Desalination and Water Treatment
www.deswater.com

32 (2011) 79–85
August

1944-3994 / 1944-3986
doi: 10.5004/dwt.2011.2681

Simulated annealing based simulation-optimization approach for identification of unknown contaminant sources in groundwater aquifers

Manish K. Jha[a,b], Bithin Datta[a,b]*

[a]*School of Engineering and Physical Sciences, James Cook University, Townsville, QLD, 4811, Australia*
[b]*Co-operative Research Centre for Contamination Assessment and Remediation of Environment, Salisbury South, SA, 5106, Australia Tel. +61 (74) 7815380; Fax +61 (74) 7816788; email: bithin.datta@jcu.edu.au*

Received 23 September 2010; Accepted in revised form 18 December 2010

ABSTRACT

The exact location and release history of groundwater pollutant sources is often unknown. Identification of unknown release histories is usually carried out by inversion of the equations governing flow and transport over time and space. This is an ill posed problem. Solution of this ill-posed inversion is complicated due to the inherent non-uniqueness of solutions, uncertainties in modelling the flow and transport processes in the aquifer and unavoidable concentration measurement errors. Several methods to solve the ill posed inversion problem have been suggested in past. The simulation-optimization approach using global heuristic search optimization methods has been found to be the most effective with regards to accuracy of solutions. However, these methods are computationally intensive. A linked simulation-optimization based methodology using a variant of simulated annealing (SA) algorithm is linked to the numerical models used to simulate flow (MODFLOW) and transport processes (MT3DMS). The objective function minimizes the difference between observed and simulated contaminant concentration for optimal values of the decision variables representing the unknown source flux magnitude, duration and timing. The developed methodology is tested for an illustrative study area. The SA based source identification methodology is demonstrated to perform more efficiently compared to other methods based on genetic algorithm.

Keywords: Source identification; Groundwater pollution; Genetic algorithm; Adaptive simulated annealing

1. Introduction

Groundwater is the primary source for irrigation and drinking water in many parts of the world and hence its sustainable use is vital for ensuring long term water security. Growing anthropogenic activities on the surface and improper management of their impacts on groundwater quality has resulted in widespread contamination of groundwater worldwide. Some of the major sources that pollute groundwater are storage tanks, septic systems, hazardous waste sites, landfills, petroleum extraction and refining waste products, agro-chemicals and mining wastes etc. In the last decade, remediation of groundwater contaminated by various chemical species has been perceived as a great concern for ensuring sustainable water supply to communities dependent on groundwater.

In order to develop methodologies for effective and economical remediation of groundwater contamination, it is necessary to locate the source of such pollution and predict the future course of groundwater contamination. The first step should be to identify the sources of contamination and characterize the fluxes released as a function

* Corresponding author.

Presented at the Third International Conference on Challenges in Environmental Science & Engineering, CESE-2010 26 September – 1 October 2010, The Sebel, Cairns, Queensland, Australia

of time. A simulated annealing (SA) based methodology is developed and evaluated for optimal identification of unknown groundwater pollution sources. Reliable information about source of pollution in terms of its location and release history is highly important in planning effective remediation strategies. It is also important for assessing and assigning the liability of pollution. Identification of groundwater sources and their characterization is complicated because of lack of information and due to uncertainty in the available information. Most often, only information available is the contaminant concentration measured in one or more affected well, average porous media properties and some possible guesses about the location of the contaminant source. Characterizing the groundwater sources from such information is an inverse problem and is prone to non-unique solutions, instability of solutions and even non-existence of a solution. [1] Non-reactive groundwater flow and transport processes can be represented in 3-dimensions by mathematical equation widely known as advection dispersion equation (ADE). [2] Solution of ADE has been solved by finite-difference methods and implemented in simulation software packages. [3] When groundwater system characteristics are known, these simulations models could be used for predicting contaminant concentration at various spatial locations. Contaminant source is a component of groundwater system and the process of ascertaining the characteristics of system components from observational data is classified as an inverse problem. Further, this inverse problem is ill-posed because the mathematical equations governing flow and transport processes are not reversible in time. Mathematical and numerical methods to solve the inverse problem have been found to be very sensitive to errors/uncertainties in the observed data, thereby limiting their practical use at field level. This is because multiple source release scenarios might fit the observed data within experimental errors, leading to non-unique solutions. Details of various methods used in identification of unknown pollutant sources in groundwater can be found in literature [4–6].

A more direct and efficient approach is to use an optimization approach. However, any optimum decision based on inadequate simulation of the physical processes in the groundwater system is almost meaningless. Therefore, a proper optimization based methodology for groundwater pollution source identification should incorporate a simulation of the physical process. This method is known as coupled simulation-optimization approach. Earlier implementations of this approach used linear programming and response matrix along with forward simulations. Recently, however evolutionary algorithms such as genetic algorithm and simulated annealing have been used for optimization with significant success in non–reactive transport scenarios in 2-dimensions [7–11]. Use of evolutionary algorithms makes it computationally efficient to link the optimization algorithm with a simulation model [12–14]. The primary objective of this study is to develop a simulated-annealing based simulation-optimization methodology for unknown groundwater contamination source identification in 2-dimensional field conditions.

2. Methodology

The first step required is to link the simulation models as well as optimization algorithms to be used in this study such that the optimization algorithm is able to evaluate candidate solutions by means of forward simulations. The scenario considered is this study assumes that the contaminant concentration has been measured in more than one downstream well at specific time intervals. In the development phase, the measured observation data is synthetically generated by forward flow and transport simulation runs. It is also assumed that the hydrogeological parameters are known without uncertainties. The measurement data obtained from wells, and the hydro-geological parameters are used for characterizing the sources of contamination in terms of their location and release history. Only point sources of contaminants are considered, and it is assumed that initial guesses for potential source location are available.

Groundwater flow simulation model to be used in this research is MODFLOW-2000 [15]. 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 partial differential equation for groundwater flow that MODFLOW solves is given by Eq. (1):

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right)+W=S_s\frac{\partial h}{\partial t} \quad (1)$$

where K_{xx}, K_{yy}, and K_{zz} represent the values of hydraulic conductivity (L/T) along the x, y, and z coordinate axes, which are assumed to be parallel to the major axes of hydraulic conductivity; h is the potentiometric head (L); W is a volumetric flux per unit volume representing sources and/or sinks of water, with $W < 0.0$ for flow out of the ground-water system, and $W > 0.0$ for flow in (T^{-1}); S_S is the specific storage of the porous material (L^{-1}); and t is time (T).

When combined with boundary and initial conditions, the above equation describes transient 3-dimension groundwater flow in a heterogeneous and anisotropic medium, provided the principal axes of hydraulic conductivity are aligned with the coordinate directions. MODFLOW used a finite difference method which divides the groundwater system into a grid of cells. The potentiometric head is calculated at a point (called node) within this cell.

Contaminant transport simulation model to be used in the study has been chosen as MT3DMS.

The partial differential equation describing three-dimensional transport of contaminants in groundwater can be written as follows [16]:

$$\frac{\partial C}{\partial t}=\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C}{\partial x_j}\right)-\frac{\partial}{\partial x_i}(v_i C)+\frac{q_s}{\theta}C_s+\sum_{k=1}^{N}R_k \tag{2}$$

where C is the concentration of contaminants dissolved in groundwater, ML^{-3}; t is time, T; x_i is the distance along the respective Cartesian coordinate axis, L; D_{ij} is the hydrodynamic dispersion coefficient, L^2T^{-1}; v_j is the seepage or linear pore water velocity, LT^{-1}; q_s is the volumetric flux of water per unit volume of aquifer representing sources (positive) and sinks (negative), T^{-1}; C_s is the concentration of the sources or sinks, ML^{-3}; θ is the porosity of the porous medium, dimensionless; $\sum_{k=1}^{N}R_k$ is a chemical reaction term, $ML^{-3}T^{-1}$.

The MT3DMS transport model uses a mixed Eulerian–Lagrangian approach to 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 [17].

MT3DMS is to be used with flow models such as MODFLOW. The transport model (MT3DMS) utilizes the flow field generated by the flow model (MODFLOW) to compute the velocity field used by the transport simulation model.

In source identification problem, contaminant source fluxes, represented by the term $q_s.C_s$ in the transport equation is unknown. Generally, in the forward run situations, C_s and q_s are the inputs required to run the transport model, the solution being values of contaminant concentration at various spatial locations at various points in time. However, in source identification problems, C (spatial and temporal contaminant concentration) is known at specific locations at various point of time but source fluxes are unknown. This is why the source identification problem is considered as an inverse problem.

The strategy for estimating unknown source characteristics is to generate candidate unknown variables in optimizations algorithm, use these values for forward simulations, compute the difference between simulated and observed (measured) contaminant concentrations and finally obtain an optimal solution that minimises the difference between observed and simulated values. The objective function for the optimization problem is defined as:

$$\text{Minimize } F1=\sum_{k=1}^{nk}\sum_{iob=1}^{nob}(cest_{iob}^{k}-cobs_{iob}^{k})^2.w_{iob}^{k} \tag{3}$$

where $cest_{iob}^{k}$ = concentration estimated by the identification model at observation well location iob and at the end of time period k; nk = 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; w_{iob}^{k} = weight corresponding to observation location iob, and the time period k.

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

$$w_{iob}^{k}=\frac{1}{(cobs_{iob}^{k}+n)^2} \tag{4}$$

where n is a constant and it should be sufficiently large so that errors at low concentrations do not dominate the solution [10].

Linking of simulation models to optimizations algorithms essentially means the code modifications required to pass the variables generated through optimization algorithms as input to the transport simulation models. Optimization methods used in this research are generally categorized as global heuristic search approaches. The strength of these approaches comes from their ability to search the entire solutions space based on an ordered scheme for improvement of solutions in each iteration. Genetic algorithms [18,19] have been widely used as the optimization algorithm in most of the recent studies on unknown pollutant source identification. Simulated annealing [20] is another global heuristic search approach which has been used to a wide variety of optimization problems in the recent past. Although, its application to the unknown pollutant source identification has been limited, it certainly poses as a very good alternative to population based optimization algorithms because it can produce comparable results by using far less computational time. Being inherently robust is dealing with uncertainties due to its unique annealing algorithm it is expected to converge to a global optimal solution more efficiently.

Simulated annealing is inspired by the physical process of annealing in metallurgy which involves heating and controlled cooling of a material to reduce defects in crystal structure. The atoms are excited by heat and they become agitated while getting into higher energy states. The slow cooling allows a better chance for these atoms to achieve lower energy states than the ones they started with. In simulated annealing, a current solution may be replaced by a random "neighbourhood" solution chosen with a probability that depends on the difference between corresponding function values and on a global parameter T (called temperature) that is gradually decreased in the process. Of the various simulated annealing implementations, it is evident in literature that the adaptive simulated annealing [21,22] algorithm converges faster while maintaining the reliability of results and hence it was preferred over traditional Boltzmann annealing implementation. Fig. 1 shows the schematic representation of implementation of a simulated annealing based linked simulation optimization model.

Errors in observation of contaminant concentrations are expected to affect the predicted source characteristics. In real life situations, the observed values of contaminant concentration will invariably have measurement errors. Also, it is very likely that the observation is available only for a short period of time, or that there are missing values for some periods. The linked simulation-optimization model would be practically effective only when it can accommodate these errors by reducing their impact on the predicted source characteristics.

In order to test the model sensitivity to errors during the development stage, varied amounts of synthetically generated statistical noise has been introduced in the observed concentration values to simulate measurement errors. Competing simulation-optimization solutions have been evaluated based on the deviations in predicted source characteristics from actual ones while incorporating measurement errors.

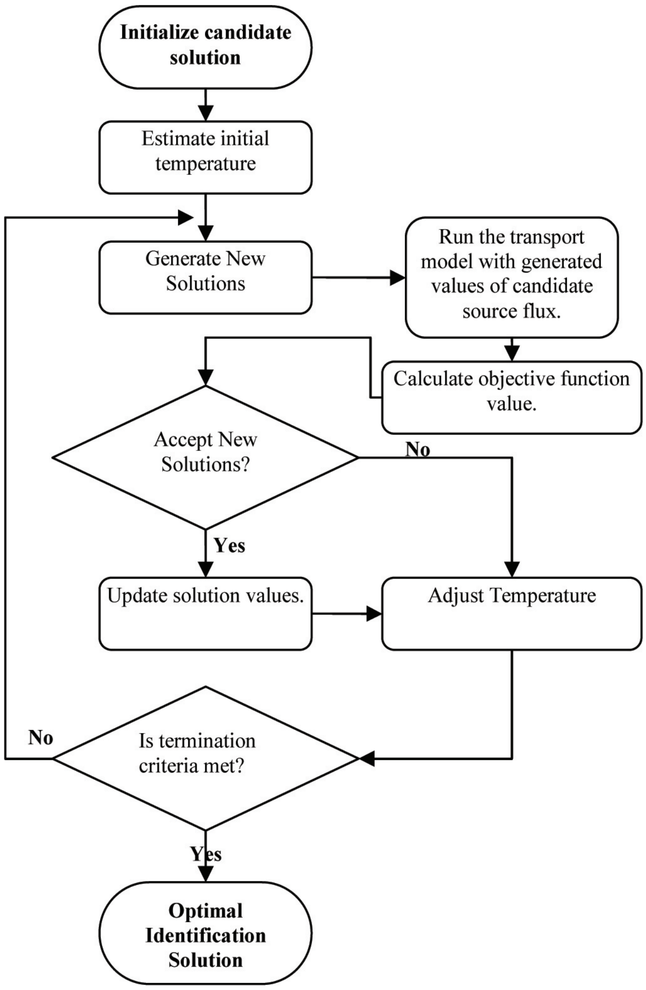

Fig. 1. Schematic representation of linked simulation-optimization model using SA.

3. Performance evaluation with numerical example

In order to evaluate the performance of the proposed approach using simulated annealing based simulation-optimization, the source identification model was implemented for an illustrative study area, with four potential source locations. Synthetic data are used for this evaluation as it is difficult to evaluate the methodology using real data obtained from the field. Real life data has inherent errors associated with it, and if the evaluation results are unsatisfactory it might be difficult to resolve whether the methodology performs poorly or the evaluation reflects the inherent unquantified errors in the field data used.

The study area is 2000 m × 1500 m in size having four potential contaminant sources. The contaminant fluxes from each of these sources are unknown. For simplicity, sources have been assumed to be continuously leaking into the aquifer. Fig. 2 shows the study area used as example while Table 1 lists the relevant parameters.

Two scenarios, based on the errors in observed contaminant concentration values, have been onsidered. In the first scenario, it is assumed that observation values are error free and all hydro-geological parameters are known with full certainty. The efficiency of simulated annealing has been tested against genetic algorithm, another widely used optimization method. Since both the optimization algorithms use different approaches, they have to be compared on some common ground. In this case the common ground is the restriction on number of groundwater transport simulations. Since simulated annealing is able to predict the source fluxes fairly accurately from error-free observation values in about 40,000 trans-

Table 1
Model parameters

Parameter	Values
Length of study area (m)	2000
Width of study area (m)	1500
Saturated thickness, b (m)	10
Grid spacing in x-direction, Δx (m)	40
Grid spacing in y-direction, Δy (m)	30
Hydraulic conductivity in x-direction, K_{xx} (m/d)	15
Hydraulic conductivity in y-direction, K_{yy} (m/d)	15
Effective porosity, θ	0.3
Longitudinal dispersivity, αL (m)	30
Transverse dispersivity, αT (m)	6
Initial contaminant concentration (mg/l)	0.00
Source fluxes (kg/d)	
Injection well 1	75
Injection well 2	54
Injection well 3	46
Injection well 4	60

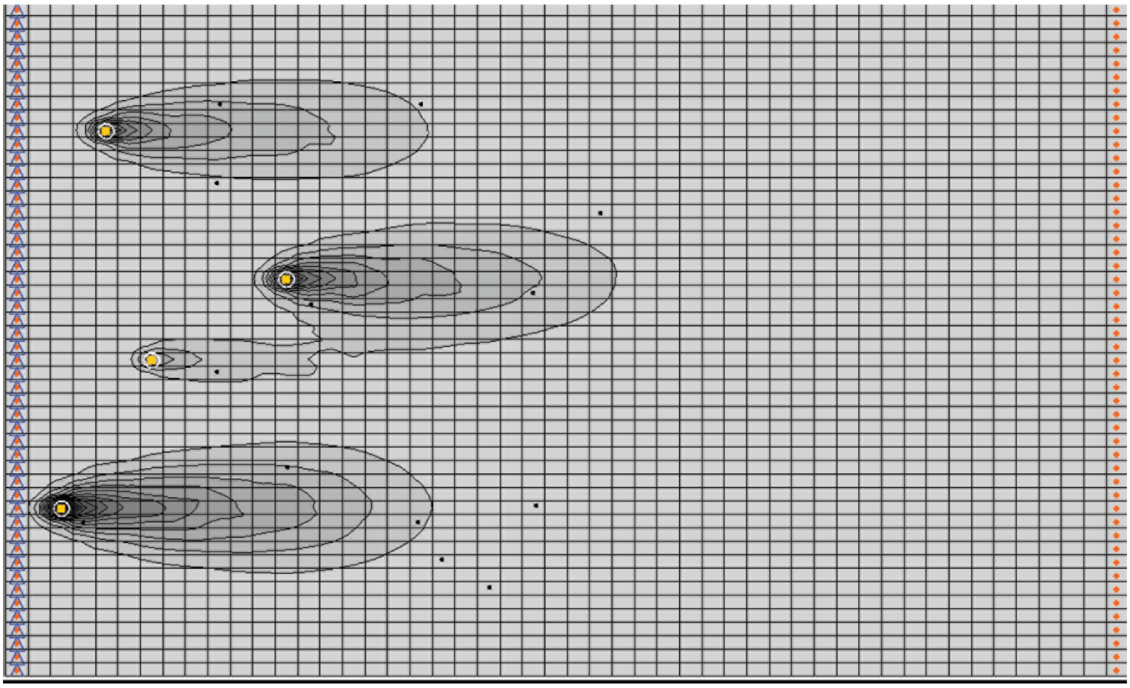

Fig. 2. The study area.

port simulation runs, genetic algorithm based solutions were restricted to this limit as well. In order to identify the most efficient population size to be used within the limitations, different population sizes viz. 10, 20, 40, 60 and 100 were considered. The number of generation considered for each population size was 4,000; 2,000; 1,000; 700 and 400 respectively.

The basis of comparison between simulated annealing and genetic algorithm is the consumed CPU time as well as the accuracy of results. The results obtained from various optimization based solutions are presented in Figs. 3, 4 and 5. Results obtained using simulated annealing are denoted by SA whereas those obtained using genetic algorithm are denoted by GA10, GA20, GA40, GA60 and GA100; where suffixes denote the population size. Fig. 3 shows the value of objective function with respect to time for genetic algorithm and simulated annealing based solutions. It is clearly shown that the convergence

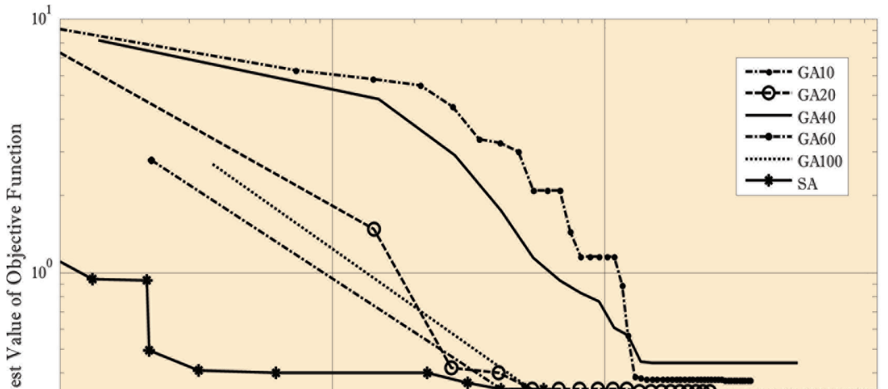

Fig. 3. CPU time vs. convergence of GA as well as SA.

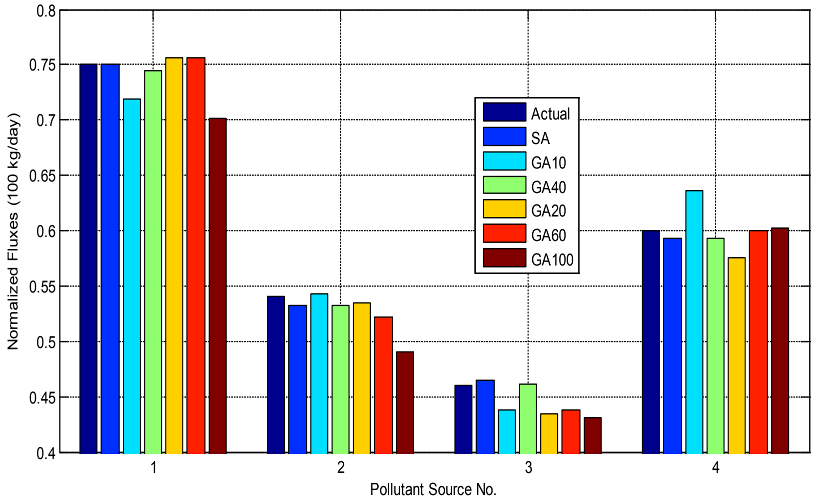

Fig. 4. Regenerated normalized flux using unperturbed data.

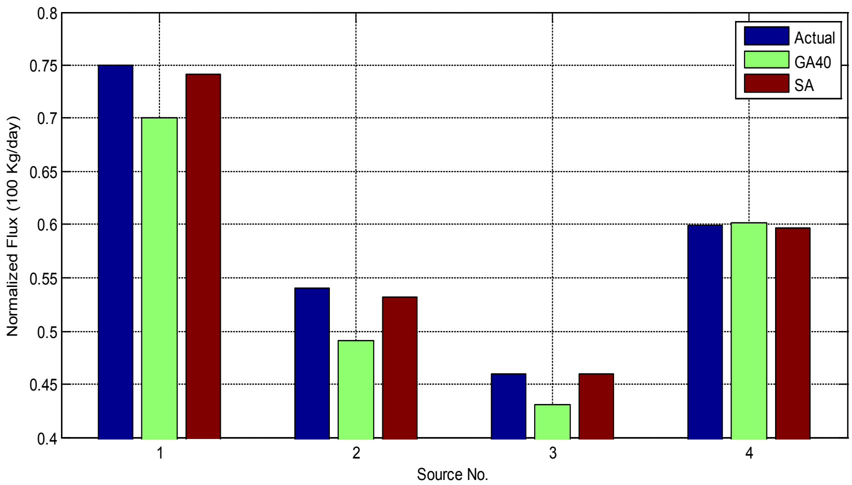

Fig. 5. Regenerated normalized fluxes using perturbed data.

of simulated annealing based simulation-optimization solution is better compared to any genetic algorithm based solution.

The source fluxes obtained by assuming error free observations are shown in Fig. 4. It is evident that among the genetic algorithm based solutions, the one with a population size of 40 predicts the source fluxes more accurately than other population sizes considered. The limited evaluation results do not show any consistent improvement in GA performance by increasing the population size while keeping the number of maximum simulation runs to 40,000. The number of maximum simulations is based on the best performance of SA. These evaluations formed the basis of selecting the optimal population size for GA.

In the next step, measured contaminant concentration data in the observation wells were assumed to have uncontrolled as well as controlled errors. Errors were added to the synthetically generated observation data as per the following formulation:

$$C_{pert} = C_{ns} + S_{nd} \cdot a \cdot C_{ns} \tag{5}$$

where C_{pert} = perturbed concentration values; C_{ns} = simulated concentration; S_{nd} = standard normal deviate; a = a fraction.

The source fluxes generated by simulated annealing based strategy as well as the best of genetic algorithm based strategy have been shown in Fig. 5. It is clearly evident that simulated annealing identifies the original source fluxes more closely than genetic algorithm.

4. Conclusion

A simulation-optimization approach for source identification was developed based on simulated annealing.

It was applied to an illustrative study area. The results obtained were compared with those obtained using genetic algorithm, a more commonly used optimization approach. It is evident from the limited numerical experiments that simulated annealing based identification solutions converge to the actual source fluxes faster than genetic algorithm based approaches. The source fluxes identified by simulated annealing based methodology are closer to actual fluxes when compared to the results obtained using genetic algorithm. Therefore, the potential for using SA based optimal source identification methodology is demonstrated. The pollutants addressed in this study refer to one or more conservative pollutants which don't react or get adsorbed with each other or with the soil. The proposed approach is currently being extended to reactive pollutants. However, this exercise is not yet complete, as it is a much more challenging issue, and not addressed in this paper.

References

[1] N.Z. Sun, Inverse Problems in Groundwater Modeling. Theory and Applications of Transport in Porous Media, Kluwer Academic, 1994.
[2] R. Freeze and J. Cherry, Groundwater, Prentice-Hall, 1979, 604 p.
[3] C. Zheng, MT3D: A modular three-dimensional transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. Report to the US Environmental Protection Agency, 1990.
[4] J. Atmadja and A. Bagtzoglou, State of the art report on mathematical methods for groundwater pollution source identification, Environ. Forensics, 3 (2001) 205–214.
[5] M. Aral, J. Guan and M.L. Maslia, Identification of contaminant source location and release history in aquifers, J. Hydrol. Eng., 3 (2001) 225–234.
[6] M. Snodgrass and P. Kitanidis, A geostatistical approach to contaminant source identification, Water Resources Res., 4 (1997) 537–546.
[7] S. Gorelick, B. Evans, M. Steven and R. Irwin, Identifying sources of groundwater pollution: an optimization approach, Water Resources Res., 3 (1983) 779–790.
[8] B. Datta, J. Beegle, M.L. Kavvas and G.T. Orlob, Development of an Expert System Embedding Pattern Recognition Techniques for Pollution Source Identification, University of California, 1989.
[9] P.S. Mahar and B. Datta, Optimal monitoring network and ground-water-pollution source identification, J. Water Resources Plan. Manage., 4 (1997) 199–207.
[10] P.S. Mahar and B. Datta, Identification of pollution sources in transient groundwater systems, Water Resources Manage., 3 (2000) 209–227.
[11] P.S. Mahar and B. Datta, Optimal identification of ground-water pollution sources and parameter estimation, J. Water Resources Plan. Manage., 1 (2001) 20–29.
[12] G. Mahinthakumar and M. Sayeed, Hybrid genetic algorithm-local search methods for solving groundwater source identification inverse problems. J. Water Resources Plan. Manage., 1 (2005) 31–45.
[13] R. Singh and B. Datta, Identification of groundwater pollution sources using GA-based linked simulation optimization model, J. Hydrol. Eng., 2 (2006) 101–109.
[14] R. Singh and B. Datta, Identification of unknown groundwater pollution sources using artificial neural networks, J. Water Resources Plan. Manage., 6 (2004) 506–514.
[15] M.G. McDonald and A.W. Harbaugh, A modular three-dimensional finite-difference ground-water flow model, US Geological Survey, 1988.
[16] I. Javandel, C. Doughty and C.F. Tsang, Groundwater Transport: Handbook of Mathematical Models, American Geophysical Union, 1984.
[17] C. Zheng and P. Wang, A field demonstration of the simulation optimization approach for remediation system design, Ground Water, 3 (2002) 258–265.
[18] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms, Wiley Publishing, 2001.
[19] J. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.
[20] S. Kirkpatrick, Optimization by simulated annealing – quantitative studies, J. Statist. Phys., 5 (1984) 975–986.
[21] L. Ingber, Adaptive Simulated Annealing: Global Optimization C-Code, Caltech Alumni Association, 1993.
[22] L. Ingber, Adaptive simulated annealing (ASA): Lessons learned. Control Cybernetics, 2 (1996) 33–54.