Int. J. of GEOMATE, March, 2014, Vol. 6, No. 1 (Sl. No. 11), pp. 757-762
Geotech., Const. Mat. and Env., ISSN:2186-2982(P), 2186-2990(O), Japan

IDENTIFICATION OF POLLUTANT SOURCE CHARACTERISTICS UNDER UNCERTAINTY IN CONTAMINATED WATER RESOURCES SYSTEMS USING ADAPTIVE SIMULATED ANEALING AND FUZZY LOGIC

Mahsa Amirabdollahian[1,2] and Bithin Datta[1,2]
[1]School of Engineering and Physical Sciences, James Cook University, Australia; [2]CRC for Contamination Assessment and Remediation of Environment, Australia

ABSTRACT: Effective environmental management and remediation strategies are required to remediate contaminated water resources. Accurate characterizing of unknown contaminant sources is vital for selection of appropriate environmental management plan and reduction of long term remedial costs. In order to characterize the sources of contamination, the aquifer boundary conditions and hydrogeologic parameter values need to be estimated or specified. In real life contaminated aquifers, often there are sparse and inaccurate information available. On the other hand, extensive collection of data is very costly. The uncertain and highly variable natures of water resources systems affect the accuracy of contaminant source identification models.

In this study, an optimal source identification model incorporating Adaptive Simulated Annealing optimization algorithm linked with the numerical flow and transport simulation models, is designed to identify contaminant source characteristics. The fuzzy logic concept is used to identify the effect of hydrogeological parameter uncertainty on groundwater flow and transport simulation. The fuzzy membership values incorporate the reliability of specified parameter values in to the optimization model. An illustrative study area is used to show the potential applicability of the proposed methodology. The incorporation of fuzzy logic in source identification model increases the applicability of contaminant source detection models in real-life contaminated water resources systems.

Keywords: Pollution Detection, Aquifer Contamination, Groundwater, Source Identification, Uncertainty.

1. INTRODUCTION

Groundwater as the major source of potable, agricultural and industrial water is subject to various sources of contamination. The first signs of the presence of contaminated underground water may be detected from the water extracted from current extraction wells. Often surface water quality like rivers or lakes, is affected by contamination of underground water. The spread of pollution in underground water increases the necessity to develop efficient techniques for remediation of contaminated aquifer. The effectiveness of a remediation strategy depends on how efficiently the contamination source characteristics are identified. Accurate identification of the contaminant sources and reconstructing their release history plays an important role in modeling of subsurface flow and transport processes, and help to reduce the long-term remedial costs.

The source identification problem deals with the spatial and temporal variations of the location, activity duration, and the injection rate of the pollutant source and is mostly inferred from the available sparse and sometimes erroneous concentration measurements at the site. Mainly source identification includes a simulation problem, like groundwater flow and pollutant transport models, which is used to estimate past phenomena or predict future scenario. The estimated values are then compared with observed values. Availability of adequate data is vital regarding the structure of source identification procedure. However, acquiring data is a cost and time intensive task.

The source identification model requires an accurate flow and transport model to estimate the contaminant concentration distribution in aquifer. The lack of adequate geographic and parameter information, results in uncertain groundwater and solute transport model. The solution of source identification model is highly sensitive to measurement errors either in the observation data or model parameters [1].

Various techniques has been proposed in literature to characterize the contamination sources including stochastic methods [2]; response matrix [3], [4]; embedded optimization [5] and linked simulation-optimization [6], [7]. Amirabdollahian and Datta [8] presented a comprehensive overview on different source identification methods.

Most of the techniques represented in literature are designed for the case where the aquifer hydrogeological parameter values are known. Jha and Datta [7] tested their proposed method under uncertain hydrogeological parameter values. Their method is able to manage moderate degree of uncertainty. By increasing the degree of uncertainty the accuracy and reliability of source identification decreases.

In this study the linked simulation-optimization model is used to characterize the sources of contamination. The Adaptive Simulated Annealing (ASA) is utilized as the optimization method. In contrast to deterministic viewpoint where all the model parameters were considered known, in the proposed methodology the uncertainty in hydrogeologic zonation and parameter values are considered using Fuzzy Logic concept. In this way the available sparse model parameter data, in addition to available uncertain information about hydrogeologic zones are utilized to increase the reliability of pollutant source characteristics identification.

2. METHODOLOGY

The source identification model consists of an optimization algorithm linked to flow and transport process simulation models. The optimal solution is obtained by minimizing the differences between observed and simulated values. The objective function for the optimization problem is defined as follow.

$$Min\, F_1 = \sum_{k=1}^{nk} \sum_{iob=1}^{k} \mu_{iob}^{k} (cest_{iob}^{k} - cobs_{iob}^{k})^2 w_{iob}^{k} / (cobs_{iob}^{k})^2 \tag{1}$$

Subject to:

$$cest = f(x,y,z,D,q,R) \tag{2}$$

The above constraint essentially represents the flow and transport process simulation models.

Where:

$cest_{iob}^{k}$ =Concentration estimated by the simulation model at observation well location iob and at the end of time period k.

k =Total number of concentration observation time periods;

nob =Total number of observation wells;

$cobs_{iob}^{k}$ =Observed concentration at well iob and at the end of time period k;

μ_{iob}^{k} =The fuzzy reliability factor for well iob and at the end of time period k;

w_{iob}^{k} =Weight corresponding to observation location iob, and the time period k.

The weight w_{iob}^{k} can be defined as follows:

$$w_{iob}^{k} = 1 / (cobs_{iob}^{k} + n)^2 \tag{3}$$

Where n is a constant and it should be sufficiently large so that errors at low concentrations do not dominate the solution [5]. The objective function is constrained by the flow and transport simulation models.

2.1. Optimization Method

In this paper the ASA optimization method is utilized to characterize the sources of contamination. Simulated Annealing (SA) optimization starts from a feasible solution and a specific objective function. A new solution is randomly selected from its neighbors and the objective function is evaluated for the new selected solution. If the new solution has a better objective function value, the most recent solution is definitely better than the old one. Therefore, it is then accepted and the search moves to a new point and continues from there. On the other hand, if the new solution is not better than the current one, the new solution may be or may not be accepted depending on the acceptance probability. The acceptance probability is strongly influenced by the choice of a parameter temperature (T). ASA is a variant of SA in which the algorithm parameters that control temperature schedule and random selection are automatically adjusted as the algorithm progresses. This makes the algorithm more efficient and less sensitive to the user defined parameters than SA [9].

2.2. Simulation Model

The ASA based source identification model estimates the source fluxes using a linked simulation-optimization model. The simulation model evaluates the contaminant concentration values at monitoring locations corresponding to candidate contaminant source characteristics. Groundwater Flow and contaminant transport simulation models used in this study are, MODFLOW-2000 [10] and MT3DMS [11], respectively.

MODFLOW is a computer program that numerically solves the three-dimensional ground water flow equation for a porous medium by using a finite-difference method. The MT3DMS transport model uses a mixed Eulerian-Lagrangian approach for the solution of the three-dimensional advective-dispersive-reactive equation. The Lagrangian part of the method, used for solving the advection term, employs the forward tracking Method of Characteristics (MOC), the backward-tracking Modified Method Of Characteristics (MMOC), or a hybrid of these two methods. The Eulerian part of the method, used for solving the dispersion and chemical reaction terms, utilizes a conventional block centered finite-difference method [11].

2.3. Fuzzy Reliability Factor

When there is inadequate available information about the hydrogeologic zonation and parameter values, the contaminant source identification model is subjected to uncertainty. The simulation model needs exact aquifer parameter information to accurately estimate contaminant concentration at monitoring locations. In practice usually an interpolation method is utilized to calculate hydrogeologic parameter values for entire aquifer using sparse data. Substantial variation of hydrogeologic parameter values in heterogeneous and non-uniform aquifers impose high level of uncertainty to the contaminant source identification model.

The hydraulic conductivity values are used by the flow simulation model to generate the head distribution in the study area. This head information is used by the simulation model to calculate contaminant concentration at different monitoring locations.

In this study the inverse Distance Weighting (IDW) interpolation technique is used to evaluate hydraulic conductivity value for entire aquifer using available sparse information. The IDW is a type of deterministic method for multivariate interpolation with a known scattered set of points. The assigned values at unknown points are computed using a weighted average of the known values at the known points. A general form of finding an interpolated value u at a given point x based on samples $u_i=u(x_i)$ for $i=1,2,..,N$ using IDW is as below.

$$u(x)=\sum_{i=0}^{N}\frac{v_i(x)u_i}{\sum_{j=0}^{N}v_j(x)} \tag{4}$$

Where:

$$v_i(x)=\frac{1}{d(x,x_i)^p} \tag{5}$$

u(x) =Interpolated value at given location x;
N =Number of nearest points which are utilized to estimate interpolated value;
u_i =known value corresponding to point i;
v_j =Weight corresponding to point i;
$d(x,x_i)$ =distance between point x and x_i;
p = power;
Power p determines the contribution of values at known points in estimation of value at unknown points with respect to distance between points with known and unknown values. This means that higher value of p results in using more near points with known values by the IDW interpolation method.
The uncertainty in aquifer zonation with respect to hydraulic conductivity value is studied using different p values in IDW interpolation method. To calculate the fuzzy reliability factors, the fuzzy membership function is utilized. The reliability membership function is calculated using Eq. (6).

$$\mu_{iob}^{k}=1-(cest_{iob,1}^{k}-cest_{iob}^{k})^2/(cest_{iob}^{k})^2 \tag{6}$$

Where $\mathrm{cest}_{iob,1}^{k}$ is estimated concentration values obtained using simulation model at observation well location iob and at the end of time period k, with $p=M_1$ for the IDW interpolation method (Eq. (5)). cest_{iob}^{k} is estimated concentration values obtained using simulation model at observation well location iob and at the end of time period k, with $p=M_2$ for the IDW interpolation method (Eq. (5)). In the optimal source identification model (Eq. (1)) the concentration estimations are based on $p=M_2$. When the available information about field hydraulic conductivity zonation is not adequate, the concentration values estimated using simulation model will not be precise. On the other hand, there would be large discrepancy between simulated concentrations using two different p values by the IDW interpolation technique. Therefore, the fuzzy reliability factor will be less than one.
The fuzzy reliability factor for each observation location and at the end of each time period is estimated progressively using optimization generated candidate source characteristics.

3. PERFORMANCE EVALUATION

Performance of the developed methodology is evaluated using synthetic hydrogeologic data. The advantage of using synthetic data is that the true source characteristics are known for evaluation. This allows testing of the source identification methodology. Figure 1 shows the hypothetical aquifer study area. There are three candidate locations for contamination sources. Two of them are active sources and the third one is dummy source. There are two active extraction wells and 9 monitoring locations. The study area is 1500 meter by 1000 meter and is 40 meter deep. The three-dimensional model of aquifer includes two layers, each of 30 rows and 20 columns. The hydraulic conductivity is generated using truncated Latin hypercube and lognormal distribution.

Three different zones with respect to hydraulic conductivity value are considered. The mean hydraulic conductivity values for low, moderate and high permeable zones are 3, 10 and 20 m/day, respectively. The standard deviation is 0.1 times mean value for all zones. These hydraulic conductivities represent actual field condition and are generated at 50 meter intervals of the study area. The simulation model requires hydraulic conductivity values at each node. These values were obtained by interpolation using Kriging method. The resulting filed is utilized by the simulation model to estimate the spatial and temporal concentration values. Figure 2 shows the generated actual hydraulic conductivity field.

To evaluate the performance of the proposed methodology the hydraulic conductivity is assumed to be uncertain. Therefore the actual hydraulic conductivity values are assumed to be known at every 300 meters distance. By this assumption the known actual hydraulic conductivity values are reduced from 1303 values to 60 values. Using these 61 values, the uncertain hydraulic conductivity field is generated by interpolating the available 60 values to entire aquifer using IDW technique. The selected p value is 2.

To conduct the fuzzy ASA source identification, the fuzzy reliability factor needs to be calculated using candidate source characteristics. The fuzzy membership function is estimated using Eq. (6). The selected M_1 and M_2 values are 4 and 2, respectively. Figures 3a and 3b show the hydraulic conductivity generated fields using IDW method. In Figs 3a and 3b, the p value (Eq. (5)) is set to 2 and 4, respectively. Since the hydraulic conductivity data used to generate both fields is the same, the illustrated difference is due to uncertainty and lack of adequate amount of data.

4. RESULTS AND DISCUSSION

Table 1, 2 and 3 show the results of source identification model for sources 1, 2 and 3, respectively. In these tables, the first row shows the actual source fluxes during each stress period. The fluxes corresponding to source 2 is zero which means that it is a dummy source.

Second rows show the result of crisp source identification model solution where there is unaccounted uncertainty in the estimation of hydraulic conductivity field. For the crisp source identification the ASA optimization model uses the same objective function (Eq. (1)). However, in this case the fuzzy reliability factor is considered to be always equal to one.

The huge discrepancy between the resulting source fluxes and actual ones shows the deficiency of the crisp source identification model in estimating of source characteristics while there is inadequate available data for the hydraulic conductivity field.

The third rows show the result of fuzzy ASA source identification model. Comparing the results show that the fuzzy model is able to increase the accuracy of the source identification.

Results show that the use of fuzzy logic can improve the source identification results by 47.3 and 25.6 percent for sources one and three, respectively.

Although both models are not able to specify zero fluxes for source two (dummy source), the estimated source fluxes by both models are small comparing to the magnitude of flux for other sources. However, the performance of both models will improve if more precise information about hydraulic conductivity field or more monitoring data are actually available. Since the inputs for the flow simulation model have associated high level of uncertainty, limited accuracy in source identification results is expected. However, by using the fuzzy model, the error in estimation of source fluxes has been reduced. Table 4 shows the estimation error using crisp and fuzzy ASA source identification models, and demonstrates improvement in estimation of source fluxes using fuzzy logic concept.

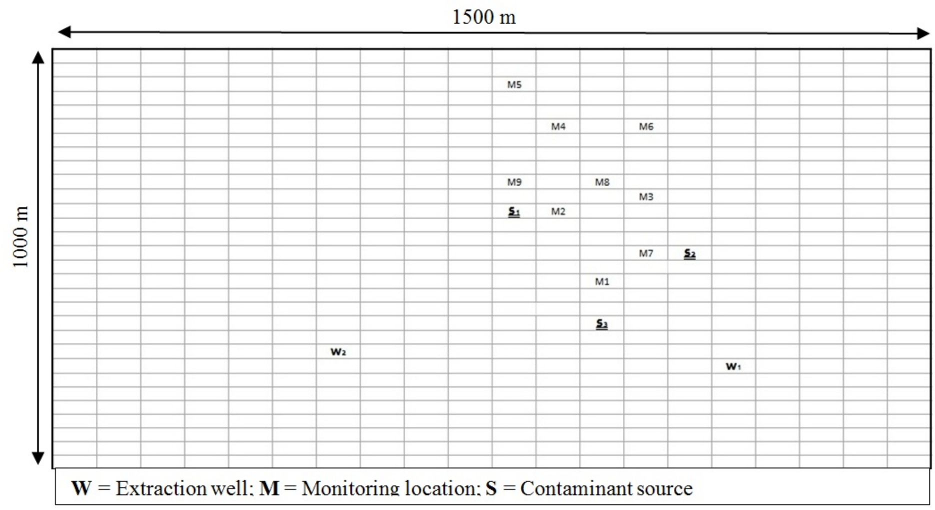


Fig. 1 Study area

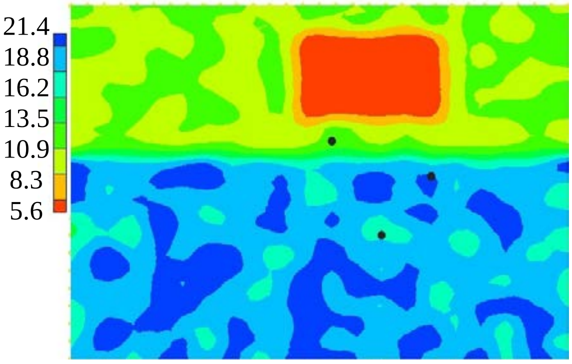


Fig. 2 Actual hydraulic conductivity values (m/day) (Dot points show the candidate source locations)

Table 1 Estimated contamination fluxes for source one (g/s)

Method	Stress Period 1	Stress Period 2	Stress Period 3
a	70	90	35
b	5.7	40.4	55.1
c	46.9	59.7	52.3

Note: a: Actual source fluxes
b) Result of crisp source identification.
c) Result of fuzzy ASA source identification.

Table 2 Estimated contamination fluxes for source two (g/s)

Method	Stress Period 1	Stress Period 2	Stress Period 3
a	0	0	0
b	0.3	9.55	15.6
c	1.2	11.5	19.2

Note: a: Actual source fluxes
b) Result of crisp source identification.
c) Result of fuzzy ASA source identification.

Table 3 Estimated contamination fluxes for source three (g/s)

Method	Stress Period 1	Stress Period 2	Stress Period 3
a	95	85	75
b	39.3	65.1	52.1
c	40.68	78.8	62.3

Note: a: Actual source fluxes
b) Result of crisp source identification.
c) Result of fuzzy ASA source identification.

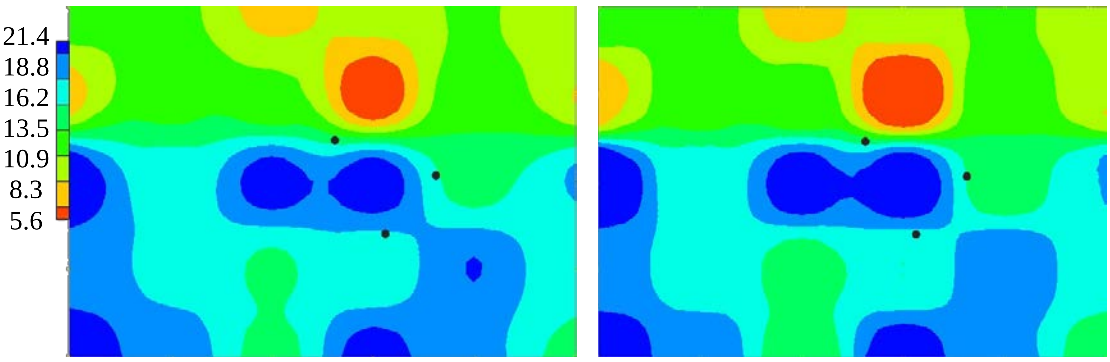
Table 4 Error in estimation of contamination fluxes (g/s) and the improvement percentage

Source	Source 1	Source 2	Source 3
b	134	25.45	98.5
c	70.7	31.9	73.22
Improvement (%)	47.3	-*	25.7

Note: b) Result of crisp source identification.
c) Result of fuzzy ASA source identification.
* Source two is dummy.

5. CONCLUSION

A new methodology for unknown groundwater pollution source identification, using Adaptive Simulated Annealing (ASA) linked simulation-optimization, and fuzzy logic is proposed. Uncertainties in hydrogeologic parameters are incorporated by using fuzzy membership function. Limited performance evaluations show the improvement in source identification using proposed methodology.



(a) (b)

Fig. 3 Uncertain hydraulic conductivity fields (m/day). Generated using IDW interpolation method where a) p=2; and b) p=4

6. ACKNOWLEDGEMENTS

We acknowledge the financial support for this work provided by CRC for Contamination Assessment and Remediation of Environment (CRC-CARE), Australia, and by James Cook University, Australia.

7. REFERENCES

[1] Sun NZ, Inverse problems in groundwater modeling. Boston: Kluwer academic, 1994.
[2] Gelhar LW, "Stochastic subsurface hydrology from theory to applications", Water Resource Research, Vol. 22, No. 9, 1986, pp. 135-145.
[3] Gorelick, SM, Evans B, and Remson, I, "Identifying Sources of Groundwater Pollution: An Optimization Approach", Water Resource Research, Vol. 19, No. 3, 1983, pp. 779-790.
[4] Datta B. et al., Development of an Expert System Embedding Pattern Recognition Techniques for Pollution Source Identification. Virginia: National Technical Information Service, 1989.
[5] Mahar, PS, and Datta, B, "Optimal identification of ground-Water pollution sources and parameter estimation", Water Resource Planning and Management- ASCE, Vol. 127, No. 1, 2001, pp. 20-29.
[6] Chadalavada, S, Datta, B, and Naidu, R, "Optimal identification of groundwater pollution sources using feedback monitoring information: a case study", Environmental Forensics, Vol. 13, No. 2, 2012, pp. 140-153
[7] Jha, M, and Datta, B, "Three dimensional groundwater contamination source identification using adaptive simulated annealing", J. of Hydrologic Engineering, Vol. 18, No. 3, 2012, pp. 307-313.
[8] Amirabdollahian, M, and Datta, B, "Identification of contaminant source characteristics and monitoring network design in groundwater aquifers: an overview", J. of Environmental Protection, Vol. 4, No. 5A, 2013, pp. 26-41.
[9] Zheng C, and Bennett, GD, Applied contaminant transport modeling. New York: Wiley-Interscience, 2002.
[10] Zheng C, Hill, MC, and Hsieh PA, MODFLOW-2000, the U.S. Geological Survey Modular Ground-Water Model- User guid to the LMT6 Package, the linkage with MT3DMS for Multi-Species Mass Transport Modeling. Denver, Colorado: U.S.G. Survey, 2001.
[11] Zheng C, and Wang, PP, A Modular Three-Dimensional Multispecies Transport Model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems. Documentation and user's guide. Denver, Colorado: U.S.G. Survey, 1999.

Int. J. of GEOMATE, March, 2014, Vol. 6, No. 1 (Sl. No. 11), pp. 757-762.
MS No. 3258 received on May 24, 2013 and reviewed under GEOMATE publication policies.

Pertinent discussion including authors' closure, if any, will be published in the March. 2015 if the discussion is received by Sept, 2014.
Corresponding Author: Masha Amirabdollahian