Advances in Water Resources 88 (2016) 68–79

Contents lists available at ScienceDirect

Advances in Water Resources

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

Estimation of river pollution source using the space-time radial basis collocation method

Zi Li[a,b], Xian-Zhong Mao[a,*], Tak Sing Li[c], Shiyan Zhang[d]

[a] Division of Ocean Science and Technology, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China
[b] School of Civil Engineering, University of Queensland, Brisbane QLD 4072, Australia
[c] Open University of Hong Kong, Hong Kong, China
[d] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 19 May 2015
Revised 24 November 2015
Accepted 24 November 2015
Available online 17 December 2015

Keywords:
Contaminant source identification
Global space-time collocation model
K-fold cross-validation
Inverse problem
Radial basis function

ABSTRACT

River contaminant source identification problems can be formulated as an inverse model to estimate the missing source release history from the observed contaminant plume. In this study, the identification of pollution sources in rivers, where strong advection is dominant, is solved by the global space-time radial basis collocation method (RBCM). To search for the optimal shape parameter and scaling factor which strongly determine the accuracy of the RBCM method, a new cost function based on the residual errors of not only the observed data but also the specified governing equation, the initial and boundary conditions, was constructed for the *k-fold* cross-validation technique. The performance of three global radial basis functions, Hardy's multiquadric, inverse multiquadric and Gaussian, were also compared in the test cases. The numerical results illustrate that the new cost function is a good indicator to search for near-optimal solutions. Application to a real polluted river shows that the source release history is reasonably recovered, demonstrating that the RBCM with the *k-fold* cross-validation is a powerful tool for source identification problems in advection-dominated rivers.

1. Introduction

Rivers and streams are quite vulnerable to pollution because they are naturally open, easily accessible and substantially used in agricultural, industrial and municipal processes. The release of hazardous contaminants into rivers, for example, sewage treatment plant emissions, accidental spills or intentional dumping of toxic materials, poses great threats to the residents and ecosystems exposed to the river water. According to the China Statistical Yearbook [16], a total of 9339 water pollution accidents took place in China from 1997 to 2008, some of which triggered serious consequences. For example, an explosion occurred in a benzene plant maintained by a Petrochemical Company in Jilin Province, China in November 2005, resulting in approximately 100 tons of benzene being spilled into Songhuajiang River with the fire-fighting water causing severe ecological problems downstream, and millions of residents living downstream suffered from a shortage of water supply for up to four days. This incident demonstrates that it is essential to develop reliable techniques to identify the missing contaminant source information in water pollution accidents, and in turn to enable proper emergency response, post-event remediation and liability apportionment. Such problems are generally referred to as river contaminant source identification, and can be formulated as inverse models to estimate missing source release history from the detected contaminant plume at monitoring sites.

Source identification problems are inherently ill-posed and difficult to solve as a result of the irreversibility of contaminant fate and diffusive nature of transport processes, which smooths the contaminant plume gradually and causes the loss in measurements of information concerning the source release. These ill-posed problems are characterized by high sensitivity to the perturbation, number and location of measurement data, and are expected to be solved by stable and efficient inverse models, such as the classic optimization methods.

In the past two decades, many innovative inverse models, including geostatistical and probabilistic modeling and deterministic direct methods, have been proposed for contaminant source identification, especially in groundwater owing to its substantial significance in engineering practices. An extensive review of many sound methods for recovery of the historical contaminant distribution, source location and release history has been presented by Atmadja and Baqtzoglou [3], Michalak and Kitanidis [32] and Bagtzoglou and Atmadja

* Corresponding author. Tel.: +86 75526036710.
E-mail address: maoxz@sz.tsinghua.edu.cn (X.-Z. Mao).

http://dx.doi.org/10.1016/j.advwatres.2015.11.019
0309-1708/

[9]. Commonly employed techniques include the Tikhonov regularization method [38], the nonlinear optimization model [2], particle-based methods [1,7], the backward beam equation method [3,4,8] and geostatistical based methods [6,32]. Some other newly emerged techniques have also been under intensive investigation, such as artificial neural networks [37], a global search algorithm with a constrained least squares estimator [40], geostatistical kriging [36] and iterative regularization [23].

As well as extensive studies to enable groundwater pollutant source identification, a few investigations have been also conducted for river pollution which tends to be more advective than in groundwater. Because of strong advection in rivers, pollutants are transported faster and farther due to the dispersion effect, leading to difficulties in detecting the contaminant plume for the recovery of source release information and to computational instability in numerical solutions. Badia et al. [5] established identifiability and stability results for identification of pollution point sources. Boano et al. [12] applied geostatistical method to estimate a spatially distributed source and independent point sources taking into account the influence of dead zones on transport processes. Cheng and Jia [14] established a computational method for identification of the locations and release time of point sources based on the backward location probability density function method. Other studies made attempts to control contaminant releases into rivers and estuaries [26,33] and in a channel flow [25], and to identify contaminant sources in atmospheric advection-dominated transport process [10,11].

The radial basis collocation method (RBCM) has recently emerged as a new inverse model. RBCM employs radial basis functions (RBFs) to approximate multivariate functions based on scattered data [21] and numerical solution of partial differential equations (PDEs) using methods by Kansa [24] and Hon et al. [22]. The RBCM is a domain-type numerical method involving the same single-step solution procedures as in direct problems and can be efficiently applied to the inverse problem. Cheng and Cabral [13] and Mao and Li [31] employed the inverse multiquadric (MQ) as a RBF to solve ill-posed inverse boundary problems for the Laplace equation. As for the time-dependent inverse problems, the global space-time RBCM was then constructed for inverse and backward heat conduction problems [28] and for estimations of the spatial plume distribution and source release history in groundwater [29]. Owing to the complexity of the source identification problem, sensitivity analysis associated with the number, location and error level of observations and model parameters was also performed by Li and Mao [28,29] The advantages of the space-time RBCM mainly lie in its efficiency due to the truly mesh-free and non-iterative property, super-convergence and insensitivity to the direction of time. However, no studies have applied the RBCM to solve contaminant source identification problems in advection-dominated rivers to date.

There is still a pending problem in the conventional RBCM: convergence of the numerical solution is sensitive to the free shape parameter determining the shape of the RBFs. As the shape parameter increases (the RBFs tend to be flatter), the error associated with the numerical solutions is reduced exponentially because of more collocation points effectively involved in the interpolation. However, if the shape parameter exceeds a critical value, the coefficient matrix becomes highly ill-conditioned resulting in the error increasing sharply. The optimal value of the shape parameter is strongly problem-dependent [18,34]. Several empirical formulations for selection of the shape parameter have been suggested in other investigations [18,20,21,24,34,35,41]. For problems of interpolating scattered data, Foley [18] improved the RBF shape parameter based on the root mean square deviation between the MQ and inverse MQ models at a set of test points, and Rippa [34] proposed instead a cross-validation technique for computing the RBF shape parameter by minimizing the cost function based on the residual errors of only the observation data; for solving PDE systems, Roque and Ferreira [35] built the cost function based only on the residual errors of the governing equations, but Tsaia et al. [41] applied the golden search algorithm to choose a good shape parameter based on the residual errors of both the governing equation and Dirichlet boundary conditions. Li and Mao [28,29] investigated convergence of the global space-time MQ solution with respect to the shape parameter and the scaling factor. The results showed that a reasonable (although not necessarily the best) numerical solution exists within a wide range of both parameters, with a V-shaped or U-shaped error profile with respect to the shape parameter and scaling parameter. Accordingly, it is still necessary to search for the best performance of the RBCM in solving the river contaminant source identification problem based on selection of optimal model parameters.

The present study was conducted to extend the global space-time RBCM [29], used for groundwater, to identify space- and time- dependent source releases causing river pollution. To search for the optimal values of the model parameters, the *k-fold* cross-validation technique based on a new cost function was developed. Also, the idea of the global space-time RBF was extended to Hardy's multiquadric (HMQ), inverse multiquadric (IMQ) and Gaussian (GS). This paper is organized as follows. Sections 2 and 3 describe the governing equations for source identification problems in rivers and the numerical procedures employed by the RBCM for the inverse model, respectively. Section 4 introduces the new cost function designed for choosing the model parameters. Section 5 investigates the sensitivity of the cost function to the parameters coupled with the numerical results of two hypothetical cases, and Section 6 shows application to a real polluted river. Finally, conclusions are given in Section 7.

2. Governing equations

Let us consider a river region Ω surrounded by an entire boundary $\delta\Omega$, in which the pollutant concentration $C(x, t)$ (M•L^{-3}) at time t (T) and position x (L) is governed by a two-dimensional depth-averaged advection-dispersion equation with a linear reaction term. We assume (or know) that the source release $s(x, t)$ is spatially distributed on the boundary Σ and therefore formulated as a part of the boundary conditions, and the initial concentration is specified to be constant. Also, the flow field can be derived from a two-dimensional depth integrated free surface flow model.

The mathematical model for predicting the fate and transport of the contaminant plume can then be described as:

$$\begin{cases} \mathbf{L}[C(\mathbf{x},t)] = \dfrac{\partial C(\mathbf{x},t)}{\partial t} - \nabla\cdot[D\nabla C(\mathbf{x},t) - \mathbf{u}C(\mathbf{x},t)] \\ \quad + KC(\mathbf{x},t) = 0,\ (\mathbf{x},t)\in \mathbf{\Omega}\times(t_0,t_{\max}) \\ C(\mathbf{x},t_0) = 0, \quad \mathbf{x}\in\Omega \\ \mathbf{F}[C(\mathbf{x},t)] = \mu\dfrac{\partial C(\mathbf{x},t)}{\partial n} \\ \quad + \beta C(\mathbf{x},t) = s(\mathbf{x},t),\ (\mathbf{x},t)\in\Sigma\times(t_0,t_{\max}), \\ \mathbf{G}[C(\mathbf{x},t)] = \eta\dfrac{\partial C(\mathbf{x},t)}{\partial n} \\ \quad + \gamma C(\mathbf{x},t) = 0, \quad (\mathbf{x},t)\in\partial\Omega\backslash\Sigma\times(t_0,t_{\max}) \end{cases} \tag{1}$$

where, L, F and G are the differential operators; μ, β, η and γ are the coefficients determining the type of boundary conditions (i.e., Dirichlet, Neumann or Robin); D (L^2•T^{-1}), $\mathbf{u}$ (L•T^{-1}) and K (T^{-1}) are the dispersion coefficient, flow velocity, and decay factor, respectively; t_0 and $t_{\max}$ are the initial and end time of the contaminant transport process, respectively.

If the source release $s(x, t)$ is prescribed and the other boundary conditions and initial conditions are properly defined, Eq. (1) leads to a well-posed contaminant transport problem. The concentration $C(x, t)$ can be determined within the domain at any time and location using conventional procedures such as a finite difference method or finite element method.