Engineering Applications of Artificial Intelligence 75 (2018) 102–113

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Source term estimation of hazardous material releases using hybrid genetic algorithm with composite cost functions

Yan Wang [a], Hong Huang [a,*], Lida Huang [a], Xiaole Zhang [a,b]

[a] Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing, China
[b] Institute for Nuclear and Energy Technologies, Karlsruhe Institute of Technology, Karlsruhe, D-76021, Germany

ARTICLE INFO

Keywords:
Source term estimation
Gaussian plume model
Hybrid genetic algorithm
Cost function
Nemenyi test

ABSTRACT

Source term estimation (STE) of atmospheric dispersion plays an important role in public safety, environmental protection and many other application fields. In this paper, several new composite cost functions for STE using hybrid genetic algorithm are proposed and compared using Nemenyi test based on 68 STE tasks from Prairie Grass field experiment. Results show one of the new composite cost functions, named as WSD, has outstanding performance in estimating both source location and emission rate. Then the patterns in STE results using different cost functions are analyzed based on the 68 tasks mentioned above, which provides further insights into what to expect from STE. At last, the relationship between composite cost functions and multi-objective optimization is analyzed to facilitate the understanding of composite cost functions. To summarize, composite cost functions such as WSD has the potential to achieve a better balance between sensitivity and robustness of cost functions applied in STE, providing the most accurate estimates. Statistical algorithm comparison techniques like Nemenyi test can help us better understand the characteristics and performance of specific settings in STE methods.

1. Introduction

Artificial Intelligence (AI) techniques like evolutionary algorithms and swarm intelligence have been applied in many engineering problems (Ahmadi et al., 2014a, b, 2015). In this paper, we apply a hybrid genetic algorithm to atmospheric source term estimation problems. The releases of hazardous materials can be due to industrial accidents or terrorist attacks (Gupta, 2002), both of which are big threats to public safety and health. Source term estimation (STE) is to estimate the location and strength of emission sources by available concentration and meteorological observations. Information about the emission source can help enhance the situation awareness of hazardous material releases (Zheng and Chen, 2011; Zhang et al., 2017; Zhang and Huang, 2017). Atmospheric STE has also been used to estimate fugitive releases from oil and gas production (Albertson et al., 2016), municipal solid waste landfills (Kormi et al., 2016) and urban natural gas pipelines (von Fischer et al., 2017) to better understand the economic and climatic impact of methane leakage. Therefore, STE is important for public safety, economic development as well as environmental protection.

In a recent review by Bieringer et al. (2017), STE methods were summarized into three general categories: forward modeling, inverse modeling, and nonlinear optimization. Forward modeling method determines the release rate by scaling the unit release from atmospheric transport and dispersion (AT&D) models using concentration observations (Flesch et al., 2005). Inverse modeling method reverses some of the physical processes in AT&D models to obtain source parameters (Bady et al., 2009). As the most popular one in the literature, nonlinear optimization based methods search for the best estimate of the source parameters based on multiple runs of a forward AT&D model, where model predictions are compared with sensor observations via a discrepancy measure (i.e. cost function or likelihood function). In this context, optimization method (Haupt et al., 2006; Thomson et al., 2007; Cervone and Franzese, 2010; Ma et al., 2013; Li and Zhang, 2017) and Bayesian inference method (Senocak et al., 2008; Keats et al., 2010; Wang et al., 2015, 2017; Ristic et al., 2017) have many commonalities and can be viewed as two subcategories of the general nonlinear optimization approach. More elaborate discussion about different kinds of STE methods along with their variations and applications is available in several review articles (Rao, 2007; Redwood, 2011; Singh et al., 2015; Hutchinson et al., 2017; Bieringer et al., 2017).

The optimal choice or theoretical justification of discrepancy measure in nonlinear optimization method is an important and challenging task. The difficulty lies in the complex patterns contained in the discrepancy between AT&D model predictions and sensor observations. These complex patterns arise from randomness and uncertainties in

* Correspondence to: Institute of Public Safety Research, Department of Engineering Physics, Tsinghua University, Beijing, 100084, China.
E-mail address: hhong@tsinghua.edu.cn (H. Huang).

https://doi.org/10.1016/j.engappai.2018.08.005
Received 19 June 2017; Received in revised form 23 March 2018; Accepted 6 August 2018

Nomenclature

n	Number of gas sensors
$\mathbf{z}$	Sensor observations, vector of length n: $[z_1, z_2, \dots, z_n]$, mg/m^3
$\bar{\mathbf{z}}$	Mean value of observation vector $\mathbf{z}$, mg/m^3
$\mathbf{b}$	Predictions by forward AT&D model at the location of n gas sensors: $\mathbf{b} = [b_1, b_2, \dots, b_n]$, mg/m^3
$\bar{\mathbf{b}}$	Mean value of prediction vector $\mathbf{b}$, mg/m^3
$\boldsymbol{\theta}$	Target parameter vector which we want to estimate, including source location and emission rate
$D(\mathbf{b}, \mathbf{z})$	Cost function describe the difference between observations $\mathbf{z}$ and predictions $\mathbf{b}$
(x, y, z)	Location of a point in 3D space, m
(x_0, y_0, z_0)	Source location, m
$C_{x,y,z}(\boldsymbol{\theta})$	Concentration at point (x, y, z) predicted by forward AT&D model using parameter $\boldsymbol{\theta}$, g/m^3
Q_0	Source emission rate, g/s
U	Wind speed, m/s
ϕ	Mean wind direction, degree
σ_y	Horizontal plume dispersion coefficient, m
σ_z	Vertical plume dispersion coefficient, m

turbulent dispersion processes, imperfect AT&D models or parameters and noises in sensor measurements. Ideally, the discrepancy measure should be both sensitive to target parameters of STE tasks and relatively robust to the discrepancy caused by sensor noises, outliers and model imperfectness.

In Wang et al. (2017), the impact of likelihood functions in Bayesian STE method was investigated empirically. Approximate Bayesian Computation (ABC) method was also investigated, where the discrepancy measure is distance measure instead of likelihood function. Results show that every likelihood function or distance measure has pros and cons on different source parameters (e.g. lognormal likelihood function is superior in estimating source emission rate but inferior in estimating source location compared with normal likelihood function). In this context, an integration of STE results using different likelihood functions and distance measures was encouraged to provide more robust estimates. On the other hand, the importance and impact of cost functions in optimization method has long been observed and investigated. Haupt et al. (2006) studied the sensitivity of STE using GA-coupled receptor-dispersion model to six cost functions based on synthetic data, finding that the best cost function is problem dependent while the variability due to the choice of cost functions is relatively small. Thomson et al. (2007) presented STE results using three cost functions with different regularization terms, where the impact of additive Gaussian noises and concentration offsets are investigated to test each cost function, showing that the best results are achieved by using a multiplicatively regularized cost function which minimizes total emission rates. Cervone and Franzese (2010) evaluated eight cost functions, where the properties of each cost function are analyzed in detail. Ma et al. (2013) pointed out that in optimization method, satisfactory STE results can be obtained with a good selection of forward AT&D models and cost functions.

The key contribution of this work is the idea of composite cost functions, especially those loose ones, which is general and useful in optimization problems with many noisy observations. Although there was an attempt to use composite cost functions (i.e. a combination of two or more cost functions) in the literature (Haupt et al., 2006), the idea of balancing the sensitivity and robustness of cost functions using loose composite method has never been considered. Also, there is no systematic investigation of composite cost functions based on multiple STE tasks before.

In our work, several new composite cost functions are constituted through the combination of two or more cost functions. To provide a more robust evaluation of different cost functions, their performances are compared systematically using all the 68 trials from the Prairie Grass field dispersion experiment (Barad, 1958). Nemenyi test (Nemenyi, 1963) is used to integrate the STE results for different trials in the Prairie Grass experiment and produce a comprehensive comparison in a statistical manner. Then the patterns in the STE results using those cost

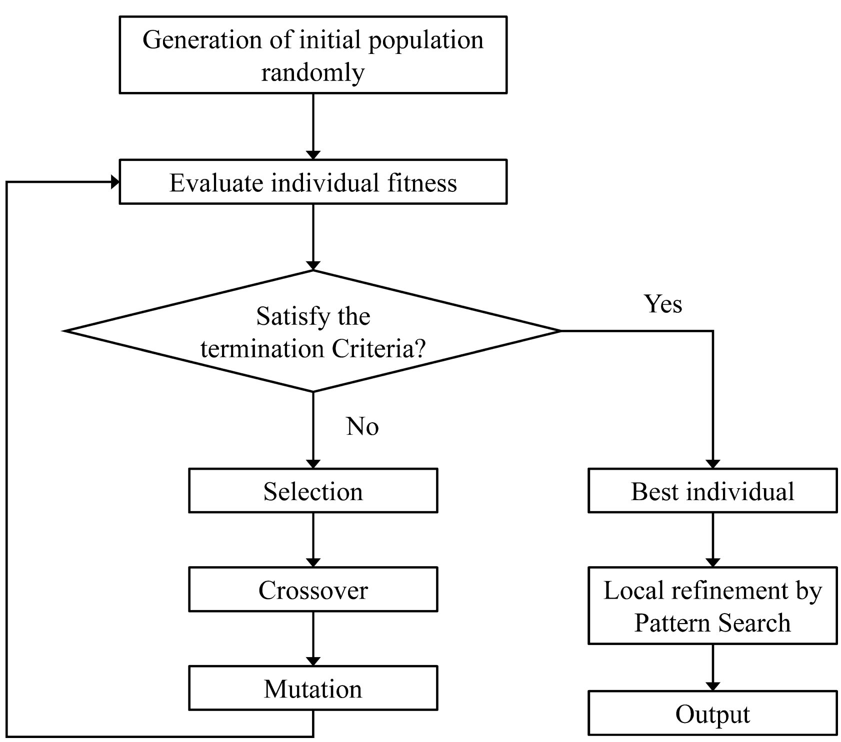

Fig. 1. The flow diagram of GA–PS method.