



Application of a link simulation optimization model utilizing quantification of hydrogeologic uncertainty to characterize unknown groundwater contaminant sources

Mahsa Amirabdollahian^{1,2} · Bithin Datta^{1,3} · Peter H. Beck⁴

Received: 6 June 2018 / Accepted: 29 September 2018 / Published online: 4 October 2018
© Springer Nature Switzerland AG 2018

Abstract

In existing groundwater contamination source characterization methodologies, simulation models estimate the contamination concentration in the study area. In order to obtain reliable solutions, it is essential to provide the simulation models with reliable hydrogeological properties. In real-life scenarios often high level of uncertainty and variability is associated with the hydrogeological properties. This study focuses on quantifying the hydrogeological parameter uncertainty to enhance the accuracy of identifying contamination release histories. Tracer experiment results at the Eastlakes Experimental Site, located in Botany Sands Aquifer, in New South Wales, Australia, are utilized to examine the performance and potential applicability of the methodology. In the selected study area, the hydrogeological heterogeneity in the microscopic scale, specifically the hydraulic conductivity, has substantial effect on the transport of pollutants. Among available tracer information, Bromide is studied as a conservative contaminant. Using possible realizations of the flow field, a coefficient of confidence (COC) is calculated for each field monitoring locations and times. Higher COC implies that the result of simulation models at that specific monitoring location and time is more reliable than other contaminant concentration data. Therefore, the optimization model should emphasise matching the corresponding estimated and observed contamination concentrations to accurately identify the contaminant release locations and histories. The linked simulation–optimisation method is utilised to optimally characterise the Bromide sources. Performance evaluation results demonstrate that the proposed methodology recovers pollution source characteristics more accurately compared to the methodology which does not consider the effect of hydrogeological parameter uncertainty.

Keywords Hydrogeology · Parameter uncertainty · Source characterization · Groundwater · Tracer test · Optimization

Introduction

In polluted groundwater systems, the contaminant release by past activities needs to be identified and characterized. The characterization process includes finding the source locations among potential sources and retrieving their pollutant release histories. This approach allows “Potential Responsible Parties” (PRVs) to use the identified contaminant source properties along with flow and transport simulation modules to (1) convince regulators that the existing contamination does not exceed the regulation thresholds and no remediation is required; or (2) develop effective remedial plans that satisfy the cost constraints and has appropriate reliability and risk consideration. Therefore, the accuracy in characterization of contamination sources is of great interest to PRVs and regulators. However, groundwater systems are subjected to various sources of uncertainty, and acquiring

✉ Mahsa Amirabdollahian
Marcia.amir@gmail.com

Bithin Datta
Bithin.datta@jcu.edu.au

Peter H. Beck
Peter.Beck@ghd.com

¹ College of Science and Engineering, James Cook University, Townsville, QLD 4811, Australia

² NSW Department of Industry, Locked Bag 5123, Parramatta, NSW 2124, Australia

³ CRC for Contamination Assessment and Remediation of Environment (CRC-CARE), University of New Castle, Callaghan, NSW 2308, Australia

⁴ GHD Pty. Ltd., Melbourne, VIC 3000, Australia

data in polluted aquifers is a costly and time consuming procedure. The main focus of this study is on quantifying the hydrogeological uncertainty and its impact on the accuracy of the groundwater simulation models. The aim is to improve the accuracy of the estimation of the contaminant release histories in existing polluted aquifers.

The uncertainty in contaminated groundwater systems is classified into three groups, (1) model, (2) parameter uncertainties, and (3) inherent uncertainty associated with the measurement data. The model uncertainty stems from incomplete knowledge or inability in accurately modelling the natural process using mathematical tools. However, the parameter uncertainty stems from incomplete knowledge of parameter values, sparsity of parameter measurements, and their spatial and temporal variability. Finally, the measurement uncertainty relates to the errors associated with the information collected from the field (Beven 2006). Wu and Zeng (2013) and Guillaume et al. (2016) presented overviews of the various sources of uncertainty in groundwater numerical simulation processes.

The model uncertainty is classified as an aleatory uncertainty and is the variation inherited by the physical system or the environment under consideration, and is studied by the stochastic approach (Freeze et al. 1992; Tiedeman and Gorelick 1993). The model uncertainty is presented as random imprecision and commonly defined by a probability distribution. However, parameter and measurement uncertainties are not random, and cannot be objectively quantified. These types of uncertainties are called fuzziness or epistemic uncertainty due to imprecise data, and subjectivity of judgment. Non-random uncertainty or fuzziness can be reduced with the acquisition of additional information (Oberkampf et al. 2004; Ross 2005).

Amirabdollahian and Datta (2013) published an overview on techniques to recover characteristics of pollutant sources using spatial and temporal contaminant concentration measurements. The linked simulation–optimization is one of the effective techniques to identify the contamination sources by linkage between flow and contaminant transport simulation modules and the optimization technique (Mahar and Datta 2001; Aral et al. 2001; Datta et al. 2009). Utilizing the evolutionary optimization algorithms has several advantages which include relative ease in solving linked simulation–optimization models. Genetic algorithm (GA) (Singh and Datta 2006), Tabu search (TS) (Yeh et al. 2007), adaptive simulated annealing (ASA) (Amirabdollahian and Datta 2014) are some of the evolutionary algorithms. The linked simulation–optimization method eliminates the mathematical limitations associated with large scale contaminated aquifer study areas, and makes it possible to solve the source contamination problems for such large scale study areas. The evolutionary optimization algorithm generates potential source characteristics. Using these potential characteristics, the flow and transport simulation modules

calculate contaminant concentrations at monitoring bores at different times. The optimization algorithm finds the contaminant source characteristics by minimizing the variation between estimated and observed temporal and spatial contaminant concentrations collected at monitoring locations. The simulation–optimization method is computationally intensive due to need to run the complex groundwater simulation model multiple times. Prakash and Datta (2015), Esfahani and Datta (2016) and Hazrati-Yadkooi and Datta (2017) used surrogate models to reduce the computational time. Using machine learning methodologies such as support vector machine and multivariate adaptive regression splines, can render simulation models less computational intensive (Kisi and Parmar 2016; Kisi et al. 2017).

The hydrogeological properties of a polluted aquifer have significant influence on the temporal and spatial properties of contamination plume. Therefore, this study focuses on hydrogeological parameter value uncertainty in polluted aquifers. In most of the previous contamination source identification methods (Mahar and Datta 2001), the effect of hydrogeological parameter uncertainty has been incorporated in the performance evaluation data. However, these sources of uncertainty were not considered and quantified explicitly in the methodology.

The main approach in quantifying parameter uncertainty in scientific models, for many years has been the probabilistic approach. Bayesian updating (Freeze et al. 1992), regression analysis (Tiedeman and Gorelick 1993), sequential Gaussian simulation (Mugunthan and Shoemaker 2004), and Monte-Carlo simulation (Dokou and Pinder 2009, 2011) are among these methodologies. O'Hagan and Oakley (2004) argued that although probability theory is robust and defensible for applications to aleatory uncertainty, it is not adequate for characterization of fuzziness. Considering the nature of probabilities, the elicitation of fuzziness using probability distribution function (PDF) tool introduces more uncertainty to the system. Therefore, in this study a new methodology is presented to quantify the hydrogeological parameter uncertainty in the contamination source identification procedure by using an averaging method among possible distributions of field parameter values. The collected hydrogeological field data are utilized and interpolated to find possible variation in the field properties. Then based on the estimated confidence in the spatial distribution of field parameter values, the worth of the modelled contaminant concentration values is estimated. The ASA based linked simulation–optimisation technique is utilized to find contaminant source characteristics. The optimization algorithm minimizes the variation between estimated and observed concentrations. However, all the estimated values do not have equal contribution and the more reliable estimated concentration data has more weightage in recovering contamination source histories.

The methodology performance evaluation is carried out using data from an experimental site set up in Botany Sands

Aquifer (BSA) located in New South Wales (NSW), Australia (Beck 2000). Bromide was utilized as the conservative tracer for measuring the resulting spatial and temporal concentrations.

Methodology

The identification of pollutant sources comprise of defining: (1) source locations; (2) source release history (time), and (3) source contaminant release flux. The proposed methodology for characterisation of unknown groundwater contamination sources using uncertain hydrogeological parameter values has two major components: Linked simulation–optimization contamination source identification model, and uncertainty analysis module. These components are described below.

Linked simulation–optimization contamination source identification model

The optimal decision model for source identification is defined by an objective function, and a set of constraints. An effective and efficient optimization model requires careful definition of the objective function and the constraints, and selection of an appropriate optimization algorithm. Therefore, first the objective functions and the constraints are defined. Then, the selected optimization algorithm and the linked simulation–optimization method are discussed.

Contamination source identification formulation

A methodology for the recovery of unknown contamination source release histories generally involves examining a set of candidate source characteristics. The final aim is to find the set of contaminant source characteristics which results in the best fitting pollutant plume and observed contaminant concentrations. This process is formulated as an optimization model, in which the objective function is defined by Eq. (1), and the constraints are defined by Eqs. (2) and (3):

$$F_1 = \text{Minimize} \sum_{k=1}^{nk} \sum_{iob=1}^{nob} \mu_{iob}^k \times \frac{(Cest_{iob}^k - Cobs_{iob}^k)^2}{(Cobs_{iob}^k + \eta)^2}. \quad (1)$$

Subject to

$$Cest_{iob}^k = f(D, HC, \theta, x_i, y_i, z_i, q_i^t) \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (2)$$

$$q_{\min} \leq q_i^t \leq q_{\max} \quad i = 1, \dots, N; \quad t = 1, \dots, T \quad (3)$$

where variables are defines as below: nk : total number of observation time periods, nob : total number of available monitoring locations, N : total number of candidate source locations, T : total number of pollutant injection stress periods.

$Cest_{iob}^k$ and $Cobs_{iob}^k$: simulated and observed contaminant concentration at monitoring location iob and at the end of time period k .

D dispersion coefficient, HC : hydraulic conductivity, θ : porosity, x_i, y_i, z_i : Cartesian coordinates of candidate source i , q_i^t : release concentration for candidate location i at stress period t .

q_{\max}, q_{\min} : upper and lower bounds for contaminant release. μ_{iob}^k : coefficient of confidence (COC) assigned to the monitoring location iob at the end of time period k .

η : a constant which is selected based on the available observed concentrations. It should be large enough, to prevent any indeterminate term in Eq. (1) due to the value of zero/ close to zero observed concentration. The selected η value should not be larger than any observed concentration. Otherwise, as a result of large denominator value, the effect of observed concentrations which are smaller than η could substantially be reduced in the optimization process.

The first constraint (Eq. 2) calculates the contaminant concentrations using flow and transport simulation modules. In the current study, for the groundwater flow simulation purpose, MODFLOW (Zheng et al. 2001) is utilized. MODFLOW is a simulation module that solves the numerical three-dimensional transient groundwater flow equation. The contaminant transport process is simulated using MT3DMS (Zheng and Wang 1999). MT3DMS is a simulation module that solves the numerical three-dimensional transient multi-spacious contaminant transport equation in ground water systems. Readers are referred to Amirabdollahian and Datta (2013) for more information on the governing equations describing the flow and transport processes. Equation (3) demonstrates the second constraint which limits the source release concentrations to upper and lower bounds. If the locations of sources are unknown, the source release concentration lower band should be set to 0. Therefore, the non-actual or dummy and inactive sources are identified by zero release fluxes for all or a number of stress periods.

Linked simulation–optimization algorithm (SOA)

The SOA is solved using the ASA. ASA is a variant of simulated annealing (SA) which is less sensitive to the user defined parameters. In ASA the reannealing temperature schedule and random step selection are automatic and adaptive. ASA eliminates the need to adjust the parameters manually (Ingber 1996).

In the current study, the flow and transport simulation modules are externally connected to the optimization algorithm, forming linked SOA (Datta et al. 2009). First, based on the available site information, potential contamination source locations are selected. Second, the optimization algorithm generates the candidate contaminant concentrations associated with each potential source location and each stress period. Third, the simulation models estimate

the contaminant concentration ($Cest_{iob}^k$) at monitoring bores (iob) and time steps (k) (Moo-Young et al. 2004). Then, for each set of candidate source characteristics, the estimated contaminant concentration values are transferred back to the optimization model to calculate the corresponding objective function. Finally, this process evolves to reach the optimal contaminant release histories for the potential source locations (Amirabdollahian and Datta 2013).

Hydrogeological parameter uncertainty analysis

Generally, a predesigned sampling network is utilized to collect field hydrogeological data for groundwater management purposes. Spatial variations in HC values are a critical factor controlling the contaminant mass transport (Datta et al. 2009). Therefore, this study focuses on the HC fuzziness in contaminated groundwater aquifers.

The spatial fuzziness (variability) in measured HC values is required to be quantified, in order to accurately simulate mass transport process. The confidence in the simulated spatial and temporal contaminant concentration values can be evaluated considering the variation and availability of the field HC values.

In order to properly quantify this fuzziness, the complex relationship between the fuzziness in transport predictions, and the variation and availability of HC measurements should be identified. In this process the followings items need to be addressed.

1. Number and spatial distribution of available field measurements.
2. The HC field variability and spatial correlation between available measurements.
3. The flow field properties including boundary and initial conditions and properties of available sinks and sources.
4. The location, initial time, activity duration, and release concentrations of contaminant sources.

The flow and transport simulation modules are the appropriate tools for estimating the relation between fuzziness in estimated contaminant concentration values and the above mentioned four items. In contamination source identification procedure, the characteristics of contaminant sources are unknown (item 4 in the above list). Therefore, a coupled procedure is required to simultaneously quantify the fuzziness in contaminant concentration prediction, and search for optimal source release histories.

Hydraulic conductivity realisations

The HC values are generally known at limited number of locations. An interpolation algorithm is required to calculate the spatial distribution of parameter values for the entire

area. The geostatistical interpolation algorithms are computationally demanding, and also require statistical properties of the HC field. For example, Kriging (Goovaerts 1997) is a geostatistical interpolation method which needs a properly selected sample variogram and log-transformation of the data. This study focuses on situations where the available HC values are uncertain, therefore accurate estimation of statistical properties of data is not possible. Therefore, in this study, the Inverse Distance Weighting (IDW) interpolation method is utilized. IDW is a deterministic method and performs multivariate interpolation for a known scattered set of points.

Using IDW, the value of HC at un-sampled location x_0 , $HC^*(x_0)$, is estimated using parameter values at surrounding locations, $HC(x_i)$, using Eq. (4):

$$HC^*(x_0) = \sum_{i=1}^n w_i HC(x_i), \quad (4)$$

where w_i are the weights for each $HC(x_i)$ value and n is the number of the closest sampled data points used for the interpolation purpose. The w_i is estimated using Eq. (5):

$$w_i = \frac{1/d_i^2}{\sum_{i=1}^n 1/d_i^2}, \quad (5)$$

where d_i is the distance between the estimated and the sample point. Using IDW method, the quantity and spatial variability of available sample data are considered in the interpolation method without any prejudgment about the statistical relation between sampled data. The value of n is selected considering the spatial correlation among available scattered set of points, or the decision maker judgment.

In the case of uncertainty, making decision on appropriate n value is difficult. Multiple realizations of HC field can be estimated and utilized from a given set of sampling locations using different n values.

Coefficient of confidence

In crisp (certain) contamination source identification model, the model parameters are known without any uncertainty (Sun 1994). Therefore all the estimated concentrations ($Cest_{iob}^k$ in Eq. 1) are assumed to be accurate and can have the same contribution to the estimation of the contamination source characteristics. Therefore, in the crisp source identification model μ_{iob}^k is always one. However, uncertain hydrogeological input values result in less accurate concentration estimations. Therefore, when the hydrogeological parameter uncertainty exists, μ_{iob}^k is the COC. It represents

the degree of confidence assigned to the simulated contaminant concentration.

For the purpose of COC estimation, multiple realizations of HC field are generated. The generation of field realizations involves utilizing the IDW interpolation with different number of closest neighbouring points (n). Then for each realisation (r), the HC field is utilised to estimate possible contaminant concentrations ($Cest_{iob}^{k,r}$) using flow and transport simulation models. The averaged estimated contaminant concentration is calculated as below:

$$Cest_{iob,averaged}^k = \sum_{r=1}^R Cest_{iob}^{k,r} / R, \tag{6}$$

where $Cest_{iob,averaged}^k$ and $Cest_{iob}^{k,r}$ are the averaged estimated concentration and the estimated contaminant concentration using the r 'th realisation for monitoring location iob and time k , respectively. The total number of realisations is called R . The fuzziness in simulated contaminant concentration $Cest_{iob}^k$ is estimated by Eq. (7):

$$F_{iob}^k = \left| \frac{Cest_{iob}^k - Cest_{iob,averaged}^k}{Cest_{iob}^k + nn} \right|, \tag{7}$$

where $Cest_{iob}^k$ is the simulated concentration at monitoring location iob and time period k . This is the same value utilized in the objective function (Eq. 1) as the estimated concentration. nn is a specified very small number which prevents division by zero in Eq. (7). F_{iob}^k is the coefficient of fuzziness for monitoring location iob and time k . The larger the difference between HC realizations, the larger is the coefficient of fuzziness.

When the coefficient of fuzziness is larger, there is less confidence in the estimated concentration for that specific monitoring data, as the result of HC uncertainty. Therefore, less weightage should be assigned to that specific monitoring information in the objective function. This weightage is defined as COC by (8):

$$\mu_{iob}^k = \begin{cases} 0.1 & \text{if } (1 - U \times F_{iob}^k \leq 0.1) \\ 1 - U \times F_{iob}^k & \text{else} \end{cases}, \tag{8}$$

where μ_{iob}^k is the COC for estimated concentration at monitoring location iob and time t . U is a relaxation factor. When U is very small, the estimated COC values are always one or very near to one. Therefore, the objective function (Eq. 1) would be the same as the crisp situation and the effect of parameter uncertainty would not be properly considered. On the other hand, when U is large, the optimization algorithm would try to find the source release concentrations which result in the larger coefficient of fuzziness values and smallest COC (more emphasis on fuzziness). In other words the priority of the optimization algorithm would not be the

matching of the estimated and observed concentrations. Thus the suitable value for U should be selected based on the range of estimated coefficients of fuzziness values and it is site specific. The process of selecting appropriate U is discussed in details in “Discussion”. Using U enables this methodology to be applicable to the sites with various levels of groundwater contamination.

If only one equation is used for the μ_{iob}^k estimation, the COC would be very small or negative when the fuzziness is very large, or the predicted concentration is very small. Since, it is useful to use all the available measurement data for the purpose of characterization of sources, in Eq. (8) the minimum possible COC is considered to be 0.1.

Performance evaluation

Performance evaluation is carried out using data from a contaminated experimental site set up in the Botany Sands Aquifer located to the south of Sydney CBD, New South Wales (NSW), Australia (Beck 2000).

Although the BSA is considered to be homogenous on a regional scale, on the microscopic scale the aquifer can be highly heterogenous. In contaminated sites, detailed HC and contaminant distribution information are vital to investigate solute movement (Fu et al. 2015). Heterogeneity becomes more important as the scale of the system reduces and more attention must be given to the microscale hydrogeological studies.

Various hydrological, hydrogeological and hydrochemical data were collected during an experimental study in a artificially contaminated aquifer site. The purpose of this experiment was to study the effect of heterogeneity on the solute transport (Jankowski and Beck 2000). The details of the study area and the tracer test experiment are described below. However, few inconsistencies in the recorded measurement data were modified (in this study) before the raw data were used for this evaluation purpose.

Study area

The northern section of the BSA is located 3–5 km south of the Sydney’s CBD, Australia (Fig. 1). The BSA was the earliest groundwater resources used to supply water for Sydney and has been utilized since the nineteenth century. In the early twentieth century commercial and industrial developments on the northern shore of the Botany Bay commenced. These developments which include oil storage and refinery facilities, airport and variety of chemical, industrial and commercial manufacturing and storage facilities, resulted in contamination of the groundwater from these industrial land use activities. As the result of present and past land uses and lack of effluent management, treatment, and disposal statutory controls (in early years after World War II), a long



Fig. 1 Location of Botany Sands Aquifer and ELE Site (Jankowski and Beck 2000)

history of contamination exists in this area. Since detection of groundwater contamination various investigation, management, remediation and groundwater consumption restrictions have been utilized to control the negative effects of contamination in this area.

The BSA is mainly formed of unconsolidated sands, clays and peaty sediments. The sediments thickness are from zero at the northern rim of the basin, in Centennial and Moore parks (Fig. 1), to approximately 80 m to the southern regions. HC values vary from 1.8 to 50 m/day. Hydraulic properties variations are related to lithological units including quartz sand, silty/peaty sand, and sandy/peaty clay.

Experimental site

Heterogeneity in site physical parameters, especially HC, dominates the solute and contaminant transport processes. Therefore, numerical simulation, tracer tests and laboratory experiments are required to study the effect of the HC heterogeneity in microscopic scale. A tracer test was completed at the Eastlakes Experimental Site (EES), located next to the Lachlan Ponds in the northern part of the BSA, in Daceyville, NSW, Australia by Beck (2000). The site was first established in 1992 with a three-dimensional piezometers network located on a 7 m × 11 m grid. The hydrogeological and chemical heterogeneity was measured at 815 sampling locations (Evans 1993) with 1 m horizontal and 150 mm–200 mm vertical spacing (Fig. 2).

The geological investigations of the ELE site were carried out by two continuous sand cores. Using the geological

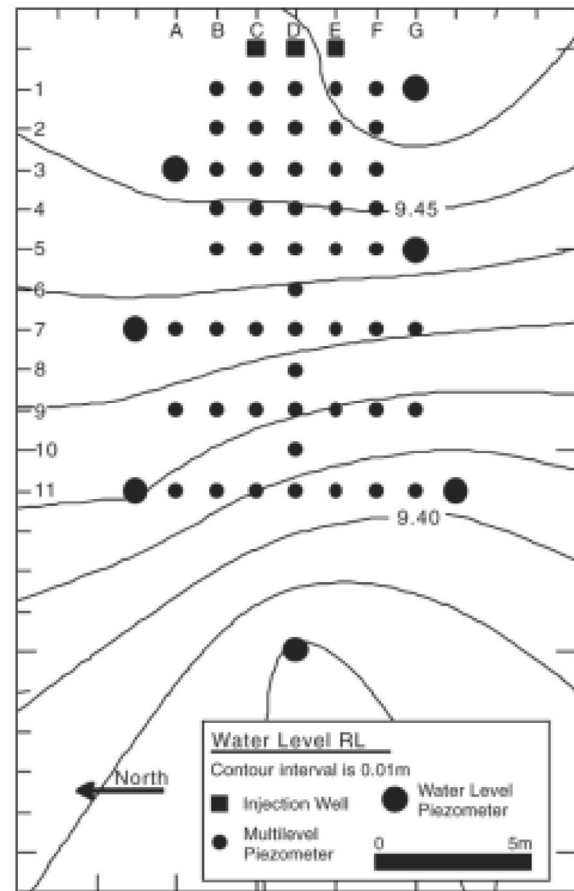


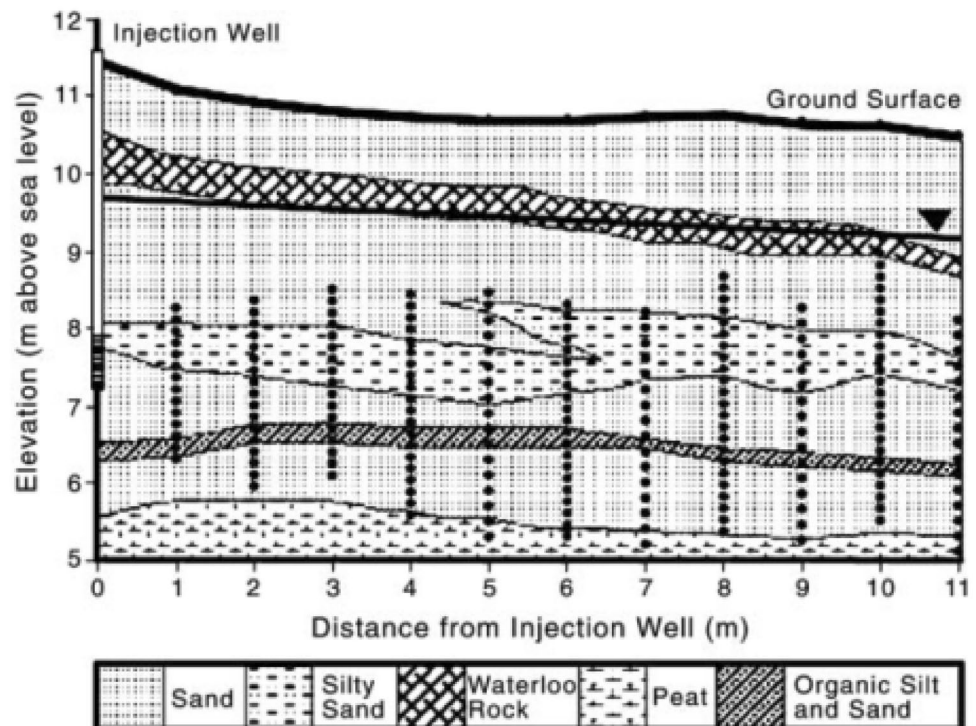
Fig. 2 Location of piezometers, injection wells and sampling points in the Eastlakes Experimental Site (Jankowski and Beck 2000)

information gathered during 1992–1993 by Evans (1993), the composite geological section through the ELE site was prepared. Figure 3 shows the composite geological formation (cross section) in line D (Fig. 2). Five distinct lithological units were discovered as Sand, Waterloo Rock, Organic Silt and Sand Bands, Peat, and Silty Sand. Since the low permeable peat layer is located between 5 and 6 m Above Sea Level (asl), the ELE area is a confined aquifer. Existence of various soil types and grain size distribution should result in variation in the HC. Moreover, the deposition environment also has significant influence over the HC distribution.

The aquifer is hydraulically connected to the pond (fixed head boundaries). The east (at injection wells) and west boundaries are fixed head and the north and south ones are variable head boundaries. The initial head distribution follows the contours in Fig. 2 with the regional hydraulic gradient of 1:240 (from east to west).

Rainfall is the main contributor to the recharge in the ELE site. The average annual rainfall at Centennial Park and Sydney Airport are 1236 mm and 1083 mm, respectively. Yu (1994) estimated the monthly average recharge for the

Fig. 3 Composite geological cross-section along line D through the ELE site (Jankowski and Beck 2000)



northern part of the BSA between 1986 and 1993. These values vary between 0 mm and 515 mm/year with a mean average monthly recharge of 68.4 mm. Therefore, for the ELE site flow model the average recharge is considered to be 68.4 mm/month.

The variability of HC was measured by multilevel piezometers in lines C, D and E with vertical spacing of 150 mm–200 mm (Evans 1993). HC was recorded at 455 locations and the minimum, mean and maximum values are 1.5 m/day, 14.6 m/day, and 50 m/day, respectively. The collected data is not normally distributed and is skewed towards the lower conductivity values. Variation of more than 50 m/day over less than 0.5 m, exist in some parts of the study area.

In order to find the spatial HC distribution in the study area, the IDW interpolation method is utilized. n (number of closest sampled data points) is selected using the spatial correlation of data. Moran’s I (Moran 1950) is an indicator of spatial auto-correlation which is applied to locations with continuous variable parameters (Eq. 9):

$$I = \frac{m \sum_{i=1}^m \sum_{j=1}^m W_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{\left(\sum_{i=1}^m \sum_{j=1}^m W_{i,j}\right) \left(\sum_{i=1}^m (X_i - \bar{X})^2\right)}, \quad (9)$$

where X_i is the variable value at location i . X_j , \bar{X} and m are the variable value at location j , the mean of all available data, and total number of available data, respectively. $W_{i,j}$ is the weight estimated using Eq. (10):

$$W_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 1/d_{ij} & \text{if } i \neq j \end{cases}, \quad (10)$$

where $d_{i,j}$ is the distance between locations i and j . Similar to correlation coefficient, the spatial auto-correlation varies between -1.0 and $+1.0$. If nearby areas are more alike, the spatial auto-correlation is positive. If nearby areas are dissimilar, the spatial auto-correlation is negative, and random patterns exhibits zero spatial auto-correlation.

The Moran’s I coefficient for the available HC values from the ELE site is 0.021 which is small and close to zero. Therefore, although BSA is a homogenous in regional scale, the ELE site is a highly heterogeneous system in the microscopic scale. The data are not spatially auto-correlated, thus $n = 3$ (Eq. 4) is selected for the IDW interpolation method. Figure 4 shows the HC distribution in the third layer of the ELE site. This distribution is estimated using IDW method with $n = 3$.

Evans (1993) had measured the HC using constant head method at three locations within the ELE site. Comparing the averaged value measured by the multilevel piezometers and the constant head method demonstrates that the multilevel piezometers underestimate the true HC as they were measured in non-equilibrium conditions. Therefore, the flow model was calibrated to better represent the flow field. The HC values shown in Fig. 4 are increased by 20% for the calibration purposes. Note, in the calibration process all the HC values are increased uniformly without changing the heterogeneity pattern of the study area. The

resulting HC field is used to estimate the head distribution, and ultimately simulate the solute concentrations at different monitoring bores and times ($Cest_{iob}^k$). In Fig. 4 the heterogeneity is presented along lines C, D, and E, since the HC measurements were taken just along these lines and there is no HC measurement out of the central region of the site. Therefore, the distribution shown in Fig. 4 may not be a good representative of the real field conditions where no measurement is available. The background hydrochemical conditions of the natural groundwater prior to tracer injection was recorded using 88 groundwater samples collected in line D.

Tracer test

This section briefly outlines the details of the tracer test conducted in the experimental site. A total of 300 L of tracer solution was prepared which included Boron (B), Bromide (Br), Chloride (Cl), and lithium (Li) as the conservative tracers and Cadmium (Cd), Lead (Pb), Potassium (K), Copper (Cu), Nickel (Ni), and Zinc (Zn) as the reactive solutes. Table 1 summarizes the injection concentrations in the tracer solution.

Three injection wells (sources C, D, and E) were developed for 1 day using a combination of pumping, surging and recharging methods, to ensure that a good interaction between the wells and the aquifer was achieved. Five 20-L batches of solution were injected in each well over a half hour period starting at 13:00 on 2 July 1996. Care was taken to maintain low injection flow rates into the wells to ensure significant increases in the hydraulic heads of the injection wells did not occur. Excessive hydraulic heads in the injection wells would force some of the tracer up-gradient of the injection wells and cause higher hydraulic gradient than would occur under natural gradient conditions.

The contaminant concentrations sampling was started 2 days after tracer injection (4 July 1996). This was followed by sampling solute concentrations every 2 days after injection. The samples were collected in lines B, C, D, E and F (Fig. 2) at various elevations.

Bromide (Br) transport simulation model

In this study, the data related to Bromide (Br) are utilized to validate the proposed methodology. Br is a conservative solute and the relative consistency and accuracy of the collected data deemed it to be appropriate for this evaluation

study. However, the presented methodology is suitable to be used for any conservative contaminant.

The Br concentrations were analyzed using ion selective electrodes and ion chromatography (Beck 2000). Analyzing the chemical condition of the natural groundwater prior to tracer test, did not show any detectable background concentration of Br. Therefore, the background concentration was zero.

Br is classified as a conservative tracer. Therefore, the two important components of the transport simulation are advection and dispersion. The dispersion coefficient was estimated by Beck (2000) using graphical method based on the concentration versus time plots. In this study, an averaged value of 0.03 m is selected for longitudinal dispersivity and the horizontal transverse dispersivity to longitudinal dispersivity ratio, and vertical transverse dispersivity to longitudinal dispersivity ratio are 0.4, and 0.1, respectively. Layers 1–4 porosities are 0.39, 0.41, 0.36, and 0.41, respectively.

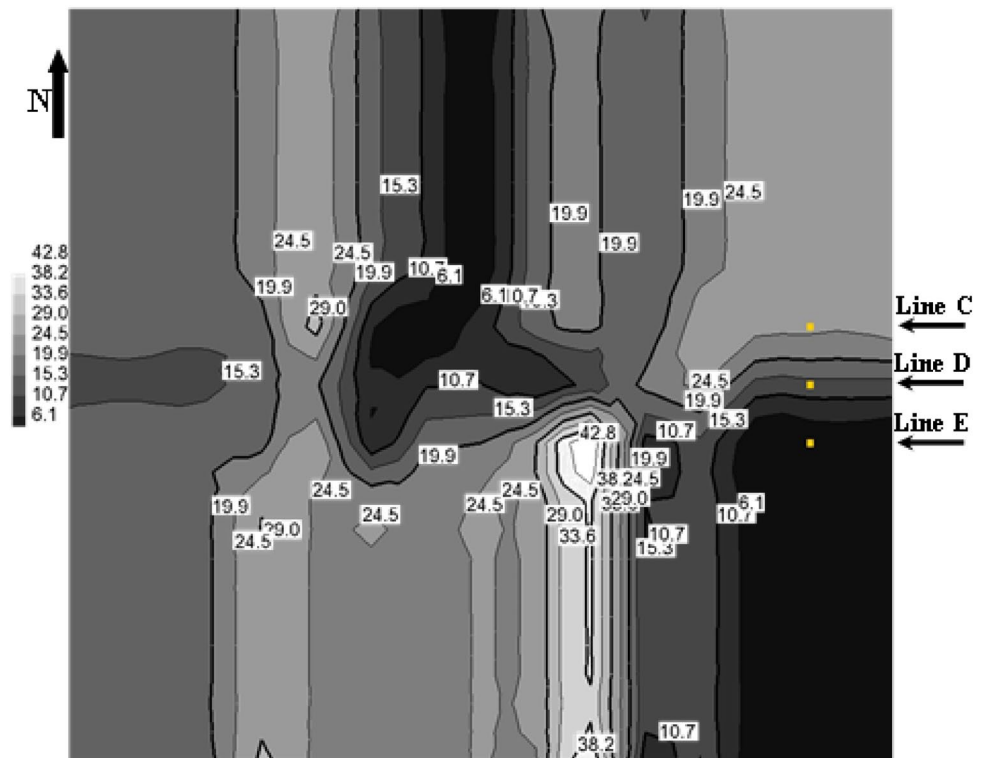
In the model conceptualization stage, the study area is divided into four layers. Layer one is between the groundwater level and top of the silty sand layer (Fig. 3). Layer two is from the top of silty sand soil layer to 7.6 m asl. Layer three is between 7.6 and 7 m asl. Layer four is between 7 m asl and the study area bed which is specified by the peat layer. Figure 5 shows the three-dimensional view of the ELE site model. The three injection sources are located in layer three. Owing to the small scale of this study area, the parameter and numerical errors have noticeable negative effect on the accuracy of simulation results. In order to reduce these negative effects in this study, layer 3 is dedicated to the contaminant monitoring procedure. The injection sources are located within this layer and the observed concentrations located within this layer are utilized for the source identification procedure.

In MODFLOW the Layer Property Flow (LPF) package is utilized. Unlike the Block Center Flow (BCF) method, the LPF model calculate the conductance for a water table bearing cell based on the water table level instead of the cell center. Dynamically following the water table can substantially increase the accuracy of flow calculations (Clemo 2003). The advection term in MT3DMS is solved using method of characteristics (MOC). The MOC method uses the particle tracking technique. MOC uses mixed Eulerian–Lagrangian method for solving the advection term, whilst finite difference method is used to solve the dispersion and sink/source mixing terms. MOC is free of numerical dispersion which is its main advantage. Other methods such

Table 1 Injection concentration of the tracer solution solutes

| Solute | B | Br | Cd | Cl | Cu | K | Li | Ni | Pb | Zn |
|----------------------|-------|-------|------|-------|------|-------|-------|------|------|------|
| Concentration (mg/L) | 110.1 | 186.0 | 53.2 | 741.8 | 56.4 | 106.9 | 123.7 | 54.8 | 51.3 | 51.4 |

Fig. 4 Hydraulic conductivity distribution in ELE site, unit is m/day



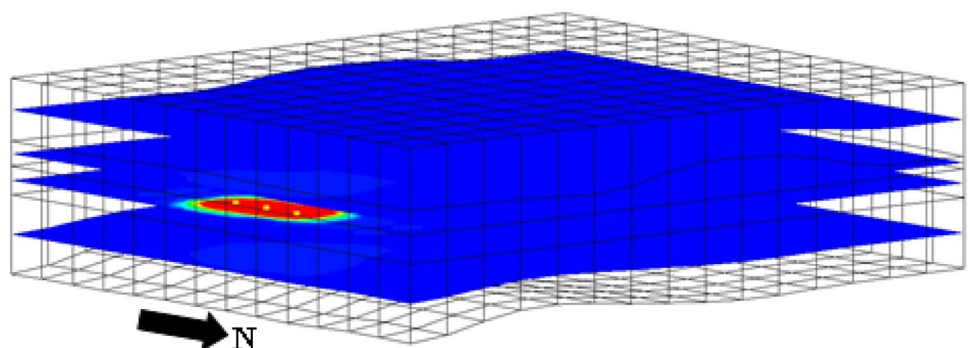
as third-order total variation diminishing (TVD) method exhibits minimal numerical dispersion and minimal oscillation in the contaminant plume fronts (Schlumberger Water Services 2011). As the result of the small scale of the ELE site, minimizing the numerical dispersion has substantial effect on the accuracy of estimated contaminant concentrations in the site.

Recovering the contamination source histories

In this section the proposed method is used to find the injection concentration at contaminant sources using the Br observed concentrations collected 2, 4, 6, and 8 days after the tracer injection. In total 19 Br concentration measurements are used for the performance evaluation. The 100 L of tracer is injected in each well over a half hour period.

Therefore, in the flow model, the injection locations are specified as flow injection wells with 4.8 m³/day flow rate. Although, care was taken to maintain very low flow rate during tracer injection, because of the small scale of the study area, still this flow rate changes the hydraulic field, and this change needs to be considered in the flow model. One candidate source location is added to the source identification model which is not actual (dummy), and is located along line G aligned with other three actual sources. Adding a dummy source location is to examine the performance of the methodology to find the location of actual sources among available potential locations. The injection flow rate in the other three source locations changes the flow system; therefore for the dummy source smaller flow rate (1 m³/day) is specified in the MODFLOW. This will diminish the effect of non-actual source on the hydraulic gradient of the study area.

Fig. 5 The discretized model and contaminant injection locations for the ELE site



The decision variables in the source identification problem are the four Br injection rates at four potential locations, and the study period is 8 days. The upper and lower bounds for contaminant release concentrations, q_{max} and q_{min} in Eq. (3), are 1000 mg/L and 0 mg/L, respectively.

This experimental tracer test seems to be simple comparing the real contamination problems. However, since the study area is small (7 m × 11 m) and the study period is short (8 days), the HC uncertainty has a substantial negative effect on the accuracy of the contaminant source identification procedure. To demonstrate this, the crisp source identification procedure is used to retrieve the source release histories. In the crisp method, it is assumed that there is no uncertainty associated with the HC distribution. In the objective function, μ_{iob}^k is one for all the monitoring locations and periods. The retrieved Br injection concentrations are presented in Table 2.

In order to quantify the HC uncertainty, multiple realisations of the flow field is required. In this study, three more realisations (in total R=4) are generated. The IDW method with n=6, 9, and 12 is utilized to interpolate the available measured HC values and generate realisations. The number of realisations and corresponding n values are selected based on the nature of available uncertain hydrogeological parameter values and computational resources. In this study area, the spatial auto-correlation of measured HCs is small. The scale of the study area is small too. Therefore, the number of selected n values to generate realisations is small, compared to the number of available HC measurements (455). For each realisation the transport simulation model needs to be executed to find the contaminant concentrations ($C_{iob}^{k,r}$) in each optimization iteration. Thus, the selected number of realisations depends on the available computational resources and also the degree of heterogeneity and uncertainty in the system. More realisations will result in better quantification of the available uncertainty while increasing computational time. The selected mn value (Eq. 7) is 0.001. The estimated Br injection rates at four potential source locations using fuzzy source identification methodology with different U values are presented in Table 2.

Discussion

The performance of the proposed methodology in recovering accurate source injection histories is tested using Normalized Absolute Error of Estimation (%NAEE). %NAEE is presented in Eq. (11):

$$\%NAEE = \frac{\sum_{i=1}^N \sum_{t=1}^T |(q_i^t)_{est} - (q_i^t)_{org}|}{\sum_{i=1}^N \sum_{t=1}^T (q_i^t)_{org}} \times 100, \tag{11}$$

where $(q_i^t)_{est}$ and $(q_i^t)_{org}$ are the estimated and original contaminant concentrations at potential source location i and stress period t, respectively. Smaller %NAEE values demonstrate that the utilized source identification algorithm is able to recover source injection concentration histories with less associated error. The %NAEE values corresponding to each source identification method are presented in Table 2. Without considering the effect of HC uncertainty, the %NAEE associated with the crisp methodology is 54%. Therefore, the crisp method exhibits the largest error compared to the fuzzy models. This demonstrates the necessity of considering the effect of hydrogeological uncertainty for recovering source injection contaminant concentration.

In the fuzzy methodology, in order to compute the COC values Eq. (8) is utilized. In this equation the appropriate U value should be selected based on the simulated concentrations. Therefore, the fuzzy method with various U values is utilized to find source injection histories. In Table 2, the smallest estimated %NAEE values are associated with U=0.3–1. Therefore, utilising the fuzzy source identification methodology with appropriate U value results in 37% error and 32% improvement in accuracy compared with the crisp methodology. All methods were able to find the non-actual (dummy) source location.

In this study, the actual injection location and injection histories $((q_i^t)_{est})$ are known and the estimated %NAEE values using Eq. 11 can be utilized to find the appropriate U value. However, in real field groundwater contamination problems, the source locations and the associated injection rates are unknown and it is not possible to estimate and use %NAEE values. In real fields, the suitable U value is

Table 2 Retrieved Br injection concentration at potential source locations

| | Br injection concentration (mg/L) | | | | | | | | | | |
|----------|-----------------------------------|-------|-------------|-------------|-------------|-------------|-------------|-------------|-----------|-----------|------------|
| | Actual | Crisp | Fuzzy U=0.1 | Fuzzy U=0.3 | Fuzzy U=0.5 | Fuzzy U=0.7 | Fuzzy U=0.8 | Fuzzy U=0.9 | Fuzzy U=1 | Fuzzy U=5 | Fuzzy U=10 |
| Source C | 186 | 127 | 111 | 8.2 | 11 | 8 | 9 | 10 | 65 | 126 | 137 |
| Source D | 186 | 89 | 135 | 163 | 160 | 171 | 170 | 160 | 179 | 91 | 93 |
| Source E | 186 | 39 | 40 | 162 | 170 | 171 | 170 | 186 | 128 | 30 | 40 |
| Source G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1 | 0 | 0 |
| %NAEE | – | 54% | 48% | 40% | 39% | 37% | 37% | 36% | 33% | 55% | 51% |

identified using estimated COC (μ_{iob}^k) values. Figure 6 presents the estimated COC associated with different U values. For each U , 19 COCs are estimated corresponding to 19 Br concentration measurements (monitoring data). The source injection concentrations, estimated as the optimal solutions of the fuzzy source identification procedure, presented in Table 2 are used to estimate COC using Eqs. (6–8). When $U=0.1$, large number of COC values are close to 1. Therefore, 0.1 seems too small for this study area. On the other hand, when U is large, in this study area $U=5$ and 10, the optimization algorithm tries to find source histories which can minimize the estimated COCs. Therefore, in Fig. 6 for $U=5$ and 10, large number of COC values are 0.1, which corresponds to the smallest possible value in Eq. (8). Therefore, these U values are too large for the fuzzy quantification purposes.

U values between 0.3 and 1 result in relative non-biased estimation of COC which expects to improve the contaminant source identification process. Results presented in Table 2 shows 36–40% improvement in accuracy of the recovered source characteristics. Results prove the effectiveness the selected U value and the fuzziness quantification in improving the accuracy of the recovered contaminant source characteristics compared to the crisp methodology.

The Br injection histories for sources D, and E were recovered with high level of accuracy by the fuzzy source identification model. However, large error is associated with the Br release history estimated for source C. Moreover, the crisp method found better estimate of the injection concentration at this source location, compared with the fuzzy method. The reason could be related to the HC uncertainty. The difference between the average HC values of the three realizations and the HC values shown in Fig. 4, are presented in Fig. 7. Figure 7 is a representative of the HC uncertainty

in this field. The counters show that there is some level of uncertainty associated with the area around source C. The contaminants move along the natural gradient which is from east to west, thus the uncertainty on the left side of the sources has effect on the simulation accuracy. As the result of this uncertainty, the COCs estimated for the monitoring locations along line C, are lower than other monitoring locations. Therefore, matching estimated and observed Br concentrations along lines D and E have higher contribution to the fuzzy source identification objective function.

The area surrounding source C is a high permeability area with large HC values (Fig. 4) compared with the neighboring areas. Therefore, in this area the number of closest sample data points used in the interpolation has noticeable effect on the estimated HC field. As fuzziness can be reduced with the acquisition of additional information (Oberkampff et al. 2004; Ross 2005), additional HC measurements in this area can reduce the associated uncertainties.

Conclusions

This study presents a methodology to quantify the hydrogeological parameter uncertainty to accurately estimate the contaminant release histories in polluted aquifers. Actual measurement data from an experimental contaminated aquifer is utilized to test the performance of the presented methodology. Characterization of contaminated aquifers requires accurate identification of the contamination source locations and their release histories. A set of potential source characteristics are evaluated using groundwater flow and pollutant transport simulation modules. Then the estimated concentrations are compared with the actual observed concentrations collected as monitored data. The linked simulation–optimization framework is utilized to search for the

Fig. 6 Estimated coefficients of confidence (COC)

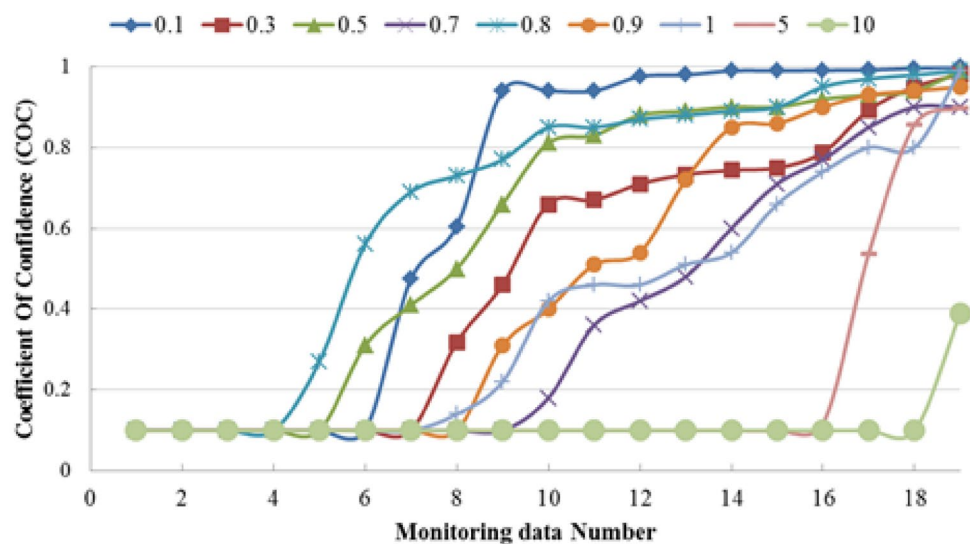
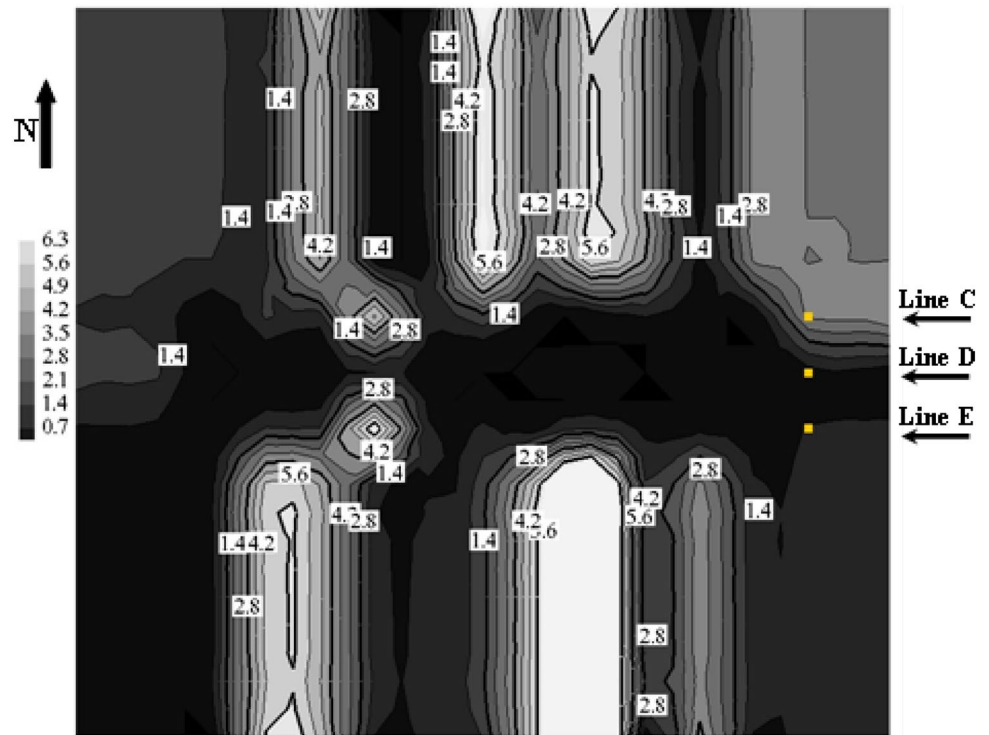


Fig. 7 Hydraulic conductivity uncertainty in ELE site, unit is m/day



pollution sources which exhibit the best match between the estimated and observed contaminant concentrations in the site. The simulation models need to be provided with reliable hydrogeological parameter values to obtain reliable source characteristics. In real life, high level of uncertainty and variability is associated with the available hydrogeological parameter values. This study quantifies the effect of uncertainty in hydrogeological parameter values on the accuracy of flow and transport models' estimates.

The proposed methodology provides insight in to the relation between the errors in groundwater simulation modules and variability and reliability in hydrogeological measurements. Tracer test results at the EES, located in BSA, Australia, are used to conduct the performance evaluation. In EES the hydrogeological heterogeneity in the microscopic scale, specifically the hydraulic conductivity, has substantial effect on the transport of pollutants. Ten tracers were injected into the groundwater system. Their movement under natural gradient were monitored by measuring concentrations in the groundwater at various locations and times after injection. Among available tracer information, Bromide is studies as a conservative element.

The adaptive simulated annealing linked simulation optimization was utilised to characterise the pollution sources using concentrations measured after tracer injection. The solution results demonstrate that the proposed methodology recovered pollution source characteristics more accurately compared with the methodologies which do not consider the effect of hydrogeological parameter uncertainty.

The developed methodology enables the decision makers to incorporate the hydrogeological parameter value uncertainty in identification of field contamination release histories. Moreover, results can be used to find the locations where available field data does not sufficiently characterise the flow field. Therefore, this methodology can also help in identifying locations where additional hydrogeologic information is required to be collected to reduce uncertainty in the flow and transport simulation models.

Funding B. Datta thanks CRC for Contamination Assessment and Remediation of Environment (CRC-CARE), University of New Castle, NSW, Australia for providing financial support for this research through Project: no. 5.6.0.3.09/10(2.6.03), CRC-CARE-Bithin Datta (JCU) which also funded the Ph.D. scholarship of the first author. M. Amirabdollahian also acknowledges the financial support by CRC-CARE, and James Cook University, Australia.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest. It is to be noted that NSW Department of Industry is not associated or endorsing any component of this research material or its results.

References

- Amirabdollahian M, Datta B (2013) Identification of contaminant source characteristics and monitoring network design in groundwater aquifers: an overview. *J Environ Prot* 4(5A):26–41

- Amirabdollahian M, Datta B (2014) Identification of pollutant source characteristics under uncertainty in contaminated water resources systems using adaptive simulated annealing and fuzzy logic. *Int J Geomate* 6(1):757–762
- Aral MM, Guan J, Maslia ML (2001) Identification of contaminant source location and release history in aquifers. *J Hydrol Eng* 6(3):225–234
- Beck PH (2000) Transport of conservative and reactive inorganic elements in the saturated part of a heterogeneous sand aquifer, Botany Basin. University of New South Wales, Sydney
- Beven K (2006) A manifesto for the equifinality thesis. *J Hydrol* 320(1–2):18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>
- Clemo T (2003) Improved water table dynamics in block-centered finite-difference flow models. In: *MODFLOW and more 2003: understanding through modeling*, Golden, Colorado, USA 11–14 September 2003
- Datta B, Chakrabarty D, Dhar A (2009) Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters. *J Hydrol* 376(1–2):48–57
- Dokou Z, Pinder GF (2009) Optimal search strategy for the definition of a DNAPL source. *J Hydrol* 376(3):542–556. <https://doi.org/10.1016/j.jhydrol.2009.07.062>
- Dokou Z, Pinder GF (2011) Extension and field application of an integrated DNAPL source identification algorithm that utilizes stochastic modeling and a Kalman filter. *J Hydrol* 398(3):277–291. <https://doi.org/10.1016/j.jhydrol.2010.12.029>
- Esfahani HK, Datta B (2016) Linked optimal reactive contaminant source characterization in contaminated mine sites: case study. *Water Res Plan ASCE* 142(12):04016061
- Evans DJ (1993) A physical and hydrochemical characterisation of a sand aquifer in Sydney. University of New South Wales, Sydney (**unpubl.**)
- Freeze RA, James B, Massmann J, Sperling T, Smith L (1992) Hydrogeological decision analysis: 4. The concept of data worth and its use in the development of site investigation strategies. *Ground Water* 30(4):574–588. <https://doi.org/10.1111/j.1745-6584.1992.tb01534.x>
- Fu T, Chen H, Zhang W, Nie Y, Wang K (2015) Vertical distribution of soil saturated hydraulic conductivity and its influencing factors in a small karst catchment in Southwest China. *Environ Monit Assess* 187(3):1–13. <https://doi.org/10.1007/s10661-015-4320-1>
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, New York
- Guillaume JHA, Hunt RJ, Comunian A, Fu B, Blakers R (2016) Methods for exploring uncertainty in groundwater management predictions. In: Jakeman AJ, Barreteau O, Hunt RJ, Rinaudo JD, Ross A (eds) *Integrated groundwater management*. Springer, Cham. https://doi.org/10.1007/978-3-319-23576-9_28
- Hazrati-Yadkooi S, Datta B (2017) Adaptive surrogate model based optimization (ASMBO) for unknown groundwater contamination source characterizations using self-organizing maps. *Water Res Prot* 9(02):193–214
- Ingber L (1996) Adaptive simulated annealing (ASA): lessons learned. *Control Cybern* 25(1):33–54
- Jankowski J, Beck P (2000) Aquifer heterogeneity: hydrogeological and hydrochemical properties of the Botany Sands Aquifer and their impact on contaminant transport. *Aust J Earth Sci* 47(1):45–64 (**copyright © Geological Society of Australia, reprinted by permission of Taylor & Francis Ltd, http://www.tandfonline.com on behalf of Geological Society of Australia**)
- Kisi O, Parmar KS (2016) Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *J Hydrol* 534:104–112
- Kisi O, Parmar KS, Soni K, Demir V (2017) Modeling of air pollutants using least square support vector regression, multivariate adaptive regression spline and M5 model tree models. *Air Qual Atmos Health* 10:873–883
- Mahar PS, Datta B (2001) Optimal identification of ground-water pollution sources and parameter estimation. *Water Res Plan ASCE* 127(1):20–29
- Moo-Young H, Johnson B, Johnson A, Carson D, Lew C, Liu S et al (2004) Characterization of infiltration rates from landfills: supporting groundwater modeling efforts. *Environ Monit Assess* 96(1–3):283–311. <https://doi.org/10.1023/B:EMAS.0000031734.67778.d7>
- Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37(1/2):17
- Mugunthan P, Shoemaker CA (2004) Time varying optimization for monitoring multiple contaminants under uncertain hydrogeology. *Bioremediat J* 8(3–4):129–146. <https://doi.org/10.1080/10889860490887509>
- O'Hagan A, Oakley JE (2004) Probability is perfect, but we can't elicit it perfectly. *J Reliab Eng Syst Saf* 85(1–3):239–248. <https://doi.org/10.1016/j.res.2004.03.014>
- Oberkampf WL, Helton JC, Joslyn CA, Wojtkiewicz SF, Ferson S (2004) Challenge problems: uncertainty in system response given uncertain parameters. *J Reliab Eng Syst Saf* 85:11–19
- Prakash O, Datta B (2015) Simulation-optimization of pollutant sources in contaminated aquifers by integration sequential-monitoring-network design and source identification: methodology and an application in Australia. *Hydrogeol J* 23(6):1089–1107
- Ross TJ (2005) *Fuzzy logic with engineering applications* (vol. book, whole). Wiley, Chichester
- Schlumberger Water Services (2011) Visual MODFLOW help. http://www.swstechnology.com/help/vmod/index.html?vm_ch5_run5.htm. Accessed 6 Feb 2018
- Singh RM, Datta B (2006) Identification of groundwater pollution sources using GA-based linked simulation optimization model. *J Hydrol Eng* 11(2):101–109
- Sun NZ (1994) *Inverse problems in groundwater modeling*. Kluwer Academic, Boston
- Tiedeman C, Gorelick SM (1993) Analysis of uncertainty in optimal groundwater contaminant capture design. *Water Resour Res* 29(7):2139–2153. <https://doi.org/10.1029/93wr00546>
- Wu JC, Zeng XK (2013) Review of the uncertainty analysis of groundwater numerical simulation. *Chin Sci Bull* 58(25):3044–3052. <https://doi.org/10.1007/s11434-013-5950-8>
- Yeh HD, Chang TH, Lin YC (2007) Groundwater contaminant source identification by a hybrid heuristic approach. *Water Resour Res* 43(9):1–16. <https://doi.org/10.1029/2005wr004731>
- Yu XW (1994) Study of physical and chemical properties of groundwater and surface water in the northern part of the Botany Basin, Sydney. University of New South Wales, Sydney (**unpubl.**)
- Zheng C, Wang PP (1999) A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems. Documentation and User's. Citeseer
- Zheng C, Hill MC, Hsieh PA (2001) MODFLOW-2000, the U.S. geological survey modular ground-water model-user guide to the LMT6 package, the linkage with MT3DMS for multi-species mass transport modeling. In: U. S. G. SURVEY (ed) *Open file report 01-82*. Denver, Colorado