Stochastic source term estimation of HAZMAT releases: algorithms and uncertainty

Yan Wang
Institute of Public Safety Research,
Department of Engineering Physics,
Tsinghua University, Beijing, China
wangyan14@mails.tsinghua.edu.cn

Hong Huang*
Institute of Public Safety Research,
Department of Engineering Physics,
Tsinghua University, Beijing, China
hhong@tsinghua.edu.cn

Wei Zhu
Beijing Research Center of Urban System Engineering,
Beijing Academy of Science and Technology, Beijing, China
zhuweianquan@126.com

ABSTRACT

Source term estimation (STE) of hazardous material (HAZMAT) releases is critical for emergency response. Such problem is usually solved with the aid of atmospheric dispersion modelling and inversion algorithms accompanied with a variety of uncertainty, including uncertainty in atmospheric dispersion models, uncertainty in meteorological data, uncertainty in measurement process and uncertainty in inversion algorithms. Bayesian inference methods provide a unified framework for solving STE problem and quantifying the uncertainty at the same time. In this paper, three stochastic methods for STE, namely Markov chain Monte Carlo (MCMC), sequential Monte Carlo (SMC) and ensemble Kalman filter (EnKF), are compared in accuracy, time consumption as well as the quantification of uncertainty, based on which a kind of flip ambiguity phenomenon caused by various uncertainty in STE problems is pointed out. The advantage of non-Gaussian estimation methods like SMC is emphasized.

Keywords

Bayesian inference, emergency response, hazardous material releases, source term estimation, uncertainty

INTRODUCTION

The release of hazardous material (HAZMAT) is an enormous threat to public safety. The occurrence of these incidents can be due to various accidents in industries as well as the increasingly rampant terrorist acts. In this type of events, specific government agencies like office of emergency management are in charge of tracking and forecasting HAZMAT transport, as well as taking effective measures to mitigate damages, where accurate information about the source location and release rate plays an irreplaceable role.

Source term estimation (STE) is the process of recovering the source parameters like location and release rate according to some measurement data, which can be integrated into information systems for crisis response and management as a function module to help enhance situation awareness and assessment of

HAZMAT releases. In addition to the accuracy of measurement concentration and meteorological data, another two key factors of the success of STE are the forward dispersion model and the back-calculation methods, to both of which a lot of research has been devoted. In most cases, these two factors are independent of each other and thus can be combined in different ways. Currently, there exist several types of back-calculation approaches to solve the STE problem, among which optimization methods and probabilistic methods are the most popular.

Optimization methods use various optimization algorithms to minimize the difference between the output of forward model and the measurement data. Thomson et al. (2007) used simulated annealing algorithm with three different cost functions to locate a gas source. Zheng and Chen (2010) compared the pattern search method with a gradient-based algorithm and an intelligent optimization algorithm, achieving optimal solutions in a relatively shorter time.

Probabilistic methods based on Bayesian inference treat the source parameters and measurement data as random variables and can incorporate the error of forward dispersion model and the noise of measurement data. The most widely used probabilistic approach is the Markov chain Monte Carlo (MCMC) method (Keats, Yee and Lien, 2007; Kim, Jang and Lee, 2011; Senocak, Hengartner, Short and Daniel, 2008). Recently, the sequential Monte Carlo (SMC) approach, also known as particle filter (PF), is used to solve the STE problem (Hofman, Šmídl and Pecha, 2013; Šmídl and Hofman, 2013). Another notable probabilistic method is the ensemble Kalman filter (EnKF) (Zhang, Su, Yuan, Chen and Huang, 2014). Details about the development of Bayesian state-estimation algorithms can be found in Ching, Beck and Porter (2006).

The quantification of uncertainty can yield a deeper insight into the capabilities and limitations of the STE process. According to Rao (2005), uncertainty in atmospheric dispersion modelling is related to: (a) uncertainty in data and parameter including initial and boundary conditions; (b) model uncertainty caused by simplified treatment of complex physical or chemical processes and approximate numerical solutions; and (c) stochastic uncertainty resulting from the turbulent nature of the atmosphere as well as unpredictable human activities.

In STE problems, another uncertainty, uncertainty in inversion algorithms caused by different assumptions during the modelling process and various parameter settings in the computational process, should also be considered. In this paper, three stochastic algorithms for STE, namely MCMC, SMC and EnKF, are compared in accuracy, time consumption as well as the quantification of uncertainty. The Bayesian formulation of the STE problem is given in a way where the three methods can be implemented under the same framework. The advantage of non-Gaussian estimation methods like SMC is emphasized, constituting a new trend in STE problem.

THE BAYESIAN FORMULATION OF THE PROBLEM

From a Bayesian perspective, both the source parameters and the measurement data are considered as random variables. Let $\mathbf{X}^t = \{X_1^{(t)}, X_2^{(t)}, \dots, X_n^{(t)}\}$ denote the n source parameters including location and release rate at time t, $\mathbf{Y}^t = \{Y_1^{(t)}, Y_2^{(t)}, \dots, Y_k^{(t)}\}$ denote the measurement concentration at time t, and $\mathbf{F}^t = \{F_1^{(t)}, F_2^{(t)}, \dots, F_k^{(t)}\}$ denote the predicted concentration at time t through running a forward dispersion model where k is the number of gas sensors.

According to Guo, Yang, Zhang, Weng and Fan (2009), if the measurement error and forward model error at any sensor at any time are assumed to be independent and satisfy the gaussian distribution with zero mean and a known standard deviation, the likelihood function can be written as

$$p(\mathbf{Y}\,|\,\mathbf{X}) = \prod_{t=1}^{m}\prod_{i=1}^{k} p(Y_i^{(t)}\,|\mathbf{X}) \propto \exp\left\{-\sum_{t=1}^{m}\sum_{i=1}^{k}\frac{[F_i^{(t)}(\mathbf{X}) - Y_i^{(t)}]^2}{2(\sigma_{y,i}^{(t)2} + \sigma_{f,i}^{(t)2})}\right\}$$

and the posterior distribution of source parameters can be written as

$$p(\mathbf{X}\,|\,\mathbf{Y}) \propto p(\mathbf{X})\exp\left\{-\sum_{t=1}^{m}\sum_{i=1}^{k}\frac{[F_i^{(t)}(\mathbf{X}) - Y_i^{(t)}]^2}{2(\sigma_{y,i}^{(t)2} + \sigma_{f,i}^{(t)2})}\right\} / \; p(\mathbf{Y})$$

where $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^m\}$ and $\mathbf{Y} = \{\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^m\}$ denote the source parameter history and measurement data history respectively, $\boldsymbol{\sigma}_y^t = \{\sigma_{y,1}^{(t)}, \sigma_{y,2}^{(t)}, \dots, \sigma_{y,k}^{(t)}\}$ and $\boldsymbol{\sigma}_f^t = \{\sigma_{f,1}^{(t)}, \sigma_{f,2}^{(t)}, \dots, \sigma_{f,k}^{(t)}\}$ denote the standard deviation of measurement error and forward model error at time t respectively.

From a more general perspective, the discrete-time state space model of STE problem can then be expressed as

$$\mathbf{X}^t = M_t(\mathbf{X}^{t-1}, \mathbf{v}^t)$$
$$\mathbf{Y}^t = H_t(\mathbf{X}^t, \boldsymbol{\omega}^t)$$

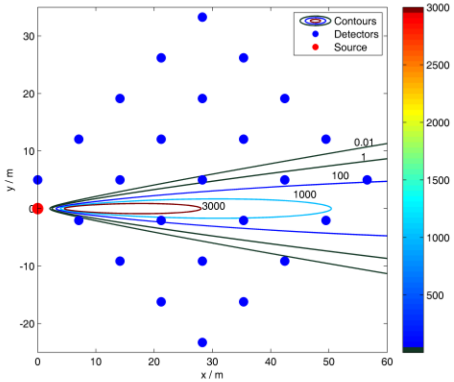

Figure 1. Schematics of the experiment setup and contours of the concentrations on the ground level

where the vectors $\mathbf{v}^t$ and $\boldsymbol{\omega}^t$ represent the system noise and the observation noise at time t respectively. In case of time-invariant uncertain parameter estimation in this study, the forecast operator M_t represents the variation of system parameters while the observation operator H_t includes the forward dispersion process and the measurement process.

THE SYNTHETIC EXPERIMENT

In common with much previous work (Thomson et al., 2007; Zheng and Chen, 2011), the Gaussian plume model was employed, which can be expressed as

$$C(x,y,z,H_e) = \frac{Q}{2\pi\mu\sigma_y\sigma_z}\exp(-\frac{1}{2}\cdot\frac{y^2}{\sigma_y^2})\cdot\{\exp[-\frac{1}{2}\cdot\frac{(z-H_e)^2}{\sigma_z^2}]+\exp[-\frac{1}{2}\cdot\frac{(z+H_e)^2}{\sigma_z^2}]\}$$

Simulation parameters are set the same as those in Zheng and Chen (2011), where Q equals 10434.78 g/s.

In our synthetic experiment, sensors were uniformly dispatched following a 5×5 grid and the real source location (x, y) is (0, 0). The wind was along the positive direction of X-axis. All the source parameters and the wind field do not change with time. The experiment setup and concentration contours are shown in Figure 1.

Considering the existence of various uncertainty, measurement data at the 25 sensors are generated by adding a maximum of 50% random noise to the output of the Gaussian plume model, namely, $\mathbf{Y}^t$ satisfies the uniform distribution U (0.5$\mathbf{F}^t$, 1.5$\mathbf{F}^t$). In this study, a flat prior on the location and strength of the source is assumed, namely the source location (x,y) satisfies the uniform distribution (U[-10,60], U[-20,30]) and the strength Q satisfies the uniform distribution U[500,20000].

DESCRIPTION OF THE THREE STOCHASTIC METHODS

The basic idea of the MCMC method is to generate a Markov chain whose stationary distribution is the posterior distribution of source parameters. The Metropolis-Hastings algorithm (Johannesson, Hanley and Nitao, 2004) was used to generate realizations from the posterior distribution $p(\mathbf{X}|\mathbf{Y})$. The starting value $\mathbf{X}_0$ is (50,-15, 1000), which means an initial release of 1000 g/s at (50,-15). The proposal distribution and the acceptance ratio are defined according to Wu, Yang, Zhang and Qiao (2013), which means the candidate realization $\mathbf{Z}_{j+1}$ at (j+1) th iteration are generated by the distribution $q(\mathbf{X}_j, \mathbf{Z}_{j+1}) = Gau(\mathbf{X}_j, \boldsymbol{\sigma}_j{}^2)$ and the acceptance ratio can be calculated by the following equation

$$\alpha(\mathbf{X}_j, \mathbf{Z}_{j+1}) = \min[1, \frac{p(\mathbf{Z}_{j+1})p(\mathbf{Y}|\mathbf{Z}_{j+1})q(\mathbf{Z}_{j+1}, \mathbf{X}_j)}{p(\mathbf{X}_j)p(\mathbf{Y}|\mathbf{X}_j)q(\mathbf{X}_j, \mathbf{Z}_{j+1})}]$$

The basic idea of SMC (Johannesson et al., 2004) is to fully take advantage of the sequential nature of the posterior distribution. The proposal distribution was taken equal to the prior distribution, the same as $q(\mathbf{X}_j, \mathbf{Z}_{j+1})$, then the calculation of importance weights for time-invariant uncertain parameters can be expressed as

$$w_{j+1} \propto w_j \frac{p(\mathbf{Y}|\mathbf{X}_{j+1})\, p(\mathbf{X}_{j+1}|\mathbf{X}_j)}{q_{j+1}(\mathbf{X}_{j+1}|\mathbf{X}_j)} = w_j\, p(\mathbf{Y}|\mathbf{X}_{j+1})$$

The EnKF method (Evensen, 2003) is a sequential data assimilation technique for nonlinear system models and observations. As is illustrated above, the STE problem was formulated into a state estimation problem of discrete-time dynamical random system. In order to compare the three algorithms in a fair manner, the forecast operator M_t representing the variation of system parameters was also chosen to be the same as $q(\mathbf{X}_j, \mathbf{Z}_{j+1})$ mentioned above.

RESULTS AND DISCUSSION

The posterior distribution of source parameters estimated by MCMC, SMC and EnKF is shown in Figure 2 and Figure 3.

Since parameters like proposal distribution and measurement errors were all chosen to be the same for the above three methods, it makes sense to compare their performance horizontally. In order to acquire a more quantitative understanding of the performance and properties of the three methods, the estimation results and some key indexes are summarized in Table 1and Table 2 respectively.

From Figure 2-3 and Table 1, it can be seen that the estimation results of MCMC and SMC are very similar in accuracy and the quantification of uncertainty while the results of EnKF show a different pattern.

Firstly, the EnKF method tends to overestimate the release rate, and has a larger variance on the downwind location x but a smaller variance on the cross wind location y than MCMC and SMC. Secondly, the posterior distribution of the cross wind location y estimated by MCMC and SMC is bimodal while EnKF results in a unimodal distribution. This is because that EnKF tend to fail to capture the non-Gaussian features of the posterior (Mandel and Beezley, 2009), proving that the capacity for quantifying uncertainty in STE are not the same for algorithms based on different theories and assumptions.

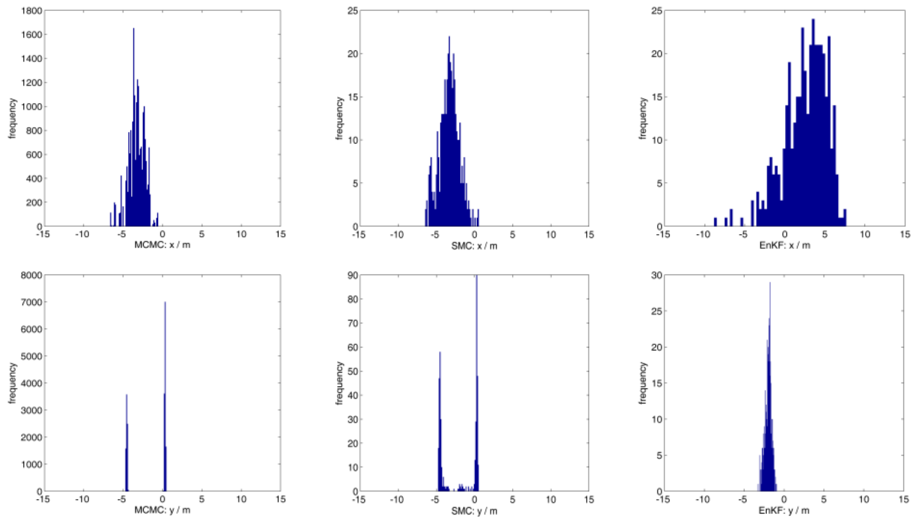

Figure 2. Posterior distribution of source location estimated by MCMC, SMC and EnKF.

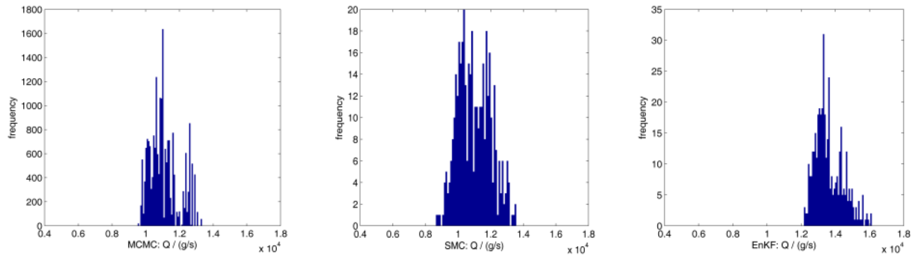

Figure 3. Posterior distribution of source strength estimated by MCMC, SMC and EnKF.

		MCMC	SMC	EnKF
Location x (m)	Mean	-3.0	-3.0	1.2
	Std	1.0	1.5	4.3
	95% CIs	[-4.9,-1.1]	[-5.9,0.0]	[-7.8,9.5]
Location y (m)	Mean	-1.0	-1.4	-2.2
	Std	1.8	2.2	0.9
	95% CIs	[-4.1,0.4]	[-4.7,0.4]	[-3.6,-0.5]
Release rate (g/s)	Mean	10755.7	10644.3	14079.1
	Std	786.1	996	1064.2
	95% CIs	[9336.2, 12322.8]	[8822.3, 12707.8]	[12015.5, 16195.2]

Table 1. Source term estimation results by the three stochastic algorithms based on 20 repeated calculations

	MCMC	SMC	EnKF
Convergence step	17325	138	173
Number of running the forward model per iteration	1	100	101
Particle number or ensemble size per iteration	1	100	100
Total running of the forward model until convergence	17325	13800	17473

Table 2. Key indexes concerned with the performance and properties of the three stochastic algorithms based on 20 repeated calculations

The appearance of bimodal posterior distribution is a kind of flip ambiguity which much research effort has been paid to in the Network Localization problem (Severi, Abreu, Destino and Dardari, 2009). The ambiguity phenomenon in STE problem can be caused by various uncertainty and remains to be further investigated. In this paper, the flip ambiguity is related to the axial symmetry of the experiment setup (eg. the collinearity of sensors with relatively larger measured values and the axial symmetry of the concentration field) as well as the random noise added to the synthetic concentration data to incorporate the presence of various uncertainty.

From Table 2, it can be seen that the total running of the forward model until convergence is almost the same for the three methods. But SMC and EnKF are not Markovian and inherently parallel. They run many forward models at the same time, and thus reach convergence in much less iteration steps and time compared with MCMC. This advantage is more obvious if the forward dispersion model is time-consuming, which is often the case.

CONCLUSION

With the presence of various uncertainty, the STE problem is very challenging. The quantification of uncertainty can yield a deeper insight into the capabilities and limitations of the STE process.

In this study, three stochastic methods for STE are implemented under a unified Bayesian framework, and are compared in accuracy, time consumption as well as the quantification of uncertainty. The estimation results of MCMC and SMC are similar while EnKF tends to overestimate the release rate and fails to capture the non-Gaussian features of the posterior. SMC and EnKF are inherently parallel and cost much less time to convergence. With all these factors being considered, the non-Gaussian SMC method is regarded as a better choice for solving steady-state STE problems.

Based on the analysis of the synthetic STE experiment, the flip ambiguity phenomenon is pointed out and considered to be caused by the axial symmetry of the experiment setup as well as the random noise added to the synthetic concentration data.

Notably, EnKF has its advantage in solving large-scale time-varying problems. In further research, multi-scale and non-stationary dispersion scenarios are to be investigated and different methods can be integrated to better quantify the uncertainty in STE problems. Analyzing the relative importance of various uncertainty is also a meaningful research topic.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 71473146, 71173128) and the Ministry of Science and Technology of the People´s Republic of China under Grant No. 2015BAK12B01.

REFERENCES

1. Ching, J., Beck, J. L. and Porter, K. A. (2006) Bayesian state and parameter estimation of uncertain dynamical systems, *Probabilistic engineering mechanics*, 21, 1, 81-96.
2. Evensen, G. (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean dynamics*, 53, 4, 343-367.
3. Guo, S., Yang, R., Zhang, H., Weng, W. and Fan, W. (2009) Source identification for unsteady atmospheric dispersion of hazardous materials using Markov Chain Monte Carlo method, *International Journal of Heat and Mass Transfer*, 52, 17, 3955-3962.
4. Hofman, R., Šmídl, V. and Pecha, P. (2013) Development of Real-Time Bayesian Data Assimilation System for Off-Site Consequence Assessment.
5. Johannesson, G., Hanley, B. and Nitao, J. (2004) Dynamic Bayesian models via Monte Carlo–an introduction with examples. Lawrence Livermore National Laboratory, UCRL-TR-207173.
6. Keats, A., Yee, E. and Lien, F. S. (2007) Bayesian inference for source determination with applications to a complex urban environment, *Atmospheric environment*, 41, 3, 465-479.
7. Kim, J. Y., Jang, H. K. and Lee, J. K. (2011) Source reconstruction of unknown model parameters in atmospheric dispersion using dynamic Bayesian inference, *Progress in Nuclear Science and Technology*, 1, 460-463.
8. Mandel, J. and Beezley, J. D. (2009) An ensemble Kalman-particle predictor-corrector filter for non-Gaussian data assimilation, *In Computational Science–ICCS 2009*, 470-478, Springer Berlin Heidelberg.
9. Rao, K. S. (2005). Uncertainty analysis in atmospheric dispersion modeling. Pure and applied geophysics, 162(10), 1893-1917.
10. Senocak, I., Hengartner, N. W., Short, M. B. and Daniel, W. B. (2008) Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference, *Atmospheric Environment*, 42, 33, 7718-7727.
11. Severi, S., Abreu, G., Destino, G. and Dardari, D. (2009) Understanding and solving flip-ambiguity in network localization via semidefinite programming. *In Global Telecommunications Conference, 2009. GLOBECOM 2009.* IEEE (pp. 1-6). IEEE.
12. Šmídl, V. and Hofman, R. (2013) Efficient Sequential Monte Carlo Sampling for Continuous Monitoring of a Radiation Situation. Technometrics, (just-accepted).
13. Thomson, L. C., Hirst, B., Gibson, G., Gillespie, S., Jonathan, P., Skeldon, K. D. and Padgett, M. J. (2007) An improved algorithm for locating a gas source using inverse methods, *Atmospheric Environment*, 41, 6, 1128-1134.
14. Wu, N., Yang, R., Zhang, H. and Qiao, L. (2013) Decentralized inverse model for estimating building fire source location and intensity, *Journal of Thermophysics and Heat Transfer*, 27, 3, 563-575.
15. Zhang, X. L., Su, G. F., Yuan, H. Y., Chen, J. G. and Huang, Q. Y. (2014) Modified ensemble Kalman filter for nuclear accident atmospheric dispersion: Prediction improved and source estimated, *Journal of hazardous materials*, 280, 143-155.
16. Zheng, X. and Chen, Z. (2010) Back-calculation of the strength and location of hazardous materials releases using the pattern search method, *Journal of hazardous materials*, 183, 1, 474-481.

17. Zheng, X. and Chen, Z. (2011) Inverse calculation approaches for source determination in hazardous chemical releases, *Journal of Loss Prevention in the Process Industries*, 24, 4, 293-301.
18.