

Article

Data assimilation in the air contaminant dispersion using particle filter and expectation-maximization algorithm with UAV observations

Rongxiao Wang ¹, Bin Chen ^{1,*}, Sihang Qiu ¹, Zhengqiu Zhu ¹ and Xiaogang Qiu ¹

¹ College of Information System and Management, National University of Defense Technology, 109 Deya Road, Changsha 410073, China; pandaself@testbot.cn (R.W.); qiusihang@gmail.com (S.Q.); admin@steven-zhu.me (Z.Z.); michael.qiu@139.com (X.Q.)

* Correspondence: nudtc9372@gmail.com; Tel.: +86-137-8714-8795

Abstract: The precise prediction of air contaminant dispersion is essential to the air quality monitoring and the emergency management of the contaminant gases leakage incidents in the chemical industry park. The conventional atmospheric dispersion models can seldom give precise prediction due to inaccurate input parameters. In order to improve the prediction accuracy of dispersion model, two data assimilation methods (i.e. one is merely based on the typical particle filter while the other is a combination of particle filter and expectation-maximization algorithm) are proposed to assimilate the UAV observations into the atmospheric dispersion model. Two emission cases are taken into consideration, the difference between which is the different dimensions of state variables. To test the performances of the proposed methods, experiments corresponding to the two emission cases are designed and implemented. The results show that the particle filter can effectively estimate the model parameters and improve the accuracy of model prediction when the dimension of state variables is low. In contrast, when the dimension of state variables becomes higher, the method of particle filter combining expectation-maximization algorithm performs better in the parameter estimation accuracy and warm-up time. Therefore, the data assimilation methods are able to effectively support the air quality monitoring and emergency management in chemical industry parks.

Keywords: air contaminant dispersion; data assimilation; particle filter; expectation-maximization algorithm; UAV

1. Introduction

The air contaminant emission and contaminant gases leakage incidents in the chemical industry park are significant events with threat to public health. Therefore, understanding the dispersion of air contaminant is essential to air quality monitoring and emergency response to the gases leakage incidents. The conventional way of predicting the atmospheric dispersion relies on the atmospheric dispersion models (e.g. the Gaussian models and Lagrangian models) with some given input model parameters. However, due to the dynamic and stochastic nature of the atmospheric dispersion, it is impractical to precisely measure these model parameters especially the meteorological data (e.g. the wind field). Further, the source terms (i.e. source location and release rate) are often unknown. These inaccurate or unknown parameters result in the inaccuracy of model prediction of air contaminant concentration.

Data assimilation (DA) provides a way of estimating model parameters dynamically and improving the accuracy of model prediction. This technique combines model and observations, in which the observations are assimilated into the model to produce a time sequence of estimates of system states. With the precise model parameters, the accuracy of model prediction is consequently improved. Therefore, data assimilation has been widely used in various fields, especially in the numerical weather forecast, meteorological pre-processing, and wild fire spread [1,2] which demand

high accuracy of prediction. Common methods of data assimilation include variational approach [3], Kalman filter [4] and its variants (e.g. extended Kalman filter [5] and Ensemble Kalman filter [6]), and particle filter [7,8]. Among these methods, particle filter is one of the most suitable approaches for highly nonlinear and non-Gaussian models. Using a series of weighted random sampling particles to approximate the posterior probability density function of system state, particle filter is able to estimate arbitrary probability densities with few assumption constraints of model. Therefore, particle filter is applied as the data assimilation method in the air contaminant dispersion in this paper. As for the air contaminant emission, we focus on two cases in the chemical plants. The one is the daily emission occurring every day, in which the source terms are known in advance because the emission is under control. The other case is the contaminant gases leakage incident, in which the release rate is hard to measure for the sake of safety. Additionally, the variation of release rate also increases the difficulty of measuring this source term. In the contaminant gases leakage incidents, with the release rate to be estimated, the accuracy of particle filter might decrease due to the higher dimension of the state vector. To deal with this problem, the particle filter is combined with expectation-maximization (EM) algorithm [9]. EM algorithm is a parameter estimation method for incomplete-data problem, which divides the estimation of complex parameters into iterations with two steps (i.e. the Expectation Step (E-step) and the Maximization Step (M-step)). Further, EM algorithm is combined with particle filter to improve the performance of parameter estimation [10] in some fields like the target tracking [11].

Data assimilation demands the support of observed data of air contaminant. In chemical industry parks, the conventional way of obtaining these observations mainly depends on the static ground monitoring station. Although the station has high accuracy of detecting, its fixed location limits the coverage of data collection especially when the station is located in the upwind of the emission source. Fortunately, the emergence of unmanned aerial vehicle (UAV) has initiated a revolution in this research. Providing the flexible mobility in space, the UAV becomes a flexible and efficient tool of data collection in the atmospheric environment monitoring. Yang and Huang [12] utilized a sensory system based on an unmanned helicopter to monitor the SO_2 , NO , and CO in a chemical industry park. The UAV has been also used in the source term estimation and boundary tracking of the atmospheric dispersion [13,14].

In this paper, the Gaussian plume model is applied to describe the air contaminant dispersion. In order to improve the accuracy of model prediction, the data assimilation based on particle filter is utilized to assimilate the observations into the atmospheric dispersion model in the two emission cases. To acquire observations efficiently in chemical industry parks, a UAV-based air contaminant sensory system is developed. In the first case, four coefficients of dispersion in Gaussian plume model are selected as the state variables to be estimated. In the second case, the release rate is added into the state variables. Further, to deal with the high dimension of state variables in the second case, the particle filter is combined with EM algorithm. The performances of particle filter and the method of particle filter combining EM algorithm are tested by two experiments. The rest of this paper is organized as follows. Section 2 introduces the atmospheric dispersion model and data assimilation methods using typical particle filter and particle filter combining EM algorithm. Section 3 illustrates the architecture and workflow of the UAV-based air contaminant sensory system. Section 4 describes the experiments. Some preliminary results are given in Section 5. Discussions and conclusions are made in the Section 6 and Section 7, respectively.

2. Model and methods

2.1. Atmospheric Dispersion Model

The modeling of air contaminant dispersion is the basis of the dispersion prediction. The Gaussian models are widely used in the atmospheric dispersion. Requiring only a few input parameters, Gaussian models are simpler compared to some complex models like the Lagrangian model. Further, the results of Gaussian models are trustworthy for near-field dispersion cases. Consequently, the Gaussian models are suitable for the modeling of air contaminant dispersion in

the data assimilation which requires fast computing of the dispersion model. In this paper, the Gaussian plume model is applied to model the continuous release of the point source in the chemical industry park. In this model, the air contaminant concentration of a given point (x, y, z) is expressed as follows:

$$C(x, y, z) = \frac{q}{2\pi u \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y^2}} \left(e^{-\frac{(z-H)^2}{2\sigma_z^2}} + e^{-\frac{(z+H)^2}{2\sigma_z^2}} \right) \quad (1)$$

where x, y, z are the coordinates of downwind, crosswind and vertical directions respectively. H and q represent the height and release rate of the source. Parameter of u is the wind velocity. In the Gaussian plume model, the air contaminant concentration in axis y and z are considered to follow the Gaussian distribution. Therefore, the key parameters of the model are σ_y and σ_z which represent the standard deviations that describe the crosswind and vertical mixing of air contaminant. The standard deviations can be described by empirical formulas:

$$\sigma_y = ax^b \quad (2)$$

$$\sigma_z = cx^d \quad (3)$$

where x represents the downwind distance. The parameters of a, b, c, d are dispersion coefficients closely related to the environment conditions, such as atmospheric stability and terrain. Several derivations of these dispersion coefficients exist where a popular approach is based on the Pasquill's atmospheric stability class [15]. The empirical formulas of σ_y and σ_z illustrate that the standard deviations increase with the downwind distance. The Gaussian plume model can be applied to model the continuous release of the point source in the chemical industry park. However, relying solely on this model may fail to make accurate prediction, since these model parameters are usually variant with the environment conditions and hard to measure precisely. For example, we can only get their empirical values of the four dispersion coefficients, but no precise value is available. There is no specific formula to describe the relationship between these coefficients and atmospheric stability class. As for the wind field parameters (i.e. wind direction and velocity), they are also difficult to precisely measure. Further, the release rate q is usually unknown during the contaminant gases leakage incident. Without precise parameters, in conventional modeling of atmospheric dispersion, it is common practice to bring imperfect estimations of these parameters into the model for calculation, which brings errors to the model prediction inevitably. Thus, there is an urgent need to dynamically estimate these parameters to improve the accuracy of model prediction.

2.2. Data Assimilation Using Particle Filter

For the modeling of air contaminant dispersion, in order to diminish the errors of input parameters and get precise prediction, a data assimilation model based on particle filter is developed. Particle filter, also called Sequential Monte Carlo (SMC) method, is a sample-based method that uses Bayesian inference and stochastic sampling techniques to recursively estimate the state of dynamic system from some given observations. The core idea of particle filter is using a series of weighted random sampling particles to approximate the posterior probability density function of the system state. A typical particle filter algorithm includes four steps (i.e. initialization, importance sampling, weight update, and resample) and goes through multiple iterations. In order to apply particle filter to the Gaussian plume model, the state space model of atmospheric dispersion needs to be developed. Usually, a dynamic system can be described and formulated as a discrete state space model:

$$s_{t+1} = f(s_t, t) + \gamma(t) \quad (4)$$

$$m_t = g(s_t, t) + \omega(t) \quad (5)$$

where (4) and (5) represent the system state transition model and measurement model, respectively. s_t and m_t are system state variables and measurement variables at time step t respectively. The function f describes the transition of the system states with time. The function g defines the relationship between state variables and measurement variables. The parameters of γ and ω are two independent random variables representing the state noise and the measurement noise, respectively.

The state transition model is based on the state variables. Therefore, the selection of the state variables is the key of the construction of the state transition model. For Gaussian plume model, there are several choices of the state variables. One of the common practices is dividing the area of dispersion into numerous grids and choosing the concentrations by grid as the state variables. This choice describes the atmospheric dispersion directly, however, the vast region of the chemical industry park means the high dimension of the state vector, which results in high computation cost. In this paper, the dispersion coefficients a , b , c , d in (2) and (3) as well as the release rate q are selected as the state variables in the second case, while the state variables in the first case only include four dispersion coefficients. Selected as the state variables, the four dispersion coefficients play important roles in Gaussian plume model. The standard deviations σ_y and σ_z , which are derived from the four dispersion coefficients, describe the crosswind and vertical mixing of air contaminant. Further, correlated closely with various environment conditions, these coefficients are hard to measure precisely. As for the release rate, it is difficult to identify for the sake of safety during the contaminant gases leakage incident. Therefore, the coefficients and release rate need dynamic update by data assimilation. Additionally, the wind field, which is hard to measure precisely, is assumed to be known for simplicity. In the following construction of the data assimilation model, only the second case is discussed because the data assimilation model in the first case is similar to that in the second case. The system state vector and state transition model can be described as follows:

$$s_t = \{a_t, b_t, c_t, d_t, q_t\} \quad (6)$$

$$s_{t+1} = s_t + \gamma_t \quad (7)$$

Due to the flat variation of the meteorological condition within limited duration of a time step in this paper (i.e. 1 min), the dispersion coefficients almost stay constant during a time step. In terms of the release rate, the variation of it is not specific but also flat within a time step. Therefore, the state transition function is defined as an identically equal function. The variation of state vector is provided by γ_t , which is considered to be a Gaussian white noise vector of which all elements follow the Gaussian distribution $(0, \sigma_1^2)$. The measurement model describes the relationship between the state variables and observations. In this paper, the observations are air contaminant concentrations collected by the UAV at the trajectory points p ($p = \{X, Y, Z\}$), where X, Y, Z are coordinates vectors. Therefore, the measurement function $g(s_t, t)$ is defined as the Gaussian plume function in (1):

$$g(s_t, t) = C(X, Y, Z) = \frac{q}{2\pi u \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y^2}} \left(e^{-\frac{(z-H)^2}{2\sigma_z^2}} + e^{-\frac{(z+H)^2}{2\sigma_z^2}} \right) \quad (8)$$

where the $\omega(t)$ is assumed to be Gaussian white noises following the distribution $(0, \sigma_2^2)$ that describes measurement errors.

Figure 1 shows the structure of particle filter and the procedure of the data assimilation. In this figure, the rectangle boxes represent the major components in one step of the algorithm. The rounded rectangles represent the data or variables. The data assimilation runs in a stepwise fashion. At time step t , the system states in time step $t-1$ (denoted as $s(t-1)$ in Figure 1) are fed into the system state transition model. Then this model performs the transition function in (7) to produce a

sample for each particle in $s(t-1)$. The resulting system state set of the transition model is denoted as $s(t)'$. To compute the importance weights of the particles, the concentration vector (denoted as $C(t)'$) corresponding to each particle is computed according to the measurement model described in (8). The locations corresponding to the concentration vectors ($C(t)'$) depend on the sensor deployment schema. In this paper, the sensor deployment schema is the trajectory of UAV-based sensory system which will be introduced in section 4.1. Then the importance weights of particles are calculated according to the likelihood between $C(t)'$ and the observed concentrations $M(t)$. After normalizing the weights of all particles, a resampling algorithm is applied to generate $s(t)$, which is the input for the next step. To observe the performance of model prediction, the prediction concentration $C(t)$ can be calculated from $s(t)$ by the measurement model.

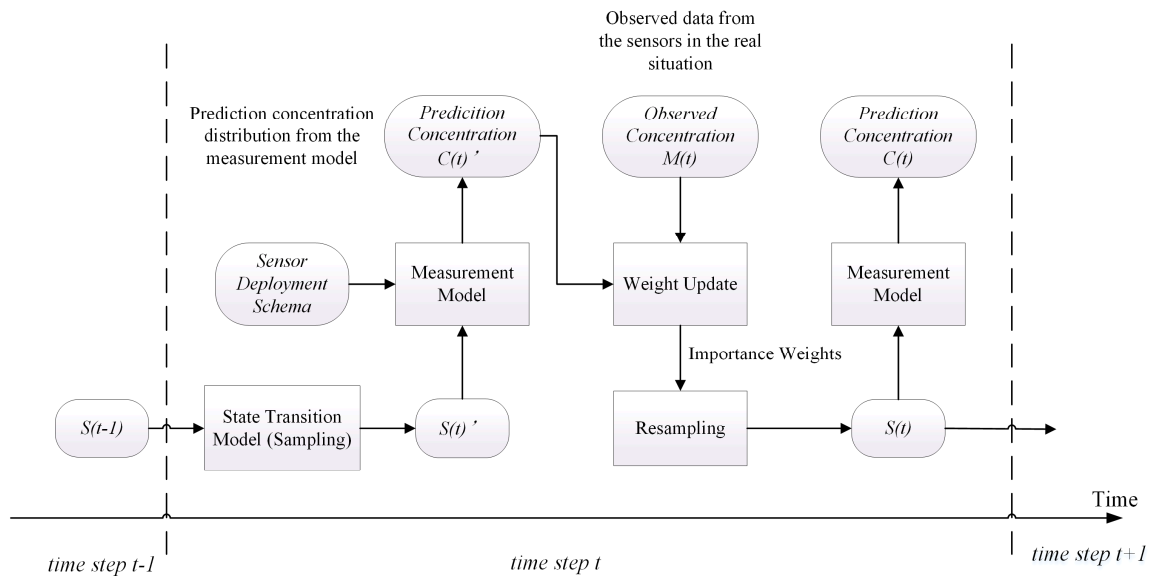


Figure 1. Data assimilation based on the particle filter.

2.3. Particle Filter Combining EM algorithm in the Second Case

In the last section, the particle filter is applied to estimate the release rate and four dispersion coefficients in the second case. However, due to the higher dimension of state vector and the complexity of the atmospheric dispersion, the particle distribution might not converge to a satisfactory result during the process of particle filter. In order to enhance the accuracy of estimation, the particle filter is combined with EM algorithm to iteratively estimate these parameters (i.e. the release rate and four dispersion coefficients). The EM algorithm is a generic method for computing the MLE of parameter θ in an incomplete-data problem. In the incomplete-data problem, the estimation of the unknown parameter θ depends on the hidden variable z , so θ is hard to estimated directly. To deal with the problem, the EM algorithm divides the estimation process of θ and z into two steps (i.e. the Expectation Step (E-step) and the Maximization Step (M-step)) and runs iteratively. Specifically, in the E-step, the posterior probability of the hidden variable, which can also be considered as their expectation, is calculated from initial values of the parameters or the model parameters in the last iteration. Further, the expectation is regarded as the estimation of the hidden variable. Based on the estimation of hidden variable, the MLE of θ is calculated by maximizing the likelihood function in the M-step. Therefore, by reducing the complexity of parameter estimation, EM algorithm has an excellent performance in the incomplete-data problem.

However, there seems to be no apparent hidden variable and incomplete-data in the parameter estimation of the Gaussian plume model because the release rate and four dispersion coefficients are all included in the parameters θ . In order to apply EM algorithm to this parameter estimation problem, the problem is adjusted to an incomplete-data one. The four dispersion coefficients and the release rate at time step t are regarded as the hidden variables z_t and parameter θ_t , respectively.

Therefore, four dispersion coefficients and the release rate are estimated in E-step and M-step, respectively. In the E-step, the hidden variables are estimated by calculating the posterior probability density function $p(z_t / x_t; \theta_t)$, which can be approximated by particle filter with observations x_t . In the M-step, the MLE of release rate is calculated by maximizing the likelihood function through Particle Swarm Optimization (PSO). The method of particle filter combining EM algorithm can be expressed as follows. At time step t , the MLE of release rate θ_t ($\theta_t = q_t$) depends on the hidden variables z_t , $z_t = \{a_t, b_t, c_t, d_t\}$. The likelihood of θ_t and z_t is:

$$L(\theta_t, z_t) = p(x_t; (\theta_t, z_t)) \quad (9)$$

Then the z_t and θ_t are estimated in the two steps, respectively. In the E-step, the posterior probability (expectation) of z_t is calculated using particle filter with the assumption that the θ_t is known:

$$Q(z_t) := p(z_t / x_t; \theta_t) \approx \sum_{i=1}^N \omega_t^{(i)} \delta(s_t - s_t^{(i)}) \quad (10)$$

where N is the number of particles which represent the system states s_t . The $\omega_t^{(i)}$ is the weight of the i^{th} particle, and δ is the Dirac delta function. The observations are assimilated into the dispersion model by the particle filter in the E-step. Then in the M-step, based on the estimation of z_t in the E-step, the MLE of θ is computed by maximizing the likelihood using PSO:

$$\theta_t = \arg \max L(\theta_t, z_t) \quad (11)$$

The E- and M-steps are alternated repeatedly until convergence, which is determined by a stopping rule:

$$|\theta^{(k+1)} - \theta^{(k)}| < \varepsilon, \quad \varepsilon > 0 \quad (12)$$

By the method of particle filter combining EM algorithm, both of the hidden variables z and parameter θ are iteratively estimated in each time step.

3. UAV-based air contaminant sensory system

In chemical industry parks, the observations of air contaminant are usually collected by static monitoring stations. However, due to the limited number and fixed locations, these facilities may fail to obtain valid data especially when they are located in the upwind of the emission source. In order to overcome this shortage, a UAV-based air contaminant sensory system is developed for efficient data acquisition. Shown in Figure 2, this system consists of an aerial platform and sensory devices. The aerial platform, which is selected from mini-UAV DJI M100, provides aerial mobility and accommodation for sensory devices. The sensory devices integrated on the UAV are composed of micro control unit, gas sensors, GPS and network connection modules. In terms of the workflow, as can be seen in Figure 3, the sensory devices achieve data collection (i.e. contaminant concentration, geo-location and time) and then transmit data to the database of a management system in time. Achieving the data collection, transmission and storage, this system can give researchers the access to obtain the observations in time.

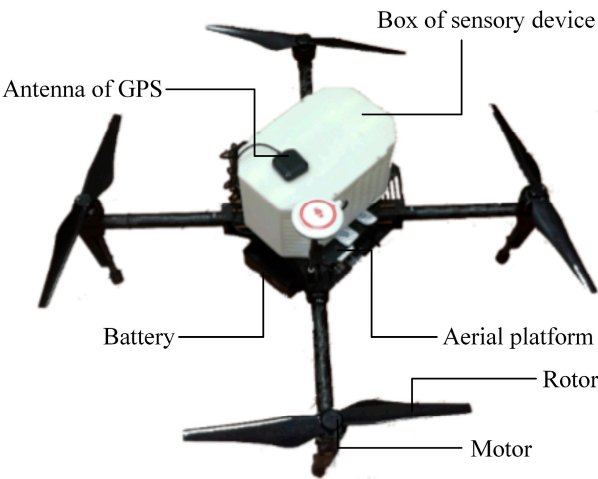


Figure 2. The UAV-based air contaminant sensory system.

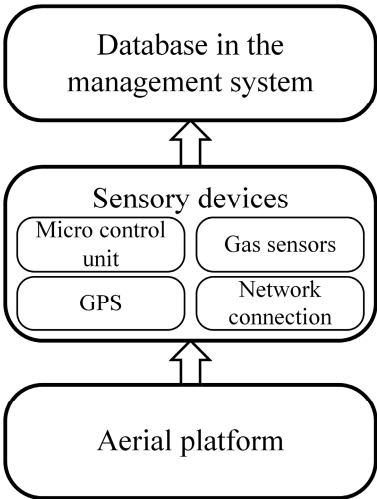


Figure 3. The architecture and workflow of the UAV sensory system.

4. Experiments

Displayed in the Table 1, two experiments are designed and implemented to test the performances of proposed data assimilation methods in the two emission cases. In the Experiment One which is used to test the effect of particle filter in the first case, the conventional atmospheric dispersion model which is only based on the Gaussian plume model without data assimilation is used as the control group. In contrast, the particle filter is applied as the data assimilation method to improve the model prediction in the treatment group. In the Experiment Two, the particle filter (section 2.2) and the EM algorithm with particle filter (section 2.3) are both implemented for data assimilation to make comparison. Before experiments, the “true” concentration distribution and observations are generated by a “real” dispersion. The “real” dispersion is not a practical air contaminant dispersion in the chemical industry park but a simulated dispersion based on the Gaussian plume model because of the lack of field experiment condition. The “real” dispersion uses default meteorological parameters (e.g. dispersion coefficients) and release rate to generate the “true” concentration, which provides a reference for the “true” situation. Further, the observations, which are to be assimilated into the model, are produced from the “true” concentration by the measurement model in (8) and the simulated trajectories of UAVs. In our experiments, an atmospheric dispersion during a period of 100 min on a square area of 1000 * 1000 m² is simulated and the concentration distribution at 30 m AGL (above ground level) is calculated. For smaller

computational cost, the region is divided into grids with the size of $10 * 10 \text{ m}^2$. As for the emission source, it is assumed to be located at (0, 0) with a height of 70 m AGL, which releases air contaminant during the whole experiment duration. The wind direction and velocity are 220° and 4 m/s, respectively. The trajectories of UAVs and parameter configuration are introduced in section 4.1 and 4.2, respectively.

Table 1. The design of experiments.

Case	Experiment Name	Group	Method	Note
The release rate is known	Experiment One	Experiment 1A	Gaussian plume model	Control group: compared to Experiment 1B
		Experiment 1B	Gaussian plume model & particle filter	Treatment group: test the particle filter
The release rate is unknown	Experiment Two	Experiment 2A	Gaussian plume model & particle filter	Treatment group: test the particle filter
		Experiment 2B	Gaussian plume model & EM with particle filter	Treatment group: test the EM algorithm with particle filter

4.1. Trajectories of UAVs

Ideally, the observations of air contaminant dispersion should be collected in the true situation by UAVs. However, due to the lack of field experiment condition, the observations are generated from the “true” concentrations by the measurement model and simulated UAVs' trajectories in our experiments. To effectively acquire observations, the trajectories of UAVs are designed according to the wind direction. According to the set of sensors in the Indianapolis experiment [16], the trajectories of UAVs are set as arcs through the plume. For example, when the wind direction is 220° , trajectories are showed in Figure 4. There are 10 simulated arc trajectories through the plume, each of which corresponds to an UAV (only one UAV is showed in the figure). It is assumed that each UAV moves along its own trajectory and collects the concentration data at the same time. The time interval of data collection is according to the characteristic of gas sensors on the UAV. In our experiments, it is set to be 1 min.

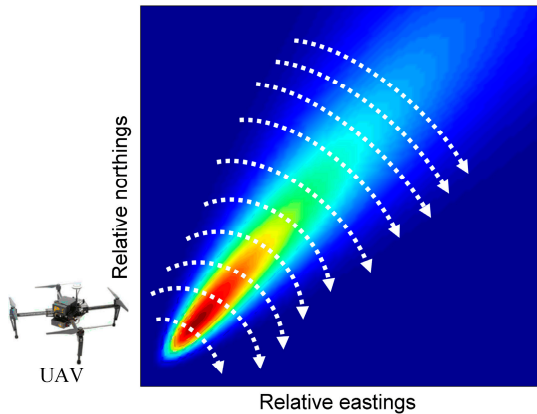


Figure 4. Sketch of UAVs' Trajectories when the wind direction is 220° .

4.2. Parameter Configuration

This section illustrates the parameter configuration of the two experiments and “real” dispersions. These parameters include release rate and four dispersion coefficients. The release rate,

which significantly influences the size and shape of the plume, is difficult to measure accurately. In the Experiment One, the known release rate is not the parameter to be estimated. Therefore, the release rates in the “real” dispersion and two groups of Experiment One are all set to be constant at 30 g/s. In the contaminant gases leakage incidents, the release rate is one of the state variables to be estimated. Thus, in the “real” dispersion it is considered to be variant to test the data assimilation method. An exponential decay curve is applied to describe the variation of release rate because in some cases the release rate decreases with the reduction of the amount of residual contaminants [17].

As for the four dispersion coefficients, the “real” dispersions use variant dispersion coefficients to simulate the “true” situation. The four dispersion coefficients are closely correlated with the environment conditions, such as the atmospheric stability and terrain. The Pasquill-Gifford-Turner (henceforth P-G-T) [15,18,19] approach is widely used in the classification of the atmospheric stability. Based on the approach, the P-G-T curves are developed to identify the four dispersion coefficients according to the atmospheric stability class. Unfortunately, this approach takes few terrain conditions into consideration. Furthermore, there are many other researchers having provided empirical values of the four coefficients under different atmospheric stability classes and terrain conditions, such as Briggs [20] and Vogt [21]. Carrascal [22] compared these empirical values under identical atmospheric conditions. In the “real” dispersions of two experiments, the parameter scheme of Vogt is adopted to describe the variation of four dispersion coefficients in chemical industry parks. The values of four coefficients according to Vogt are shown in Table 2. This table gives empirical values of a, b, c, d under different atmospheric stability classes for open country. In this table, Class A to F represents different atmospheric stability classes. Class A represents the most active class, and Class F is the most stable one. As for the terrain condition, with few obstacles like huge building, the open country is similar to the terrain of chemical industry park. Therefore, the parameter scheme of Vogt is suitable for the identification of the four dispersion coefficients in chemical industry parks. In the “real” dispersion of the Experiment One, the atmospheric stability is considered to vary following the order of A, B, C, and D during the 120min (only the duration of 0 to 100 min is simulated). Thus, the values of a, b, c, d under class A, B, C, and D are selected as their values at 0, 40, 80, 120 min in the “real” dispersion, respectively. In addition, the variations of four dispersion coefficients between two classes of atmospheric stability are assumed to be linear. Therefore, the variations of four dispersion coefficients are piecewise linear in the “real” dispersion (black solid lines in Figure 5). In the “real” dispersion of Experiment Two, the four dispersion coefficients follow the similar variation rule instead that the variation is slowed down (black solid lines in Figure 7) to lower the difficulty of the data assimilation. Some key values of coefficients in “real” dispersions of two experiments are shown in Table 3. In contrast, the four dispersion coefficients in Experiment 1A are initialed to the values under Class A and considered to stay constant before the updates to the “true” values of the “real” dispersion every 40 min (blue dash lines in Figure 5). The update is designed to enhance the prediction accuracy of conventional dispersion model. Additionally, in the data assimilation models of Experiment 1B and Experiment Two, the particles representing four dispersion coefficients are initialed with ranges covering the variation range mentioned above.

Table 2. Values of a, b, c, d under different atmospheric conditions according to Vogt [21].

Atmospheric Stability Class	Coefficients of Dispersion			
	a	b	c	d
A	0.23	1.00	0.10	1.16
B	0.23	0.97	0.16	1.02
C	0.22	0.94	0.25	0.89
D	0.22	0.91	0.40	0.76
E	1.69	0.62	0.16	0.81
F	5.38	0.58	0.40	0.62

Table 3. Some key values of a , b , c , d in the “real” dispersions.

Time (min)	“Real” dispersion of Experiment One				“Real” dispersion of Experiment Two			
	a	b	c	d	a	b	c	d
0	0.23	1.00	0.10	1.16	0.23	1.00	0.10	1.16
40	0.23	0.97	0.16	1.02	0.23	0.985	0.13	1.09
80	0.22	0.94	0.25	0.89	0.23	0.97	0.16	1.02
120	0.22	0.91	0.40	0.76	0.225	0.955	0.205	0.955

These experiments are implemented as follows. Before the two experiments, the “real” dispersions with the variant parameters are firstly executed to generate the “true” concentrations. After that, the observations are produced from the “true” concentrations by the simulated trajectories of UAVs and the measurement model. Then in the two experiments, the observations are assimilated into the Gaussian plume model by the proposed data assimilation methods to generate concentration distribution. Additionally, in the Experiment One, the Experiment 1A is implemented as the control group to generate the concentration prediction without data assimilation.

5. Results and Discussions

5.1. Experiment One

Figure 5 shows four dispersion coefficients a , b , c , d in Experiment One.

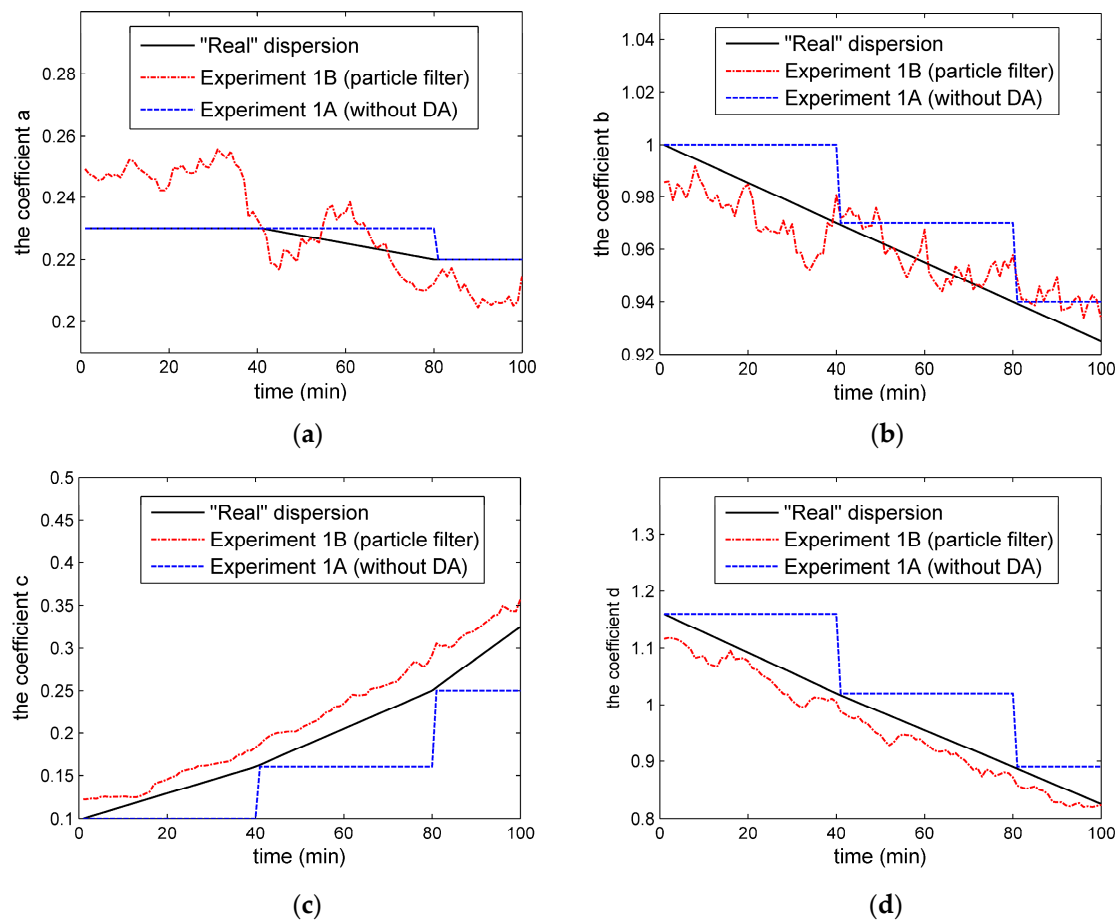


Figure 5. Four coefficients of dispersion in the Experiment One: (a) the coefficient a ; (b) the coefficient b ; (c) the coefficient c ; (d) the coefficient d .

As can be seen from the figure, the dispersion coefficients in the “real” dispersion follow the piecewise-linear variation according to the Table 3. As for the dispersion coefficients in Experiment

1B, they are close to the “true” values of the “real” dispersion. With the particle filter, these coefficients follow the trend of the “true” values closely. In comparison, without data assimilation, the coefficients in Experiment 1A stay invariant before the updates to the “true” values. However, it is worth to mention that the particle filter's performance in estimating coefficient a is less effective than performances in others. Indeed, the coefficient a in the Experiment 1A is even closer to the “true” value than that of the Experiment 1B. This might be because of the flat variation of coefficient a .

To observe the effect of particle filter on the concentration prediction, the root mean square errors (RMSE) between the model predictions and the “true” concentrations are calculated. Figure 6 shows the RMSE at trajectory points. It can be seen from the Figure 6(a) that the RMSE of the Experiment 1B is significantly lower than that of the Experiment 1A in most of the duration. The ascendant RMSE curve of Experiment 1A in Figure 6(a) comes from the errors of four dispersion coefficients. As shown in Figure 5, without data assimilation, four dispersion coefficients stay constant before the updates in Experiment 1A. Therefore, the errors of the four coefficients grow over time, causing the ascendant RMSE curve in Figure 6(a). In comparison, with the corrected dispersion coefficients by particle filter, the RMSE of Experiment 1B is reduced to a relatively low level. In most of the duration, the RMSE with particle filter is lower than 0.008 mg/m^3 . Figure 6(b) shows the “true” concentrations and model prediction concentrations with data assimilation at trajectory points. It can be seen that most of the concentrations fit the line “ $y = x$ ” well, which shows the model predictions are close to the “true” concentrations. Further, with the high correlation (0.99102) and small RMSE (0.0056039), the result illustrates the good performance of particle filter.

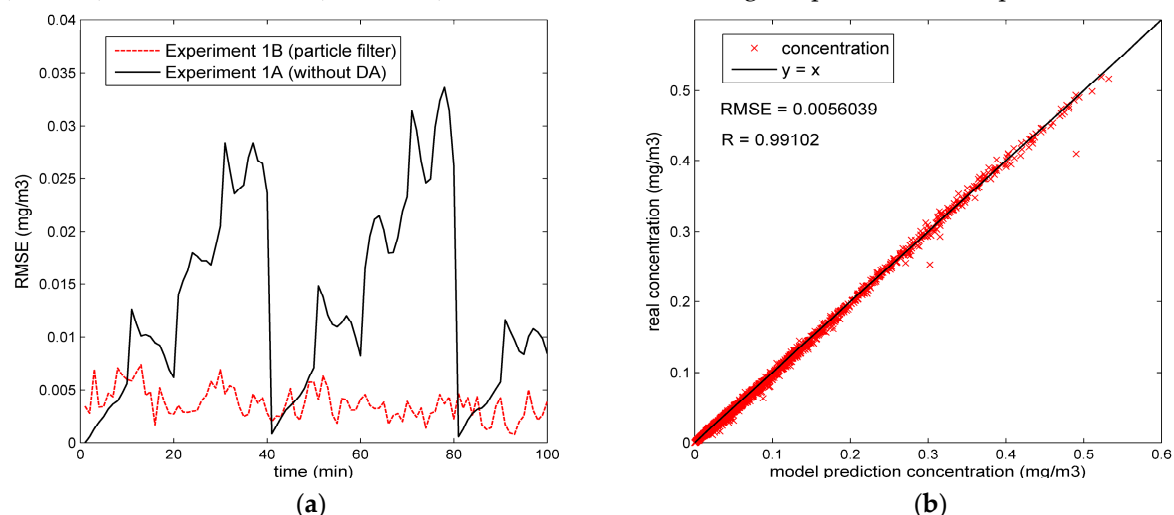


Figure 6. The assimilation results at trajectory points in Experiment One: (a) RMSE of concentration at the trajectory points in Experiment 1A and 1B; (b) the true concentration and model prediction concentration in the Experiment 1B.

5.2. Experiment Two

The Experiment Two compares the performances of particle filter and the particle filter combining EM algorithm. Figure 7 shows the four dispersion coefficients in Experiment Two. As can be seen from the figure, the variations of four coefficients in the “real” dispersion are flatter than the Experiment One. However, the performance of particle filter is unsatisfactory, the estimations of coefficients are far from the “true” values especially the coefficient c , which may result from the high dimension of state variables. With high dimension of state variables, it is difficult for particle filter to estimate the a , b , c , d and q simultaneously. In contrast, the estimation precision of EM algorithm with particle filter is slightly better than particle filter because EM algorithm decreases the complexity of parameter estimation by dividing the parameter estimation process into two steps. Further, as can be seen from the figure, the warm-up time of EM with particle filter is shorter than particle filter, which means the estimations in Experiment 2B is more likely to follow the “true”

values in the several steps after the start. As for the release rate and RMSE, the similar performances of the two data assimilation methods can be seen in the Figure 8. The error of EM algorithm with particle filter is slightly smaller than that of typical particle filter.

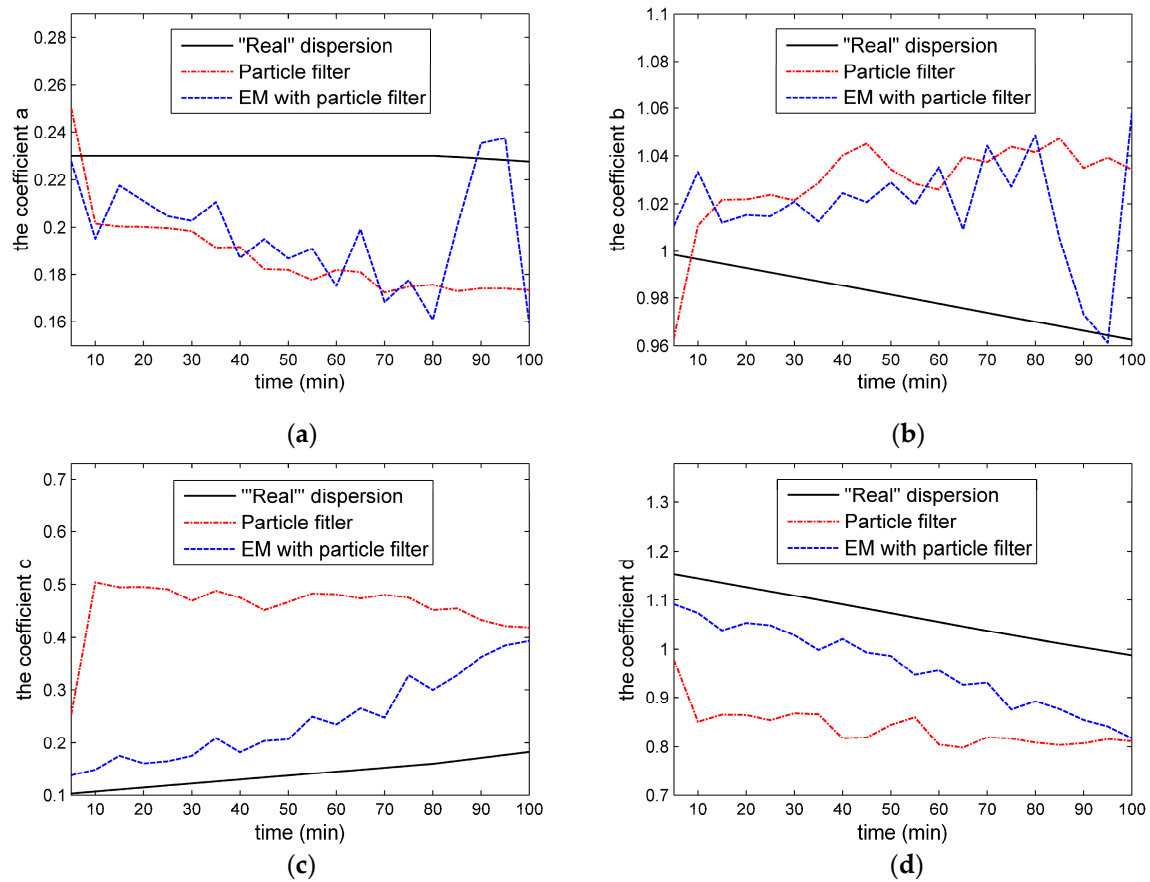


Figure 7. Four coefficients of dispersion in the Experiment Two: (a) the coefficient a; (b) the coefficient b; (c) the coefficient c; (d) the coefficient d.

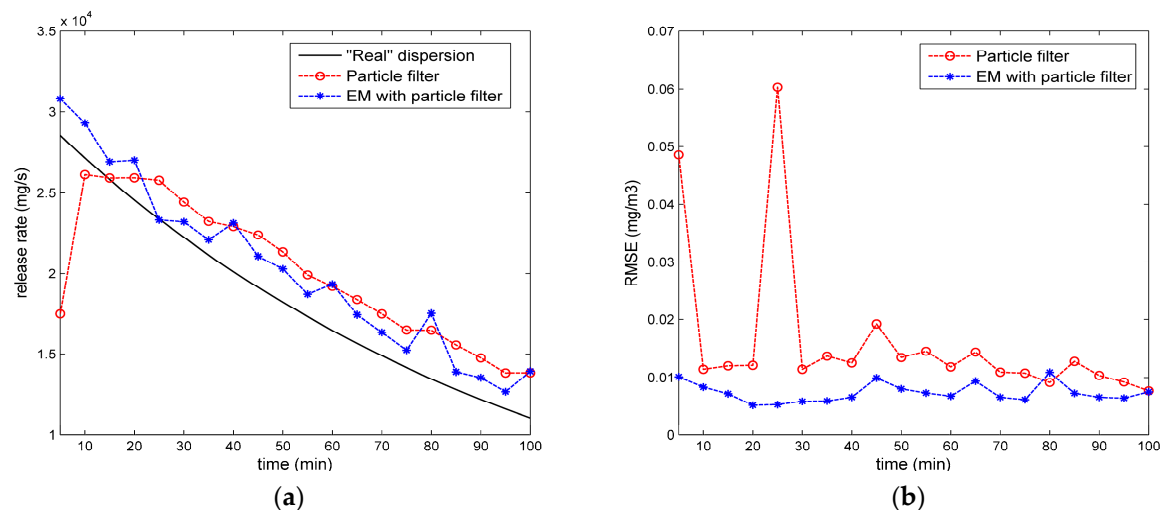


Figure 8. The release rate and RMSE of the prediction concentration in Experiment Two: (a) release rate; (b) RMSE of prediction concentration on the trajectory points.

6. Discussions

In the two experiments, the performances of particle filter and the method of particle filter combining EM algorithm are tested. The results illustrate that the dimension of state variables

significantly affects the quality of data assimilation. In Experiment One where the state variables are only four dispersion coefficients, the data assimilation based on the particle filter is effective to estimate state variables and improve the model prediction. However, when the dimension of state variables becomes higher in Experiment Two, the estimation accuracy of particle filter decreases because the particles with high dimension are hard to converge to a satisfactory result. In contrast, the method of particle filter combining EM algorithm performs slightly better on the estimation precision and warm-up time in the Experiment Two. By iteratively computing the a , b , c , d , and q in the E-step and M-step respectively, the EM algorithm with particle filter reduces the complexity of estimation and improves the estimation results. Another factor influencing the results is the number of particles. On the one hand, each particle represents the system state in our work. Thus, the larger number of particles means the more diverse dispersion states, which is beneficial to the estimation of system state. On the other hand, the large number of particles will increase the computation cost because the transition and measurement model are calculated for each particle. In our work, 150 and 250 particles are used in the typical particle filter in the two cases, respectively. High accuracy of results and acceptable efficiency prove that these particle numbers are feasible in our experiments.

As for the drawbacks of these experiments, the one is the wind field, which significantly influences the dispersion, is considered to be stable and uniform in our work. However, the actual wind field is relatively dynamic and complex. Therefore, the dynamic modeling of wind field is a research focus in the future works. Another drawback is from the EM algorithm. The method of particle filter combining EM algorithm performs better in the second case, but EM algorithm is likely to converge to a local optimum, which may lead to miss the global optimum.

7. Conclusions

In this paper, two data assimilation methods using particle filter and the method of particle filter combining EM algorithm are developed to improve the accuracy of air contaminant dispersion prediction based on the Gaussian plume model. The architecture of the data assimilation model is presented. Further, a UAV-based sensory system is developed to obtain observations effectively. Then, experiments corresponding to two emission cases are designed and implemented to test the performances of proposed data assimilation methods. The results show that the particle filter can effectively improve the accuracy of model prediction when the dimension of state variables is low. In contrast, when the dimension of state variable becomes higher (the second case), the method of particle filter combining EM algorithm performs better than the typical particle filter in the estimation accuracy and warm-up time. Therefore, these proposed data assimilation methods provide strong support to the prediction of air contaminant dispersion and emergency management in chemical industry parks. Future works include implementing the field experiment in the chemical industry park to verify the data assimilation methods in the true situation and dynamic modeling the wind field for more accurate prediction of atmospheric dispersion model.

Acknowledgments: This study is supported by National Key Research & Development (R&D) Plan under Grant No. 2017YFC0803300 and the National Natural Science Foundation of China under Grant Nos. 71673292, 61503402 and Shanghai Special Foundation of Software and Integrated Circuit under Grant No. 150312.

Author Contributions: Rongxiao Wang and Sihang Qiu conceived and designed the experiments; Rongxiao Wang performed the experiments under the guidance by Bin Chen; Zhengqiu Zhu analyzed the data; Xiaogang Qiu gave important suggestions for data analysis; Rongxiao Wang wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yan, X.; Gu, F.; Hu, X.; Guo, S. In A dynamic data driven application system for wildfire spread simulation, Proceedings of the 2009 Winter Simulation Conference (WSC), 13-16 Dec. 2009, 2009; pp 3121-3128.
2. Xue, H.; Gu, F.; Hu, X. Data assimilation using sequential monte carlo methods in wildfire spread

- simulation. *ACM Transactions on Modeling and Computer Simulation* **2012**, 22, 1-25, DOI:10.1145/2379810.2379816.
3. Krysta, M.; Bocquet, M.; Sportisse, B.; Isnard, O. Data assimilation for short-range dispersion of radionuclides: An application to wind tunnel data. *Atmospheric Environment* **2006**, 40, 7267-7279, DOI:http://dx.doi.org/10.1016/j.atmosenv.2006.06.043.
4. Kalman, R.E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **1960**, 82, 35-45, DOI:10.1115/1.3662552.
5. Pastres, R.; Ciavatta, S.; Solidoro, C. The extended kalman filter (ekf) as a tool for the assimilation of high frequency water quality data. *Ecological Modelling* **2003**, 170, 227-235, DOI:http://dx.doi.org/10.1016/S0304-3800(03)00230-8.
6. Evensen, G. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics* **2003**, 53, 343-367, DOI:10.1007/s10236-003-0036-9.
7. Reddy, K.V.U.; Yang, C.; Tarunraj, S.; Scott, P.D. In Data assimilation in variable dimension dispersion models using particle filters, 2007 10th International Conference on Information Fusion, 9-12 July 2007, 2007; pp 1-8.
8. Gordon, N.J.; Salmond, D.J.; Smith, A.F.M. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, 1993; Vol. 140, pp 107-113.
9. Ng, S.K.; Krishnan, T.; McLachlan, G.J. The em algorithm. In *Handbook of computational statistics: Concepts and methods*, Gentle, J.E.; Härdle, W.K.; Mori, Y., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, **2012**; pp 139-172.
10. Zhao, Z.; Huang, B.; Liu, F. Parameter estimation in batch process using em algorithm with particle filter. *Computers & Chemical Engineering* **2013**, 57, 159-172, DOI:https://doi.org/10.1016/j.compchemeng.2013.03.024.
11. Kim, H.-D.; Komatani, K.; Ogata, T.; Okuno, H.G. Real-time auditory and visual talker tracking through integrating em algorithm and particle filter. In *New trends in applied artificial intelligence: 20th international conference on industrial, engineering and other applications of applied intelligent systems, iea/aie 2007, kyoto, japan, june 26-29, 2007*. Proceedings, Okuno, H.G.; Ali, M., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 280-290.
12. Yang, H.; Huang, Y.; Center, S.E. Evaluating atmospheric pollution of chemical plant based on unmanned aircraft vehicle(uav). *Journal of Geo-Information Science* **2015**, 17, 1269-1274, DOI: 10.3724/sp.j.1047.2015.01269.
13. Hirst, B.; Jonathan, P.; González del Cueto, F.; Randell, D.; Kosut, O. Locating and quantifying gas emission sources using remotely obtained concentration data. *Atmospheric Environment* **2013**, 74, 141-158, DOI:http://dx.doi.org/10.1016/j.atmosenv.2013.03.044.
14. White, B.; Tsourdos, A.; Ashokaraj, I.; Subchan, S.; Zbikowski, R. Contaminant cloud boundary monitoring using uav sensor swarms. In *Aiaa guidance, navigation and control conference and exhibit*, American Institute of Aeronautics and Astronautics: 2007.
15. Pasquill, F. The estimation of the dispersion of windborne material. *Australian Meteorological Magazine* **1961**, 90, 33-49,
16. Steven Hanna, J.C., Helge R.Olesen. Inianapolis tracer data and meteorological data. National Environmental Research Institute: 2005.
17. Qiu, S.; Chen, B.; Zhu, Z.; Wang, Y.; Qiu, X. Source term estimation using air concentration measurements during nuclear accident. *Journal of Radioanalytical and Nuclear Chemistry* **2017**, 311,

- 165-178, DOI:10.1007/s10967-016-4941-z.
18. Turner, D.B. A diffusion model for an urban area. *Journal of Applied Meteorology* **1964**, 3, 83-91, DOI:10.1175/1520-0450(1964)003<0083:ADMFAU>2.0.CO;2.
 19. Gifford, F.A., Jr. Use of routine meteorological observations for estimating atmospheric dispersion. *Nuclear Safety* **1961**, 2, 47-51,
 20. Briggs, G.A. Diffusion estimation for small emissions. Preliminary report; United States, 1973-05-01, 1973.
 21. Vogt, K.J. Empirical investigations of the diffusion of waste air plumes in the atmosphere. **1977**.
 22. Carrascal, M.D.; Puigcerver, M.; Puig, P. Sensitivity of gaussian plume model to dispersion specifications. *Theoretical and Applied Climatology* **1993**, 48, 147-157, DOI:10.1007/BF00864921.