Journal of Hazardous Materials 280 (2014) 143–155

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Modified ensemble Kalman filter for nuclear accident atmospheric dispersion: Prediction improved and source estimated

X.L. Zhang, G.F. Su, H.Y. Yuan*, J.G. Chen, Q.Y. Huang

Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing 100084, PR China

HIGHLIGHTS

- A modified ensemble Kalmen filter data assimilation method is proposed.
- The method can consider four main uncertain parameters in the puff model.
- The prediction of radioactive material atmospheric dispersion is improved.
- The source release rate and plume rise height are successfully reconstructed.
- It can shorten the time lag in the response of ensemble Kalmen filter.

ARTICLE INFO

Article history:
Received 7 February 2014
Received in revised form 28 June 2014
Accepted 24 July 2014
Available online 10 August 2014

Keywords:
Atmospheric dispersion
Nuclear power plant accident
Ensemble Kalman filter

ABSTRACT

Atmospheric dispersion models play an important role in nuclear power plant accident management. A reliable estimation of radioactive material distribution in short range (about 50 km) is in urgent need for population sheltering and evacuation planning. However, the meteorological data and the source term which greatly influence the accuracy of the atmospheric dispersion models are usually poorly known at the early phase of the emergency. In this study, a modified ensemble Kalman filter data assimilation method in conjunction with a Lagrangian puff-model is proposed to simultaneously improve the model prediction and reconstruct the source terms for short range atmospheric dispersion using the off-site environmental monitoring data. Four main uncertainty parameters are considered: source release rate, plume rise height, wind speed and wind direction. Twin experiments show that the method effectively improves the predicted concentration distribution, and the temporal profiles of source release rate and plume rise height are also successfully reconstructed. Moreover, the time lag in the response of ensemble Kalman filter is shortened. The method proposed here can be a useful tool not only in the nuclear power plant accident emergency management but also in other similar situation where hazardous material is released into the atmosphere.

1. Introduction

Atmospheric dispersion models play an important role in nuclear power plant accident management. At the early phase, a reliable estimation of the radioactive material distribution in short range (about 50 km) is in urgent need for population sheltering and evacuation planning [1,2]. But the inaccurate meteorological data and uncertain source term (*i.e.*, source release rate, plume rise height) lead to erroneous simulations [3–5]. Moreover, if the source term, an important parameter for the long-range transport model, is not successfully reconstructed, it will become a large source of error for the environmental impact assessment [4,6]. As a result, to improve the model prediction and reconstruct the source term becomes an urgent task.

The measurements from monitoring network can help achieve the twofold goal. The combination of model predictions and measurements, known as data assimilation (DA), has been widely used in numerical weather prediction (NWP) over the past two decades [7]. The DA method has been introduced in the nuclear emergency response system during the past decade. In the real time on-line decision support system (RODOS) for European off-site nuclear emergency management, Kalman filter [8] method was used to improve the predictions of Gaussian plume model under quasi-steady state [9–11], and some efforts [12,13] were made to reconstruct the source term. Variational method [14–18] and particle filter method [19,20] in conjunction with plume or puff model were also used to improve the model predictions. Moreover, in order to reconstruct the source term, many studies

* Corresponding author. Tel.: +86 10 62792856; fax: +86 10 62792863.
E-mail address: hy-yuan@outlook.com (H.Y. Yuan).

http://dx.doi.org/10.1016/j.jhazmat.2014.07.064
0304-3894/

[4,21–25] have been conducted using long range transport model in conjunction with continental or global monitoring data, but the great uncertainties of the meteorological data, sparse distributed stations and poor representativeness of the measurements can lead to the failure of reconstruction [26]. For the emergency response, the key time and space scales are in the order of a few hours and tens of kilometers (around 30 km) [15], and the monitoring network near the plant is much denser, so new methodologies for short range transport (usually under about 50 km) have been developed to simultaneously improve the model prediction and reconstruct the source release rate, including the Monte Carlo dispersion model combined with ensemble Kalman filter (EnKF) DA method [27–29], and Lagrangian puff-model combined with variational DA method [30–32]. In these studies, the uncertainties in both meteorological and source data have been considered. However, it takes several hours for the DA method to minimize the uncertainties of *a priori* parameters and converge to the true values (in [27]). And the uncertainty of plume rise height, which is a sensitive and uncertain parameter at the early phase [33] is not considered.

The objective of this study is to develop a modified ensemble Kalman filter [34] data assimilation method in conjunction with a Lagrangian puff-model, to simultaneously improve the model prediction and reconstruct the source terms for short range (about 50 km) atmospheric dispersion. According to the studies [19,35], four most sensitive uncertain parameters in case of a radioactive release are considered: source release rate, plume rise height, wind speed and wind direction. Twin experiments are conducted to evaluate the effectiveness of the proposed method.

2. Theoretical background

For the atmospheric dispersion system, the state and observation can be expressed as

$$\boldsymbol{x}_k = M(\boldsymbol{x}_{k-1}) + \boldsymbol{\eta}, \quad \boldsymbol{x}, \boldsymbol{\eta} \in \mathbf{R}^n \tag{1}$$

$$\boldsymbol{y}_k^o = H(\boldsymbol{x}_k) + \boldsymbol{\varepsilon}, \quad \boldsymbol{y}^o, \boldsymbol{\varepsilon} \in \mathbf{R}^m \tag{2}$$

where $\boldsymbol{x}$ is the state vector of the dynamic system, n is the length of the state vector, and the subscript k or $k-1$ represents the time step of data assimilation. $M(\boldsymbol{x})$ is the dispersion model, $\boldsymbol{\eta}$ is the model error, $\boldsymbol{y}^o$ is the observation vector, m is the number of the observations, $H(\boldsymbol{x})$ is observation model, and $\boldsymbol{\varepsilon}$ is the measurement error. In this study, the dispersion model is Lagrangian puff-model. The state vector includes radioactive content Q and the central positions (x, y, z) of the puffs, so there are four parameters for each puff, and the length of the state vector is $n = 4N_{puff}$, where N_{puff} is the number of the released puffs. The measurements are the ground level volumetric activity concentrations, so the observation model $H(\boldsymbol{x})$ is nonlinear.

2.1. Dispersion model

The dispersion model used here is a Lagrangian puff-model, a widely used standard approach for studying the transport of airborne contaminants due to turbulent diffusion and advection [36,37]. It has advantages in predicting atmospheric dispersion close to source [38]. The model introduced in [14] is used, and the Doury's parameterization is adopted to estimate the standard deviation due to the turbulence. The expression of the total standard deviation is the sum of the spread due to turbulence (σ_{xturb}, σ_{yturb}, σ_{zturb}), the additional spread due to plume rise (σ_{ypr}, σ_{zpr}), and the initial spread due to the diameter of d_s of the source [39]:

$$\sigma_x^2 = \sigma_y^2 = \sigma_{yturb}^2 + \sigma_{ypr}^2 + d_s^2/4 \tag{3}$$

$$\sigma_z^2 = \sigma_{zturb}^2 + \sigma_{zpr}^2 \tag{4}$$

The initial spread due to plume rise is estimated following [40] for the crosswind spread and [41] for the vertical spread:

$$\sigma_{ypr} = \Delta h/3.5 \tag{5}$$

$$\sigma_{zpr} = \Delta h/2 \tag{6}$$

where Δh is the plume rise. The modifications of radioactive decay and deposition are also added to the model.

2.2. Ensemble Kalman filter

Ensemble Kalman filter (EnKF) [34] is a sequential data assimilation method, which is used recursively to produce a statistically optimal estimate of the underlying system state by merging the model prediction and the current observations. Different from the original Kalman filter, which is only applicable to linear system model, the EnKF initializes an ensemble of forecast models, each of which represents a possible state of the system and the error statistics are predicted using the collection of model states. The ensemble state vectors constitute the state matrix:

$$\boldsymbol{X} = (\,\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \ldots \quad \boldsymbol{x}_N\,) \in \mathbf{R}^{n\times N} \tag{7}$$

where the subscript i ($i = 1, 2, \ldots, N$) denotes the individual state vector of the ensemble and N denotes the size of the ensemble. EnKF is also applicable to nonlinear system models. The ensemble predictions constitute the prediction matrix, defined as

$$\boldsymbol{Y}^f = (\,H(\boldsymbol{x}_1^f) \quad H(\boldsymbol{x}_2^f) \quad \ldots \quad H(\boldsymbol{x}_N^f)\,) \in \mathbf{R}^{m\times N} \tag{8}$$

$$\overline{\boldsymbol{Y}^f} = \boldsymbol{Y}^f \mathbf{1}_N \tag{9}$$

where $\mathbf{1}_N$ is $N \times N$ matrix, with each element equaling to $1/N$, $\overline{\boldsymbol{Y}^f}$ is the mean ensemble measurement prediction matrix. For the nonlinear observation operators, the state vector can be extended to include both the original vector $\boldsymbol{x}$ and the observation predictions $H(\boldsymbol{x})$ as in [42]. The nonlinear problem is reduced to a linear one. But the analysis result is only an approximation to the optimal estimate of the state, because the valid states of system only occupy a submanifold of augmented state space instead of the whole space [42]. The performance of EnKF will deteriorate. It is found that in [27], when nonlinear parameters (wind direction and turbulence intensity) are included in the state vector, it takes several data assimilation steps to converge to the true values. The time delay between *a priori* and the converged values can be as large as 5 h.

3. Modified ensemble Kalman filter

In this study, we introduce three new modifications to improve the performance of the ensemble Kalman filter. In the original EnKF, the analysis step is conducted only once, as shown in Fig. 1A, which is insufficient for the nonlinear observation model. In the modified algorithm, an iteration cycle is introduced, which makes the posterior state gradually converge to the actual condition. As shown in Fig. 1B, there are three main modifications: convergence criterion, increment filter, and resample. In the following parts, the subscript k and i are omitted to be concise, because the iteration cycle is all conducted in the current assimilation step k, and only one ensemble member is involved. The subscript h is introduced to indicate the step of the iteration.

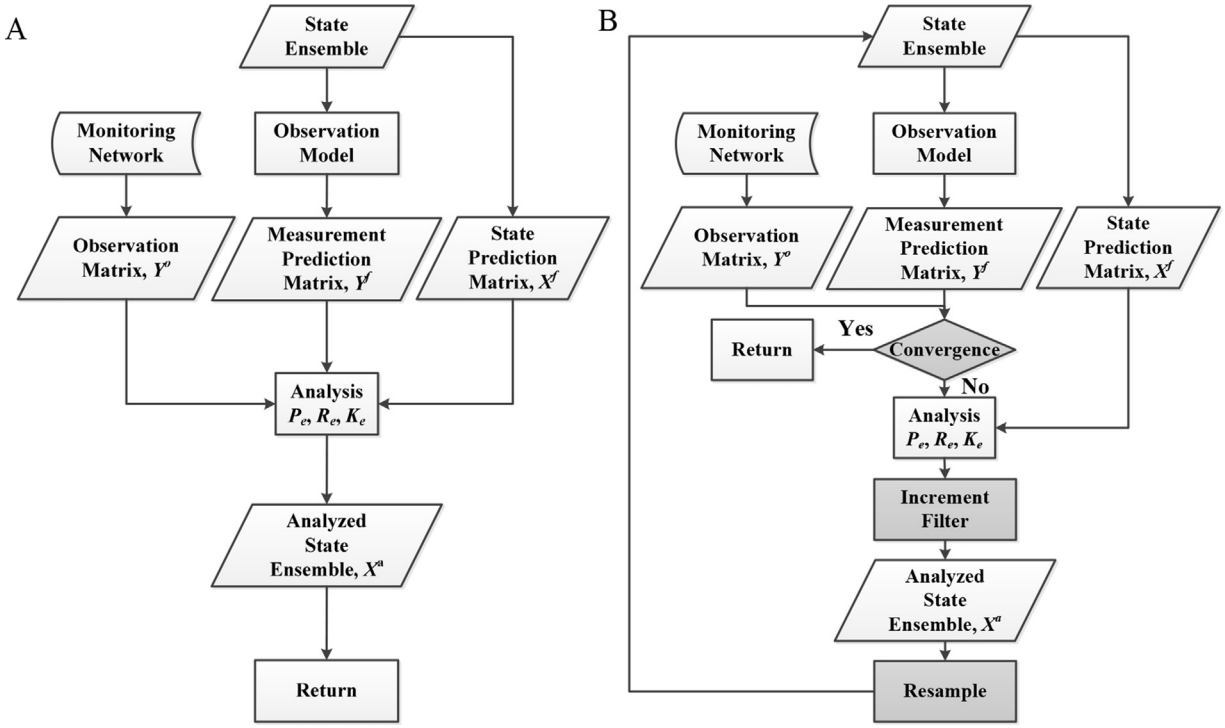

Fig. 1. The flowchart of (A) the EnKF, and (B) the modified EnKF data assimilation method.

3.1. Convergence criterion

The deviation between the predictions $\boldsymbol{y}^f$ and the measurements $\boldsymbol{y}^o$ can be calculated to judge whether the convergence is achieved. The deviation is defined as:

$$\boldsymbol{Err}_h = \left(\overline{\boldsymbol{y}_h^f} - \boldsymbol{y}^0\right) \in \boldsymbol{R}^m \tag{10}$$

where $\overline{\boldsymbol{y}_h^f}$ is one column of the mean ensemble measurement prediction matrix $\overline{\boldsymbol{Y}^f}$ defined in (9). The convergence or the stop criterion is defined as

$$\frac{\left\|\boldsymbol{Err}_h\right\|}{\left\|\boldsymbol{y}^0\right\|} \le C_{\text{factor}} \quad \text{or} \quad h = N_{\text{iteration}} \tag{11}$$

where $\left\|\boldsymbol{Err}_h\right\|$ and $\left\|\boldsymbol{y}^0\right\|$ are the L^2-norm of the error and the measurement at the h-th iteration step, and C_{factor} is the criterion for the convergence. The criterion is set as 0.1, because the relative error eventually stabilizes at around 0.1 (changes within 10% on average) in the situation of this study, as shown in Fig. 2. N_{iteraion} is the maximum number of iteration steps (50 in this study).

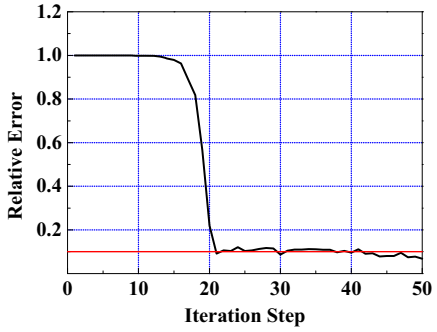

Fig. 2. Typical trajectory of the relative error during the iterations.

3.2. Increment filter

When the predicted state vectors are far away from the actual value, the analyzed state vector may be over-modified. Numerical vibration will happen, and it will be amplified during the iteration. To avoid it, the increment between the predicted state vector and the analyzed one is restricted, which is called increment filter here. The increment filter is conducted after the analysis step of the original EnKF. The content Q and the vertical position z are restricted by an increment factor:

$$Q_{j,h}^a = \max\left(f_{low} \times Q_{j,h}^f, \min\left(Q_{j,h}^{a*}, f_{high} \times Q_{j,h}^f\right)\right) \tag{12}$$

The subscript j indicates the number of this puff. The time step between two successive puffs j and $j-1$ is 30 s. $Q_{j,h}^f$ means the predicted content Q of the j-th puff at the h-th iteration step. f_{low} and f_{high} are the low and high restriction factors for Q, which are 0.5 and 2. $Q_{j,h}^{a*}$ is the analyzed content from the original EnKF, as shown in Fig. 1B, and $Q_{j,h}^a$ is the restricted content by the increment filter. The vertical position z is updated in the same way. For the horizontal position (x, y), the analyzed position is restricted in a "Restriction Cycle", as shown in Fig. 3. The radius of the restriction cycle r_j is defined as

$$r_j = v_{ref} \times \Delta t_j \tag{13}$$

$$\Delta t_j = \min(t_j, \Delta T) \tag{14}$$

where v_{ref} is the reference velocity, which is 5 m/s here. ΔT is the time between two successive assimilation moments, which is 30 min, and t_j is the travel time of puff j after release. The update process is as follows:

$$\Delta \boldsymbol{r}_{j,h}^a = \left(x_{j,h}^{a*}, y_{j,h}^{a*}\right) - \left(x_{j,h}^f, y_{j,h}^f\right) \tag{15}$$

$$\left(x_{j,h}^a, y_{j,h}^a\right) = \frac{\Delta \boldsymbol{r}_{j,h}^a}{\left\|\Delta \boldsymbol{r}_{j,h}^a\right\|} \times \min\left(\left\|\Delta \boldsymbol{r}_{j,h}^a\right\|, \boldsymbol{r}_j\right) + \left(x_{j,h}^f, y_{j,h}^f\right) \tag{16}$$

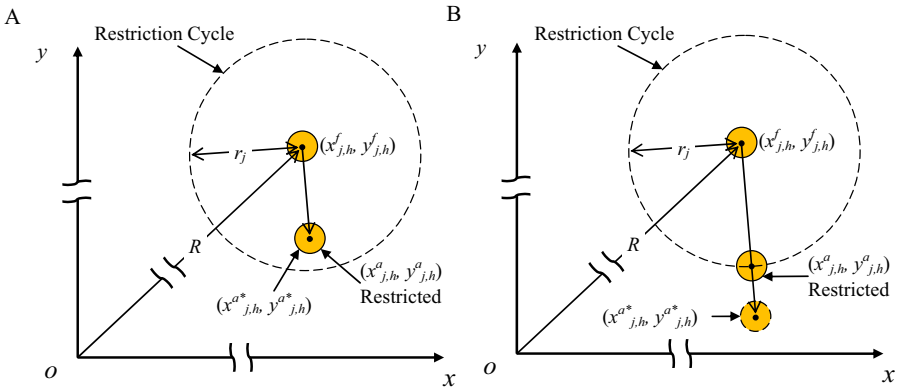

Fig. 3. The increment restriction for analyzed horizontal position (x^a, y^a). (A) The horizontal position of the puff is updated as the analyzed results, if the distance between the analyzed position $\left(x^{a*}_{j,h}, y^{a*}_{j,h}\right)$ from (25) and the predicted position $\left(x^f_{j,h}, y^f_{j,h}\right)$ is smaller than r_j; otherwise, (B) the increment is restricted as the radius of the cycle.

If the distance between the analyzed position $\left(x^{a*}_{j,h}, y^{a*}_{j,h}\right)$ and the predicted position $\left(x^f_{j,h}, y^f_{j,h}\right)$ is smaller than r_j, then the horizontal position of the puff is updated as the analyzed results, as shown in Fig. 3A. Otherwise, the increment is restricted as the radius of the cycle, and the position of the puff is updated as the restricted position $\left(x^a_{j,h}, y^a_{j,h}\right)$, as shown in Fig. 3B.

3.3. Resample

After analysis, the differences among the ensemble members will become smaller, which causes the divergence of the filter, and the performance will deteriorate. The resample procedure is introduced to keep the variety of the ensemble. The mean analyzed state vector at iteration step h is calculated as:

$$\overline{\boldsymbol{x}^a_h} = \boldsymbol{X}^a_h \times (1/N \quad 1/N \quad \dots \quad 1/N)^{\mathrm{T}} \in \mathbf{R}^n \tag{17}$$

where $\overline{\boldsymbol{x}^a_h}$ is the mean analyzed state vector at the h-th iteration step, $\boldsymbol{X}^a_h$ is the analyzed state matrix updated by the increment filter.

A new ensemble of state vectors for the next iteration is generated by adding perturbations to the mean analyzed state vector $\overline{\boldsymbol{x}^a_h}$. We use P to represent one of the four puff state parameters. Then, the parameters of the j-th puff for the next iteration step $h+1$ can be expressed as:

$$P^f_{j,h+1} = \overline{P^f_{j,h}} + F_{P,j,h} \times D_h \times P^{ref}_{j,h} \tag{18}$$

where $F_{p,j,h}$ is the scaling factor for parameter P, and D_h is the decay factor, which is defined as:

$$D_h = \frac{\left\|\boldsymbol{Err}_h\right\|}{\left\|\boldsymbol{y}^0\right\|} \tag{19}$$

When the discrepancy between the prediction and observation is large, the new ensemble spread of state vector will also be large. Otherwise, the spread will be small. The decay factor helps accelerate the convergence process. $P^{ref}_{j,h}$ is the reference value.

The scaling factor is assumed to be puff-correlated due to the assumption that the parameter should not change greatly between two successive puffs. We use the method in [29] to simulate a sequence of correlated scaling factors:

$$F_{P,j,h} = \alpha F_{P,j-1,h} + \sqrt{1-\alpha^2} w_{P,j,h} \tag{20}$$

where α determines the correlation between the scaling factors, $w_{P,j,h}$ is a white noise with mean value zero. For puff content Q and height z, the initial scaling factors follow the same uniform distribution, which is $F_{Q(z),1,h} \sim \mathrm{Un}(0.9, 1.1)$. Then, the sequence will be generated using the white noise, which is $w_{Q(z),j,h} \sim \mathrm{Un}(-0.1, 0.1)$, according to (20). Here, α is 0.9 for Q and z, and the reference values $Q^{ref}_{j,h}$ and $z^{ref}_{j,h}$ are the mean analyzed state variables after the increment filter $\overline{Q^a_{j,h}}$, $\overline{z^a_{j,h}}$.

The scaling factors for horizontal position are a little complicated. Two intermediate factors are introduced, scaling factor for angle $F_{\theta j,h}$ and scaling factor for speed $F_{v,j,h}$, with initial value $F_{\theta j,h} \sim \mathrm{Un}(-\pi, \pi)$ and $F_{v,j,h} \sim \mathrm{Un}(-1, 1)$. The white noise follows the same distribution, $w_{\theta,j,h} \sim \mathrm{Un}(-\pi, \pi)$ and $w_{v,j,h} \sim \mathrm{Un}(-1, 1)$. Here, α is 0.99 for the two factors, because the space correlation of wind field is usually very strong. Then the scaling factors for horizontal position x and y are calculated as:

$$F_{x,j,h} = F_{v,j,h} \cos(F_{\theta,j,h}) \tag{21}$$

$$F_{y,j,h} = F_{v,j,h} \sin(F_{\theta,j,h}) \tag{22}$$

The reference values $x^{ref}_{j,h}$ and $y^{ref}_{j,h}$ are all assumed to be $0.5\Delta t_j$, where Δt_j is expressed as (14).

4. Experiment setup

4.1. Twin experiment

Twin experiment is widely used to evaluate and verify the performance of a newly developed data assimilation method [43]. One advantage of twin experiments is that the true values of control variables are known, so we can test whether the calculated values converge to the true values [44,45]. During the development of the real time on-line decision support system (RODOS) for European off-site nuclear emergency management, twin experiments were conducted to evaluate the performance of the data assimilation module [9–11]. Twin experiments are also used to evaluate the performance of recently developed data assimilation methods for nuclear power plant accident management, *e.g.* in [27,29–31]. The procedure is illustrated in Fig. 4. There are two parts in the twin experiments: a control run, and a data assimilation run. In the control run, a certain set of parameters are used as the input of the model. In our study, these parameters include source release rate, plume rise height, wind speed and direction. The synthetic observation data will be generated by several virtual detectors during the control run. In the data assimilation run, *a priori* input parameter set is used, and the observations from the control run will be assimilated using the modified EnKF method. The degree of how much the state and parameters of the control run can be recovered provides a measure of the skill of the modified EnKF data assimilation scheme [46]. In our study, the twin experiments cover 10 h,

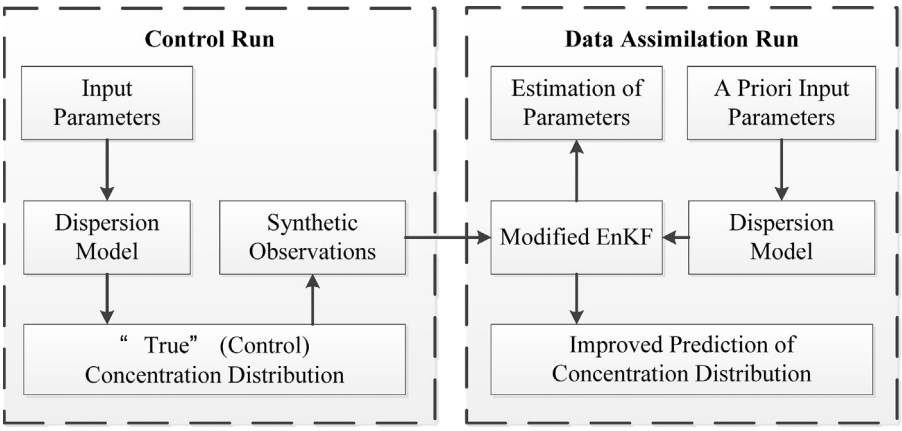

Fig. 4. Schematic diagram of the twin experiment.

and the modified EnKF data assimilation method is conducted every 30 min. The time step of the atmospheric dispersion model is 30 s.

4.2. Monitoring network

We considered a hypothetical scenario of a core meltdown accident at the Hongyan River Nuclear Power Station, which is under construction in Liaoning Province, China, as shown in Fig. 5A. Species ^{137}Cs was selected for this study, because it is responsible for most of the radiation exposure received by the general population. The break of the reactor is assumed to be 30 m high. The layout of the virtual monitoring stations network is shown in Fig. 5B. A grid of 81 monitoring stations is placed in the downwind direction of the NPP. The stations are divided into two groups. Monitoring Group I is about 1 km away from the NPP, including 36 stations spaced evenly on a square grid at 2 km apart and 2 m above the ground. Monitoring Group II is relative sparse, with 20 stations in a 2 km × 4 km grid, and 25 stations in a 4 km × 4 km grid, 2 m above the ground. The data acquisition frequency of the monitoring network is 30 min. We assume that all the detectors are homogeneous, which give ground level volumetric activity concentrations. The resolution of the detectors is 10^{-3} Bq/m^3.

There is a so-called "inverse crime" [47] in the inverse studies, which means that the models used in the control and data assimilation run are the same, and the model prediction is identical to synthetic observation. To avoid this problem, the averaged concentration between the two successive data assimilation moments t_{k-1} and t_k is used as the instantaneous observation at time t_k. And it is also disturbed by an uniform distribution error Un(-10^{-3}, 10^{-3}) Bq/m^3. The instantaneous concentrations predicted by the model and the synthetic observations are not identical at the assimilation moment, and the artificial errors of all the data are about 30% on average, as shown in Fig. 5C.

4.3. Test cases

A simplified temporal profile of ^{137}Cs release rate used in the control run is shown in Fig. 6A. The release lasts about 500 min. Two different types of changes are assumed: a discontinuous increase of release rate which happens at 100 min due to the explosion and a gradual continuous decrease of release rate which lasts from 250 to 500 min due to the countermeasures. The plume rise height follows the same form as shown in Fig. 6B. The unsteady meteorological conditions during the experiment are shown in Fig. 6C. The wind changes gradually from west wind to north wind. The wind speed increases to about 5.5 m/s at 375 min, and then it gradually decreases to about 4 m/s at 600 min.

Five different cases (Case 0–Case 4) are designed to evaluate the effectiveness of the modified EnKF, as shown in Table 1. In these cases, *a priori* release rate and plume rise height remain constant at 10 Bq/s and 30 m respectively, and the deviations of wind direction and speed between *a priori* and true values are as large as ±25 degrees, and ±2 m/s respectively.

5. Results and discussion

5.1. Improvement of the prediction

Fig. 7 gives the comparison of the concentration distribution prediction from the control run and those from the five cases without data assimilation method. The general shapes of the plume from the five cases are very different from that in the control run due to the large uncertainties of input parameters. The magnitudes also have great deviation. Note that the color bar for the control run ranges from 0 to 4 Bq/m^3, however the ranges for the five experiment cases are 0–2 × 10^{-5} Bq/m^3, because the *a priori* the release rates are 5–6 magnitudes lower than that used in the control run.

Fig. 8 presents the comparison between the control run and the reconstructed concentration distribution at 2, 4, 6 and 8 h after the accident. As shown in the first and second columns, the concentration fields reconstructed by the modified EnKF method are consistent with those in the control run. In the third column, the standard deviation fields present the extension of all the 25 reconstructed fields. Notice that the range is only from 0 to 0.8 Bq/m^3, so the differences between the 25 the converged concentration fields are small. It seems that the high value areas around the centerline are composed of a series of discrete hot spots, especially at 4 and 6 h. This is caused by the discrete puffs in the dispersion model, because a puff influences a disk area around it, which is easy to form hot spots. Shortening the time step between puffs may alleviate the discrete phenomenon. The bias between the mean reconstructed concentration field and that of the control run is shown in the last column. The bias is very small in most area. There is a positive bias near the centerline of the plume, which means the concentration is slightly overestimated in this area, and the situation is inversed on the flanks of the plume. In the study of Zheng et al. [29], the concentration was also slightly overestimated at the station on Kat O Island, which is close to the centerline. The tendency of overestimation is caused by the high concentration plume in this area, which "attracts" simulated material to accumulate to make the predicted concentration closer to the observation during the data assimilation run.

In order to quantify the effectiveness of the modified EnKF method, we first compare the DA analysis concentration against

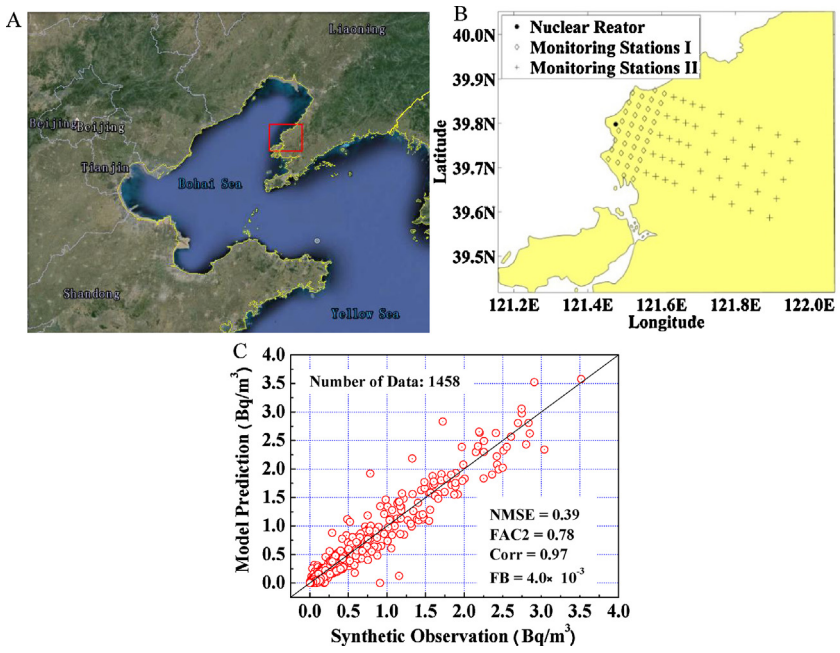

Fig. 5. (A) The location of the Hongyan River Nuclear Power Plant. (B) Layout of the virtual network of monitoring stations. (C) The comparison between the total 1458 synthetic observations and the model predictions.

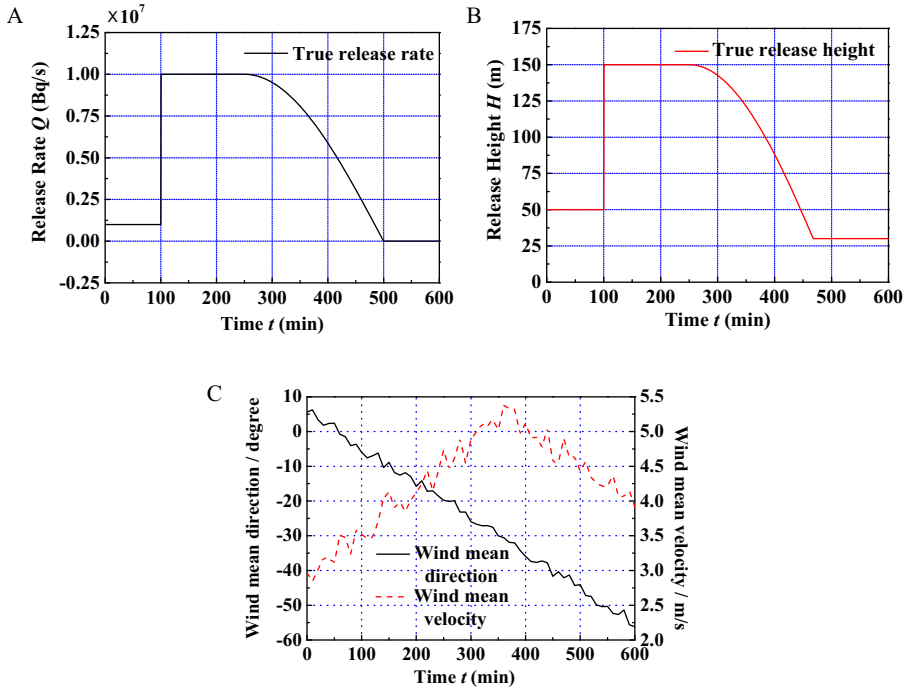

Fig. 6. Temporal profile of the "true" parameters used in the control run. (A) Release rate of the source. (B) Plume rise height. (C) Mean wind direction and speed.

Table 1
The uncertainties of *a priori* parameters in the 5 twin-experiment cases.

Case	*A priori* and true wind speed deviation (m/s)	*A priori* and true wind direction deviation (degree)	*A priori* release rate (Bq/s)	*A priori* plume rise height (m)	Times of run
0	0	0	10	30	5
1	−2	−25	10	30	5
2	−2	25	10	30	5
3	2	−25	10	30	5
4	2	25	10	30	5

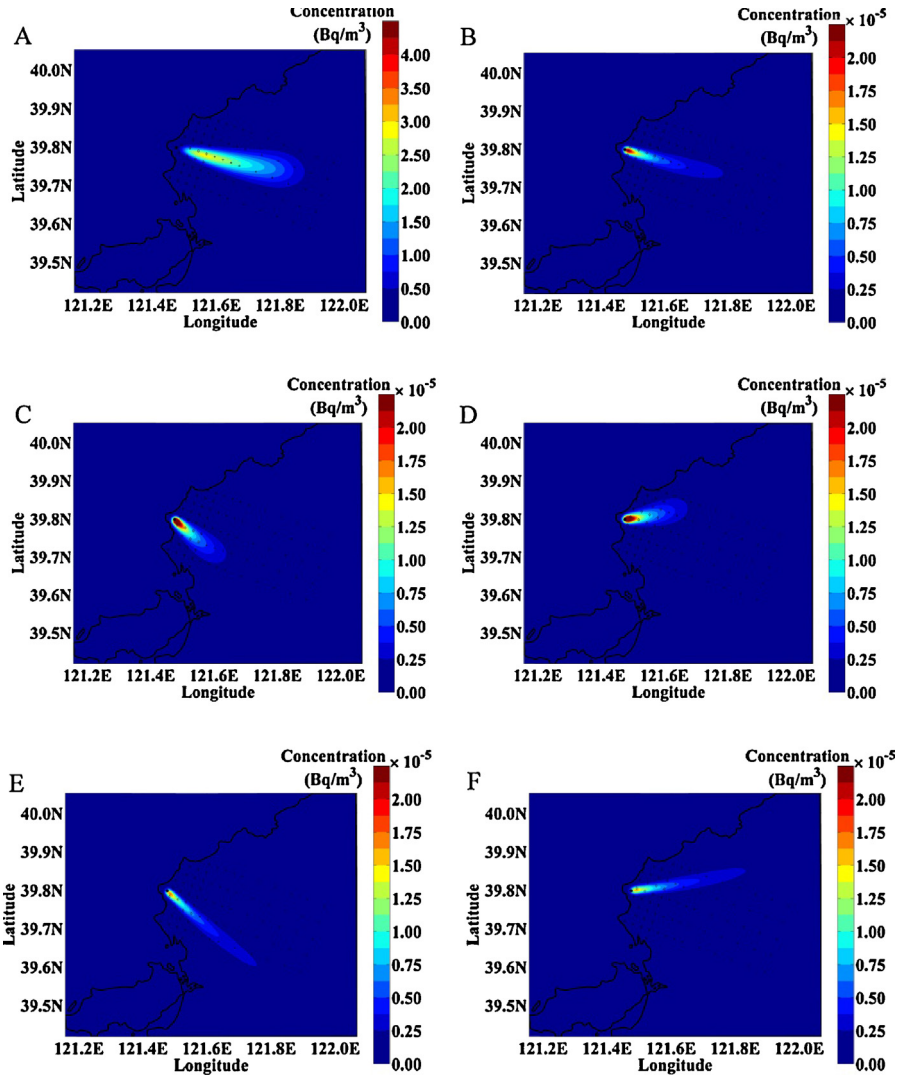

Fig. 7. Comparison of ground level concentration 4 h after the release, (A) the control run, (B) Case 0 (Table 1) without data assimilation (DA) method, (C) Case 1 without DA, (D) Case 2 without DA, (E) Case 3 without DA, (F) Case 4 without DA.

the observations that are assimilated, which is known as "sanity check". We use the comprehensive model performance measures for air quality and dense-gas dispersion models proposed by Hanna et al. [48] and American Society for Testing and Materials (ASTM) [49]. Four statistical performance measures are used: the normalized mean square error (NMSE), the fraction of predictions within a factor two of observations (FAC2), the correlation coefficient (Corr), and the fraction bias (FB), which are defined as

$$\mathrm{NMSE} = \frac{\overline{(P_{\mathrm{ctrl}} - P_{\mathrm{assim}})^2}}{(P_{\mathrm{ctrl}} P_{\mathrm{assim}})} \tag{23}$$

$$\mathrm{FAC2} = \text{fraction of data that satisfies} \quad 0.5 \le \frac{P_{\mathrm{assim}}}{P_{\mathrm{ctrl}}} \le 2.0 \tag{24}$$

$$\mathrm{Corr} = \frac{\overline{(P_{\mathrm{ctrl}} - \overline{P}_{\mathrm{ctrl}})(P_{\mathrm{assim}} - \overline{P}_{\mathrm{assim}})}}{\sigma P_{\mathrm{ctrl}} \sigma P_{\mathrm{assim}}} \tag{25}$$

$$\mathrm{FB} = \frac{(\overline{P}_{\mathrm{ctrl}} - \overline{P}_{\mathrm{assim}})}{0.5(\overline{P}_{\mathrm{ctrl}} + \overline{P}_{\mathrm{assim}})} \tag{26}$$

where P is the parameter, which can be concentration, release rate and plume rise height in this study, the subscript "ctrl" indicates the control run, and "assim" means the assimilation run, σ is the standard deviation. In the study of Chang et al. [50] and references therein, numerous applications of dispersion model and many field experiment data sets are reviewed, and it concludes that "good" performing models appear to have some typical characteristics based on comparisons, and the criteria for an acceptable model are considered as: (i) the mean bias should be with ±30% of the mean ($-0.3 \le \mathrm{FB} \le 0.3$); (ii) the fraction of predictions within a factor two of the actual values should be larger than 50% ($\mathrm{FAC2} \ge 0.5$); (iii) the random scatter should be small enough ($\mathrm{NMSE} \le 4$). The results are shown in the scatter diagrams Fig. 9A. The predictions after data assimilation are well consistent with the observations. The four statistical performance measures are better than those in Fig. 5C, which means the forcing model errors are also be successfully suppressed, and the method has positive impact on the improvement of predictions. Another set of observations from the locations 1 km to the east of the monitoring stations, which are not assimilated, are also used to examine the performance the method, as shown in Fig. 9B. Although the standard deviations (shown as error bar in the diagram) are a little larger than those in Fig. 9A, the four statistical performance measures are also very optimistic. These results show that this modified EnKF method can effectively assimilate the monitoring observations, and improve the model predictions.

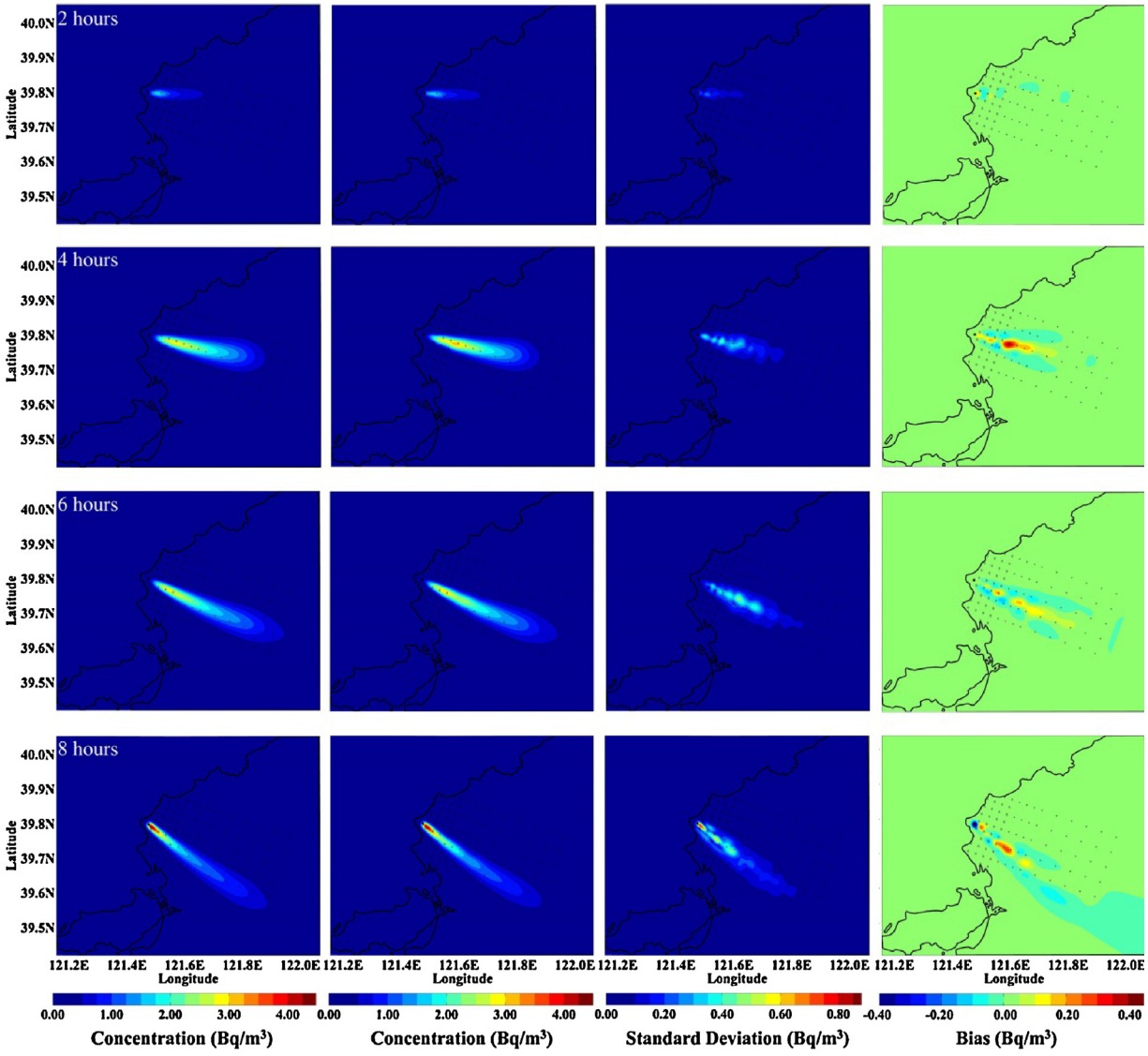

Fig. 8. Comparison between the control and the reconstructed distribution at 2, 4, 6 and 8 h after the accident happens. The first column shows the averaged concentration between two successive data assimilation in the control run, the second column shows the mean distribution of the 25 reconstructed concentration fields, the third column shows standard deviation field, and the bias between the mean reconstructed concentration field and that of the control run is shown in the last column.

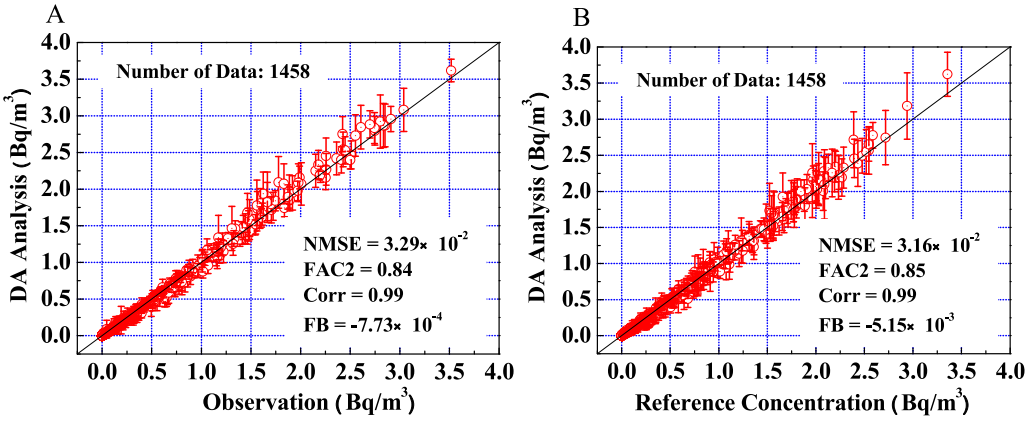

Fig. 9. Scatter plots of the reconstructed concentrations against (A) the synthetic observations from the monitoring stations used for data assimilation, (B) observations from 1 km to the east of the stations, not used for the data assimilation. The error bar is the standard deviation of the 25 runs.

5.2. Estimation of the parameters

Another purpose of the modified EnKF method is to estimate the source term. After the k-th data assimilation, the averaged content and height of the newly released puffs between the $(k-1)$-th and the k-th data assimilation moments are calculated to estimate the source release rate and plume rise height. All the five cases in Table 1 are also conducted by the original EnKF to clearly show the performance improvement of the modified method. In the original EnKF, the errors of wind field are constantly updated, following the method proposed in [27].

Fig. 10 shows the reconstructed source parameters using the original and modified EnKF method. The original EnKF fails to capture the sharp increase at 100 min after the release, and it takes about 300 min to reach the summit. The reconstructed plume rise height also significantly fluctuates before 300 min. The results of the original EnKF are comparable to those in [27], where it takes 6 data assimilation cycles (6 h) to find the true error in the wind direction. The results clearly show the substantial time delay of the original EnKF, when nonlinear parameters (wind direction and speed) are included in the state vector.

The modified EnKF successfully reconstructs the source parameters. Both of the sudden increase and the continuous decrease phase are well captured. The reconstructed source terms return to the "actual" values almost immediately after the first data assimilation cycle, in spite of the large uncertainties of the *a priori* parameters (Table 1). The time delay following the sudden increase has been also shortened to two data assimilation cycles (1 h). Table 2 shows the statistical performance comparisons between the original and modified EnKF method. The uncertainties in the parentheses are the standard deviations. The smaller uncertainties of the modified EnKF (57% reduction on average for release rate, and 73% reduction on average for plume rise height) indicate that the new method is more robust and reasonably insensitive to the errors of the wind velocities. For the release rate, the correlation coefficient is improved by 35% (from 0.54 to 0.73) by the modified EnKF, and the FAC2 is also improved by 16% (from 0.55 to 0.64). The normalized mean square error is reduced by 58% (from 0.66 to 0.28), and the fractional bias is also reduced by 62% (from 0.34 to 0.13). The modified EnKF also substantially improves the Corr, FAC2 and NMSE of the reconstructed plume rise height. The fractional bias becomes larger, but it does not mean the performance deteriorates, because FB is overall bias: the plume rise height is overestimated during the first 100 min, and then it is underestimated by the original EnKF, great biases respectively exist during the two periods, but the compensating error makes a small FB (0.01) on average.

Both of the reconstructed parameters are slightly underestimated (28% for release rate and 25% for plume rise height on average) during 100–300 min, but overestimated after 400 min. It is due to the error compensating effect of the two parameters: underestimated release rate reduces the ground level concentration, but the lower rise height increases it, and *vice versa.* The two different errors compensate each other, making small concentration error. Namely, different combinations of these parameters can give similar distribution. It will be hard for EnKF to identify which one is the optimal solution [27]. However, all the possible combinations should not be far away from the real one, as long as the monitoring stations provide enough information. The reconstructed parameters are substantially underestimated at the fourth data assimilation, which is only 20 min after the sudden increases, because the ground level concentration measurements change little due to the simultaneous increase of the two compensative parameters during such a short time. The information of the sudden changes has not been completely transported to the monitoring network, and the constraints provided by the measurements are insufficient to identify the changes. At the following data assimilation cycles, the reconstructed results are much better due to the arrivals of extra measurements.

Fig. 11 presents the four statistical performance measures for the reconstructed source terms in the five cases. The acceptable boundaries are shown as dash lines and all the measures locate within the boundaries. The correlations are beyond 0.6 for all the cases, showing strong consistency between the reconstructed and the "actual" values. The mean fraction biases are between −0.3 and 0.3, but the positive mean values indicate underestimation of the parameters. As discussed above, the underestimation is caused by the error compensating effect: lower release leads to negative concentration error, but lower rise height makes positive concentration error, which reduces the total error and makes the model predictions close to the measurements, although both parameters are underestimated. The FAC2 are all beyond 0.6 and NMSE are substantially smaller than the acceptable boundary, indicating the reconstructions are close to the "actual" values. The performance is stable for the five cases, only with a little deterioration in case 3 and case 4, where the wind speed is overestimated. Generally speaking, the modified EnKF method successfully reconstructs the source parameters in the scenarios considered here.

5.3. Influence of a priori source parameters

We use another five cases (Case I0–I4) to further investigate the effects of *a priori* source parameters. The *a priori* release rate and plume rise height are respectively changed to 10^4 Bq/s and 200 m, as shown in Table 3. All the five cases are conducted twice. Fig. 12A and B shows that the reconstructed concentrations after data assimilation are well consistent with the observations, similar to the results in Fig. 9, although the initial release rate is three magnitudes larger and plume rise height is seven times higher. Fig. 12C and D indicate that the reconstructed source parameters are better than those in Case 0–4, especially between 100 and 300 min, because the *a priori* plume rise height is 200 m in Case I0–I4, which is closer to the actual value 150 m. As a direct influence, the plume rise height is better estimated, which consequently leads to a better estimation of the release rate. After 300 min, the actual plume rise starts to decrease, and the performance becomes comparable to that in Fig. 10. After the end of the emission, the overestimations are larger than those in Fig. 10 due to the large *a priori* value of plume rise height. Generally speaking, the method successfully reconstructs the concentration field and source parameters regardless of the substantial differences of the *a priori* source parameters (10–10^4 Bq/s for *a priori* release rate and 30–200 m for *a priori* plume rise height).

6. Limitations and suggestions

Compared to the models in real-life, the model in the data assimilation run is still relatively "perfect" as shown in Fig. 4, although substantial artificial errors (about 30% on average) are added into the synthetic observations to avoid "inverse crime". From this perspective, the twin experiments only theoretically assess the algorithmical capabilities and improvements of the modified EnKF method, and it can be interpreted as optimistic since the further complications of errors in real-life scenario are ignored [51].

There are several possible errors which may influence the performance of the method in practical applications: (1) Variability of random turbulence. The standard deviation due to the turbulence is an important parameter in Lagrangian puff-model, and we adopt a relatively simple parameterization method (Doury's parameterization). In the practical applications, more sophisticated method, such as similarity theory based parameterization method,

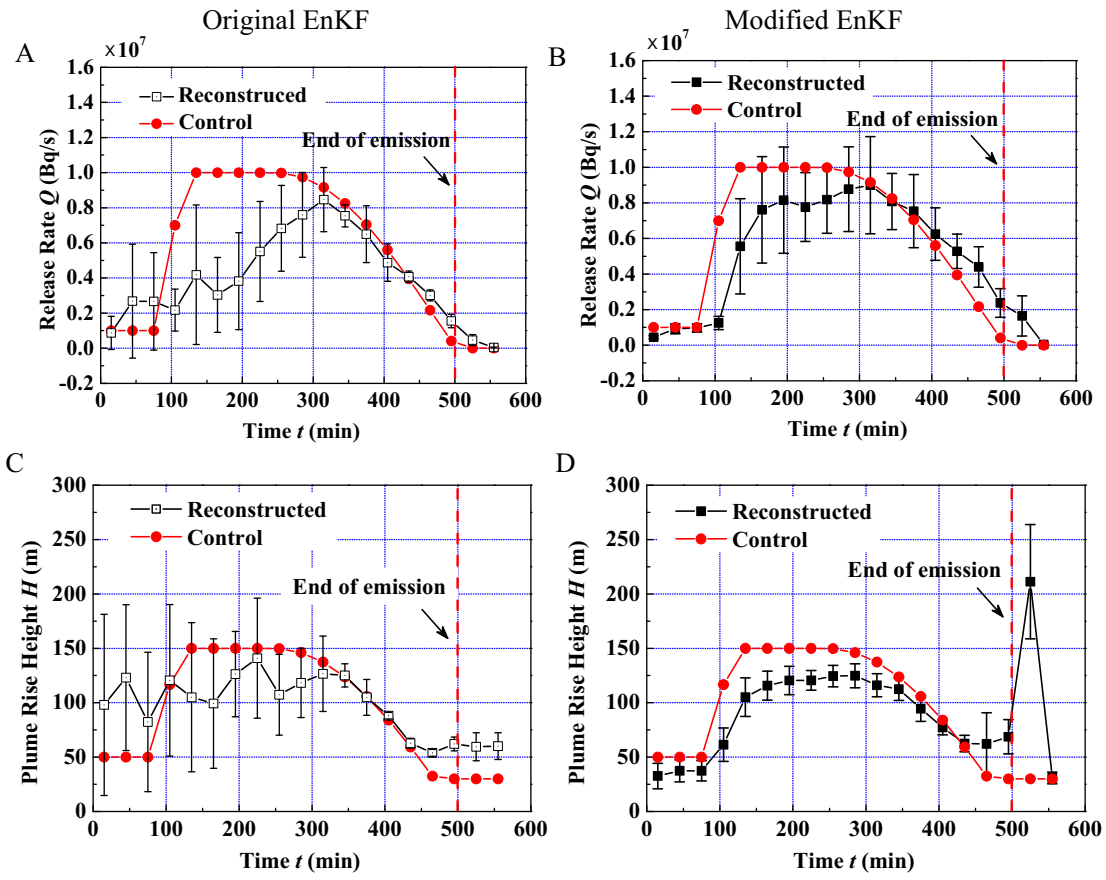

Fig. 10. Comparisons between the reconstructed and the control source terms: the first and second columns are respectively reconstructed by the original and the modified EnKF, (A) and (B) the radioactive material release rate; (C) and (D) the plume rise height. The point is the mean value of the 25 runs, and the error bar is the standard deviation. The control data is averaged by 30 min.

Table 2
Performance improvement of the modified EnKF (Uncertainties in the parentheses).

Statistical performance measures	Release rate			Plume rise height		
	Original EnKF	Modified EnKF	Performance improvement	Original EnKF	Modified EnKF	Performance improvement
Corr	0.54 (0.26)	0.73 (0.06)	35% (−77%)	0.38 (0.30)	0.78 (0.09)	105% (−70%)
FAC2	0.55 (0.11)	0.64 (0.06)	16% (−45%)	0.79 (0.10)	0.92 (0.04)	16% (−60%)
FB	0.34 (0.14)	0.13 (0.10)	−62% (−29%)	0.01 (0.21)	0.16 (0.03)	None (−86%)
NMSE	0.66 (0.40)	0.28 (0.09)	−58% (−78%)	0.31 (0.16)	0.11 (0.04)	−65% (−75%)

should be used to improve the accuracy of the model. (2) The meteorological data errors. The modified EnKF method considers the uncertainty of wind velocity, but other meteorological data, such as the heights of inverse layer, boundary layer and the friction velocity may also influence the accuracy of the model. (3) Representativeness of the monitoring network. The monitoring network should be located in downwind direction of the release, otherwise it will fail to capture the plume, and lead to the failure of data assimilation.

We are developing more sophisticated dispersion modeling system using outputs from Weather Research and Forecasting (WRF) Model. The new system will give more accurate predictions under the real meteorological condition. The modified EnKF method in conjunction with the new dispersion modeling system will be

Table 3
The uncertainties of *a priori* parameters in the 5 twin-experiment cases (I0–I4).

Case	*A priori* and true wind speed deviation (m/s)	*A priori* and true wind direction deviation (degree)	*A priori* release rate (Bq/s)	*A priori* plume rise height (m)	Times of run
I0	0	0	10^4	200	2
I1	−2	−25	10^4	200	2
I2	−2	25	10^4	200	2
I3	2	−25	10^4	200	2
I4	2	25	10^4	200	2

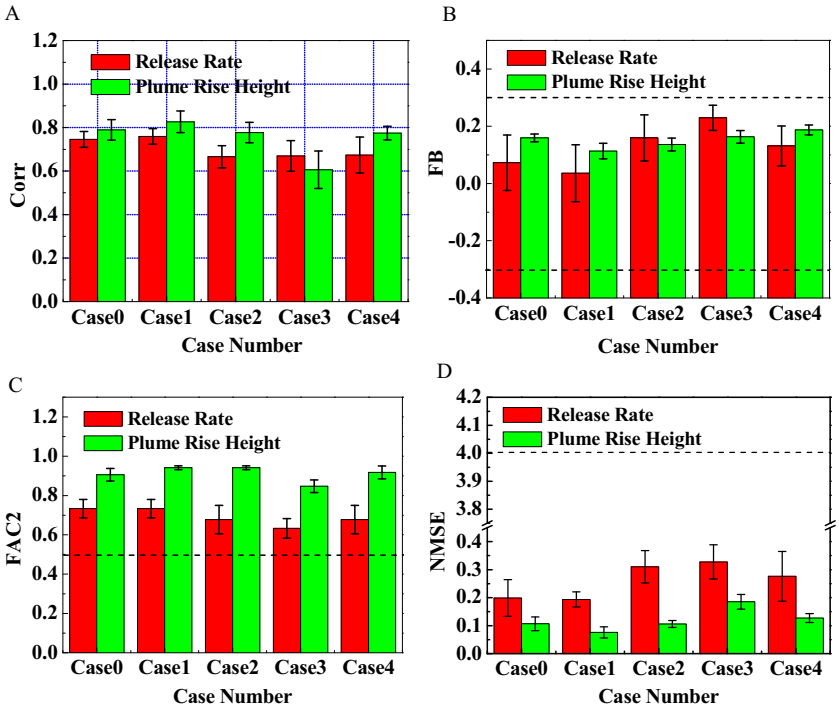

Fig. 11. Statistical performance measures for the reconstructed source release rate and plume rise height: (A) the correlation coefficient (Corr), (B) the fraction bias (FB), (C) the fraction of predictions within a factor two of controlled values (FAC2), and (D) the normalized mean square error (NMSE). The error bar is the standard deviation of 5 runs.

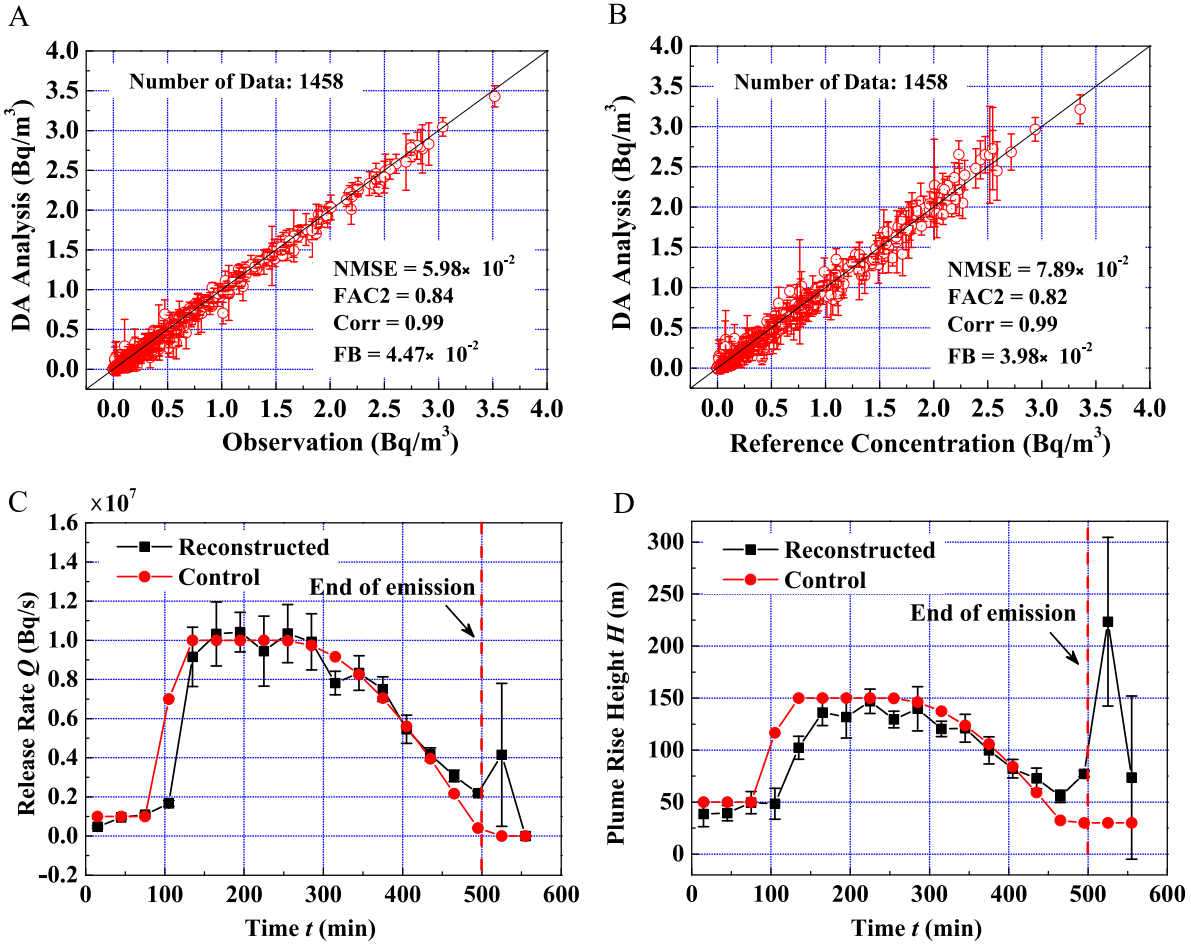

Fig. 12. Reconstructed concentrations and source parameters from Case I0 to I4. (A) and (B) scatter plots of the reconstructed concentrations, same as Fig. 9. (C) and (D) reconstructed source parameters, same as Fig. 10.

further evaluated using the field experiments data, *e.g.* Kincaid experiment data sets [41].

7. Conclusion

In this study, a modified ensemble Kalman filter data assimilation method in conjunction with a Lagrangian puff-model, is proposed to simultaneously improve the model prediction and reconstruct the source terms for short range (about 50 km) atmospheric dispersion in nuclear power plant accident. Four main uncertain parameters are considered: source release rate, plume rise height, wind speed and wind direction. Twin experiments are conducted to evaluate the effectiveness of the method. The results show that the modified EnKF method effectively improve the predicted concentration distribution with optimistic statistical performance measures: NMSE $\sim 10^{-2}$, FAC2 > 0.84, Corr $\sim$0.99, FB $\sim 10^{-3}$. The method also successfully reconstructs the source release rate and plume rise height. And the time delay of the EnKF can be greatly alleviated, which is about 30 min now. It also implies that after a simultaneous sudden change of the two source parameters, the reconstructed results are unreliable, but the problem will be solved after extra information has been gradually collected.

The method proposed here can be a useful tool not only in the nuclear power plant accident emergency management, but also in other similar situation where hazardous material is released into the atmosphere, just with a little modification in the transport model. In this study, only twin experiments are conducted to validate the method. In the further study, we will try to validate it with the measurements from field experiments.

Acknowledgements

We thank the supports provided by National Natural Science Foundation of China (Nos. 91024024, 91024017 and 91224004), and the National High-Tech Research and Development Program of China (863 Program) (No. 2012AA050907).

References

[1] NEA/OECD, The implementation of short-term countermeasures after a nuclear accident (stable iodine, sheltering and evacuation), in: Proceedings of an NEA workshop, Stockholm, Sweden, 1994.
[2] NRPB, A model for short and medium range dispersion of radionuclide releases to the atmosphere. NRPB Report National Radiological Protection Board, 1979.
[3] NRPB, The uncertainty in dispersion modelling estimates obtained from the working group models. NRPB Report National Radiological Protection Board, 1986.
[4] A. Stohl, P. Seibert, G. Wotawa, D. Arnold, J.F. Burkhart, S. Eckhardt, C. Tapia, A. Vargas, T.J. Yasunari, Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant: determination of the source term, atmospheric dispersion, and deposition, Atmos. Chem. Phys. 12 (2012) 2313–2343.
[5] C.J.W. Twenhöfel, M.M.V. Troost, S. Bader, Uncertainty analysis and parameter optimisation in early phase nuclear emergency management. RIVM Report, National Institute for Public Health and the Environment, 2007.
[6] NEA, Chernobyl Assessment of Radiological and Health Impacts – 2002 Update of Chernobyl: Ten Years On, Nuclear Energy Agency (NEA), Paris, 2002, pp. 155.
[7] E. Kalnay, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, United Kingdom, 2003.
[8] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82 (1960) 35–45.
[9] P. Astrup, C. Turcanu, R. Puch, C. Rojas Palma, T. Mikkelsen, Data Assimilation in the Early Phase: Kalman Filtering RIMPUFF, Risø National Laboratory, Roskilde, Denmark, 2004, pp. 31.
[10] C. Rojas-Palma, H. Madsen, F. Gering, R. Puch, C. Turcanu, P. Astrup, H. Müller, K. Richter, M. Zheleznyak, D. Treebushny, M. Kolomeev, D. Kamaev, H. Wynn, Data assimilation in the decision support system RODOS, Radiat. Prot. Dosim. 104 (2003) 31–40.
[11] C. Rojas-Palma, F. Gering, H. Madsen, R. Puch-Solis, K. Richter, H. Muller, Theoretical Framework and Practical Considerations for Data Assimilation in Off-site Nuclear Emergency Management, SCK·CEN, Mol, Belgium, 2001.
[12] M. Drews, B. Lauritzen, H. Madsen, Analysis of a Kalman filter based method for on-line estimation of atmospheric dispersion parameters using radiation monitoring data, Radiat. Prot. Dosim. 113 (2005) 75–89.
[13] M. Drews, B. Lauritzen, H. Madsen, J.Q. Smith, Kalman filtration of radiation monitoring data from atmospheric dispersion of radioactive materials, Radiat. Prot. Dosim. 111 (2004) 257–269.
[14] M. Krysta, M. Bocquet, B. Sportisse, O. Isnard, Data assimilation for short-range dispersion of radionuclides: an application to wind tunnel data, Atmos. Environ. 40 (2006) 7267–7279.
[15] D. Quelo, B. Sportisse, O. Isnard, Data assimilation for short range atmospheric dispersion of radionuclides: a case study of second-order sensitivity, J. Environ. Radioact. 84 (2005) 393–408.
[16] R. Abida, M. Bocquet, Targeting of observations for accidental atmospheric release monitoring, Atmos. Environ. 43 (2009) 6312–6327.
[17] V. Winiarek, J. Vira, M. Bocquet, M. Sofiev, O. Saunier, Towards the operational estimation of a radiological plume using data assimilation after a radiological accidental atmospheric release, Atmos. Environ. 45 (2011) 2944–2955.
[18] R. Abida, M. Bocquet, N. Vercauteren, O. Isnard, Design of a monitoring network over France in case of a radiological accidental release, Atmos. Environ. 42 (2008) 5205–5219.
[19] P.H. Hiemstra, D. Karssenberg, A. van Dijk, Assimilation of observations of radiation level into an atmospheric transport model: a case study with the particle filter and the ETEX tracer dataset, Atmos. Environ. 45 (2011) 6149–6157.
[20] P.H. Hiemstra, D. Karssenberg, A. van Dijk, S.M. de Jong, Using the particle filter for nuclear decision support, Environ. Model. Softw. 37 (2012) 78–89.
[21] X. Davoine, M. Bocquet, Inverse modelling-based reconstruction of the Chernobyl source term available for long-range transport, Atmos. Chem. Phys. 7 (2007) 1549–1564.
[22] M. Krysta, M. Bocquet, Source reconstruction of an accidental radionuclide release at European scale, Q. J. R. Meteor. Soc. 133 (2007) 529–544.
[23] M. Bocquet, Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var, Q. J. R. Meteor. Soc. 138 (2012) 664–681.
[24] V. Winiarek, M. Bocquet, O. Saunier, A. Mathieu, Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant, J. Geophys. Res. Atmos. 117 (2012).
[25] L. Robertson, J. Langner, Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment, Atmos. Environ. 32 (1998) 4219–4225.
[26] M. Krysta, M. Bocquet, J. Brandt, Probing ETEX-II data set with inverse modelling, Atmos. Chem. Phys. 8 (2008) 3963–3971.
[27] D.Q. Zheng, J.K.C. Leung, B.Y. Lee, Online update of model state and parameters of a Monte Carlo atmospheric dispersion model by using ensemble Kalman filter, Atmos. Environ. 43 (2009) 2005–2011.
[28] D.Q. Zheng, J.K.C. Leung, B.Y. Lee, An ensemble Kalman filter for atmospheric data assimilation: application to wind tunnel data, Atmos. Environ. 44 (2010) 1699–1705.
[29] D.Q. Zheng, J.K.C. Leung, B.Y. Lee, H.Y. Lam, Data assimilation in the atmospheric dispersion model for nuclear accident assessments, Atmos. Environ. 41 (2007) 2438–2446.
[30] I.V. Kovalets, V. Tsiouri, S. Andronopoulos, J.G. Bartzis, Improvement of source and wind field input of atmospheric dispersion model by assimilation of concentration measurements: method and applications in idealized settings, Appl. Math. Model. 33 (2009) 3511–3521.
[31] V. Tsiouri, I. Kovalets, S. Andronopoulos, J.G. Bartzis, Development and first tests of a data assimilation algorithm in a Lagrangian puff atmospheric dispersion model, Int. J. Environ. Pollut. 44 (2011) 147–155.
[32] V. Tsiouri, I. Kovalets, S. Andronopoulos, J.G. Bartzis, Emission rate estimation through data assimilation of gamma dose measurements in a Lagrangian atmospheric dispersion model, Radiat. Prot. Dosim. 148 (2012) 34–44.
[33] SCK•CEN, A European manual for 'off-site emergency planning and response to nuclear accidents, in: A. Sohier (Ed.), SCK•CEN Report, Belgian Nuclear Research Centre (SCK•CEN), Radiation Protection Division, Decision Strategy Research Department, Mol, Belgium, 2002.
[34] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, J. Geophys. Res. Oceans 99 (1994) 10143–10162.
[35] H. Eleveld, Y.S. Kok, C.J. Twenhofel, Data assimilation, sensitivity and uncertainty analyses in the Dutch nuclear emergency management system: a pilot study, Int. J. Emerg. Manage. 4 (2007) 551–563.
[36] J.M. Stockie, The mathematics of atmospheric dispersion modeling, SIAM Rev. 53 (2011) 349–372.
[37] T.J. Bauer, Comparison of chlorine and ammonia concentration field trial data with calculated results from a Gaussian atmospheric transport and dispersion model, J. Hazard. Mater. 254–255 (2013) 325–335.
[38] K.C. Nguyen, J.A. Noonan, I.E. Galbally, W.L. Physick, Predictions of plume dispersion in complex terrain: Eulerian versus Lagrangian models, Atmos. Environ. 31 (1997) 947–958.
[39] I. Korsakissok, V. Mallet, Comparative study of Gaussian dispersion formulas within the polyphemus platform: evaluation with prairie grass and Kincaid experiments, J. Appl. Meteorol. Climatol. 48 (2009) 2459–2473.
[40] J. Irwin, Scheme for estimating dispersion parameters as a function of release height. Tech. Rep., US EPA, 1979, 26 pp.
[41] S.R. Hanna, R.J. Paine, Hybrid plume dispersion model (HPDM) development and evaluation, J. Clim. Appl. Meteorol. 28 (1989) 206–224.
[42] G. Evensen, The ensemble Kalman filter: theoretical formulation and practical implementation, Ocean Dyn. 53 (2003) 343–367.

[43] L. Bengtsson, M. Ghil, E. Kallen, in: L. Bengtsson, M. Ghil, E. Kallen (Eds.), Dynamic Meteorology: Data Assimilation Methods, Springer-Verlag, New York, 1981.
[44] H. Ngodock, M. Carrier, A weak constraint 4D-var assimilation system for the navy coastal ocean model using the representer method, in: S.K.I. Park, L. Xu (Eds.), Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, Springer, New York, United States, 2009.
[45] A. Zhang, E. Wei, B.B. Parker, Optimal estimation of tidal open boundary conditions using predicted tides and adjoint data assimilation technique, Cont. Shelf Res. 23 (2003) 1055–1070.
[46] I. Fukumori, P. Malanotte-Rizzoli, An approximate Kaiman filter for ocean data assimilation: an example with an idealized Gulf Stream model, J. Geophys. Res. Oceans 100 (1995) 6777–6793.
[47] D.L. Colton, R. Kress, Integral Equation Methods in Scattering Theory, Wiley, New York, 1983.
[48] S.R. Hanna, J.C. Chang, D.G. Strimaitis, Hazardous gas model evaluation with field observations, Atmos. Environ. A 27 (1993) 2265–2285.
[49] ASTM, Standard guide for statistical evaluation of atmospheric dispersion model performance (D 6589), in: Statistical Comparison Metrics and Methods, American Society for Testing and Materials, PN West Conshohocken, PA, 2010, pp. 17.
[50] J.C. Chang, S.R. Hanna, Air quality model performance evaluation, Meteorol. Atmos. Phys. 87 (2004) 167–196.
[51] M. Santillana, Quantifying the loss of information in source attribution problems using the adjoint method in global models of atmospheric chemical transport, 2013 arXiv:1311.6315 (arXiv preprint).