Atmospheric Environment 178 (2018) 158–163

Contents lists available at ScienceDirect

Atmospheric Environment

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

Short communication

Atmospheric dispersion prediction and source estimation of hazardous gas using artificial neural network, particle swarm optimization and expectation maximization

Sihang Qiu[a,b], Bin Chen[a,*], Rongxiao Wang[a], Zhengqiu Zhu[a], Yuan Wang[c], Xiaogang Qiu[a]

[a] College of System Engineering, National University of Defense Technology, 410073 Changsha, China
[b] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 XE Delft, The Netherlands
[c] College of Territorial Resources and Tourism, Anhui Normal University, 241003 Wuhu, China

ARTICLE INFO

Keywords:
Atmospheric dispersion
Source estimation
Neural network
Particle swarm optimization (PSO)
Expectation maximization (EM)

ABSTRACT

Hazardous gas leak accident has posed a potential threat to human beings. Predicting atmospheric dispersion and estimating its source become increasingly important in emergency management. Current dispersion prediction and source estimation models cannot satisfy the requirement of emergency management because they are not equipped with high efficiency and accuracy at the same time. In this paper, we develop a fast and accurate dispersion prediction and source estimation method based on artificial neural network (ANN), particle swarm optimization (PSO) and expectation maximization (EM). The novel method uses a large amount of predetermined scenarios to train the ANN for dispersion prediction, so that the ANN can predict concentration distribution accurately and efficiently. PSO and EM are applied for estimating the source parameters, which can effectively accelerate the process of convergence. The method is verified by the Indianapolis field study with a SF_6 release source. The results demonstrate the effectiveness of the method.

1. Introduction

Hazardous gas leakage accident has brought huge damage to the society. For example, Bhopal accident caused thousands of deaths due to the methyl isocyanate gas leak accident (Varma and Guest, 1993). Consequently, it is of paramount importance to monitor industrial emission and use the monitoring data to estimate the release rate and location of emission source. To estimate the emission source, an atmospheric dispersion simulation (ADS) model and a parameter estimation algorithm with high accuracy and efficiency are necessary. The ADS model is used for predicting the concentration distribution, and the parameter estimation algorithm is used for finding the optimal source parameters to make ADS model output as close as possible to the actual measurement.

Many ADS modeling methods have been developed by researchers. Gaussian model is a typical and fast tool for atmospheric dispersion prediction, whose expression is quite simple. Usually, the Gaussian dispersion model is suitable for emergency management due to its high efficiency. However, its mechanism is too simple to give the accurate prediction, whose limitations are: it only supports low wind speed; it only supports straight-line trajectories; it assumes steady-state atmosphere; it has no memory of previous emissions. The Lagrangian model is very common in meteorological modeling tools based on random walk theory (Draxler and Rolph, 2012; Stein et al., 2015; Wilson and Sawford, 1996). It can simulate the atmospheric dispersion process in relatively complex meteorological conditions and global scale. This model is more suitable in large-scale scenarios, but the investigation area of hazardous gas leakage accident generally cannot reach that scale. Integrated model combines different dispersion model together, which is popular in commercial software for risk analysis such as PHAST (Connan et al., 2013; Hanna et al., 2008). However, the integrated models also need few minutes to calculate and the result is not always accurate. In complex environments, CFD model is currently the optimal option to obtain accurate prediction results (Hanna et al., 2009; Mazzoldi et al., 2008; Pontiggia et al., 2009). However, the CFD model needs long computation time, usually measured in hours or even days, which restricts the application of CFD in emergency management. Furthermore, a common problem of these methods is that some input parameters are quite difficult to measure and quantify. Therefore, researchers proposed the methods that can use pre-determined scenarios to train ANN for decision and bypass some troubling parameters (Krasnopolsky and Schiller, 2003; So et al., 2010). A previous study also used the integration of machine learning algorithms and traditional ADS models to predict the contaminant dispersion (Ma and Zhang,

* Corresponding author.
E-mail address: nudtcb9372@gmail.com (B. Chen).

https://doi.org/10.1016/j.atmosenv.2018.01.056
Received 3 September 2017; Received in revised form 26 January 2018; Accepted 29 January 2018
Available online 31 January 2018
1352-2310/

2016). The high accuracy of these studies represents that the ANN could be a useful tool for pollution forecasting and risk analysis.

In terms of source parameters estimation, parameters could be determined by estimating their posterior distribution or finding the maximum likelihood estimate. Thus, most source estimation methods are based on Bayesian inference or optimization algorithms (Hutchinson et al., 2017). Markov Chain Monte Carlo (MCMC) algorithm is usually used for posterior distribution estimation in source estimation problem (Borysiewicz et al., 2012; Keats et al., 2007; Tierney, 1994; Yee, 2007). Some filtering methods also apply the Bayesian theory to update the source parameters (Huber, 2014; Wawrzynczak et al., 2014; Zhang and Wang, 2013). Optimization algorithms are widely implemented to find the solution of minimum cost or maximum likelihood, whose theoretical basis is maximum likelihood estimation (MLE) principle (Qiu et al., 2016; Sharan et al., 2009). Intelligent optimization methods are usually used, such as particle swarm optimization (PSO) (Eberhart and Kennedy, 1995; Qiu et al., 2016), simulated annealing (Thomson et al., 2007) and genetic algorithm (Allen et al., 2007). In dispersion source estimation problem, the release rate and location of source should be estimated. When the source location is known, these methods could be quite effective because we only have to estimate one parameter (release rate) (Chai et al., 2015; Eslinger et al., 2014; Katata et al., 2015). However, if the source location is unknown, the problem becomes more complicated because the algorithm may be difficult to converge. Even if the algorithm can converge successfully, estimating all these parameters together is a quite time-consuming task due to the huge search space. Therefore, expectation maximization (EM) algorithm is introduced to address this problem (Do and Batzoglou, 2008). In the E-step, the expected value of source location is estimated using ANN and PSO, while in the M-step, the estimated release rate is updated on the basis of MLE.

In this paper, the proposed method is able to estimate the emission source using ANN-based dispersion prediction and PSO-EM-based parameter estimation. To verify the proposed method, SF_6 dispersion data from Indianapolis field study is applied to validate whether the method is feasible in practice.

2. Methods

2.1. Workflow

In order to predict the concentration distribution and estimate the dispersion source, the workflow of the proposed method includes several steps:

A. Obtaining a large number of release scenarios covering nearly all possibilities from gas trace experiment. If it is difficult to control the variables of field experiment, release scenarios can also be obtained from simulation experiment.
B. Extracting input and target dataset from release scenarios. To predict the concentration of the interest point, the input data should contain the information including source term, meteorological parameters and the location of interest point. The target data should be able to present the value or level of gas concentration of the interest point.
C. Training and testing of the ANN. The input and target dataset extracted from release scenarios in step B is used for ANN training and testing to construct an ANN-based ADS model.
D. Configuration of the source estimation parameters. Both temporal and spatial investigation regions are defined in this step. Furthermore, initial parameters of the source estimation algorithm should be determined before inverse calculation.
E. Estimating the source of atmospheric dispersion using PSO or the combination of PSO and EM.

Table 1
Common parameters for atmospheric dispersion model.

Parameters	Symbol	Unit
Downwind distance	D_x	m
Crosswind distance	D_y	m
Height of source	H	m
Height of interest point	z	m
Release rate	q	g s^{-1}
Atmospheric stability class	STA	/
Wind direction	d	deg
Wind speed	v	m s^{-1}
Mixing height	z_m	m
Cloud height	z_c	m
Cloud cover	p_c	%
Temperature	T	K

2.2. Structure of ANN

Generally, complicated ADS model such as CFD needs quite long time to compute the concentration distribution, while simple model can hardly give the accurate results. To address this problem, ANN is used to predict the concentration of the interest point with high efficiency and accuracy (Ma and Zhang, 2016).

To satisfy the emergency requirements, the input data of the ANN should be easily to obtain. A rough idea is using all the measured parameters shown in Table 1 as the input of the ANN. However, it is quite difficult to train the ANN if we directly put these parameters into input layer because features of the atmospheric dispersion should be extracted before training. Generally, the concentration of hazardous gas follows Gaussian distribution on crosswind direction. Moreover, the concentration of a specific point is approximately proportional to the release rate and inversely proportional to the wind speed. Due to these features, as shown in Fig. 1, we use release rate q, reciprocal of the wind speed $1/v$, and two Gaussian parameters (G_y and G_z) on y- and z-axes as ANN input. The expressions of G_y and G_z are shown in Eq. (1) according to the experience of Gaussian dispersion model.

$$\begin{cases} G_y = \exp\left(-\frac{D_y^2}{2\sigma_y^2}\right), \\ G_z = \exp\left[-\frac{(z+H)^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z-H)^2}{2\sigma_z^2}\right] \end{cases} \tag{1}$$

where D_y and z represents the crosswind distance and the height of the interest point respectively. H is the height of the emission point. σ_y and σ_z, which represent the deviation of the Gaussian distribution, are the Gaussian dispersion coefficients affected by downwind distance D_x and atmospheric stability. Gaussian parameters, wind parameters and source term parameters are inputs of traditional Gaussian dispersion model. They are very common and easy-to-measure parameters (also simple to calculate in simulation software). Moreover, Gaussian dispersion model has already been extensively used in source estimation methods. Thus, by using these parameters as the input of ANN, we can directly substitute ANN-based atmospheric dispersion model for Gaussian model in source estimation, without changing inputs.

The number of neurons in the hidden layer could be determined by evaluating some important criteria (e.g. coefficient of determination). The output layer has only one neuron, meaning the concentration of the interest point. The algorithm and detailed process of ANN training is not in our research scope, so the ANN will be directly trained by MATLAB neural network toolbox in this paper (MATLAB, 2010).

2.3. Quantifying release rate by PSO

The hazardous gas leak accidents can be classified into two categories. The category 1 is the accident that the source location is already known, while the category 2 is the accident that source location is