the non-linear least squares enhanced POD-4DVar. – Science China Earth Sciences 60: 478-490.

[25] Zheng, J., Zhang, L., Che, W., Zheng, Z., Yin, S. (2009): A highly resolved temporal and spatial air pollutant emission inventory for the Pearl River Delta region, China and its uncertainty assessment. – Atmospheric Environment 43(32): 5112-5122.