Groundwater contaminant source identification by a hybrid heuristic approach

Hund-Der Yeh,[1] Tung-Hua Chang,[1] and Yu-Chung Lin[1]

Received 16 November 2005; revised 2 June 2007; accepted 9 July 2007; published 29 September 2007.

[1] This study proposes an approach, SATS-GWT, that combines simulated annealing (SA), tabu search (TS), and the three-dimensional groundwater flow and solute transport model (MODFLOW-GWT). Our approach is used to estimate the source information: source location, release concentration, and release period. The sampling concentrations at monitoring points are simulated by the MODFLOW-GWT with an assumed release concentration and release period at a known source location. In the source estimation process, the source location is selected by TS within the suspected source area, and the trials for release concentrations and release periods are generated by SA. The MODFLOW-GWT is employed to compute the simulated concentrations at the monitoring points with the trial solution. The above mentioned procedures are repeated until the stopping criterion regarding the differences of objective function value (OFV) is met. The last trial source information which yields the best OFV is the final solution. Six studies on the homogeneous site, two studies on the heterogeneous site, and one study on the transient flow problem are conducted in this study. A suggestion regarding the optimal number of monitoring points and the condition for estimating the source information is given on the basis of the studies for homogeneous and heterogeneous aquifers. All results indicate that the proposed SATS-GWT can give good estimations, even when the sampling concentrations contain measurement errors.

Citation: Yeh, H.-D., T.-H. Chang, and Y.-C. Lin (2007), Groundwater contaminant source identification by a hybrid heuristic approach, *Water Resour. Res.*, *43*, W09420, doi:10.1029/2005WR004731.

1. Introduction

[2] Groundwater pollution problems have attracted much attention in recent years. Once a groundwater contamination site is found, the contaminant source information such as source location, release period, and release concentration should be determined. The major purposes of contaminant source identification are to recognize the source location and determine the party responsible for the groundwater contamination. Furthermore, by knowing the source information, the remedial response actions may be more efficient. *Atmadja and Bagtzoglou* [2001] divided the groundwater contaminant source identification problem into two categories. The first is the release history recovery problem which is to reconstruct the source release history at a known location. The second is the source information problem which is to determine the source location with a known release concentration and period. They also reviewed the methods that had been developed in identifying the source location and release history, including the optimization approach, probabilistic and geostatistical simulation approach, analytical solution and regression approach, and direct approach.

[3] For the release history recovery problem, *Liu and Ball* [1999] classified this problem into two types: function-fitting and full-estimation approaches. The function-fitting approach initially assumes the release history as a particular function and reformulates it to an optimization problem. Then a gradient-type or non-gradient-type method is employed to estimate the best fit parameters of the release function [*Gorelick et al.*, 1983; *Wagner*, 1992]. The full-estimation approach is used to reconstruct the release history by matching the simulated concentrations with the sampling concentrations [*Skaggs and Kabala*, 1994, 1995, 1998; *Woodbury and Ulrych*, 1996; *Snodgrass and Kitanidis*, 1997; *Woodbury et al.*, 1998; *Liu and Ball*, 1999; *Neupauer and Wilson*, 1999, 2001; *Neupauer et al.*, 2000].

[4] For the source information estimation problem, *Gorelick et al.* [1983] employed a groundwater transport simulation model incorporating linear programming and multiple regressions to estimate the source information. They defined the error as the difference between sampling concentration and simulated concentration. Then the linear programming method and multiple regression method were respectively used to minimize the sum of the absolute errors and the sum of the squared errors. Both methods could properly identify the source location, although the estimated release concentration was incorrect in the transient case. *Hwang and Koerner* [1983] employed a modified finite element model and limited monitoring well data to identify the pollution source by minimizing the sum of the squared errors between the sampling and simulated concentrations. The *National Research Council* [1990] suggested using a trial-and-error method incorporated with a forward model to solve the problem of source information estimation.

[1]Institute of Environmental Engineering, National Chiao Tung University, Hsinchu, Taiwan.

0043-1397/07/2005WR004731

Bagtzoglou et al. [1992] proposed an approach using particle methods to provide probabilistic estimates of source location and time history in a heterogeneous site. Their study indicated that the simulation with a conditional conductivity field performs as well as the simulation with a perfectly known conductivity field. *Mahar and Datta* [1997, 2000, 2001] provided a serial investigation to different types of source information estimation problems. In their study, the finite difference method was employed to approximate two-dimensional groundwater flow and the transport equations. They formulated the source estimation problem as a constrained optimization form and solved the objective function by nonlinear programming. Their study successfully identified the source information for flow in both steady and transient states. *Sciortino et al.* [2000] developed an inverse procedure based on the Levenberg-Marquardt method and a three-dimensional analytical model to solve the least squares minimization problem for identifying the source location and the geometry of a DNAPL pool. Their study showed that the result is highly sensitive to the hydrodynamic dispersion coefficient.

[5] Gradient-type approaches require giving initial guess values for the unknown variables and using an iterative scheme to solve the nonlinear equations. A proper initial guess, given by an experienced investigator, can yield good results with quick convergence. However, if an improper initial guess is made, the obtained solution may be trapped in a local optimum or even become divergent. Recently, some investigators have tried to solve the optimization problem by using heuristic approaches such as the genetic algorithm (GA), simulated annealing (SA), and tabu search (TS). In contrast to the gradient-type approaches, the heuristic approaches directly generate an arbitrary guess value as the trial solution and can still achieve optimal result for the problem. The heuristic approaches determine the optimum solution on the basis of the specified objective function value and convergence criterion. Thus these approaches do not involve taking the derivatives of the decision variables, which are generally difficult to evaluate analytically or numerically in nonlinear and nonconvex problems. *Aral and Guan* [1996] proposed an approach called improved genetic algorithm (IGA) to determine the contaminant source location, leak rate, and release period. The results obtained from IGA agreed with those obtained from linear and nonlinear programming approaches. *Aral et al.* [2001] proposed a new approach, called the progressive genetic algorithm (PGA), in which the GA is combined with the groundwater simulation model, for the source identification problem. Their results indicated that the initial guess does not influence the identified solution. *Mahinthakumar and Sayeed* [2005] and *Sayeed and Mahinthakumar* [2005] employed hybrid genetic algorithm-local search (GA-LS) methods to solve the groundwater source identification problem. In their studies, the GA was first applied to obtain the results which then were used as the initial guesses for the local search to obtain the global optimum. A total of 1800 observations were employed to locate the contaminant source location and reconstruct the release history. Their results indicated that the GA-LS methods were very effective to the groundwater source identification problem. *Mahinthakumar and Sayeed* [2006] further compared four hybrid GA-LS optimization approaches for estimating the source information. Their study also considered both single- and multiple-source releases in three-dimensional heterogeneous flow fields under the condition that the flow is in steady state. The results showed that the release history recovery problem with the known potential source locations is much easier to solve than the source location identification problem.

[6] In addition to the source information estimation problems, heuristic methods are also applied in other fields. *Zheng and Wang* [1996] treated the problem of identifying optimal parameter structure as a large combinatorial optimization problem. They employed the TS and SA to solve the combinatorial optimization problem. Their results indicated that their proposed approach performed extremely well compared to those obtained from the grid search or descent search approach. *Tung et al.* [2003] developed an optimal procedure for applying SA and the short distance method with MODFLOW to determine the best zonation of hydraulic conductivity. They determined the best zonation of hydraulic conductivity by minimizing the errors in hydraulic head. Their results illustrated that the procedure can effectively determine and delineate the hydrogeological zones. *Tung and Chou* [2004] proposed a procedure for identifying the spatial pattern of groundwater pumping rates by integrating TS and pattern classification. Their procedure was successfully applied to a simulated problem based on real conditions. They also mentioned that the procedure can also be applied to different problems of pattern classification.

[7] Three objectives are considered in this current study. The first objective is to develop a hybrid approach to solve the source information estimation problem, which is formulated as an optimization problem. This approach, called SATS-GWT, combines the merits of SA and TS and incorporates with a three-dimensional groundwater flow and solute transport model, called the modular three-dimensional finite difference groundwater flow model with groundwater transport (also called MODFLOW-GWT or MF2K-GWT). At the beginning of the source information estimation process, TS is employed to produce a trial solution for the source location within the suspected area, the so-called hot spot area, and SA is used to generate the release period and release concentration. MODFLOW-GWT is then employed, with the trial solutions generated by SA and TS, to generate the sampling concentrations at the monitoring points. The procedure of source information estimation is terminated once the difference between two consecutive objective function values (OFVs) continually satisfies the stopping criterion four times. The last trial source information which yields the best OFV is considered as the final solution. In addition, the effects of the values of initial guess and measurement error on the results are investigated when employing the proposed approach to perform source information estimation.

[8] The second objective is to suggest an effective way for optimally allocating the monitoring points in the estimation of source information on the basis of the case studies. The sampling concentration data are essential and important for source information estimation. The concentrations of contamination are detected at the monitoring points, so the locations of the monitoring points directly influence the sampling concentrations. However, the least number of monitoring points and the minimum concentration level in order to get good estimated results are often un-

known. If these criteria are available in advance, the work of source identification can be more effective and the identified results may be more accurate. The proposed approach is employed with several designed scenarios to investigate the requirements for the optimal number of monitoring points and the conditions for effectively estimating source information even if the aquifer is heterogeneous.

[9] The third objective is to test the performance of SATS-GWT on possible real-world problems in complicated aquifer systems. Three scenarios are considered herein to demonstrate the capability of our proposed approach. The first scenario deals with the problem of a heterogeneous aquifer. The second scenario is to consider a much larger suspected area representing increased difficulty in the source information estimation. Most existing works for contamination source identification assume that their groundwater flow systems are in a steady state condition. Therefore, in the last scenario the proposed method is applied to a problem with transient flow.

2. Methodology

2.1. Groundwater Flow and Transport

[10] Contaminant transport in groundwater is a complicated process. The transport processes may include advection, dispersion, diffusion, adsorption, and biodegradation. The groundwater flow field should be known before simulating the contaminant transport in groundwater. The three-dimensional groundwater flow equation may be expressed as [*Konikow et al.*, 1996]

$$\frac{\partial}{\partial x_i}\left(K_{ij}\frac{\partial h}{\partial x_j}\right) + W = S_s\frac{\partial h}{\partial t} \qquad i,j = 1,2,3, \tag{1}$$

where K_{ij} is the hydraulic conductivity tensor, h is the hydraulic head, S_s is the specific storage, t is time, W is the volumetric flux per unit volume (positive for inflow and negative for outflow), and x_i are the Cartesian coordinates. The three-dimensional contaminant transport in groundwater may be written as [*Konikow et al.*, 1996]

$$\frac{\partial(\varepsilon C)}{\partial t} + \frac{\partial}{\partial x_i}(\varepsilon C V_i) - \frac{\partial}{\partial x_i}\left(\varepsilon D_{ij}\frac{\partial C}{\partial x_j}\right) - \sum C' W = 0 \qquad i,j = 1,2,3, \tag{2}$$

where ε is the porosity, C is the contaminant concentration, V_i is the average linear velocity of groundwater flow, D_{ij} is the dispersion coefficient (a second-order tensor), and C' is the concentration of the source fluid. Equation (1) is used to predict the head distribution for a flow field. The average linear velocity in equation (2) can be determined by Darcy's law; that is,

$$V_i = -\frac{K_{ij}}{\varepsilon}\frac{\partial h}{\partial x_j} \qquad i,j = 1,2,3. \tag{3}$$

[11] The temporal and spatial concentration distribution of a contaminant being released at a specified point can be simulated by equation (2). The computer model MODFLOW-GWT, developed by the United States Geological Survey (USGS), can be used to simulate the groundwater flow and contaminant transport simultaneously. It consists of a three-dimensional, method-of-characteristics, solute transport model (MOC3D) [*Konikow et al.*, 1996] and the modular three-dimensional finite difference groundwater flow model (MODFLOW-2000) [*Harbaugh et al.*, 2000]. In addition, this model can handle discretization of space in the horizontal direction by reading the number of rows, the number of columns, and the width of each row and column, therefore allowing the use of nonuniform grid in the simulation.

2.2. Simulated Annealing

[12] The SA algorithm is based on an analogy to the physical annealing process, which is the process of heating up a solid and then cooling it down slowly until it crystallizes. Physically, when a rock is heated, the activity of molecules in the rock is increased with the temperature. Then the temperature is slowly decreased to let molecules form crystalline structures. The most stable crystalline structure of the rock only forms when the rock is properly cooled. If the cooling is carried out too fast, an irregular structure may be obtained and the system does not reach the minimum energy state.

[13] *Romeo and Sangiovanni-Vincentelli* [1991] used homogeneous and inhomogeneous Markov chain theories to prove that SA can converge to global optimal solutions. *Rayward-Smith et al.* [1996] pointed out that SA is an evolution from the descent search method which has the problem of obtaining a local optimal solution. At the beginning of SA, a control parameter analogous to the temperature in the physical world is given and an initial guess x is required to evaluate the objective function $f(x)$. An upper bound and a lower bound are given and the region between upper bound and lower bound is defined as the solution domain. All the trial solutions are generated within the solution domain. For any point x, a neighborhood trial solution x' is randomly generated and its OFV is denoted as $f(x')$. The neighborhood trial solution x' is given as

$$x' = x + (2*RD_1 - 1)*r, \tag{4}$$

where RD_1 is a random number between zero and one generated from a uniform distribution, and r is the step length vector. The variable r is first defined as the length of the difference between the upper bound and lower bound and then automatically adjusted to expect that about a half of the trial solutions can be accepted as new optima in the next temperature status. For point x, the next neighborhood trial solution x' is selected from $x - r$ to $x + r$. Notice that the trial solution must be generated between the upper bound and lower bound.

[14] To solve the minimization problem, if $f(x')$ is smaller than $f(x)$, then x' is accepted and the current optimal solution x is replaced by x'. If $f(x')$ is not smaller than $f(x)$, then the Metropolis criterion is used to test the acceptability of the trial solution. The Metropolis criterion provides a mechanism to accept inferior solutions and the acceptance of inferior solutions avoids the trial solution becoming trapped in a local minimum. To solve the minimization problem, the Metropolis criterion was given as [*Metropolis et al.*, 1953; *Pham and Karaboga*, 2000]

$$P_{SA} = \begin{Bmatrix} 1 & if \quad f(x') \le f(x) \\ \exp\left(\frac{f(x)-f(x')}{Te}\right) & if \quad f(x') > f(x) \end{Bmatrix}, \tag{5}$$

where P_{SA} is the acceptance probability of the trial solution x' and Te is the current temperature. A random number RD_2

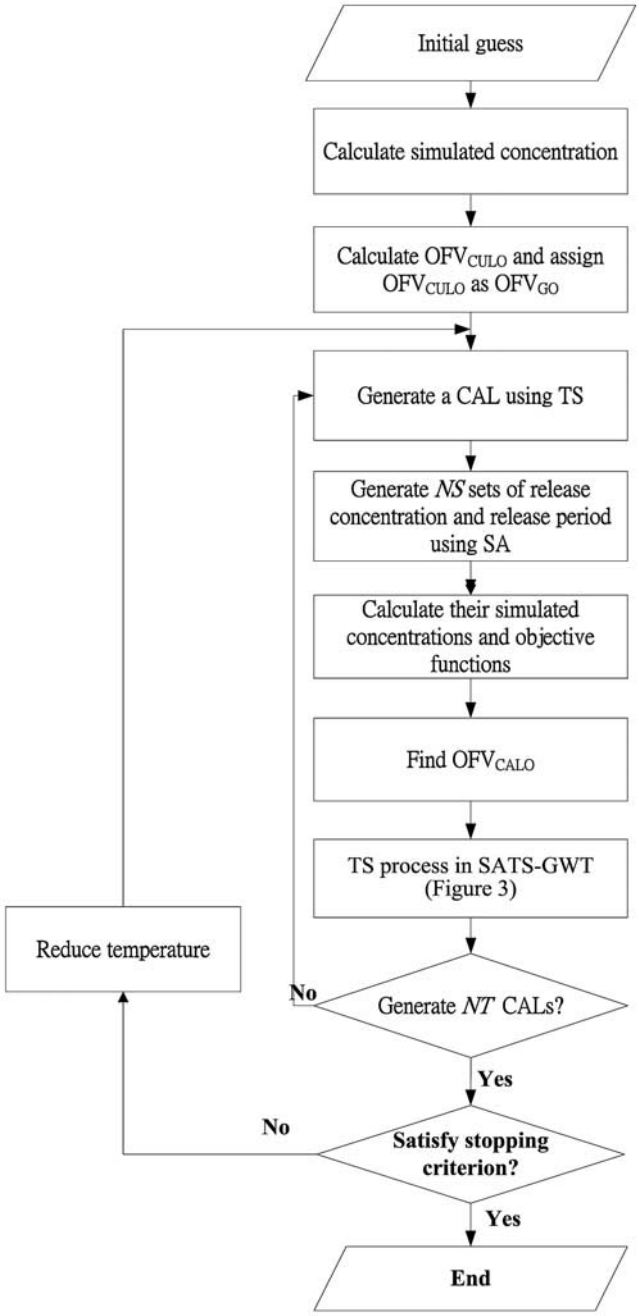

Figure 1. Flowchart of SATS-GWT. The OFV_{GO} represents the objective function value (OFV) of global optimal solution, OFV_{CULO} represents the OFV of the optimal solution at current location, OFV_{CALO} represents the OFV of the optimal solution at candidate location, and CAL represents the candidate location. TS, tabu search; SA, simulated annealing.

ranging between zero and one is generated from a uniform distribution. If RD_2 is smaller than P_{SA}, the trial solution x' is accepted and called an ascent move. Otherwise, trial solutions are continually generated from the current solution. After a series of trial solutions are generated, the current temperature is decreased by a constant, called the temperature reduction factor, and the prior steps are repeated continually. The acceptance probability is very high at the beginning of SA for expecting to explore the solution domain completely and is decreased as the temperature goes down. Comparatively, at later period of SA, the acceptance probability is much lower to anticipate obtaining the nearby global optimal solution. The algorithm is terminated when the stopping criteria are satisfied.

2.3. Tabu Search

[15] Learning and memory are the main concepts of TS proposed by *Glover* [1986]. During the processes of learning, the prior result is memorized to influence the next experiment. A worse result may cause the next experiment to be terminated and the required computation time is shortened. On the other hand, a better result may encourage the next trial to increase the accuracy of the obtained solution. According to these two ideas, TS utilizes the tabu list (TL) and aspiration criterion to interdict or to encourage some trial solutions during the iterative process. The purpose of the TL is to memorize some recently evaluated trial solutions. The intention of the aspiration criteria is to release some of the solutions memorized in the TL to avoid trapping solutions in a local optimum.

[16] The iterative process of TS contains four components: initial guess, candidate solution and movement, TL, and aspiration [*Tung and Chou*, 2004]. At the beginning, an initial guess for the unknown variables is considered as the current solution (CUS) and the guess values are used to calculate the objective value. This objective value is stored as the global optimal objective value (GOOV). Next, several adjacent candidate solutions (CASs) are generated in the neighborhood of the CUS and their objective values are also evaluated. For the minimization problem, when the best objective value (BOV) is less than GOOV, then the aspiration criterion is applied to remove the best CAS from the TL if it is in the list. At the same time, the CUS is moved to the TL. In addition, the best CAS becomes the new CUS and the BOV becomes the new GOOV. For BOV > GOOV, the next best CAS will be selected if the best CAS is in the TL; otherwise, the best CAS becomes the new CUS. The procedures are repeated continually by generating other adjacent CASs from the neighborhood of the new CUS if the stopping criterion is not satisfied.

3. Approach for Source Information Estimation

[17] This section illustrates how the hybrid SA and TS are incorporated with a groundwater flow and transport model to solve the source identification problem. An approach that combines SA and TS with MODFLOW-GWT, called SATS-GWT, is developed to solve the source information estimation problem. The objective function defined to identify the source is

$$Minimize\ f = \frac{1}{n}\sum_{i=1}^{n}\left[C_{i,sim} - C_{i,sam}\right]^2, \tag{6}$$

where $C_{i,sim}$ is the simulated concentration at the ith monitoring point, $C_{i,sam}$ is the sampling concentration at the ith monitoring point, and n is the number of monitoring points. Note that the sampling concentrations are generated by MODFLOW-GWT at the monitoring points at a specific time. The simulated concentration at the monitoring point, $C_{i,est}$, is predicted by equation (2) and equation (6) is used to calculate the OFV of the trial solution.

[18] Figure 1 shows the flowchart of SATS-GWT. In SATS-GWT, the first step is to calculate the initial OFV

on the basis of the initial guesses which include the source location, release period, and release concentration. The guess location for the source is considered as the current location. The initial OFV calculated by equations (2) and (6) with the initial guesses is considered as the optimal OFV at the current location (hereinafter referred to as OFV_{CULO}).

[19] In the second step, SATS-GWT is used to generate the trial location for the source from the candidate source locations. An upper bound and lower bound for the trial location must be specified to confine the candidate source locations. If the current location falls within the upper bound and lower bound, the criterion for assigning the trial location is

$$\begin{cases} x_{i+1} = x_i + 1 & if \quad 0 \le RD_3 < \frac{1}{3} \\ x_{i+1} = x_i & if \quad \frac{1}{3} \le RD_3 \le \frac{2}{3} \\ x_{i+1} = x_i - 1 & if \quad \frac{2}{3} < RD_3 \le 1 \end{cases}, \tag{7}$$

where x_i and x_{i+1} represent the current location and trial location, respectively, along the x coordinate and RD_3 is a random number ranging between zero and one generated by the uniform distribution function. If the current location reaches the upper bound, the criterion for assigning the trial location is

$$\begin{cases} x_{i+1} = x_i - 1 & if \ 0.5 \le RD_3 \le 1 \\ x_{i+1} = x_i & if \ 0 \le RD_3 \le 0.5 \end{cases}. \tag{8}$$

Finally, if the current location reaches the lower bound, the following criterion is applied:

$$\begin{cases} x_{i+1} = x_i & if \ 0.5 \le RD_3 \le 1 \\ x_{i+1} = x_i + 1 & if \ 0 \le RD_3 < 0.5 \end{cases}. \tag{9}$$

The trial locations in y and z coordinates can also be generated in a similar way.

[20] In the third step, SATS-GWT generates *NS* trial solutions on each release period and release concentration at each trial location. *NS* is defined as the number of generated trial solutions for each considered variable in SA. For each set of trial solutions, the MODFLOW-GWT is employed to calculate the simulated concentrations at the monitoring points. Then the simulated concentrations at the monitoring points are calculated using equation (2) and then the OFV related to each set of trial solution is calculated using equation (6) in the fourth step. The least value of the objective function among *NS* trial solutions is considered as the optimal OFV at the candidate location (hereinafter referred to as OFV_{CALO}) and the solution is considered as the local optimal solution.

[21] The fifth step is to apply the TS to find the global optimal OFV (hereafter called OFV_{GO}), which is defined as the best one among all OFVs as indicated in Figure 2. Note that the OFV of the initial guess is the first OFV_{GO} and this value is upgraded during the TS process. When $OFV_{CALO} < OFV_{GO}$, the aspiration criterion is applied to ensure that the trial location is not in the TL. Then the current location is moved to the TL and the trial location becomes the new current location. If $OFV_{CALO} > OFV_{GO}$, then the TL is checked to see whether the trial location is in the TL or not. If it is in TL, a new trial location should be generated from current location and the previous steps are repeated as indicated in Figure 2. Otherwise, the Metropolis criterion is used to test the acceptance of trial location:

$$P_L = \begin{cases} 1 & if \quad OFV_{CALO} \le OFV_{CULO} \\ \exp\left(\frac{OFV_{CULO} - OFV_{CALO}}{Te}\right) & if \quad OFV_{CALO} > OFV_{CULO} \end{cases}, \tag{10}$$

where P_L is the acceptance probability of the trial location. A random number ranging between zero and one is generated to compare with P_L. The trial location will be rejected when P_L is less than the random number.

[22] The number of trial locations generated in TS process at a specific temperature is defined as *NT*. After *NT* trial locations are generated, the temperature is reduced by the specified reduction factor. The algorithm is terminated when the difference of two OFV_{GO}s is less than 10^{-6} four times successively. The estimated source information related to the latest upgraded OFV_{GO} is considered as the final solution.

4. Case Studies for Homogeneous Aquifers

[23] This section contains two parts. A hypothetical homogeneous site is described in the first part. In the second part, six scenarios are designed to test the performance of the proposed SATS-GWT approach and suggest an effective way for estimating the source information on the basis of these case studies. The initial guess location for the source is (260 m, 500 m, −21 m) and the initial guess value for the release concentration and release period are respectively 50 ppm and 1.5 years for scenarios 2 to 6. Except for the second scenario, the upper bound values of release concentration and release period are respectively 200 ppm and 6 years.

4.1. Aquifer Site

[24] A hypothetical site with a homogeneous aquifer is used in scenarios one to seven to test the performance of the proposed approach, SATS-GWT, in source information estimation. The example assumes that a point source S1 releases the contaminant in an unconfined aquifer and the contaminant is assumed to be not biodegradable nor adsorbed onto the aquifer materials. The aquifer length and width are both 1000 m and the aquifer thickness is 24 m. The hydraulic conductivity, porosity, and the hydraulic gradient are given as 15 m/day, 0.3, and 0.005, respectively. Accordingly, the average groundwater flow velocity for a homogeneous aquifer under steady and uniform flow conditions is 0.25 m/day and the dispersion coefficients in the $x-$, $y-$, and $z-$ directions are given as 10 m^2/day, 2.5 m^2/day, and 0.25 m^2/day, respectively. The finite difference grids are block-centered and the related boundary conditions for the flow field of the unconfined aquifer system are shown in Figure 3. Both grid width and length are 40 m and the grid height is 6 m; thus the number of finite difference nodes in x, y, and z directions are 25, 25, and 4, respectively. The origin of the vertical coordinate is taken at the land surface and S1 is located at the coordinates of (220 m, 540 m, −9 m). The S1, located at 9 m below the land surface, releases a constant concentration of 100 ppm

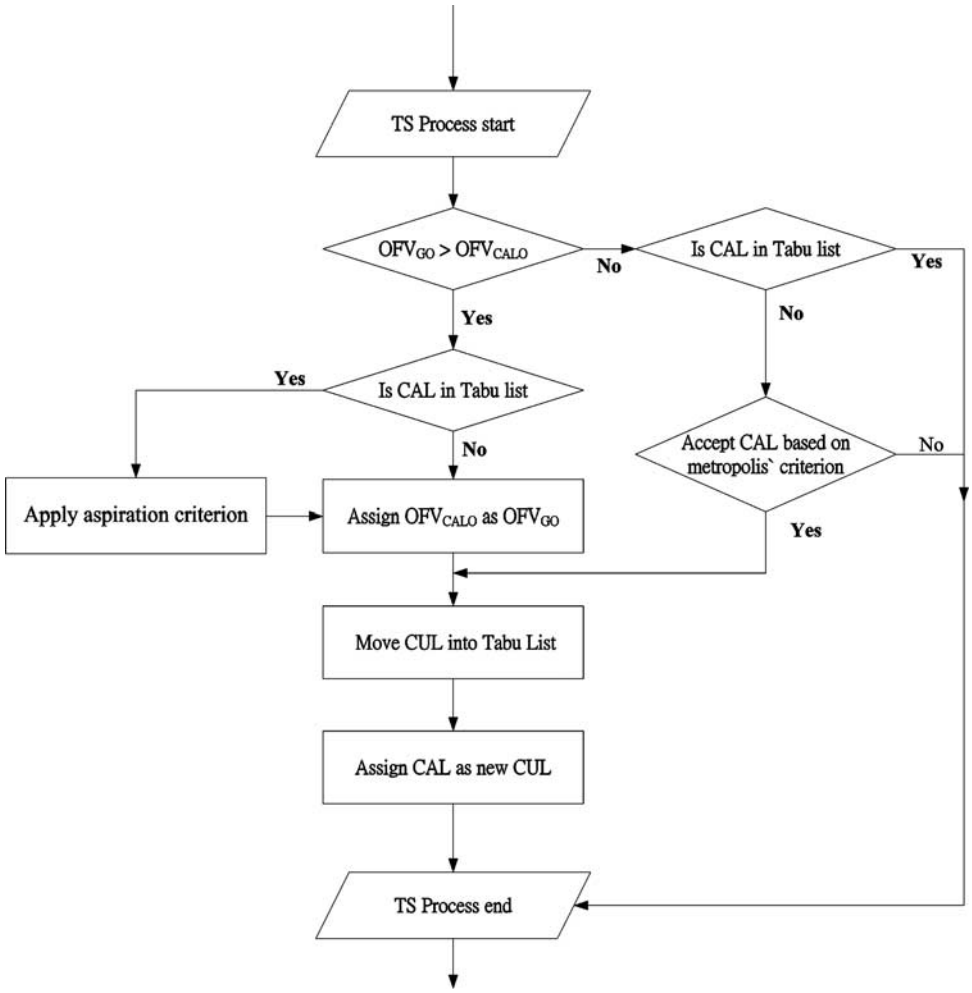

Figure 2. Flowchart of TS process in SATS-GWT. The OFV_{GO} represents the OFV of global optimal solution, OFV_{CULO} represents the OFV of the optimal solution at current location, OFV_{CALO} represents the OFV of the optimal solution at candidate location, CAL represents the candidate location, and CUL represents the current location.

with a release rate (W) of 1 m^3/day. Notice that that source information hereinafter is referred as the real solution. The sampling concentrations at 10 monitoring points, that is, wells A to J shown in Figure 3, are simulated using MODFLOW-GWT. Seventy-six sampling concentrations, after the source releases contaminant over a 3-year period, sampled from 10 monitoring wells at various depths for different contaminant sources are considered and listed in Table 1. Notice that each sampling point has only one sampling concentration. The name of the monitoring points given in Table 1 consists of the well name and layer number. For example, the sampling point A2 represents groundwater is sampled at the second layer of monitoring well A.

[25] At the beginning of the source information estimation, a total of 36 (3 rows $\times$ 3 columns $\times$ 4 layers) candidate sources, including the real source S1 and suspected sources near S1 within the suspected area, are considered. Figure 3 shows a suspected area drawn with dark solid lines for the present example problems. The *NS*, *NT*, initial temperature, and temperature reduction factor are given as 20, 10, 5, and 0.8 respectively. The maximum number of source locations that the TL memorized is three.

4.2. SATS-GWT Performance Test

[26] The first scenario is to study the effect of using different initial guesses on the estimated results of the source location, release concentration, and release period. In total, seven case studies with different initial location, release concentrations, and release period in the first scenario are chosen to test the effect of different initial guesses on the source information estimation. The second scenario explores the effect of the upper bound values of the release concentration and release period on the analyzed results. Three different upper bound values for the release concentration and release period are selected in this scenario to test their effect on the source information estimation. The third scenario examines the validity of the proposed approach when the sampling concentrations contain measurement errors with different error levels. In the fourth scenario, five cases with three to seven monitoring points are chosen to investigate the required number of monitoring points. The fifth scenario with 26 case studies is intended to explore the required number of concentration intervals and extract an effective way for allocating monitoring points for source information estimation. In the sixth scenario, the sources

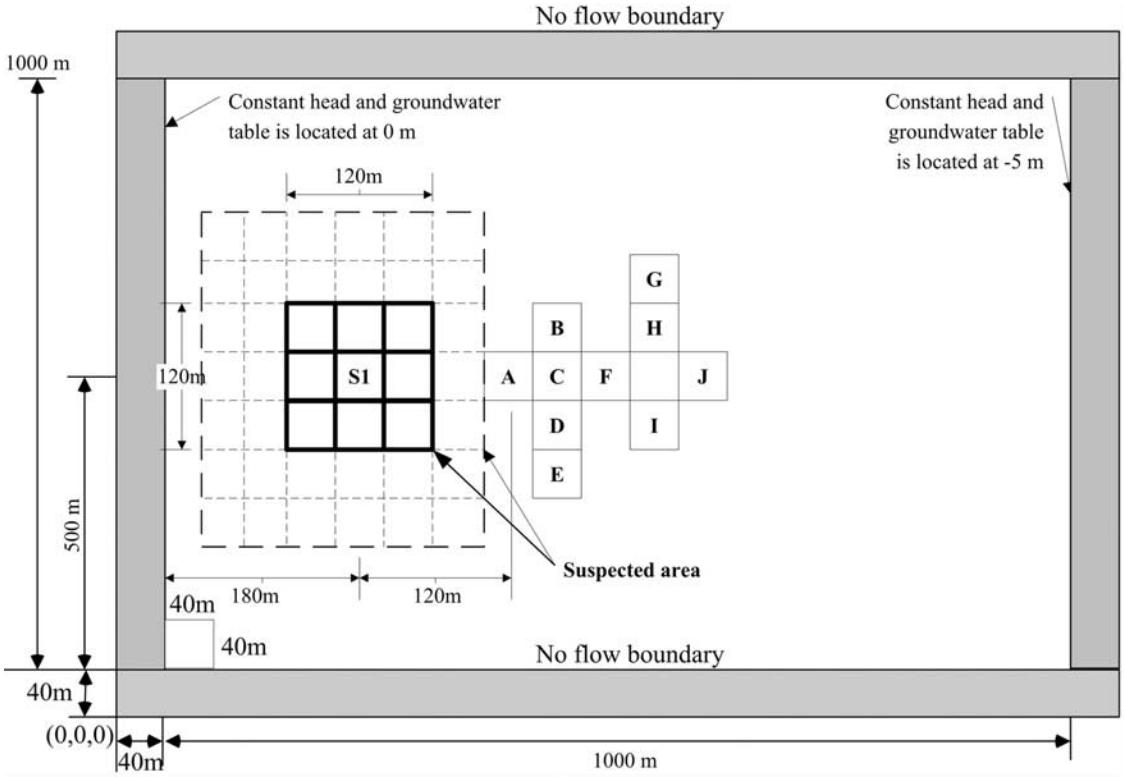

Figure 3. An aquifer system with an area of 1000 m by 1000 m (not to scale), the locations of real source S1, eight suspected sources near S1, and wells A–F.

located at four different depths, 3 m, 9 m, 15 m, and 21 m below the land surface, are considered in eight cases to explore the effect of the source depth on the allocation of monitoring points and source information estimation.

4.2.1. Examination of Initial Guess

[27] One of the advantages of applying heuristic approaches is that the initial guess values have little influence on the results of source information estimation. This scenario uses seven cases to study the effect of using different initial guesses on the estimated results of the source information. For cases 1.1 to 1.4, Table 2 shows the results when varying the guess source location with initial guesses of 50 ppm and 1.5 years for the release concentration and release period, respectively. The estimated source located at (220 m, 540 m, −9 m) is exact and the results of the release concentration and release period are

Table 1. Sampling Concentrations at Monitoring Wells A–J

	Sampling Concentration, ppm			
Sampling Point	(220, 540, −3)[a]	(220, 540, −9)[a]	(220, 540, −15)[a]	(220, 540, −21)[a]
A2	0.4577	0.4877	0.4035	0.3425
A3	0.3384	0.4029	0.4740	0.4661
B1	0.2376	0.2260	0.1892	0.1643
B2	0.2181	0.2194	0.1990	0.1821
B3	0.1826	0.1993	0.2145	0.2204
B4	0.1588	0.1828	0.2208	0.2525
C2	0.3894	0.3880	0.3498	0.3101
D1	0.2396	0.2279	0.1909	0.1657
D2	0.2189	0.2206	0.2000	0.1832
E2	0.0818	0.0815	0.0761	0.0718
F1	0.2770	0.2903	0.2537	0.2220
F2	0.2788	0.2592	0.2565	0.2425
F4	0.2144	0.2442	0.2854	0.2968
G2	0.0569	0.0564	0.0537	0.0517
H2	0.1360	0.1342	0.1265	0.1209
I1	0.1435	0.1388	0.1252	0.1153
I2	0.1355	0.1337	0.1262	0.1206
J2	0.1490	0.1528	0.1374	0.1295
J3	0.1328	0.1396	0.1512	0.1474

[a]Real source location.

Table 2. Results of Seven Cases Designed in the First Scenario for Studying the Effect of Initial Guesses on Source Information Estimation[a]

	Initial Guess Value			Results				
Case	Guess Source Location, m	Release Concentration, ppm	Release Period, years	Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-6}$	Number of Simulations
1.1	(260,580, −21)	50	1.5	(220, 540, −9)	99.86	3.00	0.31	9000
1.2	(260,500, −21)	50	1.5	(220, 540, −9)	99.17	2.99	12.1	8800
1.3	(260,500, −3)	50	1.5	(220, 540, −9)	98.36	2.97	9.19	8800
1.4	(180,500, −3)	50	1.5	(220, 540, −9)	99.36	3.00	3.27	9800
1.5	(260,500, −21)	30	0.9	(220, 540, −9)	99.17	2.99	12.1	9000
1.6	(260,500, −21)	120	3.6	(220, 540, −9)	99.17	2.99	12.1	8600
1.7	(260,500, −21)	150	4.5	(220, 540, −9)	99.17	2.99	12.1	8600

[a]Note that the real source is located at (220 m, 540 m, −9 m), real release concentration is 100 ppm, and real release period is 3 years. The monitoring points are installed at A2, B2, C2, F2, I2, and J2.

slightly different from the real solution. Case 1.3 gives the worst results, which have −1.64% relative error in the estimated release concentration and −1% relative error in the estimated release period.

[28] For cases 1.5 to 1.7, the guess source is located at (260 m, 500 m, −21 m) for these three cases, the guess values for the release concentration are given as: 30, 120, and 150 ppm, and the guess values for the release period are given as 0.9, 3.6, and 4.5 years. Table 2 also shows the estimated results for the source information, objective function values, and number of simulations in these three cases. Note that the number of simulations is defined as $2 \times NS \times NT \times NR$, where 2 represents the two unknowns of release concentration and release period, and NR represents the number of temperature reductions. The estimated source location is at (220 m, 540 m, −9 m), the estimated release concentration is 99.17 ppm, and the estimated release period is 2.99 years, indicating that the estimated source location is correct and the results of the release concentration and release period differ only slightly from the real solution. In summary, Table 2 indicates that the proposed approach is capable of estimating the source information regarding three-dimensional contaminant transport in groundwater. In addition, these results also demonstrate that the proposed approach gives good estimate results while employing various guess values for the source information.

4.2.2. Examination of Solution Domain

[29] In SATS-GWT, the upper and lower bounds need to be given to confine the trial solutions to a reasonable range. If the upper bounds for the release concentration and release period are given too low, the real solution may be beyond the range between upper and lower bounds. In contrast, if the upper bounds for the release concentration and release period are given too high, the computing time to obtain the estimated results will be large. Six cases are designed for studying the effect of the upper bounds on the analyzed results. In cases 2.1 to 2.3, large upper bound values such as 1000 ppm for release concentrations and 10 years for the release period are considered. Table 3 shows that the estimated source locations obtained in cases 2.1 to 2.3 are correct, and the estimated release concentration and period also have good accuracy. Case 2.1 gives a slightly larger relative error among those three cases, that is, 6.5% relative error in release concentration and −8.1% relative error in release period. Therefore we conclude that the effect of large upper bounds on the accuracy of source information estimation is insignificant.

[30] In cases 2.4 to 2.6, the upper bounds of the release concentration or/and release period are smaller than the real ones to explore the effect of small upper bounds on the source information estimation. The upper bounds such as 60 ppm for release concentrations or/and 2 year for release period are considered in these three cases. The estimated source locations in these three cases are all incorrect and the values of release concentration or/and release period are close to the upper bounds as indicated in Table 3. The upper bound of release concentration is set as 60 ppm in cases 2.4 and 2.6, and the estimate release concentrations obtained from these two cases are 58.12 ppm and 57.37 ppm, respectively. The upper bound of release period is set as

Table 3. Results of Six Cases Designed in the Second Scenario for the Examination of the Solution Domain[a]

	Upper Bound Value			Results			
Case	Release Concentration, ppm	Release Period, years	Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-5}$	Number of Simulations
2.1	1000	6	(220, 540, −9)	106.50	2.92	0.31	8400
2.2	200	20	(220, 540, −9)	99.91	3.01	14.7	6600
2.3	1000	20	(220, 540, −9)	98.41	2.97	0.93	6200
2.4	60	6	(260, 540, −9)	58.12	3.68	76.0	8600
2.5	200	2	(260, 540, −21)	155.54	1.99	463.9	9400
2.6	60	2	(260, 540, −3)	57.37	1.88	960.5	9400

[a]Note that the real source is located at (220 m, 540 m, −9 m), real release concentration is 100 ppm, and real release period is 3 years. The monitoring points are installed at A2, B2, C2, F2, I2, and J2.

Table 4. Results of Three Cases Designed in the Third Scenario for the Effect of Errors on Source Information Estimation[a]

Case	Error Level, %	Identified Results: Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-5}$	Number of Simulations
			Measurement Error			
3.1	1	(220,540, −9)	99.54	3.01	0.36	9400
3.2	5	(220,540, −9)	104.08	3.03	4.70	9800
3.3	10	(220,540, −9)	97.72	3.07	38.64	8800
			Model Error			
3.4	1	(220, 540, −9)	99.45	3.00	1.82	10000
3.5	5	(220, 540, −9)	101.25	3.02	41.09	9200
3.6	10	(220, 540, −9)	96.31	3.00	124.22	9400

[a]Note that the real source is located at (220 m, 540 m, −9 m), real release concentration is 100 ppm, real release period is 3 years, and the monitoring points contain A2, B2, C2, F2, I2, and J2.

2 year in cases 2.5 and 2.6, and the estimate release periods obtained from these two cases are 1.99 years and 1.88 years, respectively. The optimal OFVs obtained from these three cases are very high, that is, larger than 10^{-3}. These larger OFVs reflect that the simulated concentrations are significantly different from the sampling concentrations. Clearly, inappropriate upper bound values made in the source information estimation give wrong results.

4.2.3. Examination of Errors

[31] In this scenario, two types of random errors, normally distributed error and uniformly distributed error, are added into sampling concentrations separately. In normally distributed error analysis, the disturbed sampling concentrations are expressed as [*Mahar and Datta*, 2001]:

$$C'_{i,obs} = C_{i,obs} \times (1 + E \times RD_4), \tag{11}$$

where $C'_{i,obs}$ is the disturbed sampling concentration, E is the error magnitude, and RD_4 is a normally distributed random deviate generated by the routine RNNOF of IMSL [*Visual Numerics Inc.*, 2003]. In uniformly distributed error analysis, the disturbed sampling concentrations are expressed as [*Mahar and Datta*, 2001]

$$C'_{i,obs} = C_{i,obs} \times (1 + E \times RD_5), \tag{12}$$

where the RD_5 is a uniformly distributed random deviate generated by the routine RNUN of IMSL [*Visual Numerics Inc.*, 2003] to stand for the model error, which is not captured by the normal distribution. Six cases with the values of 1%, 5%, and 10% for E are chosen in this scenario.

[32] The estimated results shown in Table 4 indicate that the source locations are correct and the errors of the estimated release concentration and period are small, even if the error level is increased to 10%. The estimated release concentration is 97.72 ppm, with a relative error of −2.28%, when the error is normally distributed. On the other hand, the release concentration is estimated as 96.31 ppm if the error is considered uniformly distributed in case 3.6. These results indicate that the proposed approach is applicable even if the sampling concentrations contain the error levels up to 10%.

4.2.4. Number of Monitoring Points

[33] Five cases with the number of sampling points from three to seven are considered in the fourth scenario to explore the effect of the number of monitoring points on the results of source information estimation. Table 5 indicates that the estimated source location in these five cases is correctly identified. In addition, the obtained release concentration and release period have very good accuracy, except in case 4.1. The unknown variables involved in solving the problem of source information estimation include the source location, the released concentration, and release period. The source location involves three dimensions; that is, three coordinate values are needed to specify the source location. Thus five unknown variables are involved in source information estimation implying that five monitoring points are sufficient to determine these unknowns. One might expect that using more sampling concentrations would provide more information and be helpful in estimating the source information.

[34] Table 5 shows that the differences of the results obtained from cases 4.3 to 4.5 are very small, indicating

Table 5. Results for Five Cases Designed in the Fourth Scenario for the Examination of the Number of Monitoring Points[a]

Case	Number of Monitoring Points	Monitoring Points	Identified Results: Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-5}$	Number of Simulations
4.1	3	A2, F2, J2	(220, 540, −9)	81.86	3.45	4.52	9000
4.2	4	A2, B2, F2, J2	(220, 540, −9)	98.59	2.97	0.85	10400
4.3	5	A2, B2, F2, I2, J2	(220, 540, −9)	103.50	3.02	2.08	8400
4.4	6	A2, B2, F2, H2, I2, J2	(220, 540, −9)	102.32	3.04	14.39	8600
4.5	7	A2, B2, F2, G2, H2, I2, J2	(220, 540, −9)	99.89	3.00	0.29	9600

[a]Note that the real source is located at (220 m, 540 m, −9 m), real release concentration is 100 ppm, and real release period is 3 years.

Table 6. Results of 26 Cases Designed in the Fifth Scenario to Explore the Influence on the Number of Concentration Intervals[a]

Case	Monitoring Points	Concentration Intervals	Results				
			Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-6}$	Number of Simulations
5.1	B2, D2, F2, H2, I2	2	(180, 540, −9)	133.86	3.06	3.07	9800
5.2	A2, B2, F2, I2, J2	3	(220, 540, −9)	103.50	3.02	20.76	9000
5.3	A2, F2, H2, I2, J2	3	(260, 540, −9)	66.58	3.57	22.56	8800
5.4	A2, B2, D2, F2, J2	3	(220, 540, −9)	100.21	2.99	14.57	9400
5.5	A2, B2, D2, F2, I2	3	(220, 180, −9)	95.90	3.60	839.4	9600
5.6	A2, B2, C2, D2, F2	3	(220, 540, −15)	108.05	2.93	168.4	9800
5.7	A2, B2, F2, G2, J2	4	(220, 540, −9)	100.00	3.00	0.31	9800
5.8	A2, B2, C2, F2, I2	4	(220, 540, −9)	100.19	3.00	0.95	8800
5.9	A2, B2, C2, F2, J2	4	(220, 540, −9)	99.68	3.01	19.6	8800
5.10	A2, B2, C2, G2, I2	4	(220, 540, −9)	100.04	3.01	11.61	8600
5.11	A2, C2, F4, I1, J3	4	(180, 540, −9)	123.16	3.24	5.98	9000
5.12	A2, E2, F4, I1, J3	4	(220, 540, −9)	87.26	3.38	0.76	8800
5.13	A2, C2, F2, G2, J2	5	(220, 540, −9)	98.28	2.97	58.65	8800
5.14	A2, B2, F2, H2, I2, J2	3	(180, 540, −3)	96.97	3.59	10.99	9000
5.15	A2, B2, D2, F2, I2, J2	3	(220, 540, −3)	99.08	3.07	20.4	9000
5.16	A2, B2, D2, F2, G2, I2	4	(220, 540, −9)	99.92	3.00	0.06	9800
5.17	A2, B2, E2, F2, H2, I2	4	(220, 540, −9)	102.87	3.03	0.59	9400
5.18	A2, B2, D2, E2, F2, I2	4	(220, 540, −9)	103.25	3.03	0.45	8600
5.19	A2, B2, F2, G2, I2, J2	4	(220, 540, −9)	99.57	3.00	0.10	8800
5.20	A2, B2, C2, F2, I2, J2	4	(220, 540, −9)	99.17	2.99	1.21	8600
5.21	A2, B2, E2, F2, I2, J2	4	(220, 540, −9)	102.32	3.04	14.44	8600
5.22	A2, B1, F4, G2, I1, J3	4	(220, 540, −9)	99.86	3.00	0.11	9400
5.23	A2, B1, C2, F4, I1, J3	4	(220, 540, −9)	99.60	3.00	0.22	8200
5.24	A2, B1, E2, F4, I1, J3	4	(220, 540, −9)	100.27	3.01	6.56	7600
5.25	A3, B3, C2, D1, F1, G2	4	(220, 540, −9)	99.86	3.00	0.11	9400
5.26	A3, B4, D1, F1, G2, I2	4	(220, 540, −9)	105.00	3.03	0.11	7800

[a]Note that the real source is located at (220 m, 540 m, −9 m), real release concentration is 100 ppm, and real release period is 3 years.

that the sixth and seventh monitoring points employed in cases 4.4 and 4.5 provide very little improvement in the results. Note also that in case 4.1 the number of the monitoring points is three, which is insufficient to determine the problem with five unknowns. The relative error for the estimated release concentration is −18.14% and for the estimated release period is 15%, although the estimated source location is correct. Accordingly, there should be at least five monitoring points for solving the source information estimation problem in a homogeneous and isotropic site.

4.2.5. Number of Concentration Intervals

[35] Twenty six cases are proposed to explore the influence of the magnitude of the sampling concentrations on the estimation result. In cases 5.1 to 5.13, five monitoring points are chosen but the locations of the selected monitoring points are varied and six monitoring points are used to estimate the pumping source in cases 5.14 to 5.26. Table 1 shows that the sampling concentrations vary from 0.0815 to 0.4877 ppm for the real source located at (220 m, 540 m, −9 m). Such a concentration distribution may be divided into five different intervals with a concentration difference of 0.1 ppm. The sampling concentration in the range between 0 ppm and 0.1 ppm is the first concentration interval and the sampling concentration ranging from 0.4 ppm to 0.5 ppm is the fifth concentration interval. Table 6 lists the locations of the monitoring points and the number of concentration intervals, which are counted on the basis of the sampling concentrations at those five monitoring points. Two to five intervals are considered in these 26 case studies.

[36] Table 6 indicates that the estimated source information obtained from cases 5.1, 5.3, 5.5, 5.6, 5.11, 5.12, 5.14, and 5.15 are incorrect. In the first six cases, the number of monitoring points is five, but the number of sampling concentration intervals is equal to or less than three in these eight cases. In cases 5.14 and 5.15, the number of monitoring points is increased to six, however, the obtained source information is still incorrect. On the other hand, if the number of concentration intervals is equal or more than four, the obtained results are excellent. Four or more concentration intervals portray better and clear features for representing the spatial concentration distribution.

[37] In cases 5.16 to 5.21, there are six monitoring points installed at different locations. The purpose of these six cases is to study the influence if four concentration intervals are chosen from different sets of monitoring points. Table 6 shows that the identified source locations are correct and the estimated release concentration and release period are of good accuracy in these six cases. It is important to point out that case 5.14 turns to case 5.19 and the number of concentration intervals increases from three to four if the sampling point of H2 in case 5.14 is replaced by G2. Similarly, case 5.15 becomes case 5.18 and the number of concentration intervals increases from three to four if the sampling point of J2 in case 5.15 is replaced by E2. The source information estimation fails in cases 5.14 and 5.15 but succeeds in cases 5.18 and 5.19.

[38] In cases 5.22 to 5.26, case 5.22 has the same sampling location in x and y coordinates as case 5.19 but differs in depth. Cases 5.20 and case 5.23 as well as cases 5.21 and 5.24 are similarly arranged. In cases 5.22−5.24, the monitoring points installed at the first, third, and fourth layers have the same well locations, that is, wells B, J, and F. However, the monitoring points in cases 5.25 and 5.26 are installed at depths differing from cases 5.22 to 5.24. Table 6 shows that the monitoring points in these five cases,

Table 7. Results of Eight Cases Designed in the Seventh Scenario for the Examination of Source Depths[a]

Case	Real Source Location, m	Monitoring Points	Identified Results				
			Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-5}$	Number of Simulations
6.1	(220,540, −3)	A2, B1,C2, F4, I1, J3	(220, 540, −3)	99.18	3.00	0.61	8600
6.2	(220,540, −9)	A2, B1,C2, F4, I1, J3	(220, 540, −9)	100.18	3.01	10.27	9000
6.3	(220,540, −15)	A2, B1,C2, F4, I1, J3	(220, 540, −15)	84.60	3.50	8.34	9400
6.4	(220,540, −21)	A2, B1,C2, F4, I1, J3	(220, 540, −21)	101.70	3.00	1.84	7000
6.5	(220,540, −3)	A2, B2,C2, F2, I2, J2	(220, 540, −15)	107.50	2.94	150.6	9200
6.6	(220,540, −9)	A2, B2,C2, F2, I2, J2	(220, 540, −9)	99.17	2.99	12.07	8600
6.7	(220,540, −15)	A2, B2,C2, F2, I2, J2	(220, 540, −21)	110.50	2.92	92.33	8800
6.8	(220,540, −21)	A2, B2,C2, F2, I2, J2	(220, 540, −21)	99.46	3.00	0.14	9600

[a]Note that the real release concentration is 100 ppm and real release period is 3 years.

cases 5.22–5.26, are distributed at four different depths of the aquifer and the identified source locations are correct and the estimated release concentration and release period have good accuracy. Note that the case 5.11 changes to case 5.23, the number of monitoring points increases from five to six if case 5.11 has an additional sampling point B1. Likewise, case 5.12 becomes case 5.24, the number of monitoring points increases from five to six if the sampling point B1 is added to case 5.12. The source information estimation fails in cases 5.11 and 5.12 but succeeds in cases 5.23 and 5.24. In summary, the obtained source information is correct if six monitoring points with four concentration intervals are used to estimate the source information.

[39] Some researchers have considered the issues of determining the number of monitoring wells or solving the problem of monitoring network design, although the problems they dealt with were different from those of this study. For example, in the work of *Mayer et al.* [1994] and *Storck et al.* [1997] the design of monitoring networks was to provide initial detection of contamination from landfill, while in the work of *Mahar and Datta* [1997] the solution for monitoring networks was to find optimal well locations for different number of monitoring wells. Comparatively, this study provides a suggestion based on the results of case studies with different number and location of the monitoring points for effective estimating the source information.

4.2.6. Depth of Source and Well Allocation

[40] Table 1 shows the sampling concentrations obtained from the 10 monitoring wells A to J for the real source when located at the same x and y coordinates but different z coordinate. In the seventh scenario, eight cases, including four real sources located at (200, 500) in x and y coordinates and at depths of −3 m, −9 m, −15 m, and −21 m and two sets of monitoring well systems with monitoring point varied from well to well, are considered. These eight cases aim to explore the effect of the source depth and well allocation on the results of source identification. In the first four cases, these four real sources are considered for source identification with the same five monitoring points installed at various depths as listed in Table 7. In the other four cases, these four real sources are considered again but the monitoring points are installed at the same depths. The depths of the real source and the monitoring points for each case and their analyzed results are demonstrated in Table 7.

[41] Table 7 shows that the estimated source locations obtained in cases 6.1 to 6.4 are all correct and the release concentration and release period also agree with the real solution. This indicates that the proposed approach and the suggestion for effectively estimating source information both work well when the monitoring points are installed at different depths even if the source location is unknown. However, in cases 6.5 to 6.8, while all the monitoring points are installed at the same depth, only the cases 6.6 and 6.8 get good results. Therefore we suggest that it is better to install the monitoring points at different depths to obtain reliable identification results if the depth of the real source is not known. On the other hand, good results can be obtained if the monitoring points are installed at the same depth as the known source.

5. Possible Real-World Problems

[42] Aquifers in the real world may be heterogeneous and with irregular boundary conditions. Thus three scenarios, scenarios 7–9, are considered and the SATS-GWT is again used to estimate the source information for more complicated problems. In scenario 7 the aquifer formations are considered as heterogeneous. In scenario 8 the suspected area is much larger than those used in scenarios 1 to 7. Case studies given in scenarios 1 thorough 8 all assume that the aquifer boundary is regular and flow is in steady state. However, in scenario 9, the aquifer is heterogeneous with irregular boundaries and there is transient flow.

5.1. Heterogeneous Aquifer

[43] The heterogeneous aquifers used in scenarios 7 to 9 have random hydraulic conductivity fields that are spatially correlated and lognormally distributed. The heterogeneous nature of aquifers is usually characterized by three statistical parameters: the mean hydraulic conductivity, the variance in hydraulic conductivity, and the correlation length. Field aquifer tests such as slug test or pumping test is usually used to determine hydrogeologic parameters, that is, hydraulic conductivity and storativity. These aquifer parameters obtained at specific locations are used as conditional information for heterogeneous analysis study. Figure 3 shows the studied heterogeneous aquifer system, with a domain of 1000 m by 1000 m, a real source S1, eight suspected sources near S1, and monitoring wells A to F. The correlation lengths (λ) in both x and y coordinates are chosen as 100 m and the mean and standard deviation of the natural logarithm of the hydraulic conductivity ($\ln K$) are 2.7 and 0.5, respectively. The mean and standard deviation of $\ln K$ are respectively denoted as $\bar{y}$ and σ_y where $y = \ln K$. Assume that the hydraulic conductivities at wells A, B, E, F,

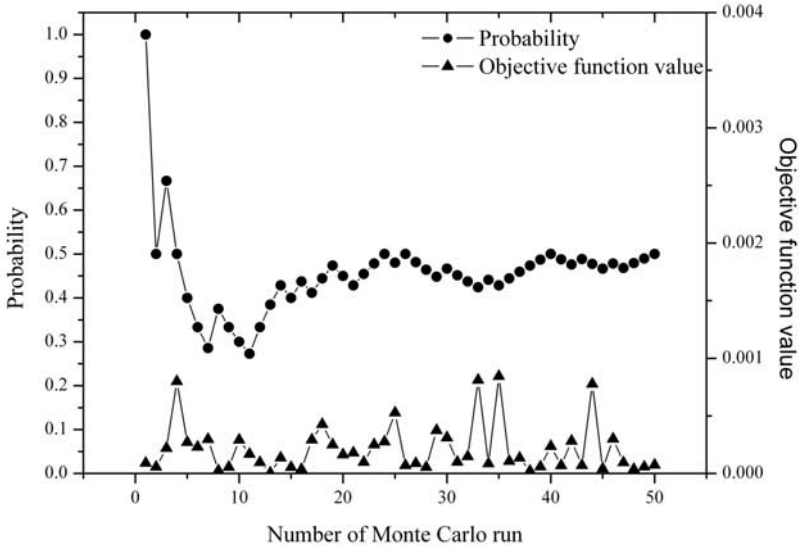

Figure 4. Fifty Monte Carlo runs and the probability of obtaining correct results when the conductivity fields are heterogeneous in scenario 8.

I, and J obtained by the aquifer test are 13.902, 14.069, 7.486, 14.629, 35.766, and 14.168 m/day, respectively.

[44] The program SASIM of the geostatistical software, GSLIB [*Deutsch and Journel*, 1998, p. 183] can produce spatially correlated conductivity fields, while preserving the known mean and variance of ln$K$ and known conductivity values at specific locations. Thus it is chosen to generate study hydraulic conductivity fields for heterogeneous aquifers with $\overline{y}$ = 2.7 m/day, σ_y = 0.5 m/day, λ = 100 m, and known conductivities at wells A, B, E, F, I, and J.

[45] The MODFLOW-GWT uses the produced conductivity field and the same hydrogeologic conditions given in section 4 to generate the sampling concentrations. Six concentrations obtained at the monitoring points of A2, B1, E2, F4, I1, and J3 are 0.5684, 0.2935, 0.1266, 0.3172, 0.2294, and 0.2548 ppm, respectively. Note that the initial guess location for the source is (260 m, 500 m, −21 m) and the initial guess values for the release concentration and release period are respectively 50 ppm and 1.5 years. The upper bound values of release concentration and release period are respectively 200 ppm and 6 years in these case studies. Those data are used in scenarios 8 to 10 for the source information estimation.

5.2. Monte Carlo Simulation

[46] The seventh scenario uses Monte Carlo simulation to test the validity of the suggestion when applying the proposed approach for the identification of contaminant source in a heterogeneous and isotropic aquifer. The Monte Carlo method is adopted and the SASIM is used again to produce 50 realizations of random conductivity fields with the same conditioning conductivity data and statistical parameters as the previous one used to generate the sampling concentrations. The SATS-GWT is then employed along with those six sampling concentrations to estimate the source information for an aquifer with each of 50 realizations of conditioning random conductivity field. Notice that in this scenario, the random conductivity fields used in identifying the source information differ from those used in generating contaminant concentrations. The field conductivities are generally unknown and only few values are known at some locations from slug test or pumping test, which are treated as conditioning data in the investigation site. The results of source estimation for those 50 realizations are shown in Figure 4, where the horizontal and vertical axes respectively represent the number of Monte Carlo runs and the probability of obtaining the correct source location. The probability is defined as the number of the correct solutions divided by the total number of Monte Carlo runs. The chance of obtaining a correct source location is 0.0278, that is, one out of 36 (3 rows × 3 columns × 4 layers) if the chance of each candidate source to be selected is equal. Figure 4 shows that a total of 25 runs can obtain the correct source location by the proposed approach from 50 Monte Carlo runs. In addition, the probability of obtaining the correct source location ranges between 0.4 and 0.5 after several Monte Carlo runs. Those results indicate that the SATS-GWT statistically has a high chance of finding the correct source location even when the conductivity field is heterogeneous with only few known data. It is noteworthy that those results imply that the suggestion for effectively estimating the source information is also applicable to heterogeneous aquifers.

5.3. Larger Suspected Area

[47] The eighth scenario is to test the performance of the proposed SATS-GWT applied in a heterogeneous aquifer with a larger suspected area, which has 168 candidate sources (7 rows × 6 columns × 4 layers) delineated by broken lines, as shown in Figure 3. Three cases, randomly selected from those 25 correct results of Monte Carlo runs, are used to test the performance for this enlarged suspected area. The estimated source locations listed in the Table 8 are also correct. Among those three cases, the relative errors in the release concentration are −6.72, 0.08, and 9.98% and in the release period are 26.33, 3.00, and 9.98%.

5.4. Transient Flow Study

[48] Another new hypothetical site, demonstrated in Figure 5, is given for testing the performance of the source

Table 8. Results of Three Cases Designed in the Ninth Scenario for the Examination of a Larger Suspected Area with Heterogeneous Conductivity Fields[a]

	Identified Results				
Case	Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-5}$	Number of Simulations
8.1	(220, 540, −9)	93.28	3.79	12.54	9800
8.2	(220, 540, −9)	100.08	3.09	3.18	11200
8.3	(220, 540, −9)	109.98	3.41	8.543	10400

[a]Note that the real release concentration is 100 ppm and real release period is 3 years.

information estimation by SATS-GWT in scenario 9. The aquifer geometry, hydrogeological properties, the location of the contaminant source, and the source release concentration and release rate are the same as those used in scenarios 1 to 8 given at the beginning of section 4, except that the specific storage is given as 10^{-4} for the transient state simulation. However, there are four differences between this scenario and scenarios 1 to 8. The first difference is that the boundary of the new hypothetical site is irregular. The second is that the hydraulic conductivity field is divided into three zones with hydraulic conductivities of 10, 20, and 30 m/day in zones 1 to 3, respectively. The third is that the groundwater flow is transient. During the simulation of SATS-GWT, the groundwater recharge in each year is divided into a wet period and a dry period. Each period contains six months and the wet period is the first period in

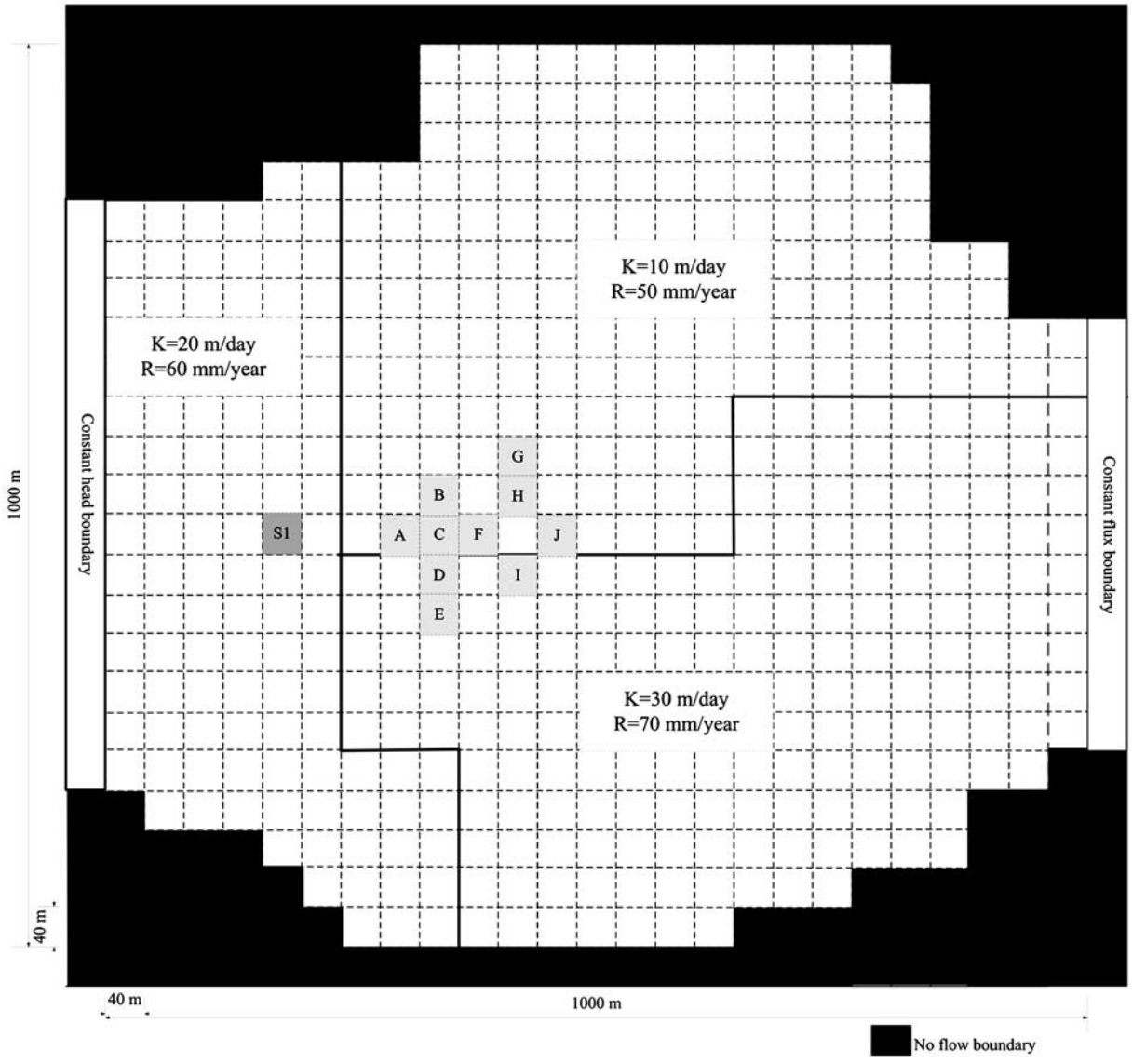

Figure 5. The plan view of the new aquifer system used in scenario 10.

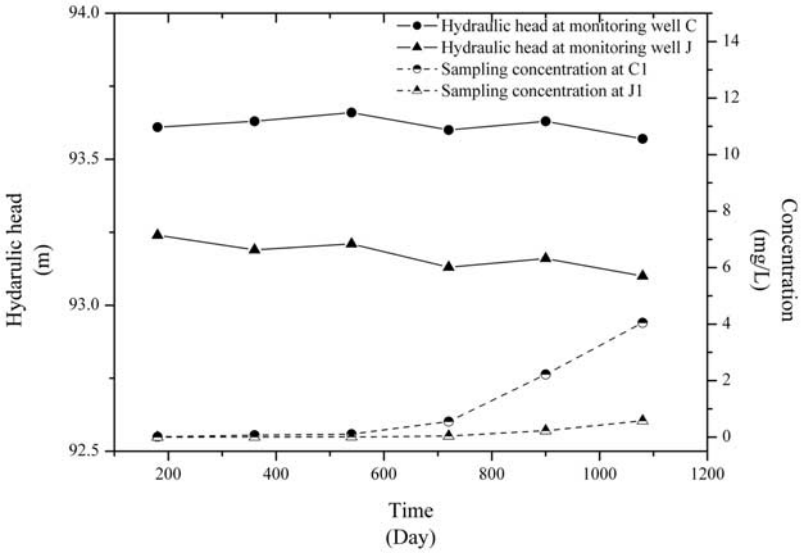

Figure 6. The variations in hydraulic heads and contaminant concentrations at points C and J.

each year. The recharge rates are all zero in the dry season and 50, 60, and 70 mm/year in the wet season during three consecutive years. The last difference is that a constant flux boundary is imposed at the left-hand side boundary with a flow rate of 1860 m^3/day. Figure 6 shows the temporal distribution of the hydraulic head and contaminant concentration at monitoring wells C and J. Figure 6 indicates that both hydraulic heads at sampling wells C and J change with time owing to the effect of wet and dry periods.

[49] Two cases for contaminant source identification are considered in this scenario. The first case considers using the contaminant concentrations sampled at A2, B1, C2, F4, I1, and J3 to estimate the source information. The concentrations at these sampling points are 8.778, 2.476, 4.016, 2.037, 3.670, and 0.572 mg/L, respectively. In the second case, A2, B2, C2, F2, I2, and J2 are the sampling points. The concentrations at these six points are respectively 8.778, 2.453, 4.106, 2.119, 3.655, and 0.579 mg/L. The source locations are correctly identified as shown in Table 9. The relative errors are −0.59 and 1.6% in the release concentration and 3.33 and −6.67% in the release period for the cases 9.1 and 9.2, respectively. The sampling points in these two case studies are selected to follow the suggestion for effectively estimating the source information. The analyzed results indicate that the suggestion is also applicable to the heterogeneous aquifer and the transient flow condition.

[50] Note that SATS-GWT takes about 5 hours to obtain the solution for homogeneous aquifers when performed on a personal computer with 2.4 G *Pentium IV* CPU and 512 MB RAM, whereas SATS-GWT requires about 7 hours for heterogeneous aquifers.

6. Conclusions and Future Recommendations

[51] A hybrid approach combining SA, TS, and a three-dimensional groundwater flow and solute transport model, MODFLOW-GWT, has been developed for solving the source identification problem. At the beginning, this approach uses TS to generate the candidate source locations and uses SA to generate the release concentration and release period at the candidate source location. Then the MODFLOW-GWT is employed to simulate the three-dimensional plume concentrations at the monitoring wells. The effects of the initial guess values and measurement errors on the results when employing the proposed approach to perform source information estimation are studied. The proposed approach is also employed to investigate the requirements for the optimal number of monitoring points and the conditions for effectively estimating source information. In addition, an effective way to optimally allocate the monitoring points for the estimation of source information is suggested on the basis of the studies for the designed cases in a homogeneous aquifer and steady state groundwater flow. Six conclusions can be drawn as follows.

[52] First, the approach we developed is capable of solving the three-dimensional groundwater source information estimation problem in both homogeneous and heterogeneous aquifers. The estimated results obtained from this study are correct, as indicated by comparison with the real solutions. Second, the identification results of the source location, release concentration, and release period are independent of the initial guess, indicating that even an inexperienced user could apply this approach to estimate the source information. Third, the effect of large upper bounds on the accuracy of the source information estimation is insignificant. In contrast, a small upper bound may give

Table 9. Results of Two Cases Designed in the Tenth Scenario for the Examination of the Transient Flow Condition and Heterogeneous Conductivity Fields[a]

	Identified Results				
Case	Source Location, m	Release Concentration, ppm	Release Period, years	Objective Function Value, $\times 10^{-5}$	Number of Simulations
9.1	(220, 540, −9)	99.41	3.1	2.53	10000
9.2	(220, 540, −9)	101.60	2.98	4.13	9400

[a]Note that the real release concentration is 100 ppm and real release period is 3 years.

wrong results if the real source is located outside the upper bound. Fourth, it is found that at least six monitoring points with four concentration intervals are required for estimating the source information. The finding is initially concluded from the case studies for homogeneous and isotropic aquifers and a point source with a uniform release concentration over a finite duration. However, it has also been demonstrated to be applicable for estimating a single source with uniform release concentration in heterogeneous aquifers under steady or transient flow conditions. Fifth, if the depth of the real source is not known, we suggest installing the monitoring points at different depths to obtain better identification results. Finally, the proposed SATS-GWT and the suggestion for estimating the source information statistically are also applicable to heterogeneous aquifers and larger suspected area.

[53] The key objectives of this study are to develop an approach for estimating the location of a contaminant source and its release concentration and release period, and to give a suggestion for optimally allocating the monitoring points in the estimation of that source information. A constant concentration of the contaminant is assumed to be released from a single point source. If a real-world problem involves an unknown source with a nonuniform release history or the source is considered to be of arbitrary size, the problem of source information estimation will then become complicated. Such a problem may be solved using the heuristic approach we used, yet it is beyond the scope of this article and deserves future study. In addition, this study uses a grid size of 40 m in spatial discretization, which may be too coarse for modeling a heterogeneous aquifer or predicting accurate concentration in the suspected area. Under this circumstances, a finer grid size or nonuniform grid may be adopted for a better source information estimation.

[54] **Acknowledgments.** This study was partly supported by the Taiwan National Science Council under the grant NSC94-2211-E-009-012. The authors thank Yi-Tsung Liu for providing several simulation results that were helpful in determining the number of concentration intervals. In addition, the authors also thank the Associate Editor Neupauer and three anonymous reviewers for their constructive comments and suggested revisions.

References

Aral, M. M., and J. Guan (1996), Genetic algorithms in search of groundwater pollution sources, in *Advances in Groundwater Pollution Control and Remediation*, pp. 347–369, Springer, New York.

Aral, M. M., J. Guan, and M. L. Maslia (2001), Identification of contaminant source location and MODFLOW-GWT release history in aquifers, *J. Hydrol. Eng.*, *6*(3), 225–234.

Atmadja, J., and A. C. Bagtzoglou (2001), State of the art report on mathematical methods for groundwater pollution source identification, *Environ. Forensics*, *2*, 205–214.

Bagtzoglou, A. C., D. E. Kougherty, and A. F. B. Tompson (1992), Applications of particle methods to reliable identification of groundwater pollution sources, *Water Resour. Manage.*, *6*, 15–23.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, 2nd ed., Oxford Univ. Press, New York.

Glover, F. (1986), Future path for integer programming and links to artificial intelligence, *Comput. Oper. Res.*, *13*, 533–549.

Gorelick, S. M., B. Evans, and I. Remson (1983), Identifying sources of groundwater pollution: An optimization approach, *Water Resour. Res.*, *19*(3), 779–790.

Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), MODFLOW-2000: The U.S. Geological Survey Modular Ground-Water Model–User guide to modularization concepts and the ground-water flow process, *U.S. Geol. Surv. Open File Rep., 00–92*.

Hwang, J. C., and R. M. Koerner (1983), Groundwater pollution source identification from limited monitoring well data: part I, Theory and feasibility, *J. Hazard. Mater.*, *8*, 105–119.

Konikow, L. F., D. J. Goode, and G. Z. Hornberger (1996), A three-dimensional method of characteristics solute-transport model (MOC3D), *U.S. Geol. Surv. Water Resour. Invest. Rep., 96–4267*.

Liu, C., and W. P. Ball (1999), Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware, *Water Resour. Res.*, *35*(7), 1975–1985.

Mahar, P. S., and B. Datta (1997), Optimal monitoring network and groundwater pollution source identification, *J. Water Resour. Plann. Manage.*, *123*(4), 199–207.

Mahar, P. S., and B. Datta (2000), Identification of pollution sources in transient groundwater systems, *Water Resour. Manage.*, *14*(3), 209–227.

Mahar, P. S., and B. Datta (2001), Optimal identification of groundwater pollution sources and parameter estimation, *J. Water Resour. Plann. Manage.*, *127*(1), 20–29.

Mahinthakumar, G., and M. Sayeed (2005), Hybrid genetic algorithm-local search methods for solving groundwater source identification inverse problem, *J. Water Resour. Plann. Manage.*, *131*(1), 45–57.

Mahinthakumar, G., and M. Sayeed (2006), Reconstructing groundwater source release histories using hybrid optimization approaches, *Environ. Forensics*, *7*(1), 45–54.

Mayer, P. D., A. J. Valocchi, and J. W. Eheart (1994), Monitoring network design to provide initial detection of groundwater contamination, *Water Resour. Res.*, *30*(9), 2647–2659.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*(6), 1087–1092.

National Research Council (1990), *Groundwater Models: Scientific and Regulatory Applications*, Natl. Acad. Press, Washington, D. C.

Neupauer, R. M., and J. L. Wilson (1999), Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant, *Water Resour. Res.*, *35*(11), 3389–3398.

Neupauer, R. M., and J. L. Wilson (2001), Adjoint-derived location and travel time probabilities for a multi-dimensional groundwater system, *Water Resour. Res.*, *37*(6), 1657–1668.

Neupauer, R. M., B. Borchers, and J. L. Wilson (2000), Comparison of inverse methods for reconstructing the release history of a groundwater contamination source, *Water Resour. Res.*, *36*(9), 2469–2475.

Pham, D. T., and D. Karaboga (2000), *Intelligent Optimisation Techniques*, Springer, New York.

Rayward-Smith, V. J., I. H. Osman, C. R. Reeves, and G. D. Smith (1996), *Modern Heuristic Search Methods*, John Wiley, New York.

Romeo, F., and A. Sangiovanni-Vincentelli (1991), A theoretical framework for simulated annealing, *Algorithmica*, *6*(3), 302–345.

Sayeed, M., and G. Mahinthakumar (2005), Efficient parallel implementation of hybrid optimization approaches for solving groundwater inverse problems, *J. Comput. Civ. Eng.*, *19*(4), 329–340.

Sciortino, A., T.-C. Harmon, and W. W.-G. Yeh (2000), Inverse modeling for location dense nonaqueous pools in groundwater under steady flow conditions, *Water Resour. Res.*, *36*(7), 1723–1735.

Skaggs, T. H., and Z. J. Kabala (1994), Recovering the release history of a groundwater contaminant, *Water Resour. Res.*, *30*(1), 71–79.

Skaggs, T. H., and Z. J. Kabala (1995), Recovering the history of a groundwater contaminant plume: Method of quasi-reversibility, *Water Resour. Res.*, *31*(11), 2669–2673.

Skaggs, T. H., and Z. J. Kabala (1998), Limitations in recovering the history of a groundwater contaminant plume, *J. Contam. Hydrol.*, *33*, 347–359.

Snodgrass, M. F., and P. K. Kitanidis (1997), A geostatistical approach to contaminant source identification, *Water Resour. Res.*, *33*(4), 537–546.

Storck, P., J. W. Eheart, and A. J. Valocchi (1997), A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional heterogeneous aquifers, *Water Resour. Res.*, *33*(9), 2081–2088.

Tung, C. P., and C. A. Chou (2004), Pattern classification with tabu search to identify spatial distribution of groundwater pumping, *Hydrogeol. J.*, *12*(5), 488–496.

Tung, C. P., C. C. Tang, and Y. P. Lin (2003), Improving groundwater flow modeling using optimal zoning methods, *Environ. Geol.*, *44*, 627–638.

Visual Numerics Inc. (2003), *IMSL Fortran Library User's Guide: Stat/Library*, vol. 2, Houston, Tex.

Wagner, B. J. (1992), Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modeling, *J. Hydrol.*, *135*, 275–303.

Woodbury, A. D., and T. J. Ulrych (1996), Minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant, *Water Resour. Res.*, *32*(9), 2671–2681.

Woodbury, A., E. Sudicky, T. J. Ulrych, and R. Ludwig (1998), Three-dimensional plume source reconstruction using minimum relative entropy inversion, *J. Contam. Hydrol.*, *32*, 131–158.

Zheng, C., and P. Wang (1996), Parameter structure identification using tabu search and simulated annealing, *Adv. Water Res.*, *19*(4), 215–224.

T.-H. Chang, Y.-C. Lin, and H.-D. Yeh, Institute of Environmental Engineering, National Chiao-Tung University, No. 75, Po-Ai Street, Hsinchu 300, Taiwan. (hdyeh@mail.nctu.edu.tw)