Trihalomethane Species Forecast Using Optimization Methods: Genetic Algorithms and Simulated Annealing

Yu-Chung Lin[1] and Hund-Der Yeh[2]

Abstract: Chlorination is an effective method for disinfection of drinking water. Yet chlorine is a strong oxidizing agent and easily reacts with both organic and inorganic materials. Trihalomethanes (THMs), formed as a by-product of chlorination, are carcinogenic to humans. Models can be derived from linear and nonlinear multiregression analyses to predict the THM species concentration of empirical reaction kinetic equations. The main objective of this study is to predict the concentrations of THM species by minimizing the nonlinear function, representing the errors between the measured and calculated THM concentrations, using the genetic algorithm (GA) and simulated annealing (SA). Additionally, two modifications of SA are employed. The solutions obtained from GA and SA are compared with the measured values and those obtained from a generalized reduced gradient method (GRG2). The results indicate that the proposed heuristic methods are capable of optimizing the nonlinear problem. The predicted concentrations may provide useful information for controlling the chlorination dosage necessary to assure the safety of water drinking.

DOI: 10.1061/(ASCE)0887-3801(2005)19:3(248)

CE Database subject headings: Algorithms; Evolutionary computation; Simulation; Disinfection; Optimization; Chlorination; Potable water.

Introduction

The basic operations in a water treatment plant include three major processes such as clarification, filtration, and disinfection. Removal of pathogenic microorganisms, tastes, odors, color, turbidity, dissolved minerals, and harmful organic materials are very important in water treatment. Water contains many microorganisms, some of which cause diseases. Typical bacterial reductions of 60 to 70% are possible by coagulation, flocculation, and sedimentation. A filtration process increases the overall bacterial removal to about 99%. Disinfection is an essential and final barrier against human exposure to disease-causing pathogenic microorganisms in water supply engineering.

Chlorination began at the start of the last century to provide an additional safeguard against pathogenic microorganisms. A number of equilibria affect the form and the effectiveness of chlorine in water. Chlorine is a very strong oxidizing agent and may combine with water to form hypochlorous and hydrochloric acids (Bitton 1999).

Hypochlorous and hydrochloric acids react with both organic and inorganic materials to produce trihalomethanes (THMs). Chloroform ($CHCl_3$), bromodichloromethane ($CHBrCl_2$), dibromochloromethane ($CHBr_2Cl$), and bromoform ($CHBr_3$), produced by chlorination of water and wastewater, are suspected mutagens/carcinogens. Because of their carcinogenic effect, the U.S. Environmental Protection Agency (USEPA) limited the concentration of four kinds of THM species in recent years. The USEPA finalized the new disinfectant/disinfection by-products rules in 1998 and limited the residual chlorine concentration to 0.08 mg/L in drinking water. A recent study by the USEPA found that women exposed to higher levels of chlorine by-products had a 15.7% risk of miscarriage, whereas those who had little exposure to THM had a lower (9.5%) risk (Elshorbagy 2000).

Elshorbagy (2000) applied a generalized reduced gradient method (GRG2), developed by Lasdon and Warren (1982), to characterize and simulate the kinetics of THM species under representative extreme conditions associated with temperature, chlorine dosage, and bromide content. The GRG2 model was employed to combine site-specific trends with stoichiometric expressions based on an average representative bromine concentration factor. The model was evaluated, tested, and validated using data from finished desalinated water collected in the United Arab Emirates. Elshorbagy (2000) obtained reasonable agreements between predicted and measured values of different species. Milot et al. (2002) applied artificial neural networks to predict the contaminant levels of total THMs in the chlorinated water under laboratory condition.

The objective of this study was to find the concentrations of THM species by minimizing the nonlinear function representing the errors between the measured and calculated THM concentrations. Results obtained from genetic algorithm (GA) and simulated annealing (SA) are compared with the measured values and those obtained from the GRG2 model. The proposed methods provide a useful tool for controlling the chlorination dosage to assure the safety of water drinking.

[1]Graduate Student, Institute of Environment Engineering, National Chiao-Tung Univ., Hsinchu, Taiwan.

[2]Professor, Institute of Environment Engineering, National Chiao-Tung Univ., Hsinchu, Taiwan.

Note. Discussion open until December 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on August 22, 2003; approved on December 9, 2004. This paper is part of the ***Journal of Computing in Civil Engineering***, Vol. 19, No. 3, July 1, 2005. ©ASCE, ISSN 0887-3801/2005/3-248–257/ $25.00.

J. Comput. Civ. Eng. 2005.19:248-257.

Studies on Trihalomethane Formation

Many researchers discovered that THM formation potential is a useful and conservative indicative parameter (Rook 1974; Bunn et al. 1975; Trussel and Umphres 1978; Minear 1980). The indicative parameter could apparently develop in the presence of total available organic precursors. Those studies also showed that THM levels increase with increasing bromide concentration. Gould et al. (1981) defined the bromine incorporation factor (BIF), which changes little with reaction time. The pH and organic precursor content, depending on the bromide concentration, have a maximum influence on BIF. Montgomery Watson (1993) also found that bromide-rich sources of water such as desalinated and/or some groundwater wells can increase the THM formation potential. Various parameters, such as total organic carbon, ultraviolet light absorbance, temperature, chlorine dosage, bromide concentration, reaction time, and chlorination pH also affect THM formation.

Clark et al. (1994) presented certain dynamic water quality studies to treat THM as a conservative substance throughout the distribution system and to approximate its kinetics by a first order, growth-limited reaction rate expression. More recently, Clark (1998) proposed an improved model based on second-order reaction kinetics for THM formation as a function of chlorine demand. In contrast, Nokes et al. (1999) proposed a reaction scheme with another kinetic model to simulate the THM formation. The mathematical model related the extent of bromination and the relative abundances of the four THM species to the $[Br^-]:[Cl^-]$ ratio.

In addition to the dynamic models mentioned previously, some empirical models were also proposed to describe the mechanisms of THM formation (e.g., Amy et al. 1987; Adin et al. 1991; Golfinopoulos et al. 1998). Among the empirical models, Amy et al. (1987) examined the relationships between several key water quality parameters and their corresponding THM formation. Based on the empirical results, a multiparameter power function, derived from linear and nonlinear multiregression, was developed to predict THM formation in drinking water.

Development and Application of the Heuristic Approaches

Recently, considerable attention has been paid on stochastic optimization techniques such as GA and SA. The predominance of these heuristic methods is that the optimization problem is allowed to be formulated as a nonlinear or mixed integer problem. The user does not need much experience in providing the initial guesses for solving nonlinear problems. This is the major difference between the heuristic algorithm and the traditional approach such as Newton's method and/or GRG2. The main advantage of using the heuristic algorithm is that it can solve the problem with arbitrary initial guesses and may give optimal results without any rules. Two heuristic methodologies, GA and SA, will be adopted in this study for the optimal determination of the concentration of THMs.

GA was developed by Holland (1975) in the 1950s and 1960s. Darwin's theory of evolution is the basic concept of GA. Using a population of guesses in GA to solve the optimization problem significantly differs from those solvers using a single guess. The main idea is that the genetic information of a good solution spreads over the entire population. Thus, the best solution can be obtained by thoroughly combining the "chromosomes" in the population. Selection, crossover, mutation, and reproduction are the essential operators in GA. The selection operator picks up the relative fitter strings according to their objective function values. Then the newborn trial solutions are generated from the relative fitter string with the crossover operator. Avoiding the trial solution trapped in a local region is the main idea behind the mutation operator.

In recent years, GA has been applied in a variety of fields. McKinney and Lin (1994) presented a groundwater management model using GA. Lippai et al. (1999) performed robust analysis and optimization of a water distribution network based on GA-based optimizers with *Win-Pipes.EXE*, a Windows program based on the *EPANET* source code. They applied this method on the New York City Water Supply Tunnel problem and the results obtained meet the multiple goals of a reliable, low-cost water distribution system, and also satisfy the maximum hour demands and fire flow demands. Gupta et al. (1999) presented a methodology for minimizing the cost of a pipe network using GA. The solution set obtained from GA and nonlinear programming (NLP) techniques for several medium size networks showed that GA provided a better solution in general, in comparison with that obtained with NLP techniques. GA is a general stochastic evolutionary algorithm with a wide range of applicability to optimization problems with good performance (Coley 1999; Pham and Karaboga 2000). Gwo (2001) used a simple GA to search for preferential flow in a structured porous media. The result showed that GA can invert the correct pictures of simple fracture networks.

Another heuristic method employed in this study is SA. Metropolis et al. (1953) first applied SA in a two-dimensional rigid-sphere system. Kirkpatrick et al. (1983) demonstrated the strengths of SA by solving large-scale combinatorial optimization problems. SA is a random search algorithm that allows, at least in theory or in probability, to obtain the global optimum of a function in any given domain (Aarts and Korst 1989). One of the advantages of SA is its ability to use a descent strategy which allows random ascent moves to avoid possible traps in a local optimum. Ease of implementation is another advantage of SA. Goffe et al. (1994) presented the results of global optimization stochastic functions with SA. Cunha and Sousa (1999) used SA to minimize the costs of a water distribution network. Pham and Karaboga (2000) demonstrated that SA can be applied for a wide range of optimization applications. Domer et al. (2003) applied SA to control the quasi-static displacement of a tensegrity structure with multiple objectives and interdependent actuator effects. Their results suggest that SA provide a rapid convergence to "good" solutions.

These two heuristic approaches could obtain the global optimal solutions. However, when the problem or the solution space is fairly complicated, heuristic approaches may have the problems of taking much computing time and effort to solve the optimization problem. Differing from the gradient type approach, the heuristic approaches should generate the trial solutions in the specified solution space. In addition, all the trial solutions require calculating the objective function values even though those solutions are incorrect. Besides, Youssef et al. (2001) pointed out that if excess population size and/or maximum evolutionary generation were specified, GA also took much time and effort to obtain global optimum solutions. Similar to SA, the local optimum solution would be obtained if the initial temperature given was too low. On the other hand, if a higher initial temperature was given, more time would be consumed for using SA.

J. Comput. Civ. Eng. 2005.19:248-257.

Methodology

In this section, the kinetics of THM and THM species, GA, and SA are described separately. An exhaustive description to simulate the THM formation and formulate the objective function is given first. The detailed general idea and algorithm of GA and SA are also presented.

Kinetics of Trihalomethane and Trihalomethane Species

THM compounds are commonly developed in chlorinated water containing organic precursors, such as humic and fulvic acids. Total trihalomethane (TTHM) is known to increase with time; however, information about the reaction mechanism of THM and its species is still limited.

Trihalomethane Simulation

Let $[THM_0]$, $[THM_1]$, $[THM_2]$, and $[THM_3]$ be the molar concentrations of the four THM species: Chloroform, bromodichloromethane, dibromochloromethane, and bromoform, respectively. The following equation describes the species mass balance:

$$[THM_0] + [THM_1] + [THM_2] + [THM_3] = [TTHM] \quad (1)$$

The TTHM formation rate is assumed to be equal to the chlorine consumption rate in a differential equation form. The decay of total residual chlorine modeled using a first-order decay reaction relationship is

$$\frac{d[Cl^-]}{dt} = -K_{cl}[Cl^-] \quad (2)$$

where $[Cl^-]$=chlorine concentration and K_{cl}=chlorine reaction rate coefficient. Thus, the TTHM formation rate may be written as

$$\frac{d[TTHM]}{dt} = -F\frac{d[Cl^-]}{dt} = FK_{cl}[Cl^-] \quad (3)$$

where F=linear proportionality constant between TTHM formation and chlorine decay and [TTHM]=molar concentration of TTHM. Eq. (3) can be expressed in the difference form as

$$[TTHM]_{(t+\Delta t)} = [TTHM]_t - F([Cl^-_{(t+\Delta t)}] - [Cl^-_t]) \quad (4)$$

where Δt=time difference.

Bromoform is assumed to be dependent on temperature. Bromoform formation is therefore modeled using a limited first-order growth relation (Montgomery Watson 1993):

$$[Br^-_t] = \{[Br^-_u] - [Br^-_0]\}(1 - e^{-k_b t}) \quad (5)$$

where k_b=bromoform reaction rate coefficient and $[Br^-_t]$, $[Br^-_0]$, and $[Br^-_u]$=respectively, the bromoform mass concentrations ($\mu g/L$) at time t, 0, and ∞. Note that $[Br^-_u]$, the ultimate growing concentration of bromoform, has a similar physical meaning to the coefficient F.

Objective Function

Gould et al. (1981) defined BIF as

$$BIF = \frac{\sum_{N=0}^{N=3} N[THM_N]}{\sum_{N=0}^{N=3}[THM_N]} \quad (6)$$

where N=number of bromine atoms in the THM compound; N=0 for chloroform and N=3 for bromoform. Thus BIF ranges between zero and three, based on the molar concentration of brominated species in the TTHM.

The bromine distribution factor, introduced to simplify Eqs. (1) and (6), is defined as

$$S_N = \frac{[THM_N]}{[TTHM]} \quad (N = 0, 1, 2, \text{ or } 3) \quad (7)$$

Thus, Eqs. (1) and (6) can be rewritten as

$$S_0 + S_1 + S_2 + S_3 = 1 \quad (8)$$

and

$$BIF = S_1 + 2S_2 + 3S_3 \quad (9)$$

The measured THM species concentrations indicate that the following relationships hold (Elshorbagy 2000):

$$S_0 \geqslant S_1, \quad S_2 \geqslant S_3 \quad (10)$$

Note that [TTHM] and $[THM_3]$ can be calculated based on Eqs. (4) and (5), respectively. These three unknown concentrations, $[THM_0]$, $[THM_1]$, and $[THM_2]$, can be solved by Eqs. (8) and (9) while being subject to the constraint of Eq. (10). The formulation of Eq. (9) as an optimization problem is

$$\text{Min}[S_1 + 2S_2 + 3S_3 - BIF]^2 \quad (11)$$

subject to the constraints represented by Eqs. (8) and (10).

Employing the Lagrange multiplier theory (Hillier and Lieberman 1990), Eqs. (8) and (9) may be expressed as

$$\text{Min}[S_1 + 2S_2 + 3S_3 - BIF]^2 + \lambda[1 - S_0 + S_1 + S_2 + S_3]^2 \quad (12)$$

where λ=Lagrange multiplier. It is also known as a penalty factor since the second term is simply a penalty term in the objective function. Additionally, Eq. (12) should be subjected to the constraint of Eq. (10).

Genetic Algorithm

Genetic algorithm is an optimization algorithm inspired by both natural selection and natural genetics (Holland 1975). Based on Darwin's theory of evolution, the better filial generation in GA will survive and generate the next generation. Naturally, the best generation will have better presentation to get with the conditions. The method can be applied to an extremely wide range of optimization problems. Rather than starting from a single point within the search space, GA initiates a population of guesses. A group of initial guesses is required at the start of GA optimization. A common way of generating the initial population of guesses is to use a random number generator.

Four major steps such as encoding, selection, crossover, and mutation, are required in GA. In the encoding step, the initial guess of each unknown is converted to binary strings, called substrings, of length l_i. The summation of each substring length is the total string length L. It should be noted that, an increasing number of GAs use "real-valued" encoding as the data structure of the problem (Pham and Karaboga 2000). Each string can be viewed as a simple data structure.

Based on an initial population of guesses, selection, crossover, and mutation operations are started. The aim of the selection procedure is to reproduce more copies of individuals whose fitness values are higher than those with lower fitness values. Poorer individuals are weeded out and better performing individuals have a greater than average chance of promoting the information they contain within the next generation.

J. Comput. Civ. Eng. 2005.19:248-257.

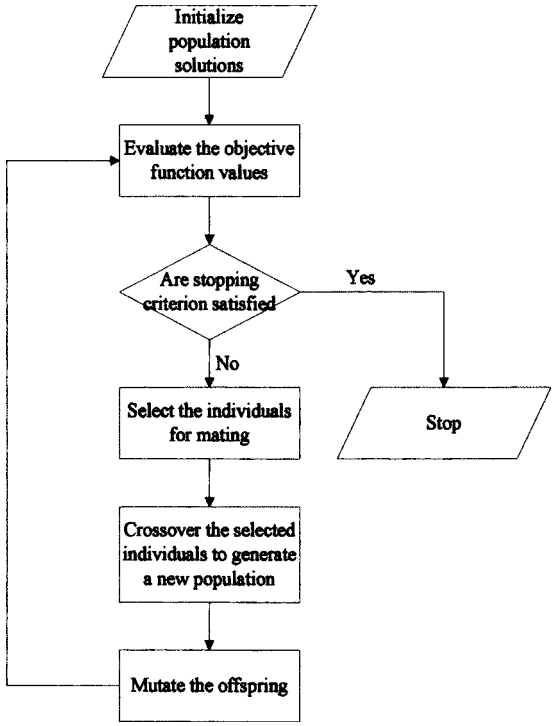

Fig. 1. Flowchart of genetic algorithm

The selection mechanism is applied twice in order to select a pair of individuals to undergo, or not to undergo, crossover. The crossover operator is considered the one that makes the GA differ from other algorithms. Crossover is used to create two new individuals (children) from the two existing individuals (parents). The pair of individuals selected undergo crossover with a probability P_c. Typical values of P_c range from 0.4 to 0.9 (Coley 1999). In general, four kinds of crossover methods such as one-point crossover, two-point crossover, homogeneous crossover, and schema theorem, are favored (Coley 1999).

Mutation is used to randomly change the value of a single bit or multiple bits within an individual string with a specified rate P_m, the probability of mutation. Coley (1999) also suggested using $1/L$ or 0.001 as the probability of mutation. Fig. 1 shows the flowchart for the GA approach. After selecting the initial guesses, the trial solution, determined by the optimal fitness function value, will proceed to the crossover step. In addition to the crossover step, each bit in the newborn trial solution, the binary string, will be mutated depending on the mutation probability. The algorithm will be terminated when the solution or the number of iterations satisfies the stopping criteria.

Simulated Annealing

The concept of SA is based on an analogy with the physical annealing process. In the beginning of the process, the temperature is increased to enhance the molecular mobility. Next, the temperature is slowly decreased to allow the molecules to form crystalline structures. When the temperature is high, the molecules have a high level of activity and the crystalline configurations assume a variety of forms. If the temperature is lowered properly, the crystalline configuration is in the most stable state; thus, the minimum energy level may be naturally reached.

At a given temperature, the probability distribution of the system energy is determined by the Boltzman probability (Pham and Karaboga 2000):

$$P(E) \propto \exp(-E/kT) \tag{13}$$

where E=system energy; k=Boltzmann's constant; T=temperature; and $P(E)$=occurrence probability. There exists a small probability that the system may have high energy even at low temperature. Therefore, the statistical distribution of energies allows the system to escape from a local minimum energy. This is the major reason why the solutions obtained from SA may not become trapped as a local optimum or result in a poor solution. The Boltzmann probability is applied in Metropolis' criterion (Kirkpatrick et al. 1983) to establish the probability distribution function for the trial solution. The Metropolis' criterion takes the place ΔE, the difference between the current solution and trial solution of E, and k being equal to one. The modified Boltzmann probability which represents the probability that the trial solution will be accepted is given as

$$P(\Delta E) = \exp(\Delta E/T) \tag{14}$$

As an iterative improvement method, an initial point x is required to evaluate the objective function value $f(x)$. Let x' assume the position as the neighbor of x and its objective function value is $f(x')$. In the minimization problem, if $f(x')$ is smaller than $f(x)$, then the trial solution (x') takes the place of the current optimal solution (x). If $f(x')$ is not smaller than $f(x)$, then one has to test Metropolis's criteria and generate a new random number D between zero and one. For solving minimization problems, the Metropolis's criterion is given as (Metropolis et al. 1953):

$$P_{\mathrm{SA}}\{\text{accept } j\} = \begin{cases} 1 & \text{if } f(j) \leqslant f(i) \\ \exp\left(\dfrac{f(i) - f(j)}{T}\right) & \text{if } f(j) > f(i) \end{cases} \tag{15}$$

where $f(i)$ and $f(j)$ are, respectively, the function value when $x = x_i$ and $x = x_j$. x_i and x_j are, respectively, the current optimal solution and neighborhood trial solution of x. Here T, a control parameter, is usually the current temperature. For solving the maximization problem, Metropolis's criterion is expressed as

$$P_{\mathrm{SA}}\{\text{accept } j\} = \begin{cases} 1 & \text{if } f(j) > f(i) \\ \exp\left(\dfrac{f(j) - f(i)}{T}\right) & \text{if } f(j) \leqslant f(i) \end{cases} \tag{16}$$

Fig. 2 displays the SA algorithm approach. The first step in SA is to initialize a solution and set the initial solution to equal the current optimal solution. The second step is to update the current optimal solution, if the trial solution generated from the initial solution within the boundary is better than the current optimal solution; otherwise, continue generating trial solutions until the algorithm satisfies the temperature decrease criterion. The algorithm will be terminated when SA obtains the optimal solution or the obtained solution satisfies the stopping criteria. In general, the stopping criteria are defined to check whether the temperature or the iteration number reaches the specified value or not.

Two modifications to the SA approach are introduced in this study to ensure that the solutions obtained from SA are the optimum solutions. In the first modification, the Metropolis criterion is expressed as

J. Comput. Civ. Eng. 2005.19:248-257.

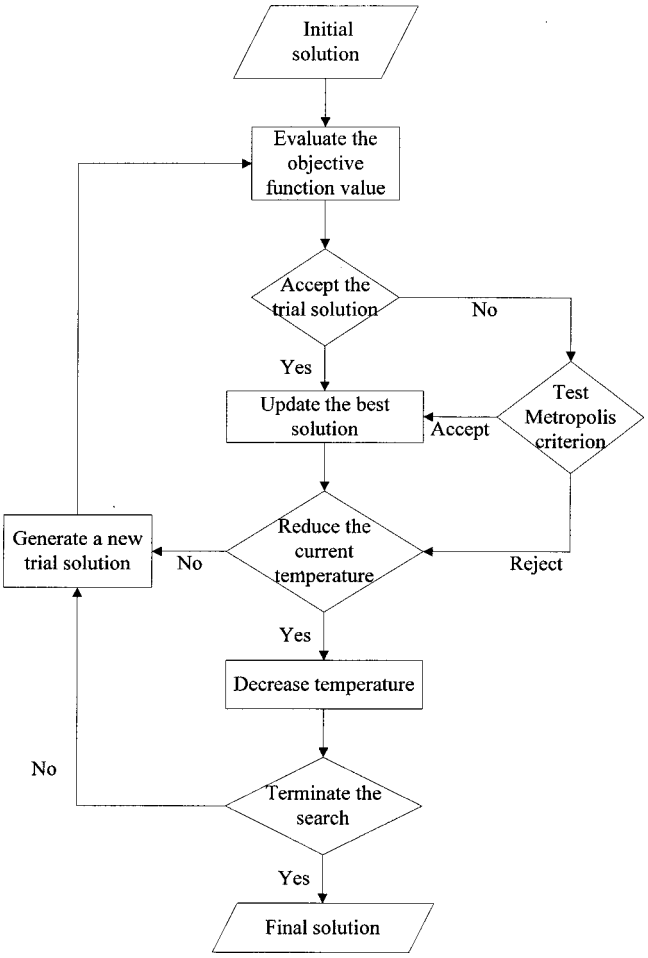

Fig. 2. Flowchart of simulated annealing [adapted from Pham and Karaboga (2000)]

$$P_{SA}\{\text{accept } j\} = \begin{cases} 1 & \text{if } f(j) \leq f(i) \\ \exp\left[\left(\dfrac{f(i)-f(j)}{f(j)}\right)/T\right] & \text{if } f(j) > f(i) \end{cases} \tag{17}$$

This modified criterion is different from the general Metropolis criterion as mentioned previously. In Eq. (17), the increment between the current best solution and the neighborhood trial solution is divided not only by parameter T, but also by the neighborhood trial solution. After the temperature decreases several times, any acceptance probability obtained from the modified Metropolis criterion will be smaller than that obtained from the general Metropolis criterion. The best solution obtained from the modified Metropolis criterion will converge much faster than that using the general Metropolis criterion because unfavorable solutions will not be accepted in the algorithm.

The second modification is to adjust the searching number with a factor β for decreasing temperature. In general, β is given as 1.1 (Cunha and Sousa 1999). Due to an increasing of the searching number, more trial solutions will be created and a much higher possibility will be achieved to obtain the optimal solution. In addition, Fig. 3 shows a brief schematic explaining how to apply GA and/or SA coupled with the kinetics of THM and THM species to forecast the THM species concentrations, and procedure is given in the following steps:

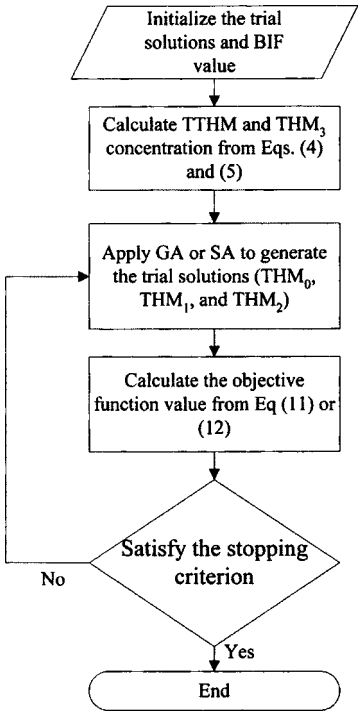

Fig. 3. Schematic diagram of the forecasting procedure

1. Initialize the initial guesses and the BIF value obtained from prior experimental results.
2. Calculate the TTHM and THM_3 concentrations based on Eqs. (4) and (5).
3. Apply GA or SA to analyze the THM_0, THM_1, and THM_2 concentrations. Note that the algorithmic parameters in GA or SA need to be specified in this step.
4. Check the obtained results. If the trial solutions do not satisfy the stopping criterion, go back to Step 3 to keep on generating the possible solutions. Otherwise, the algorithm is terminated.

Case Study and Results

The finished water system in Alain, United Arab Emirates (Elshorbagy 2000) was chosen as the study case for this experiment. Water was injected with chlorine from the pumping station to maintain a residual level from 0.3 to 0.5 ppm prior to pumping into the Alain distribution system. The bromide levels in the finished water ranged between 0.1 and 0.4 ppm. A prior experiment recorded the variation of parameters such as total organic carbon and pH with time. The experimental results indicated that those parameters are approximately constant with respect to time. Based on the experimental results, BIF was also concluded to be a constant. This assumption provides a basis for the development of the modeling approach.

In this study, K_{cl}, F, and k_b were, respectively, taken as 0.070, 0.706, and 0.121 1/h, and the concentrations of $[Br_0]$ and $[Br_u]$ were, respectively, 2.6 and 8.2μg/L (Elshorbagy 2000). In the GA approach, several parameters are needed to be assumed at the beginning. Each unknown was encoded into a 24-bit binary string. Other parameters such as population size, P_c, and P_m, were given as 20, 0.6, and $1/L$, i.e., 1/72, respectively. After 5,000 generations, the GA approach was terminated.

J. Comput. Civ. Eng. 2005.19:248-257.

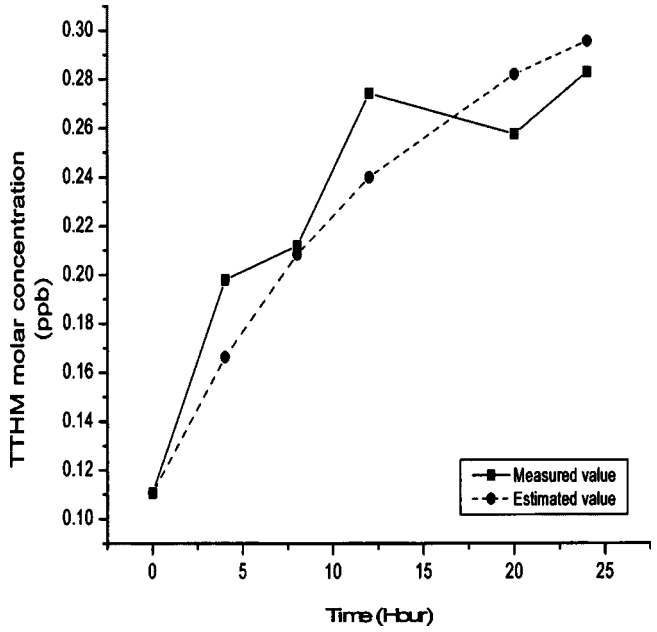

Fig. 4. Temporal concentration distributions of the measured and estimated total trihalomethanes

When using the SA approach, four different cases for determining the THM concentrations were considered. Due to the random number generator needs two seeds to generate the random number, Cases 1 and 2 employed different random number seeds for the SA. The random number seeds used in Case 1 were one and two and in Case 2 were three and four, since SA requires two seeds to generate a series of random numbers. Case 3 used the modified Metropolis criteria, Eq. (17), and Case 4 used a variable searching number, which increased with a multiplication factor 1.1 as the temperature is decreased.

In Case 3, the acceptance probability decreased with decreasing temperature. Thus the solutions obtained at the last step were better than those of the previous steps. In Case 4, the searching number increased upon decreasing the temperature and naturally resulted in increasing the number of trial solutions. The probability of obtaining the global optimum was then increased. Note that all parameters such as reducing temperature factor, initial temperature, initial guesses, and error tolerance were kept all the same in Cases 1–4. The initial temperature and reducing temperature factor were, respectively, taken as 5 and 0.8 in Cases 1–4. The algorithm was terminated when the difference between the current best solution and last step best solution became less than 10^{-6} in 12 iterations. After 1,200 iterations, the temperature was decreased in Cases 1–3.

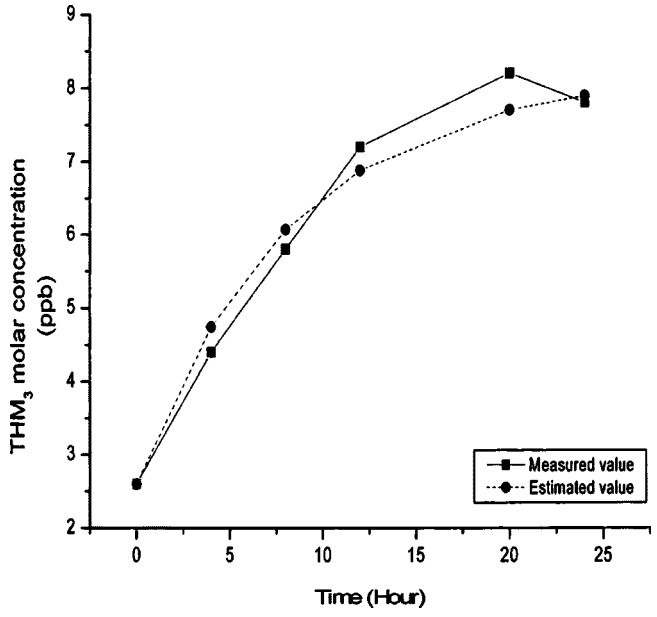

Fig. 5. Temporal concentration distributions of the measured and estimated bromoform

The calculated concentrations of TTHM from Eq. (4) and THM_3 from Eq. (5) are given in Figs. 4 and 5, respectively. The concentrations of THM_0, THM_1, and THM_2 were obtained from experiment and simulations at different times throughout one day. The temporal concentration distributions of THM_0, THM_1, and THM_2 listed in Table 1 show that the calculated values are not significantly different from the measured ones. Also, the THM_0 concentrations obtained from heuristic methods are shown to be very close to the experimental and GRG2 values. However, the concentrations of THM_1 and THM_2 obtained from GA and SA are slightly different than compared to those obtained from GRG2.

Note that the measured values of S_1 and S_3 at 8, 12, and 20 h listed in Table 2 do not satisfy the constraint of Eq. (10). This may be the major reason why the calculated concentrations of THM_2 and THM_3 differ from the measured ones. The objective function values of Eq. (12) versus time obtained from the GRG2 model, GA, and SA are demonstrated in Table 3. Note that the values at different times for SA Cases 1, 2, and 3 are roughly the same, and the values obtained from SA Case 4 are much better than those of the GRG2 model. Although the objective function values obtained from GA are not better than those obtained from GRG2, the THM species concentrations obtained from GRG2 and GA do not show a significant difference.

Fig. 6 exhibits the THM_1, THM_2, and THM_3 concentrations obtained from SA while employing values of 0.1, 1, and 10 for the Lagrange multiplier. The estimated concentrations for each species are not significantly different, implying that the first term of Eq. (12) dominates the calculation. In this study, the concentrations of THM species are estimated with a constant BIF. In reality, the values of BIF can be calculated using Eq. (6) with the measured concentrations of the THM species, indicating that the calculated BIF slightly varied with time. The BIF values calculated from Eq. (6) at 0, 4, 8, 12, 20, and 24 h are 0.7303, 0.7453, 0.7309, 0.7287, 0.723, and 0.7526, respectively (Elshorbagy 2000). However, Fig. 7 indicates that the four estimated THM concentrations based on the average BIF value and the calculated BIF values do not show a considerable difference. Thus, assuming that BIF is a constant is reasonable and practical in the THM estimations. Note that the maximum computing time is about 6.83 s for SA and 7.27 s for GA when simulating the THM concentrations on a personal computer with Intel PIII-800 CPU.

Conclusions and Future Study

Chlorination is a commonly used method for disinfection in the water distribution system. The purpose of disinfection is to remove the water-borne disease for providing healthy drinking water. However, chlorine reacts with the organic or inorganic materials to produce THM. In recent years, some studies indicate that the THM species are the carcinogens.

J. Comput. Civ. Eng. 2005.19:248-257.

Table 1. Measured and Modeled Values for Chloroform (THM_0), Bromodichloromethane (THM_1), and Dibromochloromethane (THM_2) Species Concentrations (ppb) at Different Times

	Time (h)					
Method	0	4	8	12	20	24
			(a) THM_0			
Experiment	7.4	14.2	16.2	21.2	20.4	19.4
GRG2	8.35	12.56	15.73	18.08	21.11	22.07
GA	8.26	12.26	15.51	18.12	21.04	21.48
SA Case 1	7.30	12.11	15.62	17.64	20.18	20.61
SA Case 2	7.49	12.44	15.20	17.49	19.88	21.32
SA Case 3	7.18	12.03	15.52	17.45	19.93	20.87
SA Case 4	7.86	12.49	15.38	17.98	20.96	21.97
			(b) THM_1			
Experiment	4.4	4.6	3.4	3.6	3.2	9.6
GRG2	1.69	3.08	3.94	4.56	5.43	5.73
GA	1.69	3.90	4.54	4.48	5.82	7.40
SA Case 1	4.58	4.30	4.25	5.79	8.02	9.72
SA Case 2	4.05	3.41	5.36	6.21	8.81	7.79
SA Case 3	4.89	4.52	4.52	6.33	8.65	9.03
SA Case 4	3.05	3.28	4.87	4.83	5.86	5.98
			(c) THM_2			
Experiment	2.4	7	6.8	9.6	7.2	6.4
GRG2	4.19	4.93	5.96	6.99	8.63	9.24
GA	4.33	4.40	5.55	7.05	8.27	8.12
SA Case 1	2.37	4.15	5.77	6.19	6.96	6.70
SA Case 2	2.69	4.71	5.08	5.91	6.48	7.92
SA Case 3	2.16	4.02	5.58	5.87	6.61	7.14
SA Case 4	3.33	4.80	5.39	6.81	8.35	9.09

Note: GRG2=generalized reduced gradient method; GA=genetic algorithm; and SA=simulated annealing .

Table 2. Measured Bromine Distribution Factor (BDF) at Different Times

	Time (h)					
Measured BDF	0	4	8	12	20	24
S_0	0.5599	0.6004	0.6398	0.6475	0.6635	0.5747
S_1	0.2427	0.1418	0.0979	0.0802	0.0759	0.2073
S_2	0.1042	0.1698	0.1541	0.1682	0.1343	0.1088
S_3	0.0930	0.0880	0.1083	0.1040	0.1261	0.1093

Table 3. Temporal Values of the Objective Function Calculated Based on Eq. (12)

	Time (h)					
Methods	0	4	8	12	20	24
GRG2	2×10^{-8}	2×10^{-8}	3.1×10^{-5}	5×10^{-8}	5×10^{-8}	5×10^{-8}
GA	1.4×10^{-4}	1.7×10^{-7}	1.9×10^{-6}	8.5×10^{-7}	1.2×10^{-5}	2.9×10^{-6}
SA Case 1	1.6×10^{-8}	3.9×10^{-8}	7.2×10^{-11}	2.1×10^{-8}	2.9×10^{-7}	7.6×10^{-9}
SA Case 2	3.8×10^{-8}	2.4×10^{-8}	8.0×10^{-7}	5.9×10^{-7}	8.3×10^{-9}	1.6×10^{-11}
SA Case 3	4.1×10^{-11}	9.6×10^{-8}	1.2×10^{-7}	7.9×10^{-7}	9.9×10^{-7}	3.8×10^{-9}
SA Case 4	3.3×10^{-10}	4.8×10^{-8}	9.2×10^{-7}	1.6×10^{-11}	8.1×10^{-11}	1.6×10^{-7}

Note: GRG2=generalized reduced gradient method; GA=genetic algorithm; and SA=simulated annealing.

J. Comput. Civ. Eng. 2005.19:248-257.

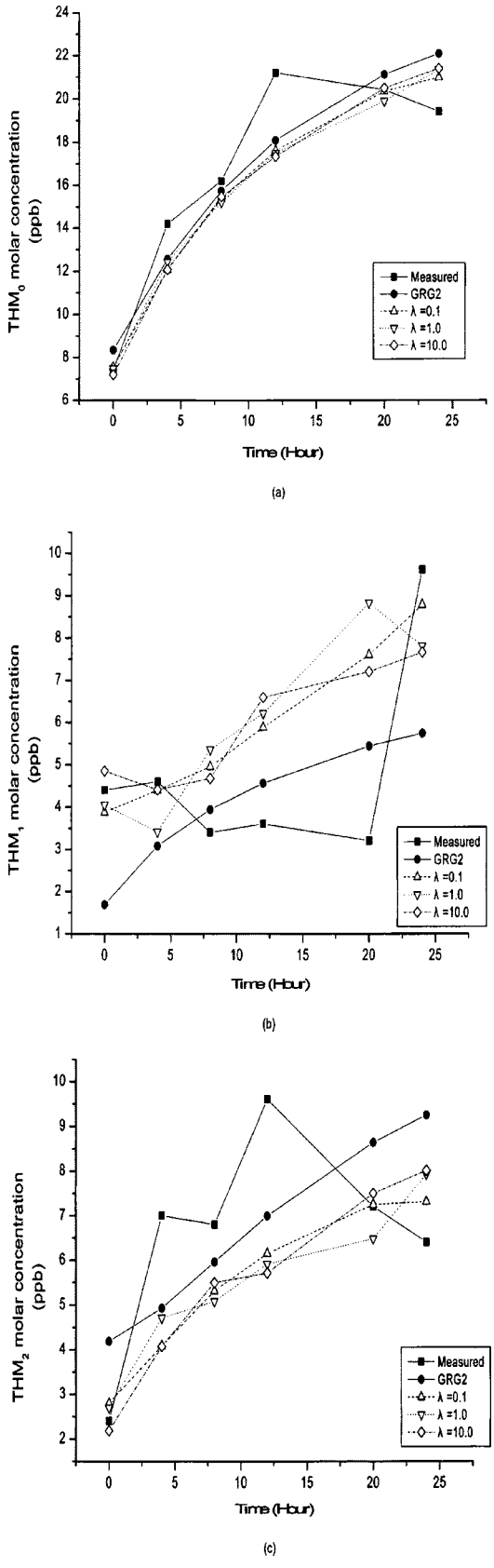

Fig. 6. (a) Chloroform concentration; (b) bromodichloromethane concentration; and (c) dibromochloromethane concentration obtained from simulated annealing when different values of the Lagrange multipliers are employed

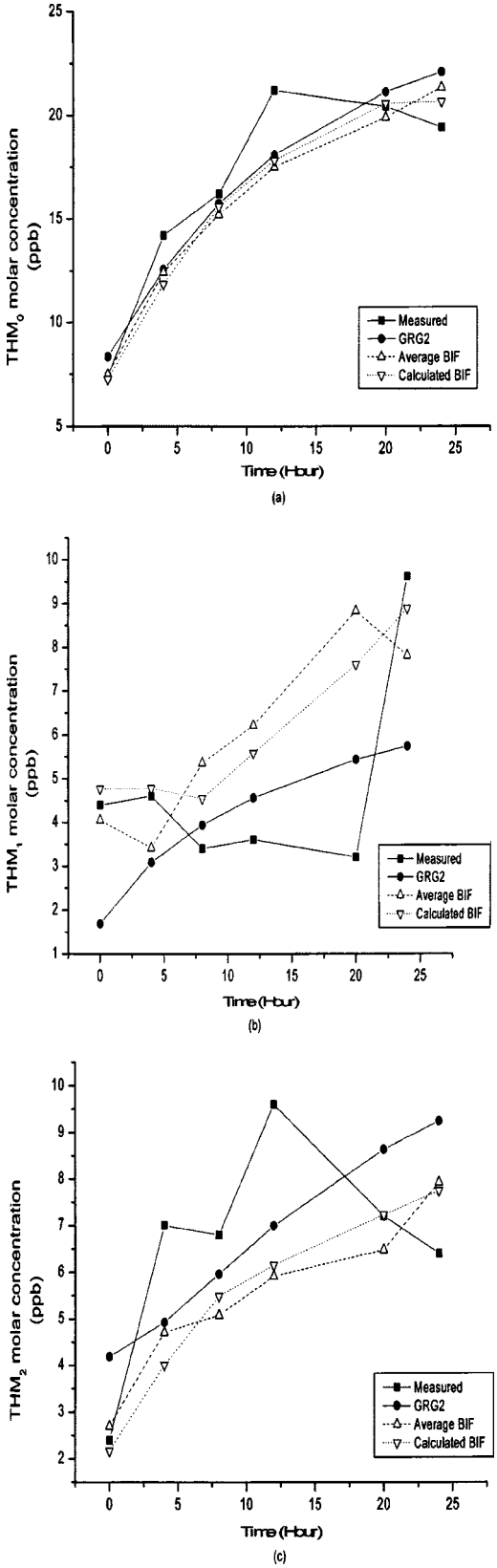

Fig. 7. Three trihalomethane species concentration calculated based on average bromine incorporation factor (BIF) and estimated calculated BIF. (a) Temporal concentration of chloroform concentration; (b) temporal concentration of bromodichloromethane concentration; and (c) temporal concentration of dibromochloromethane concentration

J. Comput. Civ. Eng. 2005.19:248-257.

Heuristic methods, developed very rapidly in recent years, are wildly employed in the research field. Both GA and SA were used to calculate the concentrations of THM species in desalinated water. The control parameters in the kinetics of THM species are the chlorine demand–[TTHM] growth coefficient, the bromoform reaction rate coefficient, and the average representative bromine incorporation. The present approach incorporates necessary constraints, such as the bromine distribution factor and BIF factor, in the optimization formulation as a means of better representation of the site-specific water quality characteristics.

The heuristic approaches of GA, general SA, and two modified SAs are employed to solve this optimization problem. These approaches have the merit of arbitrarily choosing the initial guess and, still, can obtain reasonably good results. The calculated concentrations of THM species by GA and SA differed from the measured values, yet those results satisfied the constraints of the species mass balance and are in close agreement with those results obtained from GRG2.

In this study, the objective function values obtained from GA are slightly inferior to those of GRG2. Yet, the calculated THM species concentrations obtained from GA do not differ significantly from GRG2. In Cases 1 and 2 of the SA approach, most results obtained from SA were about the same order of magnitude as those of GRG2. However, some values were better than those of GRG2. In Case 4, SA provides better results than those of GRG2 in terms of the objective function values. This indicates that increasing the searching number with decreasing temperature may yield better results.

The approach of using GA or SA coupled with the THM formation kinetic equations might provide a useful tool for analyzing the concentration of THMs and facilitate the ability of water quality management.

For problems where traditional approaches, such as gradient-type solvers or programming techniques, fail to obtain the optimal solution, the heuristic approaches might be an alternative. The future study for forecasting the quality of drinking water might include developing an online monitoring system to assure the safety of water usage. In addition, applying the heuristic approaches to identify the model parameters is also one of our ongoing researches.

Acknowledgment

This study was partially financially supported by the National Science Council of Taiwan under Grant No. NSC90-2621-009-003.

Notation

The following symbols are used in this paper:

- BIF = bromine incorporation factor;
- E = system energy;
- F = linear proportionality constant between total trihalomethane and chlorine;
- K_b = bromoform reaction rate coefficient;
- K_{cl} = chlorine reaction rate coefficient;
- k = Boltzmann's constant;
- L = total string length;
- l_i = substring length;
- P_c = crossover probability;
- P_m = mutation probability;
- P_{SA} = acceptance probability that when the trial solution doesn't better than the current best solution;
- $P(E)$ = occurrence probability when the system and temperature are E and T, respectively;
- S_0, S_1, S_2, S_3 = bromine distribution factors of chloroform, bromodichloromethane, dibromochloromethane, and bromoform, respectively;
- T = temperature;
- $[x]$ = x species concentration; and
- λ = Lagrange multiplier.

References

Aarts, E., and Korst, J. (1989). *Simulated annealing and Boltzmann machines: A stochastic approach to combinatorial optimization and neural computing*, Wiley, New York.

Adin, A., Katzhendler, J., Alkaslassy, D., and Ravacha, C. (1991). "Trihalomethane formation in chlorinated drinking water—A kinetic model." *Water Res.* 25(7), 797–805.

Amy, G. L., Chadick, P. A., and Chowdhury, Z. K. (1987). "Developing models for predicting trihalomethane formation potential and kinetics." *J. Am. Water Works Assoc.*, 79(7), 90–97.

Bitton, G. (1999). *Wastewater microbiology*, Wiley, New York.

Bunn, W. W., Haas, B. B., Deane, E. R., and Kleopfer, R. D. (1975). "Formation of trihalomethanes by chlorination of surface waters." *Environ. Lett.*, 10(3), 205.

Clark, R. M. (1998). "Chlorine demand and THM formation kinetics: A second-order model." *J. Environ. Eng.* 124(1), 16–24.

Clark, R. M., Smalley, G., Godrich, J., Tull, R., Rossman, L. A., Vasconcelos, J. J., and Boulus, P. F. (1994). "Managing water quality in distribution systems: Simulating TTHM and chlorine residual program." *J. Water SRT-Aqua*, 43(4), 182–191.

Coley, D. A. (1999). *An introduction to genetic algorithms for scientists and engineers*, World Scientific, Arbid, Mass.

Cunha, M. D. C., and Sousa, J. (1999). "Water distribution network design optimization: Simulated annealing approach." *J. Water Resour. Plan. Manage.* 125(4), 215–221.

Domer, B., Raphael, B., Shea, K., and Smith, I. F. C. (2003). "A study of two stochastic search methods for structure control." *J. Comput. Civ. Eng.* 17(3), 132–141.

Elshorbagy, W. E. (2000). "Kinetic of THM species in finished drinking water." *J. Water Resour. Plan. Manage.* 126(1), 21–28.

Goffe, W. L., Ferrier, G. D., and Rogers, J. (1994). "Global optimization of statistical functions with simulated annealing." *J. Econometr.* 60(1), 65–100.

Golfinopoulos, S. K., Xilourgidis, N. K., Kostopoulou, N., and Lekkas, T. D. (1998). "Use of a multiple regression model for predicting trihalomethane formation." *Water Res.* 32(9), 2821–2829.

Gould, J. P., Fitchhorn, L. E., and Urheim, E. (1981). "Formation of brominated trihalomethanes: Extent and kinetics." *Water chlorination: Environment impacts and health effects*, Vol. 4, R. L. Jolley et al., eds., Ann Arbor Science, Ann Arbor, Mich., 297–310.

Gupta, I., Gupta, A., and Khamma, P. (1999). "Genetic algorithm for optimization of water distribution systems." *Environ. Modeling, Software*, 14(5), 437–446.

Gwo, J. P. (2001). "In search of preferential flow paths in structured porous media using a simple genetic algorithm." *Water Resour. Res.* 37(6), 1589–1601.

Hillier, F. S., and Lieberman, G. J. (1990). *Introduction to operations research*, 5th Ed., McGraw-Hill, New York.

Holland, J. (1975). *Adaptation in natural and artificial system*, University Press, Cambridge, Mass.

Kirkpatrick, S., Gelatt, C. D., Jr., and Vecchi, M. P. (1983). "Optimization by simulated annealing." *Science*, 220(4598), 671–680.

J. Comput. Civ. Eng. 2005.19:248-257.

Lasdon, L. S., and Warren, A. D. (1982). *GRG2 user's guide*, Dept. of General Business, Univ. of Texas at Austin, Tex.

Lippai, I., Heaney, J. P., and Laguna, M. (1999). "Robust water system design with commercial intelligent search optimizers." *J. Comput. Civ. Eng.* 13(3), 135–143.

McKinney, D. C., and Lin, M. D. (1994). "Genetic algorithm solution of groundwater management models." *Water Resour. Res.* 30(6), 1897–1906.

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

Milot, J., Rodriguez, M. J., and Serodes, J. B. (2002). "Contribution of neural networks for modeling trihalomethanes occurrence in drinking water." *J. Water Resour. Plan. Manage.* 128(5), 370–376.

Minear, R. A. (1980). "Production, fate and removal of trihalomethanes in municipal drinking water systems." *Res. Rep. No. 78*, Tennessee Water Research Center, Knoxville, Tenn.

Montogomery Watson. (1993). "Mathematical modeling of the formation of THMs and HAAs in chlorinated natural waters." *Final Rep. submitted to the AWWA Disinfectant/Disinfection By-Products Technical Advisory Workgroup*, American Water Works Association, Denver.

Nokes, C. J., Fenton, E., and Randall, C. J. (1999). "Modeling the formation of brominated trihalomethanes in chlorinated drinking waters." *Water Res.* 33(17), 3357–3568.

Pham, D. T., and Karaboga, D. (2000). *Intelligent optimization techniques*, Springer, U.K.

Rook, J. J. (1974). "Formation of haloforms during chlorination of natural waters." *Water Treat. Exam.*, 23, 234–243.

Trussel, R. R., and Umphres, M. D. (1978). "The formation of trihalomethanes." *J. Am. Water Works Assoc.*, 70(11), 604–612.

Youssef, H., Sait, S. M., and Adiche, H. (2001). "Evolution algorithms, simulated annealing and tabu search: A comparative study." *Eng. Applic. Artif. Intell.* 14, 167–181.

J. Comput. Civ. Eng. 2005.19:248-257.