Development of an Efficient and Accurate Global Space-Time Radial Basis Collocation Model for Estimation of River Pollution Source

T. S. Li and S. M. Wong

Abstract—**Pollution source estimation is an active research area nowadays. Researchers are trying to estimate the pollution source based on a set of data from monitoring sites. This kind of problems is known to be ill-posted and sensitive to computational errors. In this paper, we present a method using a space-time radial basis function to estimate the pollution source. To overcome the problem of ill-conditioning, we first use the data from the monitoring site to get a continuous approximation of the data. Then, we use this data together with the governing equation, and boundary conditions to set up an over-determined system of equations. The single value decomposition is used to get the least square solution. An example is used to illustrate the method. It is found that our method is effective and accurate in estimating the pollution source.**

Index Terms—**Pollution source estimation, radial basis functions, inversion problem**

I. INTRODUCTION

Rivers and estuaries occasionally suffer from different kinds of pollution. In order to investigate the problems, monitoring sites are set up to record the level of pollution at different time. However, it would be better if we can use this limited set of data to estimate the actual profile of the pollution source.

This kind of problem belongs to a class called inversion problems. In such a problem, we need to estimate the boundary condition from a set of data in the interior of a studied region. This kind of problems is known to be ill-posted and therefore suffers from huge errors due to limited computer accuracy.

The estimation of pollution source is an active research area. Different methods have been proposed and applied to similar problem. These methods can be classified into two categories: statistical methods or numerical approximation methods. Wei *et al.* [1] used a statistical method to estimate the pollution source. Michalak *et al.* [2] used a method called the adjoint state method together with a geo-statistical approach to address forensic groundwater problems. Pan *et el* [3] attempted to improve the estimation using a special kind of statistical test called the one-end U test.

For numerical methods, Alapati *et al.* [4] used a non-linear least-squares (NLS) method to recover the release history of a

groundwater contaminant plume from its current measured spatial distribution. However, the paper only addressed one-dimensional problems. Zhang *et al* [5] used an inversion method to estimate the contaminant sources in air.

In this study, we propose a numerical scheme based on radial basic functions to get a reasonable estimation of the pollution source.

II. RIVER POLLUTION SOURCE

A. The direction problem

In the direct problem, all the initial and boundary conditions are known. We just need to use them together with the governing equation to find the values of the pollutant.

Let Ω be a river region under the study with boundary $\delta\Omega$, and $C(\mathbf{x}, t)$ be the pollutant concentration. A pollution source $s(\mathbf{x}, t)$ is distributed along Σ which is part of $\delta\Omega$.

The governing equation for the problem is given in (1).

$$\frac{\partial C(\mathbf{x},t)}{\partial t} - \nabla \cdot \left(D\nabla C(\mathbf{x},t) - \mathbf{u}C(\mathbf{x},t) \right) + KC(\mathbf{x},t) = 0 \qquad (1)$$

For $(\mathbf{x}, t) \in \Omega \times (0, t_{max}]$

The corresponding boundary conditions are given in (2) and (3).

$$\alpha \frac{\partial C(\mathbf{x},t)}{\partial n} + \beta C(\mathbf{x},t) = s(\mathbf{x},t) \text{ , for } (\mathbf{x},t) \in \Sigma \times (0, t_{max}] \quad (2)$$

$$\gamma \frac{\partial C(\mathbf{x},t)}{\partial n} + \varepsilon C(\mathbf{x},t) = 0, \text{ for } (\mathbf{x},t) \in \delta\Omega \backslash \Sigma \times (0, t_{max}] \quad (3)$$

where α, β, γ and ε in equations (2) and (3) are for different types of boundary conditions; $\mathbf{u}$, D, K are the water velocities, dispersion coefficient and decay parameter respectively. t_{max} is the end time of the study period.

The initial condition is given in (4).

$$C(\mathbf{x}, 0) = 0, \text{ for all } \mathbf{x} \in \Omega. \qquad (4)$$

This is the direct problem and can be solved easily using different numerical schemes like the finite element method or the finite different method. In this paper, we propose an approximation scheme devised from the radial basis collocation method.

B. Radial Basis Function Method

Radial basis functions (RBFs) approximation was firstly devised for the scattered geographical data interpolation by Hardy [6] in the early 1970's. He introduced a class of RBFs

called multiquadric function (MQ). A review of several types of RBFs methods for scattered data interpolation had been presented by Franke [7] in 1982. His comparison showed that MQ has the best performance in terms of accuracy among the methods tested. Micchelli [8] later gave a proof on the solvability of the RBFs interpolation. In the early 1990's Madych *et al* [9] and Wu *et al* [10] had shown that the RBFs interpolant consists of super-convergent property.

Kansa [11][12] refined the Hardy's MQ method for solving elliptic, parabolic and hyperbolic PDEs problems over irregular regions. A number of experiments of the method had been carried out and applied to solve various scientific and engineering problems.

The original RBFs approximations were derived from the idea of the interpolation using a function $\Phi: \mathbb{R} \to \mathbb{R}$. The MQ introduced by Hardy has the form as described in (5).

$$\Phi(r) = \sqrt{r^2 + \delta^2} \tag{5}$$

where $\delta^2 \in \mathbb{R}$ is called the shape parameter. This shape parameter is used to control the fitting of a smooth surface to the data. One obvious property of the MQ function is that it is infinitely differentiable. So if a function is approximated by an MQ function, it is also infinitely differentiable. This property is a desirable characteristic for solving high order PDEs.

C. Using MQ to Approximate a 2D Function

Assume that we approximate a 2D function $f: \mathbb{R}^2 \to \mathbb{R}$ so that it has the following values at n nodes:

$$f(\mathbf{x}_i) = a_i, \qquad i=1,2,\cdots,n. \tag{6}$$

This function $f(\mathbf{x})$ can be approximated by collocation using (7).

$$f(\mathbf{x}) \cong \widetilde{f(\mathbf{x})} = \sum_{i=1}^{n} b_i \Phi(\|\mathbf{x} - \mathbf{x}_i\|), \tag{7}$$

where $\|\mathbf{x} - \mathbf{x}_i\|$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_i$, and b_i's are constant to be determined by the system of equations defined in (8).

$$\sum_{i=1}^{n} b_i \Phi(\|\mathbf{x}_j - \mathbf{x}_i\|) = a_j \quad j=1,2,\cdots,n. \tag{8}$$

This system of equations forms n equations with n unknowns. After solving for b_i's, we can determine the approximated value $\widetilde{f(\mathbf{x})}$ at any $\mathbf{x}$.

D. Using MQ to solve PDE Systems

Equation (8) is used to approximate a function given a number of values on it. In order to solve a PDE, we need to replace $f(\mathbf{x})$ by $\widetilde{f(\mathbf{x})}$. For example, consider the boundary value problem PDE as defined in 9).

$$\begin{cases} \boldsymbol{L}[f(\mathbf{x})] = 0, & \text{for } \mathbf{x} \in \Omega, \\ \boldsymbol{B}[f(\mathbf{x})] = 0, & \text{for } \mathbf{x} \in \delta\Omega, \end{cases} \tag{9}$$

where $\boldsymbol{L}$ and $\boldsymbol{B}$ are the differential operators. $\boldsymbol{L}$ is for the governing equation and $\boldsymbol{B}$ is for the boundary condition. In order to solve the PDE, we replace $f(\mathbf{x})$ by $\widetilde{f(\mathbf{x})}$ as defined in (10).

$$\begin{cases} \boldsymbol{L}[\widetilde{f(\mathbf{x}_i)}] = 0, & \text{for } \mathbf{x}_i \in \Omega, \\ \boldsymbol{B}[\widetilde{f(\mathbf{x}_i)}] = 0, & \text{for } \mathbf{x}_i \in \delta\Omega. \end{cases} \tag{10}$$

Given n nodal points, some in Ω and some in $\delta\Omega$, then a system of n equations with n unknowns can be formed. Thus we can solve for b_i's.

The radial basis functions have been successfully applied in solving different boundary value problems. However, we need a space-time approximation if we extend it to solve initial and boundary value PDE problems.

E. Space-time RBFs

In order to solve initial boundary PDE problems, we need an approximation for a function $f: \mathbb{R}^2 \times T \to \mathbb{R}$. Myers *et el* [13] had discussed different kinds of space-time RBFs. He also showed that the product of two positive definite functions on lower dimensional spaces is positive definite on the product space. He also extended it into a number of other space-time RBFs which include a product-sum, integrated product, and the integrated product-sum.

In this paper, we employ a simple product form of space-time RBF. Let $\Phi: \mathbb{R} \to \mathbb{R}$ be the defined RBF, we can approximate the function $f: \mathbb{R}^2 \times T \to \mathbb{R}$ by equation (11).

$$\begin{aligned} f(\mathbf{x}, t) &\cong \widetilde{f(\mathbf{x}, t)} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} \Phi(\|\mathbf{x} - \mathbf{x}_i\|) \Phi(\|t - t_j\|) \end{aligned} \tag{11}$$

where $(\mathbf{x}_i, t_j) \in \Omega \times [0, t_{max}]$. This equation (11) is used to approximate the governing equations (1)-(4), then we have the following set of equations (12)-(15) respectively.

$$\sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} \begin{pmatrix} \Phi(\|\mathbf{x} - \mathbf{x}_i\|) \frac{\partial \Phi(\|t - t_j\|)}{\partial t} \\ -\nabla \cdot \begin{pmatrix} D\nabla\Phi(\|\mathbf{x} - \mathbf{x}_i\|)\Phi(\|t - t_j\|) \\ -\mathbf{u}\Phi(\|\mathbf{x} - \mathbf{x}_i\|)\Phi(\|t - t_j\|) \end{pmatrix} \\ +K\Phi(\|\mathbf{x} - \mathbf{x}_i\|)\Phi(\|t - t_j\|) \end{pmatrix} = 0, \tag{12}$$

for $(\mathbf{x}, t) \in \Omega \times (0, t_{max}]$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} \begin{pmatrix} \alpha\Phi(\|t - t_0\|) \frac{\partial \Phi(\|\mathbf{x} - \mathbf{x}_i\|)}{\partial n} \\ +\beta\Phi(\|\mathbf{x} - \mathbf{x}_i\|)\Phi(\|t - t_0\|) \end{pmatrix} = s(\mathbf{x}, t), \tag{13}$$

for $(\mathbf{x}, t) \in \Sigma \times (0, t_{max}]$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} \begin{pmatrix} \gamma\Phi(\|t - t_0\|) \frac{\partial \Phi(\|\mathbf{x} - \mathbf{x}_i\|)}{\partial n} \\ +\varepsilon\Phi(\|\mathbf{x} - \mathbf{x}_i\|)\Phi(\|t - t_0\|) \end{pmatrix} = 0, \tag{14}$$

for $(\mathbf{x}, t) \in \delta\Omega \backslash \Sigma \times (0, t_{max}]$

The given initial condition can also be approximated by RBF as defined in (15).

$$\sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} \Phi(\|\mathbf{x} - \mathbf{x}_i\|)\Phi(\|t - t_0\|) = 0, \qquad \mathbf{x} \in \Omega. \tag{15}$$

Given n nodes in Ω and m nodes in $[0, t_{max}]$, then we would have $n \times m$ unknowns b_{ij}.

III. AN APPLICATION

A simple example as depicted in Fig. 1 is used to illustrate how the space-time RBF can be applied to calculate the

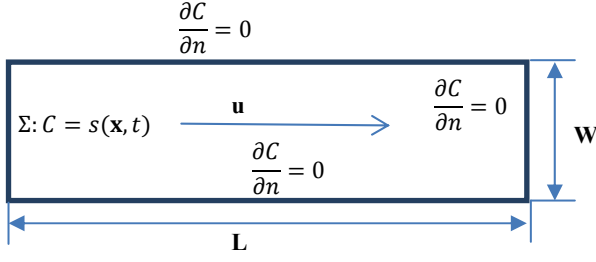pollutant in a given river region.

Fig. 1. The studied region of a river.

Fig. 1 shows the studied region in the example. The upper and lower edges are the banks of the river. The left edge has a pollution source $s(\mathbf{x}, t)$ given by equation (16).

$$s(\mathbf{x}, t) = 10\exp\left[-20\left(\frac{y - W/2}{W}\right)^2\right] \times \frac{t}{T_o}\exp\left(-\frac{t^2}{2T_o^2}\right). \quad (16)$$

The right edge is the mouth of the river. We distribute 11×21 spatial nodes in the studied region and 21 nodes in the time domain. Altogether we have 11×21×21 unknowns. There is a set of 9×19×20 equations for the governing equation at nodes in $\Omega\backslash\delta\Omega\times(0, t_{max}]$; a set of 11×21 equations for initial condition at nodes $\Omega\times\{0\}$; a set of (21+21+9)×20 equations for the boundary conditions at the two banks and the mouth; and a set of 9×20 equations for the boundary condition at the left edge.

The computation involves 4581 equations and 4581 unknowns. Fig. 2 illustrates the variation of pollutant concentrations obtained at three different times (a) $\frac{1}{4}t_{max}$, (b) $\frac{1}{2}t_{max}$ and (c) $\frac{3}{4}t_{max}$.

(a) $\frac{1}{4}t_{max}$

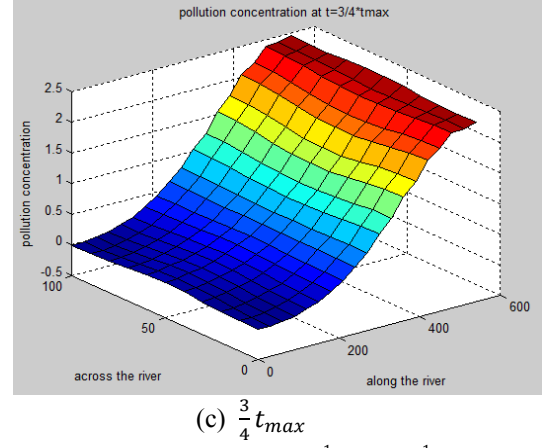(b) $\frac{1}{2}t_{max}$

(c) $\frac{3}{4}t_{max}$

Fig. 2. Pollutant concentrations at (a) $t = \frac{1}{4}t_{max}$, (b) $\frac{1}{2}t_{max}$ and (c) $\frac{3}{4}t_{max}$

The value of pollution source plotted against time and y is depicted in Fig. 3.
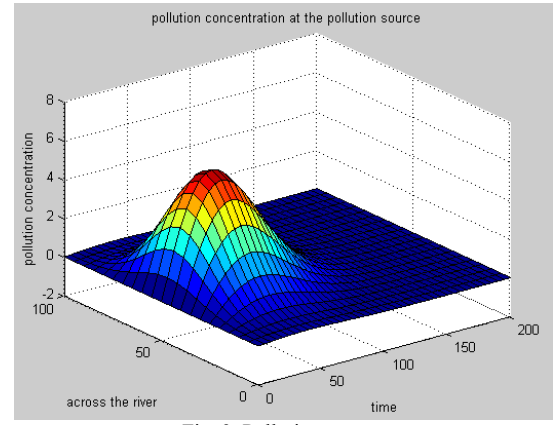

Fig. 3. Pollution source

The pollution source discharged into a water body will give serious impacts to water quality. If we have sufficient data of the influx pollutants, we can easily identify the problems and set up appropriate control systems to reduce the infuse rate.

In order to protect our water environments, the source of pollutants are important input data. This study employs the method of inversion problem to approximate the value of pollution source level. The involved algorithm will be discussed in the next section.
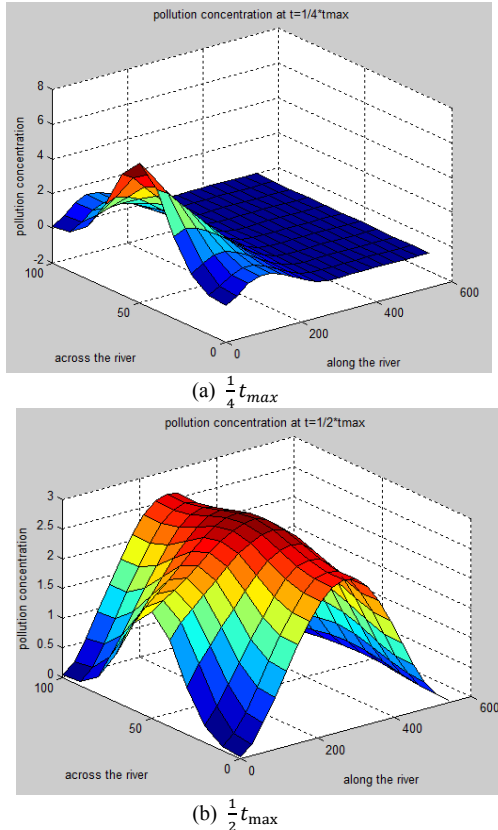
IV. THE INVERSION PROBLEM

In real-life, it is usually difficult to collect sufficient information of the pollution source accurately when such an incidence occurs. This is because we can only rely on a limited number of monitoring sites in the water region. So we may miss an important piece of information in the direct problem. The equations involved in the inversion problem are defined in equations (17) to (20).

$$\frac{\partial C(\mathbf{x}, t)}{\partial t} - \nabla \cdot \left(D\nabla C(\mathbf{x}, t) - \mathbf{u}C(\mathbf{x}, t)\right) + KC(\mathbf{x}, t) = 0, \quad (17)$$

for $(\mathbf{x}, t) \in \Omega \times (0, t_{max}]$

$$\alpha\frac{\partial C(\mathbf{x}, t)}{\partial n} + \beta C(\mathbf{x}, t) = s(\mathbf{x}, t), \text{ for } (\mathbf{x}, t) \in \Sigma \times (0, t_{max}]. \quad (18)$$

$$C(\mathbf{x}, 0) = 0, \quad \text{for } \mathbf{x} \in \Omega. \tag{19}$$

and

$$C(\mathbf{xx}_i, tt_j) = S_i(tt_j), \quad i=1, \dots, N_{sites}, j = 1, \dots, M_{mon}. \tag{20}$$

The only difference between equations (1)-(4) and equation (17)-(20) is that we now only have the value of C at points $\mathbf{xx}_i$ and at time tt_j, $j = 1, \dots, M_{mon}$. The original problem is needed to be solved, so that we would have an underdetermined set of equations. This is because we usually do not have enough data from the monitoring sites to replace the data from the source in equation (1).

In order to solve this problem, we use the data obtained from the monitoring site to interpolate the pollutant concentration around the region of the monitoring sites by equation (21).

$$S(\mathbf{x}, t) = \sum_{i=1}^{N_{sites}} \sum_{j=1}^{M_{mon}} c_{ij} \Phi(\|\mathbf{x} - \mathbf{xx}_i\|) \, \Phi(\|t - tt_j\|) \tag{21}$$

The coefficients $\{c_{ij}\}$ can be determined by using the collected data in the monitoring sites. By using this approximation, equation (20) changes to (22) for the region around the monitoring sites.

$$C(\mathbf{x}, t) = S(\mathbf{x}, t), \tag{22}$$

We would usually set more equations than the number of unknowns and then use the least square method to find the best fit solution.

A. Example with Monitoring Sites

In order to test our method, we use the solution from (12)-(15) to obtain the data of pollutant concentration from five monitoring sites as depicted in Fig. 4 at ten different times. However, to mimic real life measurement errors, we add 2% noise to the data before feeding them to our inversion problem.
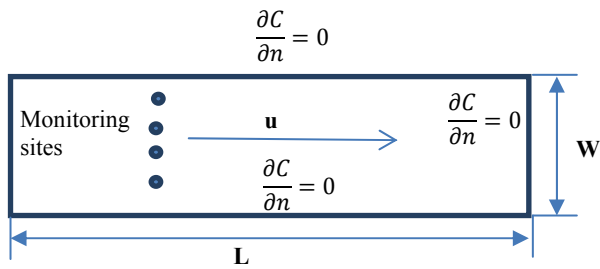

Fig. 4. The river region with monitoring sites.

The set of 50 space-time nodes is used to approximate the value of C around the monitoring sites. We can then use this approximation to set up the last equation in (21). Therefore, we would have a set of over determined equations and the solution can be found using the least square method. However, because of the ill-conditioning of the problem, we need to use the singular value decomposition to find a more accurate solution.

Fig. 5 shows the variation of pollutant concentrations obtained at times (a) $\frac{1}{4} t_{max}$, (b) $\frac{1}{2} t_{max}$ and (c) $\frac{3}{4} t_{max}$.
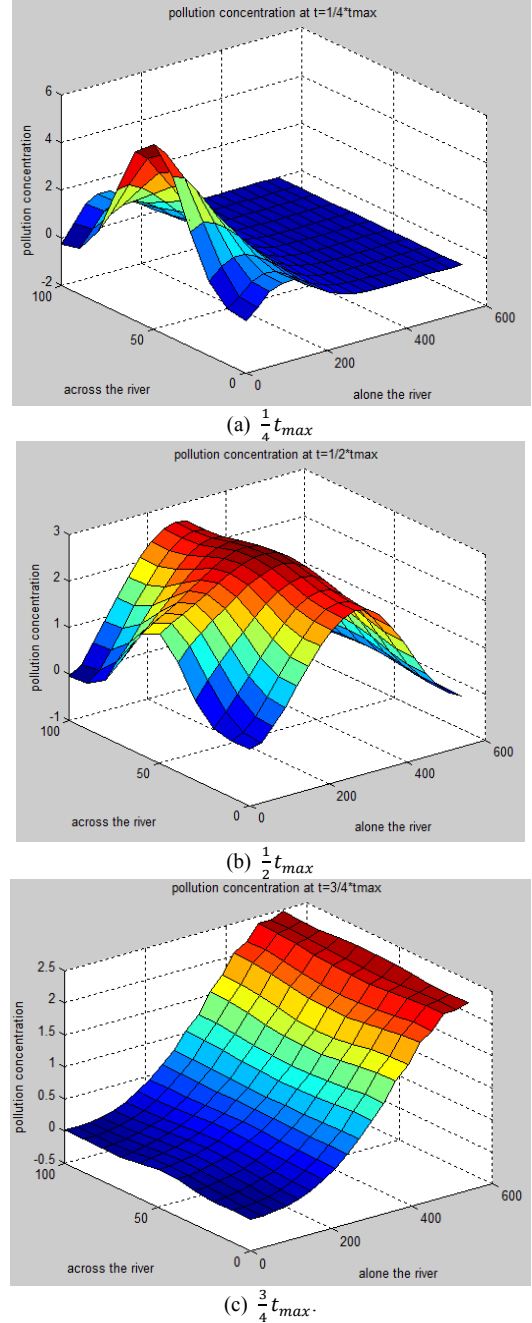

(a) $\frac{1}{4} t_{max}$

(b) $\frac{1}{2} t_{max}$

(c) $\frac{3}{4} t_{max}$.

Fig. 5. The polluting concentration at time=$\frac{1}{4} t_{max}$, $\frac{1}{2} t_{max}$ and $\frac{3}{4} t_{max}$.

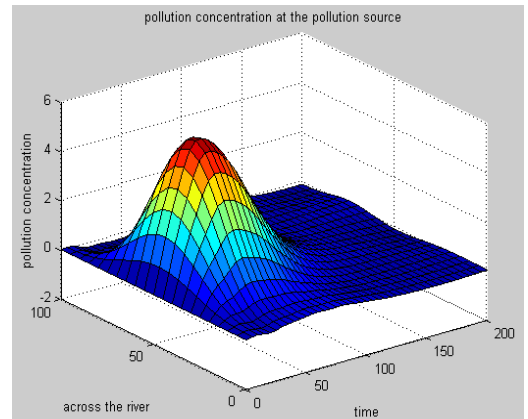The value of pollution source plotted against times and y is shown in Fig. 6.

Fig. 6. The estimated -pollution source plotted against y and time.

Compare the Fig. 3 and Fig. 6, we can see that our proposed method gives a reasonable estimation of the actual pollution source. The maximum error between the actual pollution source and the estimated one is 0.231 which is about 4.77%. Given the fact that we only have a very limited set of data to work on, this result is already remarkable agreeable.

V. Conclusions

We have successfully developed a robust method to estimate the pollution source in a river region with a small number of pollution value data obtained from some monitoring sites at some specific times.
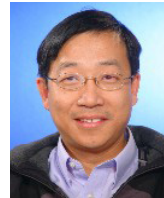
Acknowledgment

References

[1] W. P. Cheng and Y. F. Jia, "Identification of contaminant point source in surface waters based on backward location probability density function method," *Advances in Water Resources*, vol. 33, no. 4, pp. 397-410, 2010.
[2] A. M. Michalak and P. K. Kitanidis, "Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling," *Water Resources Research*, vol 40, no.8, 2004.
[3] L. P. Pan, Y. Zhang, and H. Q. Qu, "Accidental contaminant detection method based on one-tailed U test," in *Proc. International Conference on Bioinformatics and Biomedical Engineering*, 2011, pp. 1-4.
[4] S. Alapati and Z. J. Kabala, "Recovering the release history of a groundwater contaminant using a non-linear least-squares method," *Hydrological Processes*, vol. 14, no. 6, pp. 1003-1016, 2000.
[5] T. F. Zhang, Q. Chen, "Identification of contaminant sources in environments by inverse CFD modeling," *Indoor Air*, vol. 17, no. 3, pp. 167-177, 2007.
[6] R. L. Hardy," Multiquadric equations of topography and other irregular surfaces," *J. Geo-phys, Res*. 17G, pp. 1905-1915, 1971.
[7] C. Franke, "Scattered data interpolation: test of some methods," *Math. Comput.*, vol. 38, no. 157, pp. 181-200, 1982.
[8] C.A. Micchelli, "Interpolation of scattering data: distance matrices and conditionally positive definite functions," *Constr. Approx.* 2, pp. 11-22, 1986.
[9] W. R. Madych and S. A. Nelson, "Multivariate interpolation and conditionally positive definite functions II," *Math. Comp.*, vol. 54, pp. 211-230, 1990.
[10] Z. Wu and R. Schaback, "Local error estimate for radial basis function interpolation of scattered data," *IMA J. Num. Anal.*, Vol. 13, pp. 13-27, 1993.
[11] E. J. Kansa, "Multiquadrics-a scattered data approximation scheme with applications to computational fluid dynamics-I," *Comput. Math, Applic.,* vol. 19, no. 8-9, pp. 127-145, 1990.
[12] E. J. Kansa, "A strictly conservative spatial approximation scheme for the governing engineering and physics equations over irregular regions and homogeneously scattered nodes," *Comput. Math. Applic.*, vol. 24, no. 5-6, pp. 169-190, 1992.
[13] D. E. Myers, S. De Iaco, D. Posa, and L. De Cesare, "Space-time radial basis functions," *Computers and Mathematics with Applications*, vol. 43, pp. 539-549, 2002.

Tak Sing Li received a BSc (Hons) degree in Mechanical Engineering from the University of Hong Kong. He then received an MSc degree in Computer Science in 1991 and a PhD in Computer Science in 1994. Both degrees were from the Queen's University of Belfast. He joined the Open University of Hong Kong in 1995 as a lecturer in the computing team. He is now an associate professor in the same university.

Sze Mui Wong received a BSc (Hons) degree in Mathematical Studies in 1988, and received an MSc degree in Operational Research in 1989 at the University of London. She completed the PhD in Computational Modeling in 2001 in the City University of Hong Kong. She joined the Open University of Hong Kong in 1993 as a lecturer. She is now an associate professor in the same university.