Environmental Pollution 241 (2018) 759–774

Contents lists available at ScienceDirect

Environmental Pollution

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

New approach for point pollution source identification in rivers based on the backward probability method☆

Jiabiao Wang [a], Jianshi Zhao [a,*], Xiaohui Lei [b], Hao Wang [a,b]

[a] State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, 100084, China
[b] China Institute of Water Resources and Hydropower Research, Beijing, 100038, China

A R T I C L E I N F O

Article history:
Received 4 January 2018
Received in revised form
22 May 2018
Accepted 30 May 2018

Keywords:
Backward probability method
Linear regression
Multi-point pollution source identification
Parameter decoupling
River

A B S T R A C T

Pollution risk from the discharge of industrial waste or accidental spills during transportation poses a considerable threat to the security of rivers. The ability to quickly identify the pollution source is extremely important to enable emergency disposal of pollutants. This study proposes a new approach for point source identification of sudden water pollution in rivers, which aims to determine where (source location), when (release time) and how much pollutant (released mass) was introduced into the river. Based on the backward probability method (BPM) and the linear regression model (LR), the proposed LR–BPM converts the ill-posed problem of source identification into an optimization model, which is solved using a Differential Evolution Algorithm (DEA). The decoupled parameters of released mass are not dependent on prior information, which improves the identification efficiency. A hypothetical case study with a different number of pollution sources was conducted to test the proposed approach, and the largest relative errors for identified location, release time, and released mass in all tests were not greater than 10%. Uncertainty in the LR–BPM is mainly due to a problem with model equifinality, but averaging the results of repeated tests greatly reduces errors. Furthermore, increasing the gauging sections further improves identification results. A real-world case study examines the applicability of the LR–BPM in practice, where it is demonstrated to be more accurate and time-saving than two existing approaches, Bayesian–MCMC and basic DEA.

1. Introduction

River pollution from industrial waste discharge or accidental spills during transportation is a major threat to both the environment and human health (Azizullah et al., 2011). However, the number of sudden water pollution incidents is increasing (Shao et al., 2006). To protect river systems from sudden pollution damage, it is necessary to dispose of the pollutant quickly and expediently, and it is thus crucial to develop accurate and efficient tools to identify pollution sources, determining where (source location), when (released time), and how much pollutant (released mass) was introduced into the river. This type of reliable tools for point source identification would play an important role in river pollution emergency response and disposal (Shi et al., 2017). And tracking the cause of pollution quickly is also helpful to investigate the responsibilities.

However, point pollution source identification is a challenging task because of the uncertainties and nonlinearity in the transport process of pollutants. This type of identification aimed at reconstructing the pollutant transport process is an inversed and ill-posed problem (Boano et al., 2005; Sun et al., 2006). The accuracy and efficiency of this problem in rivers mainly relies on three aspects: (1) obtaining prior information about the pollution source, which can be partly achieved using monitoring systems (Ghane et al., 2016); (2) gaining complex information about the pollution incidents regarding flow simulation dimensions, the number of point sources involved, and the pollutant release process; and (3) using a method that simulates flow and pollutant transport process in a river system (Wei et al., 2016).

Several approaches for point source identification of sudden water pollution incidents have been published in literature (Ghane et al., 2016; Xu and Gomez-Hernandez, 2016; Yang et al., 2016), and

☆ This paper has been recommended for acceptance by Sarah Harmon.
* Corresponding author.
E-mail address: zhaojianshi@tsinghua.edu.cn (J. Zhao).

https://doi.org/10.1016/j.envpol.2018.05.093
0269-7491/

they can be generally classified into two types: deterministic and statistical. Early studies mainly focused on groundwater pollution using optimization approaches, such as the least squares (LS) approach (Gorelick et al., 1983; Alapati and Kabala, 2000), Tikhonov regularization (TR) approach (Skaggs and Kabala, 1994; Akçelik et al., 2003), and minimum relative entropy approach (Woodbury and Ulrych, 1998). Hydraulic and water quality simulation models are combined in optimization processes (Guan et al., 2006). Deterministic approaches are aimed at providing an optimal solution to this inversed problem; however, the solution is difficult to understand and is sensitive to observation noise. In contrast, statistical approaches were developed based on probability theory and have the better capacity to deal with noisy and incomplete prior information (Hazart et al., 2014). One of the most popular statistical approaches is the backward probability method (BPM), which was firstly derived by Liu (1995) as a backward-in-time model. This method was developed based on the use of equations governing the backward process, which are adjoint to the pollutant transport process (Neupauer and Wilson, 1999; Michalak and Kitanidis, 2004). Another commonly used stochastic approach is the Bayesian and Markov Chain Monte Carlo method (Bayesian–MCMC). Based on Bayesian inference theory and the MCMC sampling technique, this method converts point source identification into a reiterative computation of the posterior probability distribution of the pollution source parameters. A random sampling process of Gaussian or uniform distribution is needed in the Bayesian–MCMC (Hazart et al., 2014). In addition, several heuristic algorithms, such as genetic algorithms (GAs; Singh and Datta, 2006; Zhang and Xin, 2017), artificial neural network (ANN; Srivastava and Singh, 2014; Singh et al., 2004), and the differential evolution algorithm (DEA; Yang et al., 2016) have been applied within the solving process to accelerate computation.

According to previous research, the BPM is preferable for use in solving problems that are time-consuming and where poor prior information is available. When applied to a river pollution case (Ghane et al., 2016) and a lake pollution case (Cheng and Jia, 2010), the BPM showed good identification accuracy and efficiency. However, almost all previous applications using BPM have been limited to single-point source pollution incidents. When the number of pollution sources increases, problems of complexity, uncertainty, and information insufficiency will be more severe and even completely ill-pose the source identification (Amirabdollahian and Datta, 2013). Therefore, in multi-point cases, the solutions of the identification problem may be nonunique (Boano et al., 2005; Hamdi, 2016). Furthermore, the relationship between the composite pollutant concentration and corresponding location probability of more than one pollution source is currently unclear.

In practice, it is hard to determine whether pollution incidents are caused by single- or multi-point sources prior to sources being found. Therefore, an approach that is not limited by the number of point sources is more valuable and promising for the source identification of real-world pollution incidents. To address this issue, a Linear Regression–BPM (LR–BPM) approach is proposed in this paper. Based on the BPM, this new approach converts the source identification problem into an optimization model using linear regression. As this approach is independent of the number of point sources, it can thus be applied to pollution incidents with multi-point (or unknown) sources.

The remainder of this paper is organized as follows. A brief introduction to the BPM is firstly provided, and the LR–BPM is then derived and established. The DEA is subsequently applied in the operational process of the LR–BPM. Three hypothetical cases are used to validate and test the LR–BPM, and its uncertainty is analyzed. A real-world case study is then conducted to examine its applicability in practice. Finally, conclusions are drawn and presented.

2. Methodology

2.1. Introduction of BPM

In a sudden water pollution incident occurring in a river, the pollutant is instantaneously released from a point source. And it mixes and diffuses to a transverse uniform state in the river flow direction after a certain distance away from the source (Abderrezzak et al., 2015; Zhang and Xin, 2017). The laws of pollutant migration and diffusion in rivers can be expressed by a one-dimensional equation:

$$\begin{cases} \dfrac{\partial(AC)}{\partial t} + \dfrac{\partial(QC)}{\partial x} = \dfrac{\partial}{\partial x}\left(EA\dfrac{\partial C}{\partial x}\right) - k_s AC \\ C(x, t_0) = 0, \ \ x_0 < x < l \\ C(x_0, t) = 0, \ \ t > 0 \\ C(x_0, t_0) = c_0 \end{cases}, \tag{1}$$

where x_0 is the pollution source location (m); x is the location from the source (m); t_0 is the release time of pollutant (s); t is time since the pollutant released (s); C is the mean pollutant concentration (mg/L); Q is the flow discharge (m^3/s); A is the water area perpendicular to the river flow direction; E is the dispersion coefficient in the flow direction, which is generally evaluated by empirical equations or based on experiments (m^2/s); k_s is the decay coefficient (s^{-1}); c_0 is the initial mean pollutant concentration determined by released mass (mg/L); and l is the river length (m). The hydrodynamic variables Q and A in Equation (1) can be determined by a hydrodynamics model based on the Saint-Venant equation as follows:

$$\begin{cases} \dfrac{\partial A}{\partial t} + \dfrac{\partial Q}{\partial x} = 0 \\ \dfrac{\partial Q}{\partial t} + \dfrac{\partial}{\partial x}\left(\dfrac{Q^2}{A}\right) + gA\dfrac{\partial Z}{\partial x} + gAS_f = 0 \end{cases}, \tag{2}$$

where Z is the water level (m), and S_f is the energy slope.

When pollution sources are known, the pollutant transport process can be simulated by solving Equation (1). However, in most real-world cases, information about the pollutant concentration can be obtained at some downstream cross-sections, but the actual pollution sources remains unknown. Therefore, source identification is required.

According to the BPM, the river pollutant concentration and possible location of the corresponding pollution source can be described using probability density function. The pollutant transport along the river is defined as the forward process, while the determination of the pollution source location probability is defined as the backward process (Cheng and Jia, 2010). Neupauer and Wilson (2001) demonstrated the connection between these two processes and derived an equation for the backward source location probability, as follows:

$$\begin{cases} -\dfrac{\partial(AP)}{\partial t'} + \dfrac{\partial(QP)}{\partial x_s} + \dfrac{\partial}{\partial x_s}\left(EA\dfrac{\partial P}{\partial x_s}\right) - k_s AP = 0 \\ P(x_s, t_d) = 0, \ \ x_s \in (x_0, x_d) \\ P(x_d, t') = 0, \ \ t' > 0 \\ P(x_d, t_d) = 1 \end{cases}, \tag{3}$$

where x_s is the possible location of the unknown pollution source