HYDROLOGICAL PROCESSES
Hydrol. Process. **22**, 3010–3019 (2008)
Published online 23 January 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.6875

Identifying groundwater pumping source information using simulated annealing

Yu-Chung Lin and Hund-Der Yeh*
Institute of Environmental Engineering, National Chiao Tung University, No.75, Po-Ai Street, Hsinchu City 300, Taiwan

Abstract:

This study develops an approach referred to as SA-MF using simulated annealing (SA) and a three-dimensional groundwater flow model (MODFLOW-2000) to determine the pumping source information, namely the pumping source location, the pumping rate and the pumping period. Eight scenarios, each with various cases for aquifers with homogeneous or heterogeneous conductivities, are used to test the applicability and accuracy of SA-MF in determining the pumping source information. The results show that at least five measured heads should be used to analyse the pumping source identification problem and at least three observation wells are required to effectively determine the pumping source information. The SA-MF gives good estimated results in both synthetic and real problems even if the measured heads contain measurement errors. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS groundwater; inverse problem; heuristic approach; pumping source identification; simulated annealing

Received 13 November 2006; Accepted 18 July 2007

INTRODUCTION

Groundwater is an ideal freshwater resource for human usage. The merits of using groundwater include convenient availability near the point of use, excellent quality and relatively low cost of development (Todd and Mays, 2005). The groundwater resource reveals its important significance as there is a shortage of surface freshwater for domestic and industrial uses as well as for irrigating agriculture which is the main consumer of groundwater. Effective and appropriate management strategy of groundwater usage can provide fresh groundwater for human uses and environmental functions (Loáiciga, 2004). In Taiwan, the groundwater resource is not allowed to be pumped for private use without obtaining a governmental permit as the groundwater resource is a public property. Illegal pumping may have adverse effects on the groundwater flow system. Effects may include seawater intrusion into coastal aquifers, land subsidence, upwelling of poor-quality groundwater into fresh-water aquifers, decreased base flow and spring flow to support riparian and aquatic ecosystems, and aquifer dewatering and increased pumping cost (Loáiciga, 2004). Moreover, illegal pumping may also disturb the quality of the groundwater in a groundwater contamination site, change the groundwater flow pattern or agitate the capture zone at an ongoing remediation site. There is therefore a need to develop a methodology for the determination of unknown pumping source information from a practical viewpoint.

Bear (1979) indicated that the groundwater management problem could be solved by incorporating a simulation model with an optimization program. The groundwater management problems could be classified into two categories (Todd and Mays, 2005). The first is the hydraulic management problems that are aimed at managing pumping and recharge. The second is the policy evaluation problems that consider the economics of water allocations. Traditionally, linear programming, nonlinear programming or a Newton-type approach was often employed to solve the optimization problem. It is difficult to find optimal solutions for large-scale combinatorial optimization problems using gradient-type approaches or to resolve parameter estimation problems with complex search spaces (Cai *et al*., 2001). In general, two major difficulties obstruct inexperienced users. The first is that the gradient-type approaches might converge to a local solution nearest to the starting point if an improper initial guess was made (Cai *et al*., 2001). In addition, the gradient-type approaches need to compute the derivatives with respect to the decision variables and iteratively update the solution to obtain the optimal solution. The second is that those derivatives may be difficult to calculate analytically or numerically in the highly nonlinear and non-convex optimization problems (Shieh and Peralta, 2005).

Stochastic optimization methods such as genetic algorithms (GAs) and simulated annealing (SA) are global optimization solvers and have been widely used recently. Differing from the gradient-type approaches, these global optimization methods iteratively renew the trial solution by the objective function value (OFV) in determining the nearby global optimum. The first advantage of using global optimization methods is that the users don't

* Correspondence to: Hund-Der Yeh, Institute of Environmental Engineering, National Chiao Tung University, No.75, Po-Ai Street, Hsinchu City 300, Taiwan.
E-mail: hdyeh@mail.nctu.edu.tw

require much experience in providing the initial guesses for solving nonlinear problems. The initial guesses can be arbitrarily given or even generated by a random number generator, and still achieve optimal results. The second advantage is that these global optimization methods do not need to calculate the derivates to renew the trial solution.

SA is a random search algorithm that allows, at least in theory or in probability, to obtain the global optimum of a function in a given domain. The philosophy of SA is to imitate the physical annealing process of heating up a solid and then cooling it down slowly until it crystallizes. The temperature should be reduced properly to minimize the energy state and crystallize a regular structure. Comparatively, if the cooling process occurs too fast, the crystallized structure becomes irregular and its energy does not stay in minimum. Romeo and Sangiovanni-Vincentelli (1991) used homogeneous and inhomogeneous Markov chain theories to prove that SA can converge to global optimal solutions.

SA evolved from the descent search method, which has the drawback of obtaining the solution in a local optimum (Rayward-Smith *et al.*, 1996). Basically, SA starts with a single initial guess named as the current optimum. Next, a trial solution neighboured to the current optimum is randomly generated and its corresponding OFV is then calculated. Finally, the nearby global optimum is obtained by improving the current optimum iteratively based on the OFV. To avoid obtaining a local optimum, SA uses the Metropolis mechanism to control which ascent moves, i.e. inferior solution, could be accepted as a current optimum (Rayward-Smith *et al.*, 1996). The Metropolis mechanism is a criterion based on Boltzman's probability for the difference between the OFVs of current optimum and trial solution at a given state (Pham and Karaboga, 2000). Therefore, the SA has a property of going descent but allowing random ascent to avoid a possible trap in a local optimum, which prevents the SA from having the same problem as the descent method.

SA was successfully applied in a wide range of optimization applications such as capacity extension for pipe network system (Cunha and Sousa, 1999; Monem and Namdarian, 2005), parameter calibration and identification problems (Zheng and Wang, 1996; Cooper *et al.*, 1997; Li *et al.*, 1999; Gou and Zheng, 2005; Lin and Yeh, 2005; Yeh *et al.*, 2007), groundwater management problems (Doutherty and Marryott, 1991; Marryott *et al.*, 1993), groundwater remediation system problems (Kuo *et al.*, 1992; Marryott, 1996; Rizzo and Dougherty, 1996; Tung *et al.*, 2003; Shieh and Peralta, 2005) and source identification of groundwater contamination problems (Aral and Guan, 1996; Aral *et al.*, 2001; Mahinthakumar and Sayeed, 2005, 2006; Sayeed and Mahinthakumar, 2005). Shieh and Peralta (2005) mentioned that the straightforward formulation, no requirement for computing derivatives and easy implementation with groundwater simulation models, are the main advantages of employing SA to solve the optimization problem.

The aim of this work is to propose an approach referred to as SA-MF based on simulated annealing (SA) and a three-dimensional groundwater flow model (MODFLOW-2000) to determine the pumping source information, namely the pumping source location, pumping rate and pumping period, in a groundwater flow system. Three synthetic problems are designed to test the performance of the proposed approach. The problems include a hypothetical pumping well which is assumed at a specific location and withdraws the groundwater at a specified flow rate for a period of time. The temporal distribution of water level (referred to as measured hydraulic head or simply measured head) at the observation wells can be either observed at the field for real problems or simulated by MODFLOW-2000 for testing the performance of SA-MF. In the estimation of pumping source information, the trial solutions for the pumping source location, pumping rate and pumping time are generated by SA. The trial solutions are used as input to MODFLOW-2000 to generate the simulated heads. The pumping source information can then be determined by SA-MF when minimizing the sum of square errors between the measured heads and simulated heads at the observation wells.

This study is organized into six sections. Following this introduction, brief illustrations for the groundwater flow model, SA and a detailed description for SA-MF are given in the second section. Sections 3–5 contain the problems of a homogeneous aquifer with four scenarios, heterogeneous aquifers with three scenarios and a real aquifer with one scenario to test the performance of SA-MF and related requirement in the estimation of pumping source information. Finally, we summarize the findings and offer concluding remarks in the sixth section.

METHODOLOGY

Groundwater Flow Model

The three-dimensional equation describing the groundwater flow can be expressed as (Harbaugh *et al.*, 2000)

$$\frac{\partial}{\partial x_i}\left(K_{ij}\frac{\partial h}{\partial x_i}\right) + W = S_s\frac{\partial h}{\partial t} \tag{1}$$

where h is the hydraulic head (l), K_{ij} is the hydraulic conductivity tensor (l T^{-1}), S_s is the specific storage (l^{-1}), W is the volumetric flux per unit volume representing sources and/or sink (negative for flow out and positive for flow in (l T^{-1})), x_i are the Cartesian coordinates, and t is time (T).

Equation (1), together with the flow and/or head conditions at the boundaries of an aquifer system and initial head condition, constitutes the mathematical model for a groundwater flow system. The computer model MODFLOW-2000 developed based on Equation (1) can be used to simulate the head distribution for a given aquifer system. MODFLOW-2000, an extended version of MODFLOW, can simulate steady state and transient

DOI: 10.1002/hyp

flow in the flow system with an aquifer which is confined, unconfined or a combination of these (Harbaugh *et al*., 2000).

Simulated annealing

Similar to Newton's methods, the SA starts with an initial guess x_k and calculates the objective function value (OFV) $f(x)$ which is treated as the current optimum. Then one trial solution x'_k is generated near the current optimal solution to update the optimal solution based on the OFV, $f(x'_k)$. In minimization problems, if $f(x'_k) < f(x_k)$ the current optimal solution takes place with trial solution x'_k. If $f(x'_k) \geq f(x_k)$, one has to use the Metropolis criterion to determine the acceptance of the trial solution. The Metropolis criterion is given as (Pham and Karaboga, 2000):

$$P_{\mathrm{SA}} = \left\{ \begin{array}{ll} 1, & f(x'_k) \leq f(x_k) \\ \exp\left(\dfrac{f(x_k) - f(x'_k)}{Te}\right), & f(x'_k) > f(x_k) \end{array} \right\} \tag{2}$$

where P_{SA} is the acceptance probability of the trial solution and Te, a control parameter, is the current temperature. If P_{SA} is greater than the random number generated from a uniform distribution and distributed between 0 and 1, then the trial solution is accepted as the new current optimal. Otherwise, the trial solution is rejected. The algorithm will be iteratively applied and terminated when the obtained solution satisfies the stopping criterion. In general, two stopping criteria are used in SA. One is to check whether the total objective function evaluations exceed the specified maximum number of iterations. This criterion is applied to avoid the waste of large computer time if the optimal solution is too far to obtain. The second is to check whether the current optimal OFV is small enough to be considered a minimum.

SA-MF implementation

Figure 1 depicts a flowchart describing the eleven steps of SA-MF. The first step is to initialize random initial solutions within the specified solution domains. The second step is to generate the simulated heads by MODFLOW-2000 with the estimated solutions. The objective function is defined as

$$\text{Minimize } f = \frac{1}{n_{\mathrm{well}} \times n_{\mathrm{step}}} \sum_{m=1}^{n_{\mathrm{step}}} \sum_{l=1}^{n_{\mathrm{well}}} (h_{lm,\mathrm{sim}} - h_{lm,\mathrm{mes}})^2 \tag{3}$$

where n_{step} is the number of the measured heads, n_{well} is the number of observation wells, $h_{lm,\mathrm{sim}}$ is the simulated head at the mth time step at the lth observation well, $h_{lm,\mathrm{mes}}$ is the measured head at the mth time step and the lth observation well. Note that the initial solutions are considered as the current optimal solutions.

The third step is to generate a new trial solution for source location x'_k given as

$$x'_k = x_k + \mathrm{Integer}[(2 \times RD_1 - 1) \times VM_k] \tag{4}$$

and the trial solutions for pumping rate and pumping period are generated as

$$x'_k = x_k + (2 \times RD_2 - 1) \times VM_k \tag{5}$$

where RD_1 and RD_2 are random numbers ranging from 0–1 and produced from a uniformly distributed random generator and VM_k is a step length vector. The variable VM_k is first defined as the length of the difference between the upper and lower bounds and then automatically adjusted periodically so that half of the evaluations could be accepted as new optima in the next exploration domain. The goal here is to sample the function widely. The acceptance of the trial solutions below 0·4 indicates that the global optimum does not locate within the exploration domain. Comparatively, a high percentage of points accepted for x_i means that the relevant element of VM_k is enlarged. If a new trial solution does not fall within the solution domain, the SA-MF has to generate another. The fourth step is to simulate the hydraulic head distribution by MODFLOW-2000 with the trial solutions and calculate its corresponding OFV. With the OFV, the trial solution is checked to determine if whether or not this is a new optimum in the fifth step. If the OFV satisfies Metropolis criterion, the current optimal solution is replaced by the trial solution in the sixth step. Otherwise, the algorithm will continue generating a new trial solution in the seventh and eighth steps.

After NS steps through all considered variables, the step length vector VM_k is adjusted so that 50% of all moves are accepted in the ninth step. In other words, after $N \times NS$ function evaluations, each element of VM_k is adjusted so that half of all function evaluations are accepted. Note that N represents the number of considered variables and NS represents the number of steps at a specific temperature. After NT times through the above loop, the temperature Te is reduced in the tenth step. Accordingly, after $N \times NT \times NS$ function evaluations, the temperature is decreased by the temperature reduction factor R_{Te} even if no improvement in the optimum takes place. The new temperature is then

$$Te' = R_{Te} \times Te. \tag{6}$$

Note that the value of R_{Te} is generally chosen to be less than one (Pham and Karaboga, 2000). The temperature should be reduced properly to guarantee the obtained solution is the global optimal solution (Zheng and Wang, 1996). In the eleventh step, the identification process will be terminated when the obtained solution satisfies the stopping criterion, which is used to check whether the absolute value of the difference between the two OFVs obtained at two consecutive temperatures is less than 10^{-6}, four times in succession.

APPLICATION TO A SYNTHETIC HOMOGENEOUS AQUIFER

Four unknown variables, i.e. the source location of x and y coordinates, pumping rate and pumping period,

DOI: 10.1002/hyp

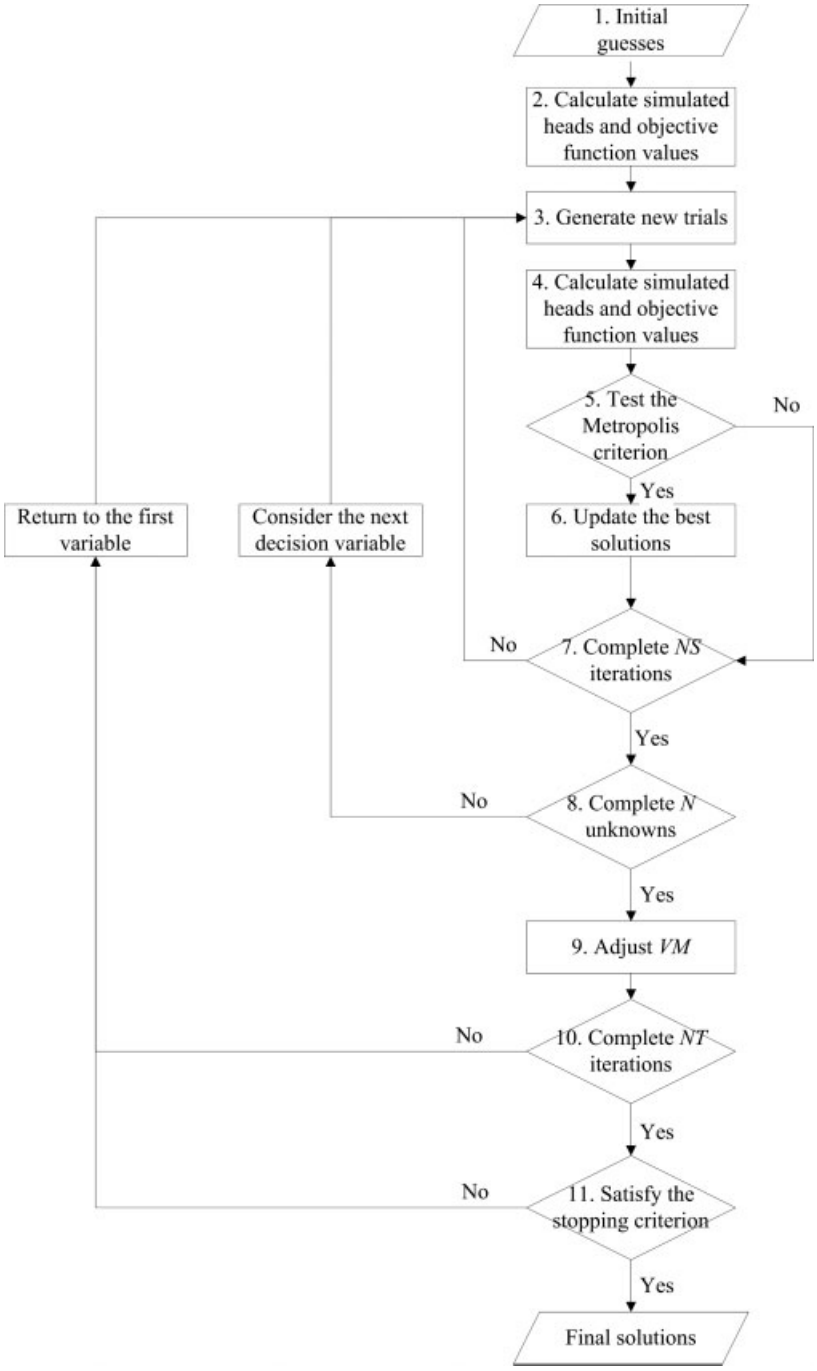

Figure 1. Flowchart depicting SA-MF processes

are considered for solving the problem of pumping source identification. Note that SA-MF determines the source location at the centre of the finite difference grid. Four scenarios are considered in this section. Scenario 1 with eight cases is designed to verify the minimum number of measured heads for effectively solving the pumping source identification problem. The purpose of scenario 2, with 30 cases, is to examine the influence of measurement error on the estimated results. Scenario 3 with eight cases is used to find an effective way to increase the accuracy of the source identification with high measurement error level. Finally, the fourth scenario with 23 cases is to determine the required number of observation wells to identify the pumping source.

Site description

A hypothetical site is used to illustrate the identification procedure for a pumping source in a homogenous and isotropic confined aquifer. The aquifer length and width are both 1000 m and the aquifer thickness is 20 m. The top and bottom elevations, hydraulic conductivity, porosity and storage coefficient are 10·0 and −10·0 m, 5×10^{-6} m s^{-1}, 0·1 and 1×10^{-4}, respectively. The finite difference grids are block-centred and the related boundary conditions for the aquifer system are shown in

DOI: 10.1002/hyp

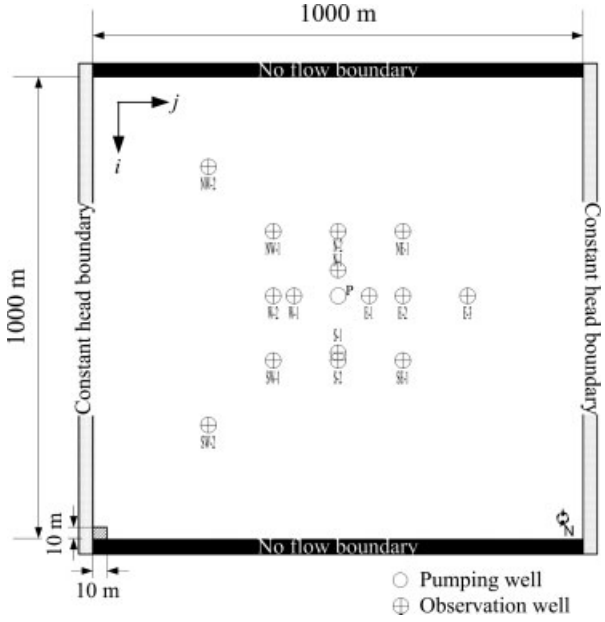

Figure 2. Plan view of hypothetical aquifer site

Figure 2. The upper and lower boundaries are specified as no-flux and the left and right boundaries are specified as constant-head. The elevations of the hydraulic head at left and right boundaries are 30·0 m and 20·0 m, respectively. The grid width and length are both 10 m; thus, the number of finite difference meshes is 100 × 100. It is assumed that the pumping well P1 is located at (495 m, 495 m) and the pumping rate and pumping period are 200·0 cubic metres per day (m^3 day^{-1}) and 60·0 min, respectively. The measured heads at 15 observation wells indicated in Figure 2 are simulated using MODFLOW-2000 and listed in Table I. Note that the drawdown at a distance of 500 m away from P1 is smaller than 0·001 m when the pumping rate is 200 m^3 day^{-1} and the pumping period is less than 170 min based on the calculation using Theis solution (Todd and Mays, 2005).

All the nodal points are considered as suspected pumping source locations except those meshes containing the observation wells. The upper bounds for the pumping rate and pumping period are set as 1000·0 m^3 day^{-1} and 120·0 min, respectively, and the lower bounds are both zero. The *NS, NT*, initial temperature and R_{Te} are chosen as 20, 10, 5 and 0·8, respectively. All initial guesses in the case studies are produced by a random number generator.

Scenario 1: Number of measured heads

The purpose of this scenario is to investigate the required number of measured heads in determining the pumping source information, namely the source location, and the pumping rates and periods, at a homogeneous and isotropic aquifer site. Mathematically, at least four measured drawdowns are needed to solve four unknowns. Eight cases, with different combinations of observation wells as indicated in Table II, are designed to study the influence of the number of measured drawdowns on the results of pumping source identification.

Table I. Location and observed hydraulic head for homogeneous and isotropic confined aquifers

Observation wells	Location	Measured heads(m)		
		Initial state	Pumping period	
			$T = 30$ (min)	$T = 60$ (min)
N-1	(465 m, 495 m)	25·041	22·621	21·275
N-2	(395 m, 495 m)	25·040	24·769	24·388
NW-1	(395 m, 395 m)	26·031	25·945	25·781
NW-2	(295 m, 295 m)	27·023	27·02	27·013
W-1	(495 m, 425 m)	25·734	25·086	24·408
W-2	(495 m, 395 m)	26·031	25·759	25·378
SW-1	(595 m, 395 m)	26·031	25·945	25·781
SW-2	(695 m, 295 m)	27·022	27·020	27·012
S-1	(585 m, 495 m)	25·040	24·680	24·216
S-2	(595 m, 495 m)	25·040	24·769	24·388
SE-1	(595 m, 595 m)	24·051	23·965	23·801
E-1	(495 m, 535 m)	24·645	22·958	21·809
E-2	(495 m, 595 m)	24·051	23·779	23·398
E-3	(495 m, 695 m)	23·062	23·043	22·995
NE-1	(395 m, 595 m)	24·050	23·964	23·800

Four measured heads are used to determine the pumping source information in cases 1·1 to 1·4 and five measured heads are employed in cases 1·5 to 1·8. The analysed results using four observations with different allocation of the observation wells are shown in Table II. The pumping locations are correctly determined at (495 m, 495 m) in cases 1·1, 1·2 and 1·4. However, the estimated location of (485 m, 505 m) in case 1·3 is incorrect. In cases 1·1, 1·2, and 1·4, the estimated pumping rates are 200·99, 216·22 and 198·92 m^3 day^{-1}, respectively, and the estimated pumping periods are 54·22, 43·98 and 59·57 min, respectively. The greatest relative errors are 8·11 % in the estimated pumping rate and 26·7 % in the estimated pumping period in case 1·2.

When the number of measured heads is increased to five, the source locations for cases 1·5 to 1·8 are all correctly predicted and the predicted pumping rates and pumping periods are all close to the correct solutions. The greatest relative errors are −1·89% in the estimated pumping rate in case 1·7 and −2·3% in the estimated pumping period in case 1·5. Obviously, one more measured head is helpful in determining pumping source information. Thus, the number of measured heads used to solve the pumping source identification problem is suggested to be at least one more than the number of unknowns.

Scenario 2: Measurement error

Scenario 2 considers that the measured head has measurement error and the disturbed measured head is expressed as

$$h'_{i,\mathrm{obs}} = h_{i,\mathrm{obs}} \times (1 + Er \times RD_3) \tag{7}$$

where $h'_{i,\mathrm{obs}}$ is the disturbed measured head, Er is the level of measurement error, and RD_3 is a random standard

DOI: 10.1002/hyp

Table II. Required number of measured heads to determine pumping source information (pumping source is located at (495 m, 495 m) and the pumping rate and pumping period are 200 m^3 d^{-1} and 60 min, respectively)

Case	Observation wells	Source location	Pumping rate (m^3 day^{-1})	Pumping period (min)
1–1[1]	N-1, W-1, S-1, E-1	(495 m, 495 m)	200·99	54·22
1–2[1]	N-2, W-2, S-2, E-2	(495 m, 495 m)	216·22	43·98
1–3[1]	NW-1, SW-1, SE-1, NE-1	(485 m, 505 m)	226·90	72·42
1–4[1]	NW-2, SW-2, E-3, E-1	(495 m, 495 m)	198·92	59·57
1–5[2]	N-1, W-1, S-1, S-2, E-1	(495 m, 495 m)	200·21	58·62
1–6[2]	N-1, N-2, W-2, S-2, E-2	(495 m, 495 m)	198·73	60·02
1–7[2]	NW-1, SW-1, SE-1, E-2, NE-1	(495 m, 495 m)	196·22	61·26
1–8[2]	NW-2, SW-2, E-2, E-3, E-1	(495 m, 495 m)	198·99	60·38

[1] Number of measured heads = 4
[2] Numbers of measured heads = 5

normal deviate generated by the routine RNNOF of IMSL (IMSL, 2003). Three different values of Er, i.e. 1, 2 and 3 cm, are chosen in this scenario and ten cases for each level of measurement error are studied. In cases 2·1 to 2·30, the wells N-1, W-1, S-1, S-2, and E-1 depicted in Figure 2 are chosen as the observation wells. Table III shows that the source location is all correctly identified when $Er = 1$ and 2 cm in cases 2·1 to 2·20. The average of estimated pumping rate and pumping period are 199·42 m^3 day^{-1} and 60·41 min, respectively. The greatest relative error is −1·24% in the estimated pumping rate and 3·72 % in the estimated pumping period in case 2·2 as $Er = 1$ cm. In addition, the average of the estimated pumping rate and pumping period are 199·24 m^3 day^{-1} and 60·64 min, respectively. The greatest relative error is −2·84% in the estimated pumping rate and 11·55 % in the estimated pumping period in case 2·20 as $Er = 2$ cm. However, when the measurement error level is increased to 3 cm, only six out of ten cases obtain the correct source location as indicated in Table III. The average of the estimated pumping rate and pumping period in those six cases are 200·09 m^3 day^{-1} and 60·20 min, respectively. The greatest relative error is 3·65 % in the estimated pumping rate and −9·28% in estimated pumping period in case 2·30 as $Er = 3$ cm.

Table III. Results for measurement with error (pumping source is located at (495 m, 495 m) and pumping rate and period are 200 m^3 day^{-1} and 60 min, respectively)

Case	Source location	Pumping rate (m^3 day^{-1})	Pumping period (min)
2–1	(495 m, 495 m)	198·42	60·54
2–2	(495 m, 495 m)	197·53	62·23
2–3	(495 m, 495 m)	200·12	60·08
2–4	(495 m, 495 m)	199·43	60·12
2–5	(495 m, 495 m)	200·63	59·93
2–6	(495 m, 495 m)	198·12	61·12
2–7	(495 m, 495 m)	198·96	60·43
2–8	(495 m, 495 m)	200·72	59·21
2–9	(495 m, 495 m)	198·43	61·73
2–10	(495 m, 495 m)	201·87	58·75
2–11	(495 m, 495 m)	197·32	61·21
2–12	(495 m, 495 m)	200·88	59·83
2–13	(495 m, 495 m)	201·32	58·76
2–14	(495 m, 495 m)	199·08	60·19
2–15	(495 m, 495 m)	203·43	56·42
2–16	(495 m, 495 m)	196·43	62·48
2–17	(495 m, 495 m)	197·83	62·13
2–18	(495 m, 495 m)	203·53	56·83
2–19	(495 m, 495 m)	198·21	61·63
2–20	(495 m, 495 m)	194·32	66·93
2–21	(495 m, 495 m)	198·50	61·42
2–22	(505 m, 505 m)	205·51	54·32
2–23	(505 m, 505 m)	199·23	60·45
2–24	(495 m, 495 m)	202·67	58·48
2–25	(495 m, 505 m)	200·68	60·43
2–26	(495 m, 495 m)	204·13	55·46
2–27	(495 m, 495 m)	193·98	66·43
2–28	(495 m, 495 m)	194·15	64·98
2–29	(505 m, 495 m)	203·12	62·30
2–30	(495 m, 495 m)	207·13	54·43

Note: Maximum error for cases 2–1 to 2–10, 2–11 to 2–20 and 2–21 to 2–30 is 1, 2 and 3 cm, respectively.

Scenario 3: Performance improvement

Based on the results in scenario 2, the estimated pumping source location may be incorrect and the pumping rate and pumping period may not be accurate if the level of measurement error is increased. Accordingly, cases 3·1 to 3·8 are designed to find an effective way to increase the accuracy of the identified result as the measurement error level is high. Cases 3·1 to 3·4, cases 3·5 and 3·6, and cases 3·7 and 3·8 use five, six and seven measured heads, respectively, to determine the pumping information. Note that all the measured heads are taken from different observation wells and are considered to have random measurement errors with $Er = 3$ cm. The estimated pumping source location, pumping rate and pumping period are correct in cases 3·1 and 3·2. However, the accuracy of the estimated results decreases with the increasing average distance from the observation wells, as indicated in cases 3·3 and 3·4. However, if the number of the measured heads is increased to six or seven, the pumping source location, pumping rate and pumping period are all correctly estimated in cases 3·5 to 3·8 indicated in Table IV. Accordingly, it is found that the best way to obtain good estimated results is to employ more measured head data in the pumping source

DOI: 10.1002/hyp

Table IV. Required number of measured heads to determine the pumping source information when the maximum measurement error is 3 cm (pumping source is located at (495 m, 495 m) and pumping rate and period are 200 m^3 day^{-1} and 60 min, respectively)

Case	Observation wells	Ave. dist. (m)	Source location	Pumping rate (m^3 day^{-1})	Pumping period (min)
3–1	N-1, W-1, S-1, S-2, E-1	66·0	(495 m, 495 m)	198·50	61·42
3–2	N-1, N-2, W-2, S-2, E-2	86·0	(495 m, 495 m)	197·26	62·38
3–3	NW-1, SW-1, SW-2, SE-1, NE-1	169·7	(505 m, 485 m)	195·18	64·29
3–4	NW-2, SW-2, E-1, E-2, E-3	181·1	(455 m, 545 m)	208·31	50·41
3–5	NW-1, SW-1, S-2,SE-1, NE-1, E-2	120·5	(495 m, 495 m)	198·76	61·18
3–6	NW-2, SW-2, S-2 E-1, E-2, E-3,	150·9	(495 m, 495 m)	202·42	58·31
3–7	N-2, NW-1, SW-1, S-2,SE-1, NE-1, E-2	123·7	(495 m, 495 m)	199·15	60·73
3–8	N-2, NW-2, SW-2, S-2, E-1, E-2,E-3	157·9	(495 m, 495 m)	202·42	57·43

Table V. Number of observation wells required to determine pumping source information (pumping source is located at (495 m, 495 m) and pumping rate and period are 200 m^3 day^{-1} and 60 min, respectively)

Case	Observation wells	Source location	Pumping rate (m^3 day^{-1})	Pumping period (min)
4–1[1]	N-1, W-1, S-1, E-1	(495 m, 495 m)	202·18	57·43
4–2[1]	N-2, W-2, S-2, E-2	(495 m, 495 m)	198·66	61·29
4–3[1]	NW-2, SW-2, E-3, E-1	(495 m, 495 m)	198·38	60·81
4–4[2]	W-1, S-1, E-2	(495 m, 495 m)	198·64	61·32
4–5[2]	N-1, W-1, S-1	(495 m, 495 m)	198·62	61·63
4–6[2]	N-1, E-2, S-1	(495 m, 495 m)	198·58	61·58
4–7[2]	N-1, W-1, E-2	(495 m, 495 m)	198·45	61·42
4–8[2]	N-2, S-2, E-2	(495 m, 495 m)	197·88	60·92
4–9[2]	N-2, W-1, S-2	(495 m, 495 m)	198·45	61·35
4–10[2]	W-1, S-2, E-2	(495 m, 495 m)	198·48	61·72
4–11[2]	N-2, W-1, E-2	(495 m, 495 m)	197·85	60·67
4–12[3]	W-1, E-2	(565 m, 385 m)	216·79	50·48
4–13[3]	S-1, E-2	(495 m, 495 m)	198·36	60·93
4–14[3]	N-1, E-2	(495 m, 495 m)	197·17	62·04
4–15[3]	W-1, S-1	(425 m, 415 m)	199·51	59·87
4–16[3]	N-1, W-1	(495 m, 495 m)	198·12	61·42
4–17[3]	N-1, S-1	(495 m, 495 m)	199·63	60·59
4–18[3]	N-2, S-2	(495 m, 475 m)	203·43	58·46
4–19[3]	S-2, E-2	(495 m, 495 m)	198·18	61·32
4–20[3]	W-1, S-2	(565 m, 385 m)	216·67	50·14
4–21[3]	N-2, E-2	(495 m, 495 m)	198·42	60·68
4–22[3]	N-2, W-2	(495 m, 495 m)	198·32	61·32
4–23[3]	W-1, E-2	(495 m, 495 m)	197·43	62·43

[1] Number of observation wells = 4
[2] Number of observation wells = 3
[3] Number of observation wells = 2

identification when the level of measurement error is high.

Scenario 4: Number of observation wells

In scenario 4, 23 cases are considered to examine the effect of using the observations measured at two different times on the determination of the pumping source information. The results are listed in Table V. In cases 4·1 to 4·3, eight observations measured at four wells are used to determine the pumping information. However, six observations measured at three wells in cases 4·4 to 4·11 and four observations measured at two wells in cases 4·12 to 4·23 are utilized. As the number of observation wells is four or three, the pumping source location, pumping rate and pumping period are correctly determined. However, when the number of observation wells is two, eight out of twelve cases yield the correct results of pumping source information. Obviously, the number of observation wells should be at least three to determine the pumping source information.

HETEROGENEOUS AQUIFER

In this section, we estimate the pumping source information for problems with heterogeneous formations representing possible real-world aquifers. We consider three scenarios, i.e. scenarios 5 to 7, with different degrees

DOI: 10.1002/hyp

of variation in random hydraulic conductivity fields and conditioning conductivity data.

Site description

Field aquifer tests such as a slug test or pumping test are usually used to determine hydrogeologic parameters, i.e. hydraulic conductivity and storativity. These aquifer parameters obtained at specific locations can be used as the conditioning data. The heterogeneous aquifers may be characterized by three statistical parameters: the mean hydraulic conductivity K, the variance in hydraulic conductivity and the correlation length. The heterogeneous formations are assumed to have random conductivity fields which are spatially correlated and log-normally distributed. The correlation length λ is 100 m and the mean of the logarithm of hydraulic conductivity $\ln K$ is $-12{\cdot}206$ m s^{-1}. The mean and standard deviation of $\ln K$ are denoted as $\overline{y}$ and σ_y, respectively, where $y = \ln K$. Three different values of σ_y, representing different levels of aquifer heterogeneity, are considered: 0·5, 1·0 and 1·5 m s^{-1}. The dimensions of the aquifer are the same as those given previously, i.e. the length and width of the aquifer are both 1000 m and the thickness is 20 m. The locations of pumping source and observation wells are as depicted in Figure 2.

For scenarios 5 to 7, monitoring wells are N-1, W-1, S-1 and E-1. Table VI lists the hydraulic conductivities, which are assumed to be obtained from the aquifer test and adopted as the conditioning data for the heterogeneous sites. All the random conductivity fields are generated by the program SASIM of the geostatistical software, GSLIB (Deutsch and Journel, 1998). SASIM can produce spatially correlated random conductivity fields by preserving the known mean and variance of $\ln K$ and known conductivities at specific locations. Therefore, the SASIM is chosen to generate random hydraulic conductivity fields for heterogeneous aquifers with those three statistical parameters. Then MODFLOW-2000 is used with the generated conductivity field and under the same hydrogeological conditions described previously to produce the measured heads. The generated measured heads by MODFLOW-2000 with the pumping rate and period of 200·0 m^3 day^{-1} and 60·0 min, respectively, are listed in Table VI.

Table VI. The hydraulic conductivity for the generated random conductivity field and the corresponding measured heads

Scenario	σ_y (m s^{-1})	Observation well	Hydraulic conductivity ($\times 10^{-6}$ m s^{-1})	Measured head (m)	
				T = 30 min	T = 60 min
5	0·5	N-1	5·40	22·847	21·506
		W-1	2·41	25·148	24·499
		S-1	6·79	24·700	24·256
		E-1	7·07	22·751	21·483
6	1·0	N-1	15·70	23·589	22·579
		W-1	1·39	25·330	24·808
		S-1	4·64	24·689	24·247
		E-1	6·18	21·978	20·602
7	1·5	N-1	10·30	23·790	23·017
		W-1	0·59	25·526	25·236
		S-1	5·58	24·790	24·483
		E-1	18·30	22·228	20·996

Monte Carlo simulation

Field conductivities, obtained from a slug test or pumping test, are generally available only at a few locations. The Monte Carlo simulation is employed to test the applicability of SA-MF. SASIM is used to produce 50 realizations of random conductivity fields with the same conditioning conductivity data and statistical parameters used in the process of generating the measured heads. The SA-MF is then employed to estimate the pumping source information for an aquifer with a realization, i.e. a conditioning random conductivity field. Note that in scenarios 5 to 7, the conductivity fields used to generate the simulated heads in identifying the pumping source information are different from the conductivity field used in generating the measured heads.

The estimated results show that the locations of pumping source are all correct in scenarios 5 to 7. In scenario 5, the average estimated pumping rate and pumping period are 197·11 m^3 day^{-1} and 61·11 min, respectively, when the value of σ_y is 0·5 m s^{-1}. The standard deviation of the estimated pumping rate and pumping period are 5·53 m^3 day^{-1} and 4·48 min, respectively. The greatest relative error is $-7{\cdot}09\%$ in the estimated pumping rate and 22·69 % in pumping period. In scenario 6, the average estimated pumping rate and pumping period are 193·74 m^3 day^{-1} and 63·37 min, respectively, and the standard deviation of the estimated pumping rate and pumping period are 9·39 m^3 day^{-1} and 8·63 min, respectively, when $\sigma_y = 1{\cdot}0$ m s^{-1}. The greatest relative error is $-11{\cdot}06\%$ in the estimated pumping rate and 38·01 % in the pumping period. If σ_y is increased to 1·5 m s^{-1}, the average estimated pumping rate and pumping period are 188·02 m^3 day^{-1} and 65·23 min, respectively. The standard deviation of the estimated pumping rate and pumping period are 10·59 m^3 day^{-1} and 7·88 min, respectively. The greatest relative error is $-16{\cdot}07\%$ in the estimated pumping rate and 38·52 % in the pumping period. The accuracy of the estimated pumping rate and pumping period obviously decreases with increasing σ_y.

ACTUAL FIELD CASE

In the last scenario (scenario 8) the data obtained from a field pumping test is used to determine the pumping source information. The field pumping test was performed in 1987 at Yenliao, a site of the fourth nuclear power plant, located at the northeast corner of Taiwan. The dimension of the groundwater flow system is chosen as 200 m in both length and width. One pumping well, denoted as P, and four observation wells, denoted as A to D, were installed for the aquifer test, as shown in

DOI: 10.1002/hyp

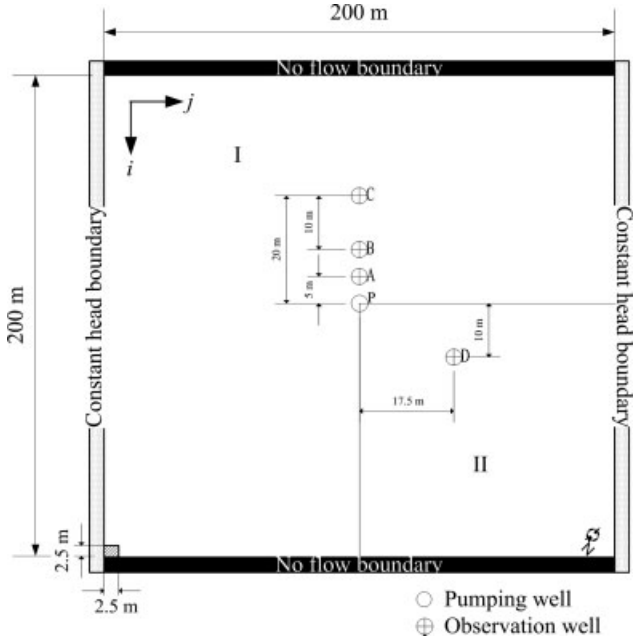

Figure 3. Plan view of Yenliao site

Figure 3. The pumping well is located at (97·5 m, 97·5 m). Observation distances from the pumping well are A: 5 m, B: 10 m, C: 20 m and D: 20 m. Wells A to D are located at (92·5 m, 97·5 m), (87·5 m, 97·5 m), (77·5 m, 97·5 m), and (92·5 m, 110·0 m), respectively. According to the drill log obtained at P, the thickness of the aquifer is about 16 m and the formation composed of medium sand with a small amount of shale.

The surface topography is flat and the gradient is about 0·01. The water level is at a depth of 1 m under the ground surface. The groundwater flows toward the northeast, the upper and lower boundaries are assumed as no-flow and the left and right boundaries are considered as constant-head. The elevations of the hydraulic head at the left and right boundaries are 17 m and 15 m, respectively. The grid width and length are both 2·5 m; thus, the number of finite difference mesh is 80×80. The aquifer is assumed as isotropic, the pumping well and observation wells are fully penetrated through the aquifer. The conductivity field is divided into two zones based on the results of pumping test data analysis. The estimated transmissivity and storage coefficient are 3·90 m^2 day^{-1} and $3{\cdot}73 \times 10^{-4}$, respectively, in zone 1 and 20·51 m^2 day^{-1} and $3{\cdot}83 \times 10^{-3}$, respectively, in zone 2 as indicated in Figure 3. The pumping rate was 10 l min^{-1} and the drawdown data were taken at the pumping times of 10 and 20 min. The drawdowns at wells A to D were 0·529, 0·130, 0·038 and 0·001 m, respectively, at pumping time of 10 min and 0·774, 0·250, 0·101 and 0·005 m, respectively, at pumping time of 20 min.

All the nodal points inside the problem domain are considered as the suspected pumping source locations except those nodes containing the observation wells. The upper bounds for the pumping rate and pumping period are set as 100·0 l min^{-1} and 60·0 min, respectively, and the lower bounds are both given as zero. The *NS, NT*, initial temperature and R_{Te} are chosen as 20, 10, 5 and 0·8, respectively. All initial estimates are produced by a random number generator.

Table VII. Identified results for the Yenliao site (pumping source is located at (97·5 m, 97·5 m) and the pumping rate and period are 10 l min^{-1} and 20 min, respectively)

Case	Observation wells	Source location	Pumping rate (l min^{-1})	Pumping period (min)
8–1	A, B, C	(97·5 m, 97·5 m)	10·38	19·29
8–2	A, B, C, D	(97·5 m, 97·5 m)	10·73	18·87
8–3	B, C, D	(97·5 m, 97·5 m)	10·34	19·21
8–4	A, B, D	(97·5 m, 97·5 m)	10·30	19·17

Four cases are used to test the performance of SA-MF in this scenario. Table VII lists the observation wells used in these four cases. The estimated source locations are all correct at (97·5 m, 97·5 m) in cases 8·1 to 8·4. The average estimated pumping rate and pumping period are 10·43 l min^{-1} and 19·14 min, respectively. The greatest relative error is 7·3 % in the estimated pumping rate and −5·65% in the pumping period in case 8·2. Therefore, the proposed SA-MF has been shown to be able to estimate the pumping source information for a real field problem.

CONCLUSIONS

This study proposes a new approach, SA-MF, based on a 3D groundwater flow, MODFLOW-2000, and SA for solving the pumping source identification problem. The SA is first employed to generate the trial solutions randomly, i.e. pumping source location, pumping rate and pumping period and MODFLOW-2000 is then employed to simulate the hydraulic heads at the observation wells. Eight scenarios were designed to test the accuracy and applicability of SA-MF in the determination of pumping source information on the homogeneous, heterogeneous aquifer systems. Four major conclusions were drawn.

First, the proposed approach can be used to determine the pumping source information for problems in either synthetic heterogeneous aquifers or a real aquifer system and the estimations generally give good results. Second, SA-MF can provide good estimated results even if the initial guesses are generated by a random number generator. Third, at least five measured heads should be used to analyse the pumping source identification problem and at least three observation wells are required for the effective determination of the pumping source location, the pumping rate and the pumping period. Fourth, if the level of measurement error is high, increasing the number of measured heads is helpful in determining the pumping source information correctly. In other words, the key for the successful determination of the pumping source information is to use more observed hydraulic heads.

DOI: 10.1002/hyp

ACKNOWLEDGEMENTS

This study was partly supported by the Taiwan National Science Council under the grant NSC94-2211-E-009-012. The authors thank two anonymous reviewers for their very detailed and insightful comments.

REFERENCES

Aral MM, Guan J. 1996. Genetic algorithms in search of groundwater pollution sources: advances in Groundwater Pollution Control and Remediation. *NATO ASI Series 2 Environment*. Kluwer Academic Publisher: Netherlands; 347–369.

Aral MM, Guan J, Maslia ML. 2001. Identification of contaminant source location and modflow-gwt release history in aquifers. *Journal of Hydrologic Engineering* **6**(3): 225–234.

Bear J. 1979. *Hydraulics of Groundwater*. McGraw-Hill: New York.

Cai X, McKinney DC, Lasdon LS. 2001. Solving nonlinear water management models using a combined genetic algorithm and linear programming approach. *Advances in Water Resources* **24**(6): 667–676.

Cooper VA, Nguyen VTV, Nicell JA. 1997. Evaluation of global optimization methods for conceptual rainfall-runoff model calibration. *Water Science Technology* **36**(5): 53–60.

Cunha MDC, Sousa J. 1999. Water distribution network design optimization: simulated annealing approach. *Journal of Water Resources Planning and Management* **125**(4): 215–221.

Deutsch CV, Journel AG. 1998. *GSLIB: Geostatistical Software Library and User's Guide*. Second edition. Oxford University Press: Oxford.

Doutherty DE, Marryott RA. 1991. Optimal groundwater management: I. Simulated annealing. *Water Resources Research* **27**(10): 2493–2508.

Guo JQ, Zheng L. 2005. A modified simulated annealing algorithm for estimating solute transport parameters in streams from tracer experiment data. *Environmental Modeling and Software* **20**(6): 811–815.

Harbaugh AW, Banta ER, Hill MC, McDonald MG. 2000. *MODFLOW-2000, The U.S. Geological survey modular ground-water model-user guide to modularization concepts and the ground-water process*. Open-File Report 00–92.

IMSL. 2003. *Fortran Library User's Guide Stat/Library*. Volume 2 of 2. Visual Numerics Inc.: Houston.

Kuo CH, Michel AN, Gray WG. 1992. Design of optimal pump-and-treat strategies for contaminated groundwater remediation system using the simulated annealing algorithm. *Advances in Water Resources* **15**(2): 95–105.

Li L, Barry DA, Morris J, Stagnitti F. 1999. CXTANNEAL: an improved program for estimating solute transport parameters. *Environmental Modelling and Software with Environment Data News* **14**(6): 607–611.

Lin YC, Yeh HD. 2005. THM species forecast using optimization method: genetic algorithm and simulated annealing. *Journal of Computing in Civil Engineering* **19**(3): 248–257.

Loáiciga HA. 2004. Analytic game-theoretic approach to ground-water extraction. *Journal of Hydrology* **297**: 22–33.

Mahinthakumar G, Sayeed M. 2005. Hybrid genetic algorithm-local search methods for solving groundwater source identification inverse problem. *Journal of Water Resources Planning and Management* **131**(1): 45–57.

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

Marryott RA. 1996. Optimal ground-water remediation design using multiple control technologies. *Ground Water* **34**(3): 425–433.

Marryott RA, Dougherty DE, Stollar RL. 1993. Optimal groundwater management: 2. Application of simulated annealing to a field-scale contaminant site. *Water Resources Research* **29**(4): 847–860.

Monem MJ, Namdarian R. 2005. Application of simulated annealing (SA) techniques for optimal water distribution in irrigation canals. *Irrigation and Drainage* **54**(4): 365–373.

Pham DT, Karaboga D. 2000. *Intelligent optimization techniques: genetic algorithms, tabu search, simulated annealing and neural networks*. Springer-Verlag: New York.

Rayward-Smith VJ, Osman IH, Reeves CR, Smith GD. 1996. *Modern heuristic search methods*. John Wiley & Sons: New York.

Rizzo DM, Dougherty DE. 1996. Design optimization for multiple period groundwater remediation. *Water Resources Research* **32**(8): 2549–2561.

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

Sayeed M, Mahinthakumar G. 2005. Efficient parallel implementation of hybrid optimization approaches for solving groundwater inverse problems. *Journal of Computing in Civil Engineering* **19**(4): 329–340.

Shieh HJ, Peralta RC. 2005. Optimal in situ bioremediation design by hybrid genetic algorithm-simulated annealing. *Journal of Water Resources Planning and Management* **131**(1): 67–78.

Todd DK, Mays LW. 2005. *Groundwater hydrology*. Third edition. Wiley: New Jersey.

Tung CP, Tang CC, Lin YP. 2003. Improving groundwater flow modeling using optimal zoning methods. *Environmental Geology* **44**: 627–638.

Woodbury AD, Ulrych TJ. 1996. Minimum relative entropy inversion: theory and application to recovering the release history of a groundwater contaminant. *Water Resources Research* **32**(9): 2671–2681.

Yeh HD, Lin YC, Huang YC. 2007. Parameter identification for leaky aquifers using global optimization methods *Hydrological Processes* **21**(7): 862–872.

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