Atmospheric Environment 164 (2017) 128–138

Contents lists available at ScienceDirect

Atmospheric Environment

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

# Bayesian identification of a single tracer source in an urban-like environment using a deterministic approach

Fei Xue [a], Xiaofeng Li [a, b, *], Weirong Zhang [c]

[a] Department of Building Science, School of Architecture, Tsinghua University, Beijing, China
[b] Beijing Key Lab of Indoor Air Quality Evaluation and Control, Beijing 100084, China
[c] Department of Architecture, Tokyo Polytechnic University, Japan

## HIGHLIGHTS

- We estimated the location and strength of an emission source in built-up area.
- A two-step deterministic method is presented in the Bayesian inference framework.
- Presented method is faster and more accurate than the existing stochastic ones.
- Presented method is evaluated with two wind tunnel cases of an urban mock-up.
- Performance indicators are tested to evaluate the credibility of certain estimations.

ARTICLE INFO

*Article history:*
Received 22 February 2017
Received in revised form
24 May 2017
Accepted 26 May 2017
Available online 29 May 2017

*Keywords:*
Source term estimation
Bayesian inference
Performance evaluation
Urban dispersion
CFD

ABSTRACT

This paper presents a two-step deterministic approach for identifying an unknown point source with a constant emission rate in built-up urban areas. The analytic form of the marginal posterior probability density function of the source location is derived to estimate the source location. The emission rate is then estimated using the conditional posterior distribution. Such a procedure deconstructs the calculation of the joint posterior distribution of the source parameters into calculations of two separate distributions and can thus be easily calculated directly and accurately without stochastic sampling. The proposed method is tested using real data obtained in two wind tunnel scenarios of contaminant dispersion in typical urban geometries represented by block arrays. Computational fluid dynamics (CFD) modeling and the adjoint equations are used to calculate the building-resolving source-receptor relationship required in the identification. The estimated source parameters in both cases are close to true values. In both cases, the source locations are identified with errors less than half of the block size, and the emission rates are well estimated, with only slight overestimation. Moreover, in this paper, we test two potential performance indicators for a posteriori evaluation of the credibility of a certain estimation. One indicator is the size of the highest probability density region, and the other is the angle between the observed and predicted concentration vectors, which is derived from the analytic form of the marginal posterior distribution of the source location. Synthetic concentration data are generated to test the validity of both indicators. It is found that the former is not appropriate for denoting the credibility of estimations but that the latter shows a strong correlation with estimation performance and is likely to be an effective performance indicator for Bayesian source term estimation.



## 1. Introduction

The release of hazardous material into the atmosphere is a major threat to public health and security. Such incidents may be either intentional attacks using chemical, biological, radiological, or nuclear (CBRN) agents or accidental events such as gas leaks in an industrial park. In either case, the rapid and accurate determination of the source location and strength is crucial for initiating an appropriate post-incident response to reduce human exposure and further impairment. Such an identification of unknown sources is a

* Corresponding author. Department of Building Science, School of Architecture, Tsinghua University, Beijing, China.
*E-mail address:* xfli@mail.tsinghua.edu.cn (X. Li).

http://dx.doi.org/10.1016/j.atmosenv.2017.05.046

typical source term estimation (STE) problem, which has recently attracted much attention (Hutchinson et al., 2017; Singh et al., 2015).

STE is an assimilation process of the observed concentration data obtained by a sensor network and the predicted concentration data calculated by an atmospheric dispersion model. However, this process is accompanied by several difficulties: (i) the sensors are sparsely distributed in space and generally outnumbered by possible source locations, so multiple releasing scenarios may exist that match the observations; (ii) the observations contain errors due to sensor noise and information loss caused by potential averaging processes, and (iii) the predicted concentration data contain errors due to uncertainty in dispersion models, such as imperfect numerical models, inadequate representation of the complex terrain or urban geometry, or inaccurate meteorological conditions. Therefore, STE is perceived as an ill-posed inverse problem (Singh et al., 2015) that is characterized by non-uniqueness and unstable solutions.

To make the STE problems solvable, multiple methods have been developed based on various approaches. One intuitive solution is to directly invert the transport equation. In this context, Kato et al. (2004) estimated the origin of the continental scale transport of airborne contaminants using backward trajectory analysis. Zhang and Chen (2007a, 2007b) employed both quasi-reversibility and pseudo-reversibility methods to locate indoor sources. Another more general idea is to minimize a cost function, which measures the discrepancies between observed and predicted concentrations. Various cost functions were constructed according to the features of different STE problems. The simplest form is the classical least squares (Matthes et al., 2005; Sharan et al., 2012; Singh and Rani, 2015). A more robust alternative is to involve a regularization term in the cost function. The main idea is to apply constraints to the solution space, thus causing the ill-posed problems to become well-posed problems with unique solutions. Different regularization terms were implemented to solve STE problems, such as retrieving the emission rates of a point source with a known location (Kathirgamanathan et al., 2004) and locating unknown point sources (Akcelik et al., 2003; Thomson et al., 2007). Another strategy is called renormalization, which can avoid spurious sources close to sensors by assigning weights according to the geometry of the sensor network (Issartel et al., 2007). This method was adopted to reconstruct point sources in flat terrain (Sharan et al., 2009; Singh et al., 2013) and built-up areas (Kumar et al., 2015, 2016).

In addition to the aforementioned approaches, STE problems can also be solved using the Bayesian inferential framework. In this context, STE problems are considered in terms of probabilistic logic. Source parameters are treated as random variables rather than constants. Correspondingly, the Bayesian inference not only yields the results of extreme points of parameters but also their probability distributions. This important feature provides a natural and rigorous way to quantify uncertainties consistent with all a priori information. Yee and Flesch (2010) derived the analytic form of the posterior probability distribution for the emission rates of known sources and used them to estimate the emission rates of four simultaneous tracer releases. For releases with unknown locations, the source parameters become multidimensional, which makes the direct calculation of posterior distribution computationally expensive. To reduce the computational cost, stochastic sampling approaches have been implemented to approximate the probability distribution, among which the Markov chain Monte Carlo (MCMC) method is the most popular. Chow et al. (2008) used a combination of Bayesian inference and the MCMC method to estimate source parameters. Those authors ran thousands of forward simulations to obtain the source-receptor relationships. Keats et al. (2007) refined this method using adjoint transport equations and reconstructed single point releases in a complex urban area. Delle Monache et al. (2008) and Yee et al. (2014) implemented similar approaches to solve STE problems on a continental scale. Hazart et al. (2014) estimated the locations of single temporal sources with different releasing profiles. Ristic et al. (2015) proposed a likelihood-free Bayesian method with three different candidate dispersion models to investigate source parameters. In addition to locating a single source, this method was further extended to multiple sources, including cases in which the number of sources is known (Yee, 2007) and unknown (Wade and Senocak, 2013; Yee, 2008, 2012). Alternative stochastic sampling approaches are available and have been investigated, such as adaptive multiple importance sampling (AMIS) (Rajaona et al., 2015). To summarize, the Bayesian inferential framework has recently been widely used to solve STE problems, especially with stochastic sampling method to approximate the joint probability distribution of source parameters.

However, when locating an unknown source, the marginal posterior distribution of location, rather than the joint posterior probability distribution of both location and emission rate, is needed to estimate the source location. In this paper, a two-step approach based on Bayesian inference is developed to locate an unknown steady releasing point source using a deterministic method instead of stochastic sampling approaches. First, the analytic form of the marginal posterior function of the point source location is derived to estimate the source location. Then, the emission rate is calculated using the conditional posterior distribution, which can also be derived analytically. Such a procedure decomposes the calculation of the joint posterior distribution of source parameters into calculations of two separate distributions and can thus be obtained directly without stochastic sampling. This method reduces the computational burden and improves accuracy compared with the conventional MCMC approaches. The proposed method is evaluated with two wind tunnel cases with an urban-like geometry. Additionally, this method derives a performance indicator for STE problems, and a large amount of synthetic concentration data is generated to test its capacity to effectively evaluate the credibility of a certain estimation.

It should be noted that although the method proposed in this paper addresses a single point source with a constant emission rate, it can be applied to multiple sources by following the algorithm presented by Xue and Zhai (2017), which decomposes the identification of multiple sources into several identifications of a single source. However, the proposed method cannot be generalized to variable emission rates or nonpoint sources such as area sources or line sources.

## 2. Bayesian framework

### 2.1. Problem formulation

We are interested in estimating the unknown characteristics $\boldsymbol{\theta}$ of a source, given the measurements, $\boldsymbol{\mu}$, obtained from a network of limited sensors. It is assumed that the source is constantly releasing and is located at ground level, thus denoted by $\boldsymbol{\theta} = (\boldsymbol{x_s}, q)$, where $\boldsymbol{x_s} = (x, y)$ is the location and $q$ is the emission rate of the source. Bayesian theory provides a rigorous way to make the inference based on all of the information presented by the problem. More specifically, the estimation result is obtained as the posterior distribution, which is presented by the Bayes' rule as

$$p(\boldsymbol{\theta}|\boldsymbol{\mu}) = \frac{p(\boldsymbol{\mu}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{\mu})} \propto p(\boldsymbol{\mu}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (1)$$

The terms in Bayes' rule embodied in Eq. (1) can be interpreted