Atmospheric Environment 122 (2015) 1–9

Contents lists available at ScienceDirect

Atmospheric Environment

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

Source area identification with observation from limited monitor sites for air pollution episodes in industrial parks

Zihan Huang, Yuan Wang, Qi Yu*, Weichun Ma, Yan Zhang, Limin Chen

Department of Environmental Science and Engineering, Fudan University, Shanghai, 200433, China

HIGHLIGHTS

- Source area analysis is designed to do source identification with limited monitors.
- Uncertainties in the source parameters were involved in the back-calculation.
- The characteristic of the deduced source area was illustrated by case studies.

ARTICLE INFO

Article history:
Received 29 December 2014
Received in revised form
14 August 2015
Accepted 18 August 2015
Available online 28 August 2015

Keywords:
Source area analysis
Abnormal release
Industrial park
Limited monitor sites
Back-calculation

ABSTRACT

Air pollution episodes of unknown origins are often detected by online equipment for air quality monitoring in industrial parks in China. The number of monitors available to provide observation data, as well as the source information, is often very limited. In such case, the identification of a potential source area is more practical than the precise back-calculation of the real source. The potential source area which can be deduced from the observation data from limited monitors was concerned in this paper. In order to do the source area identification, two inverse methods, a direct method and a statistical sampling method, were applied with a Gaussian puff model as the forward modeling method. The characteristic of the potential source area was illustrated by case studies. Both synthetic and real cases were presented. The distribution of the source locations and its variation with the other unknown source parameters were mainly focused in the case study. As a screening method, source area identification can be applied not only when the number of effective monitors is limited but also when an ideal number of monitors are available as long as the source information is almost uncertain.

1. Introduction

Air pollution episodes of unknown origins are often detected by online equipments for air quality monitoring in industrial parks in China. Such episodes are usually caused by abnormal or furtive emissions which are hard to be traced. Usually they do not last long, e.g. for several hours or less. Although they are not as serious as significant accidents which lead to acute health damage or property loss, the environmental impact is still negligible.

Inverse methods are applicable to back-track the source information with the observation data. The unknown source parameters usually includes the location (x, y, H) where x and y are the coordinates and H is the source height, the start time T, and the source strength $Q(t)$ during the emission period t. It is theoretically feasible to get a global optimal estimation of these parameters by inverse methods providing appropriate number of monitors e.g. no less than five (Crenna et al., 2008). Even more monitors were recommended by other researchers. For example, Rudd et al. (2012) and Singh and Rani (2014) suggested that the number of monitors should be at least higher than the number of unknowns because of uncertainties in the measurements, the model and the variable quality in sensor placement; Allen et al. (2007b) and Haupt et al. (2007) demonstrated that at least an 8-by-8 grid of receptors was necessary to find the combination of source location, source strength, and wind direction by a GA system. However, such demand for monitors can seldom be satisfied because the scales of the monitoring networks in industrial parks are generally smaller than required. The limitation of monitoring sites as well as source information leads to infinite solutions to such inverse problem in industrial parks.

Although the solutions with limited monitors are infinite, the

* Corresponding author.
E-mail address: qiyufd@gmail.com (Q. Yu).

http://dx.doi.org/10.1016/j.atmosenv.2015.08.048
1352-2310/

reasonable sources do not spread all over the whole domain of an industrial park due to the constraint of the transportation and dispersion laws. In such case, the identification of a potential source area was more practical than the precise back-calculation of the real source. So the potential source area which can be deduced from the observation data from limited monitors was concerned in this paper. Inverse methods for back-tracking the sources for such episodes were applied to do the source area identification.

There are many inverse methods for seeking an optimal approximation of the source parameters such as genetic algorithm (Haupt, 2005; Haupt et al., 2007; Allen et al., 2007b), simulated annealing algorithm (Thomson et al., 2007), Newton–Raphson method (Najafi and Gilbert, 2003), least square method (Singh et al., 2013; Singh and Rani, 2014), pattern search (Zheng and Chen, 2010) and etc. It is applicable to obtain a collection of optimal solutions by doing the back-calculation of source parameters by such method repeatedly. However, statistical methods based on Bayesian inference coupled with stochastic sampling (Chow et al., 2008; Guo et al., 2009; Keats et al., 2007; Neuman et al., 2006) are more preferable for the simulation of the distributions of multi-dimension variables. A typical application of Bayesian inference and stochastic models based on Markov chain Monte Carlo (MCMC) sampling for the source identification was reported by Chow et al. (2008). The source parameters including x, y and Q were sampled independently by 20 000 MCMC iterations with a burn-in phase of about 10 000 iterations. For the isolated building example with certain artificial measurement error added to the synthetic data, four monitors were placed in a diamond-shaped array in the lee of the building while the source was in the upwind of the building. The peak of the simulated distribution with a probability of about 0.2 occurred just upwind of the actual source location. The probability of the peak of the release rate was about 0.52. The peak of the release rate coincided with the actual value. For the urban environment example, 15 monitors were given. The examination of the appropriate number of monitors indicated that even as few as two monitors may be useful in an urban environment provided they are placed appropriately and the results were more sensitive to the arrangement of the monitors than to the number of monitors.

The inverse method relies on a forward model to calculate the corresponding concentrations for given source parameters. A variety of models has been adopted in literature like high-resolution CFD models (Chow et al., 2008; Keats et al., 2007), Gaussian plume models (Haupt, 2005; Haupt et al., 2007; Allen et al., 2007b), a Gaussian puff model (Haupt et al., 2009), the SCIPUFF model (the Second-Order Closure Integrated Puff model) (Allen et al., 2007a), the UDM model (an empirical puff model) (Neuman et al., 2006), a Lagrangian puff model (Najafi and Gilbert, 2003) and a 3-D numerical model (Guo et al., 2009). Najafi and Gilbert (2003) tested a Lagrangian puff model against the Desert Tortoise sensor data and found that the maximum error between the simulated and measured rate ranged from 20% to 90%. The numerical models are usually used for predicting plume evolution in complex domains. Testing with the field data (IOP3) in Oklahoma City with the UDM model (Neuman et al., 2006), three distinct peaks were both found in the probability distributions of (x, y) and Q. The highest peak for (x, y) was within 20 m of the actual source location and the lowest peak for Q corresponded with the actual mass of the source. Testing with the same field data with a CFD model by Chow et al. (2008), the peak of the (x, y) distribution was located approximately 70 m south of the actual source location and the peak of the distribution of Q fell near about one fifth of the actual Q value. It is really difficult to tell which model was more suitable for this case based on the quality of the peak values.

In order to do the source identification with limited monitors, a statistical model with a Gaussian puff model was adopted as the inverse method. Gaussian puff models as well as segmented plume models or heavy gas dispersion models are guideline models for the environmental risk assessment on projects (MEPPRC, 2004). The Gaussian puff model was adopted in our study due to its capability of simulating unsteady emissions and wind conditions as well as its simplicity. Actually the Gaussian puff model is not the most suitable approach for all the dispersion scenarios in an industrial park considering the complex terrain or the complicated emission scenarios. We did not use a refined method because running a refined model requires more details about the source otherwise a refined model would not necessarily provide better simulation than a simple model.

The statistical model was based on Bayesian inference. Gibbs sampling (Chan, 2010; Voss, 2014), a MCMC algorithm for obtaining a sequence of observations which are approximated from the joint probability distribution of several random variables, was used to do stochastic sampling. In addition to that, a non-statistical method which directly takes samples at finite source locations (hereinafter referred to as a direct method) was also taken into consideration. In the direct method, the sampling on the distribution of the horizontal location of the source was simplified aiming to reduce the computation cost. Actually we needed only one of them because their outputs would be consistent, so that the direct method was applied in all the cases. The statistical method was used in the first case to show the spatial distribution of the samples.

In the following sections, the inverse methods we used will be introduced first; then the characteristics of the source area revealed by the inverse method will be described via different case studies; finally discussions and conclusions are given.

2. Method

Source area information was drawn from the solutions obtained by back-calculation. Inverse methods with a Gaussian puff model were used in the back-calculation.

2.1. Gaussian puff model

A Gaussian puff model was employed to allow the simulation of the spatial and temporal variations of both the wind data and the concentration measurements. The model was used to produce both synthetic concentration observations and concentration predictions for each trial solution generated in the inverse section:

$$C_i = \frac{Q_i}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_z}\exp\left[-\frac{(x-x_i)^2}{2\sigma_x^2}\right]\exp\left[-\frac{(y-y_i)^2}{2\sigma_y^2}\right] \left\{\exp\left[-\frac{(z-z_i)^2}{2\sigma_z^2}\right]+\exp\left[-\frac{(z+H_i)^2}{2\sigma_z^2}\right]\right) \tag{1}$$

where C_i is the concentration contribution at site (x, y, z) from puff i at (x_i, y_i, H_i), Q_i is the total mass in a puff, t is the time after release, H_i is the effective height of the release, and $(\sigma_x, \sigma_y, \sigma_z)$ are the dispersion coefficients. The dispersion coefficients were computed as the functions of the dispersion distance according to MEPPRC (1991). Empirical functions for computing dispersion coefficients were provided for a sampling time of 30 min in the guideline (MEPPRC, 1991). In addition, modifications according to the sampling time were required for shorter or longer sampling times.

2.2. Inverse methods

Consider the variables x, y, H, T and $Q(t)$ to be back-calculated.