Pattern classification using tabu search to identify the spatial distribution of groundwater pumping

Ching-Pin Tung · Chun-An Chou

Abstract Pattern classification and Tabu Search are integrated to optimize the zonation and associated average groundwater pumping rates. A simulated problem analogous to the Choushui Creek Alluvial in Janghauh county of Taiwan is used to examine the proposed method. Three problems are tested including (1) optimizing zonation patterns with known pumping rates and number of zones, (2) optimizing pumping rates and zonation patterns with a predetermined number of zones, and (3) optimizing all pumping rates, number of zones, and zonation. The results indicate that the proposed method successfully identifies spatial patterns of groundwater pumping for all three cases. However, some initial zonations result only in local optimums. Thus, using more than one initial zonation to increase search diversification is recommended. Application to different problems such as the spatial distribution of hydraulic conductivity values and area weighting of rainfall stations, is encouraged.

Résumé La classification des patrons et la recherche Tabu sont intégrés pour optimiser la zonation et la moyenne associée du taux de pompage de l'eau souterraine. Une région analogue à l'alluvion du ruisseau Choushui de la région Janghauh en Taiwan a été recrée afin d'examiner la méthode proposée. Trois problèmes sont évalués, incluant l'optimisation de (1) la zonation avec seulement un taux de pompage et un nombre de zones donnés, (2) le taux de pompage et la zonation avec un nombre prédéterminé de zones, et (3) tous les taux de pompage, nombre de zones et la zonation. Les résultats indiquent que la méthode proposée identifie avec succès les patrons de pompage de l'eau souterraine dans les trois cas. Toutefois, certaines zonations initiales résultent seulement en optimums locaux. Par conséquent, l'utilisation de plus d'une zonation initiale afin d'augmenter la diversification de la recherche est recommandée. L'utilisation de la méthode est suggérée pour la résolution de différents problèmes tels que la distribution spatiale de la conductivité hydraulique et la pondération des stations de mesure de précipitations.

Resumen El patrón de clasificación y la búsqueda tabú están integrados para optimizar la zonificación y las tasas promedio de bombeo de aguas subterráneas asociadas. Se ha utilizado un área diseñada análoga al arroyo aluvial Choushui en la provincia Janghauh en Taiwan para examinar el método propuesto. Se han probado tres problemas que incluyen (1) la zonificación únicamente con las tasas de bombeo y número de zonas dadas, (2) tasa de bombeo y zonificación con un número predeterminado de zonas, y (3) todas las tasa de bombeo, número de zonas y zonificaciones. Los resultados indican que el método propuesto identifica exitosamente los patrones espaciales de bombeo de las aguas subterráneas en los tres casos. Sin embargo, Algunas zonificaciones iniciales solamente resultan en óptimos locales. Por esta razón se recomienda utilizar más de una zonificación inicial para incrementar la diversificación de la búsqueda. Se sugiere la aplicación a problemas diferentes tales como la distribución espacial de conductividades hidraúlicas y la ponderación de áreas de estaciones de monitoreo de precipitación.

Keywords Spatial analysis · Tabu search · Simulation · Optimization · MODFLOW

Introduction

A method for identifying spatial patterns of groundwater pumping is very necessary for groundwater management in land subsidence areas in Taiwan. Fishponds are widely distributed in the southwestern coast of Taiwan. Groundwater is the major water resource due to its stable quantity, quality, and temperature. Groundwater has been over-pumped and land subsidence has resulted in these areas. It is estimated that there are more than ten thousand wells in each county in the southwestern coast, and almost all of the wells are illegal; thus their capacities and locations are not well known. To mitigate land subsidence, groundwater simulation models are important tools for

Received: 10 November 2002 / Accepted: 19 March 2004

C.-P. Tung (✉) · C.-A. Chou
Bioenvironmental Systems Engineering,
National Taiwan University,
Taiwan 1, Sec. 4, Roosevelt Rd., 10617 Taipei, Taiwan
e-mail: cptung@ntu.edu.tw
Tel.: +886-2-23620327

DOI 10.1007/s10040-004-0344-2

evaluating management strategies, but they need the spatial pattern of groundwater pumping as input. Thus, this study proposes a method to identify the spatial pattern of groundwater pumping using the theory of pattern classification and the tabu search (TS) algorithm.

There are a variety of methods for determining hydraulic parameter values. If the values at some locations in a study site are known, a spatial interpolation method such as the kriging method or a zonation method such as the Thiessen method can be applied. The kriging method requires training. Although some software provides the kriging method to interpolate parameters, an inadequate application may be made without carefully examining whether assumptions of the algorithm are met. On the other hand, the Thiessen method results in a geometric zonation with little physical significance, and it does not guarantee obtaining the best simulation results. If there are no reference parameter values available, inverse analysis (Yeh 1975) can be applied to determine the values. Yang and Yeh (1998) assigned parameters values according to geologic characteristics and then optimized their values. However, the spatial geologic features are difficult to define precisely. Consequently this approach is uncertain. Tung and Chou (2002) applied TS to optimized spatial patterns of hydraulic conductivity values, but their method also requires reference values of hydraulic conductivity. In general, the above-mentioned methods are difficult to use for identifying the spatial pattern of groundwater pumping if no reference pumping rates are available.

The identification of the pumping rate for each well in a region is not realistic and not necessary. Instead, it is more practical to determine an average pumping rate for each zone. Thus, this study develops a procedure to optimize zonation and its associated average pumping rates. The goal of the zonation is to obtain simulated hydraulic heads that are as close to the observed ones as possible when other hydrogeologic parameter values are known. The optimized zonation and associated pumping rates represent the spatial pattern of groundwater withdrawal. The problem then is how to find the optimal zonation.

Heuristic algorithms have been widely applied to search for optimal solutions for different research areas (Dougherty 1991; McKinney and Lin 1994; Pan and Wu 1998; Zheng and Wang 1999). The most-used methods include genetic algorithms (Holland 1975), and simulated annealing (Černy 1985; Kirkpatrick et al. 1984; Metropolis et al. 1953), and tabu search (Glover 1989). These algorithms have their own characteristics for approaching the global optimum for a nonlinear problem. Zheng and Wang (1996, 1999) have applied tabu search to parameter-structure identification and remediation-system design. To make the next move, tabu search evaluates all neighboring solutions, which is more efficient when the size of the neighborhood is manageable. Thus, in this study it was decided to optimize the zonation of groundwater pumping. Both the genetic algorithms and simulated annealing methods are expected to be able to take on the role of tabu search; however, the two methods are not discussed any further in this report.

The purpose of this study is to propose a procedure for optimizing zonation and associated groundwater pumping rates by integrating tabu search and pattern classification. Both the number of zones and their average pumping rates are decision variables. Reference values as in Tung and Chou's study (2002), which will make the proposed method more applicable to the problem, are not required.

Methodology

The probability density function (PDF) of normal distribution is used as a membership function of a pumping pattern, i.e., an average pumping rate. Then, tabu search (TS) is applied to optimize parameter values of the functions to determine spatial pumping patterns. The following discussions describe how the probability density functions are used to identify spatial patterns and how TS is integrated to optimize the zonation of groundwater pumping. The following methodology and later application is considered only for two-dimensional and isotropic conditions.

Pattern Classification

Patterns can be classified based on the values of membership functions (Schürmann 1996). For example, in a one-dimensional problem with two types, two density functions of normal distribution can be used to separate two groups. The satisfaction of location X belonging to pattern i can be described as the following:

$$N_i(X, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{X-\mu_i}{\sigma_i}\right)^2\right] \quad i = \mathrm{I}\ \&\ \mathrm{II} \tag{1}$$

where X is the position of a grid point, μ_i is the mean and the location of the centroid of pattern i, and σ_i is the standard deviation. Point X belonging to pattern I or II is determined by its values of the two density functions as shown in Fig. 1a. If the first function yields a higher value, then it belongs to the first pattern. Otherwise, it is the second pattern. Different values of (μ_{I}, σ_{I}) and (μ_{II}, σ_{II}) will result in different spatial distribution of patterns I and II.

Similarly, a two-dimensional problem can use two-dimensional density function of normal distribution to determine patterns as follows:

$$N_i(X, Y, \mu_{Xi}, \mu_{Yi}, \sigma_{Xi}, \sigma_{Yi}) = \frac{1}{2\pi\sigma_{Xi}\sigma_{Yi}} \exp\left\{-\frac{1}{2}\left[\left(\frac{X-\mu_{Xi}}{\sigma_{Xi}}\right)^2 + \left(\frac{Y-\mu_{Yi}}{\sigma_{Yi}}\right)^2\right]\right\} \tag{2}$$

A grid point P(X, Y) can be classified to pattern k, if the N_k yield the maximum value among n functions as seen in the following:

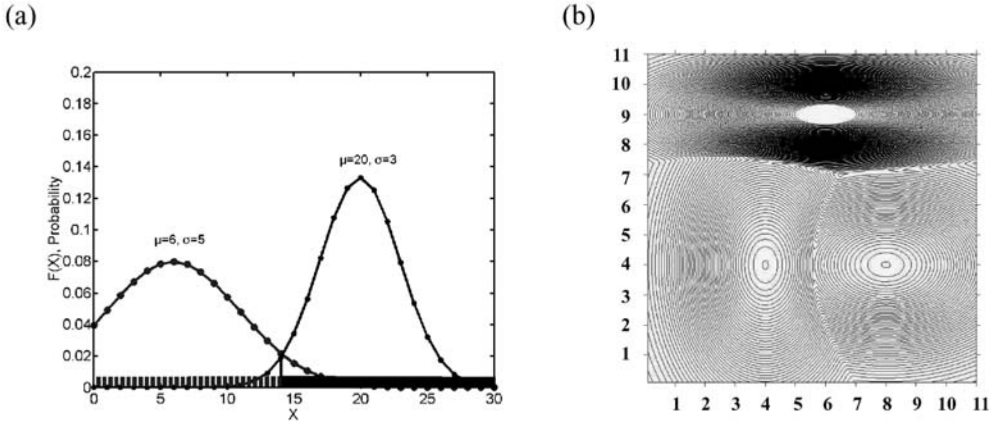


Fig. 1 Spatial patterns classified by the probability density functions of normal distribution: **a** one-dimension and **b** two-dimension

$$P(X, Y) \in \text{Pattern k, if } N_{\text{k}} = \text{Max}[N_{\text{I}},\ N_{\text{II}},\ \cdots N_{\text{k}},\ \cdots,\ N_{\text{n}}] \tag{3}$$

According to the above principles, there exists one parameter set of (μ_{Xi}, μ_{Yi},σ_{Xi} σ_{Yi}) that determines the spatial patterns for all type i. The normal PDF has an infinite tail to cover the whole area and has fewer parameters, and is enough to determine a reasonable zonation. With proper adjustments of the parameters, it can produce many combinations of spatial patterns. As shown in Fig. 1b, curvilinear boundaries can also be produced if the assigned values for σ_{Xi} and σ_{Yi} are different. Other PDFs may also produce good results, but they are not tested in this study. Future research is encouraged to establish principles to determine the best PDF to be a membership function.

Tabu Search

Tabu search (TS) is applied to optimize the parameters of (μ_{Xi}, μ_{Yi}, σ_{Xi} σ_{Yi}) and average pumping rates to determine the spatial patterns of groundwater withdrawal which can result in the best simulation of hydraulic heads. TS uses a tabu list that records better solutions of the recent iterations so that these solutions will not be used again within a given number of iterations, which can avoid being trapped by a local optimum.

Tabu search finds the optimal solution by moving from the current solution to the best solution in its neighborhood. If the best solution in the neighborhood is not prohibited by the tabu list, the current solution is replaced. TS keeps moving from one solution to its neighbor until stop criteria are met. The main components of TS include the initial solution, movement and neighborhood, tabu list, aspiration criteria, and principles for stopping the search. These components are briefly introduced here. More detailed descriptions of TS can be seen in Glover and Laguna (1997).

Initial solution

An initial solution should be given before the search process. Different initial solutions may lead to different local optimal solutions in a nonlinear problem. Initial solutions could be randomly generated. Thus, a robust algorithm needs to result in minimum deviation of the optimums which are found from different initial solutions.

Neighborhood and movement

During the search process, the current solution ***X*** is directed to move to one solution ***X'*** in its neighborhood, *N(**X**)*. A neighborhood is defined by movement, i.e., the changes of decision variables. Different definitions of neighborhood may be made by different analysts and made for different problems. After defining the neighborhood, the best solution in the neighborhood can be selected as a new current solution for the next iteration.

Tabu list

The tabu list has four types of memory, including recency, frequency, quality, and influence (Glover and Laguna 1997). The recency memory is adopted to design the tabu list in this study. The tabu list records better solutions in recent iterations and prohibits moving back to these solutions in order to avoid being trapped by local optimums. The length of the tabu list should be long enough to avoiding falling into a search cycle. On the other hand, the length of the tabu list cannot be too long because it may limit the search space to a small feasible region and could reduce its efficiency. There is no universal principle for determining the length of the tabu list. Glover (1993) suggested an arbitrary number of seven that can be used for the first try.

Aspiration criterion

The tabu list records better solutions of previous iterations, and it is normally designed to record only partial attributes of those solutions. For example, a solution set contains two decision variables X and Y. If changing $X1$ to $X2$ with Y=$Y1$ finds the best solution among the neighborhood of $N(X1, Y1)$, the tabu list will record that X=$X1$ is tabued. However, if Y changing to $Y2$ with X=$X1$ has the better value of objective function, the tabu should be relaxed and the solution is allowed to move to ($X1$, $Y2$). Thus, the aspiration criterion allows moving to a solution containing tabued attributes, if its value of the objective function is better than the current recorded best value.

Principles for stopping the search process

Different principles can be applied for stopping the search process and may include (1) having the recorded optimal solution reach the required levels, (2) having given maximum iterations for searching neighborhoods; (3) searching the number of successive iterations for those solutions which have not improved. The selection of the criterion that is most proper for a problem is a care-by-care decision.

The tabu search also records the optimal solution in the previously explored solution space. If the new selected solution is better than the recorded solution, the record is renewed. When the algorithm stops searching, the final recorded solution is the selected optimal solution.

Problem Formulation and Search Procedure

The technique of pattern classification is applied to identify the spatial distribution of groundwater pumping when all available hydrogeologic information is known. The identified spatial patterns of groundwater pumping can result in the minimal difference between simulated hydraulic heads and observed ones, and thus mean square error is used as the objective function. The mathematical model is formulated as follows:

$$\text{Min. } Z = \sqrt{\sum_{l=1}^{\mathrm{M}} \sum_{t=1}^{\mathrm{T}} \left(h_{lt} - h_{lt}^{o}\right)^2 \Big/ \mathrm{M}} \tag{4}$$

subject to

$$S\frac{\partial h}{\partial t} = \mathrm{T}_{xx}\frac{\partial^2 h}{\partial x^2} + \mathrm{T}_{yy}\frac{\partial^2 h}{\partial y^2} - \varpi$$

$$\varpi \in \{w_1,\ w_2,\ \cdots,\ w_k,\ \cdots,\ w_N\} \tag{5}$$

$$\begin{aligned} \varpi_j = w_k \quad & if\ N_k \\ & = \mathrm{Max}\{N_1,\ N_2,\ \cdots,\ N_k,\ \cdots,\ N_N\} \end{aligned}$$

$$N_i(X, Y, \mu_{Xi}, \mu_{Yi}, \sigma_{Xi}, \sigma_{Yi}) = \frac{1}{2\pi\sigma_{Xi}\sigma_{Yi}} \exp\left\{-\frac{1}{2}\left[\left(\frac{X-\mu_{Xi}}{\sigma_{Xi}}\right)^2 + \left(\frac{Y-\mu_{Yi}}{\sigma_{Yi}}\right)^2\right]\right\} \tag{6}$$

where M is the number of observation wells, T is the period within a year, T_{xx} and T_{yy} are the transmissivity (T_{xx}=k_{xx}×b, T_{yy}=k_{yy}×b, k_{xx}=hydraulic conductivity, b= thickness of the confined aquifer), S is the storage coefficient, h_{lt} and h_{lt} are simulated and observed hydraulic heads at location l and time period t, ϖ is a set of average pumping rates for the time period t=1 to T, and

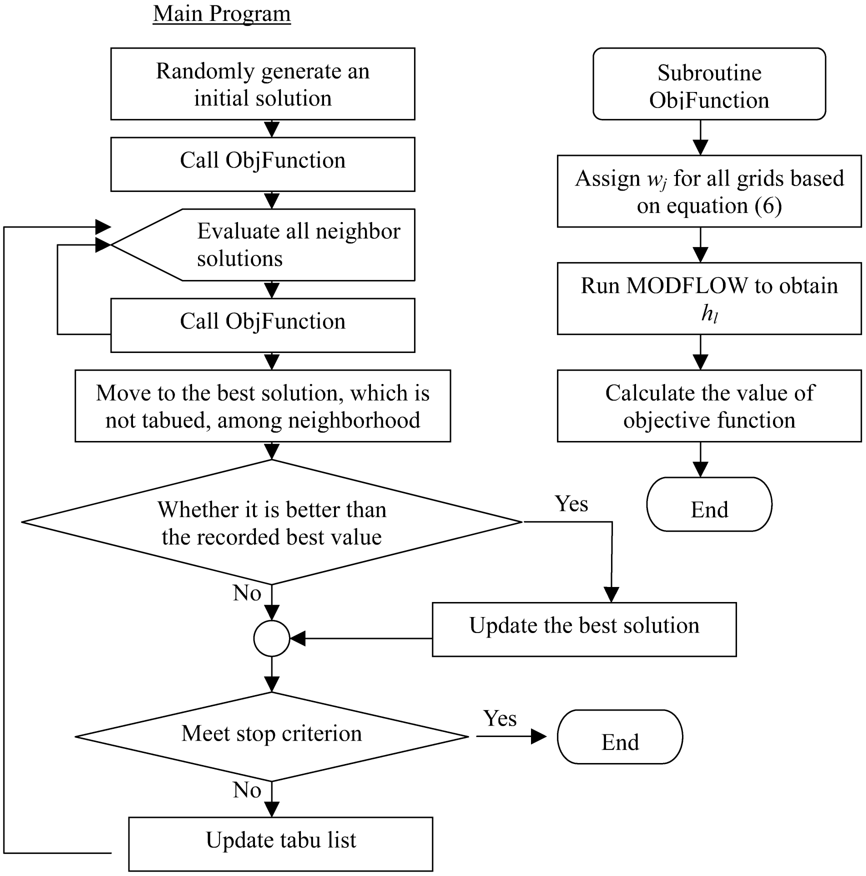


Fig. 2 Flowchart to determine optimal spatial patterns of pumping rates with tabu search and pattern classification

$\{w_1, w_2, \cdots, w_N\}$ are possible average pumping rates for N patterns. The average pumping rates are also variables since real values are always unknown. The hydraulic heads can be simulated by the model MODFLOW (McDonald and Harbaugh 1988) with given hydrogeologic parameter values and pumping rates. The average pumping rate w_k is assigned to a grid j as in Eq. (6), if the grid belongs to pattern k. The decision variables of the above model are w_i and $(\mu_{xi}, \mu_{yi}, \sigma_{xi}, \sigma_{yi})$ for pattern i=1, ..., N. TS can help to find optimal values of these variables.

For a given number of patterns, the optimal solution can be found following the flowchart in Fig. 2. However, both the number of patterns and average pumping rates are unknown in most cases. An iterative procedure is designed to solve such problems. The procedure is described as follows:

1. The number of patterns is initially set to N=2
2. Initial values of w_i and $(\mu_{xi}, \mu_{yi}, \sigma_{xi}, \sigma_{yi})$, i=1, ..., N are determined
3. TS is applied to adjust w_i and $(\mu_{xi}, \mu_{yi}, \sigma_{xi}, \sigma_{yi})$ to find the best simulated hydraulic heads following the flowchart in Fig. 2
4. Let N=N+1. Go to step 2 if there is still significant improvement, i.e., $Z_N\_Z_{N-1} \geq \xi$. The ξ is the given convergent threshold.

Design of Problem

To verify the proposed method, a designed example problem, instead of a real case, is given to avoid the uncertainty regarding the hydrogeologic information. Hydraulic conductivity and specific storage coefficient can be obtained from pumping tests and concurrent water-level measurements in observation wells, but their spatial distributions are often not well known in a field study. Such uncertainty may cause bias in the simulation. The spatial distribution of values of hydraulic conductivity and specific storage coefficient are designed as described in this study, which allows one to focus on verifying whether the proposed method can successfully identify the spatial pattern of the various pumping rates.

Example of Designed Problem

A confined aquifer for which the amount of groundwater pumping and values of the hydraulic conductivity and specific storage coefficient are known, is designed as an example. The boundary conditions for the area are analogous to the Choushui Creek Alluvial in Janghauh County of Taiwan, for which the upper, lower and left boundaries have specific heads, and the right boundary is a constant-flux boundary. The thickness of the aquifer is designed as 50 m. Values of hydraulic conductivity and specific storage coefficient of five zones are given in Table 1 and the zonation is shown as Fig. 3a. The area is divided into 121 grids and each grid represents 2 km x 2 km. The values of the hydraulic conductivity and specific storage of grids within the same zone are the same. Three zones with different average pumping rates are designed. The zonation and average pumping rates are in Fig. 3b and Table 2, respectively. Hydraulic heads simulated by MODFLOW, based on these simulated conditions, are considered to represent observations.

Next, the spatial distribution of groundwater pumping is assumed to be unknown, and the proposed procedure is applied to direct the initial zonation toward finding the optimal one with the objective function of minimizing mean square error between observed and simulated heads.

Table 1 Hydraulic conductivities and specific storage coefficients for five zones

Zone	Hydraulic conductivity (m/s)	Specific storage coefficient (1/m)
1	0.0004	0.0007
2	0.0003	0.0005
3	0.0007	0.0001
4	0.0003	0.0003
5	0.0001	0.00002

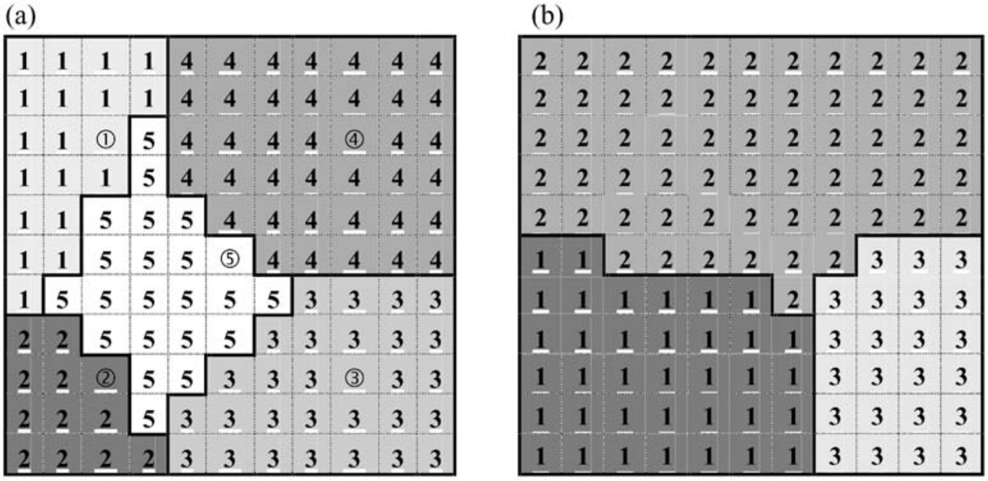


Fig. 3 a Simulated zonation of hydraulic conductivity and specific storage coefficient; *circled numbers* are the location of five observation wells. **b** Simulated zonation of average pumping rate

(a)

1	1	1	1	4	4	4	4	4	4	4
1	1	1	1	4	4	4	4	4	4	4
1	1	①	5	4	4	4	4	④	4	4
1	1	1	5	4	4	4	4	4	4	4
1	1	5	5	5	4	4	4	4	4	4
1	1	5	5	5	⑤	4	4	4	4	4
1	5	5	5	5	5	5	3	3	3	3
2	2	5	5	5	5	3	3	3	3	3
2	2	②	5	5	3	3	3	③	3	3
2	2	2	5	3	3	3	3	3	3	3
2	2	2	2	3	3	3	3	3	3	3

(b)

2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
1	1	2	2	2	2	2	2	3	3	3
1	1	1	1	1	1	2	3	3	3	3
1	1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	1	1	3	3	3	3
1	1	1	1	1	1	1	3	3	3	3

DOI 10.1007/s10040-004-0344-2

Table 2 Simulated monthly groundwater pumping rates for three zones

Month	Average pumping rate (m^3/s)		
	Zone 1	Zone 2	Zone 3
1	0.02	0.01	0.02
2	0.02	0.01	0.02
3	0.02	0.01	0.02
4	0.02	0.01	0.02
5	0.03	0.04	0.05
6	0.03	0.04	0.05
7	0.03	0.04	0.05
8	0.03	0.04	0.05
9	0.03	0.04	0.05
10	0.03	0.04	0.05
11	0.02	0.01	0.02
12	0.02	0.01	0.02

Five water-level monitoring wells located within the area are used as check points.

Application of Tabu Search

In order to apply TS to identify the optimal zonation, several types of information have to be specified first, including the initial zonation, the neighborhood, and the tabu list. The other components of TS are considered the same as general principles.

Initial solution

Several initial solutions are given for the problem. The decision variables are μ_{xi}, μ_{yi}, σ_{xi}, σ_{yi}, w_{it} for each pattern i and time period t, and their values can be assigned randomly in reasonable ranges.

Movement and neighborhood

The movement is defined to adjust one unit of a decision variable. Each variable has two changing directions. If there are N patterns, each pattern contains four parameters and T periods of pumping. Thus, there are $2\times N\times(4+\mathrm{T})$ neighbor solutions. For example, if there are three patterns and each pattern has two pumping periods, the neighborhood contains $2\times3\times(4+2)=36$ solutions.

Tabu list

The tabu list is designed to be $T_i=(\mu_x, \mu_y, \sigma_x, \sigma_y, w_1, w_2)$ for two pumping periods to record the altered pattern. The length of the tabu list initially can be given as the number of decision variables or neighborhoods. If the search process is dragged into a local optimum, a cycle of the value of the objective function can be observed and then the length of the tabu list has to be made longer than the length of the cycle.

Results and Discussions

Three simulation conditions were considered. The first case (case A) assumed only that the zonation pattern is unknown, while the second case (case B) adds pumping rates as decision variables. The last case (case C) is the most flexible in which the number of zones, the zonation pattern, and their associated pumping rates are unknown.

Case A: Known Number of Zones and Average Pumping Rates

In this case, the undetermined feature is the area of each zone. Thus, decision variables are $(\mu_{xi}, \mu_{yi}, \sigma_{xi}, \sigma_{yi})$ where

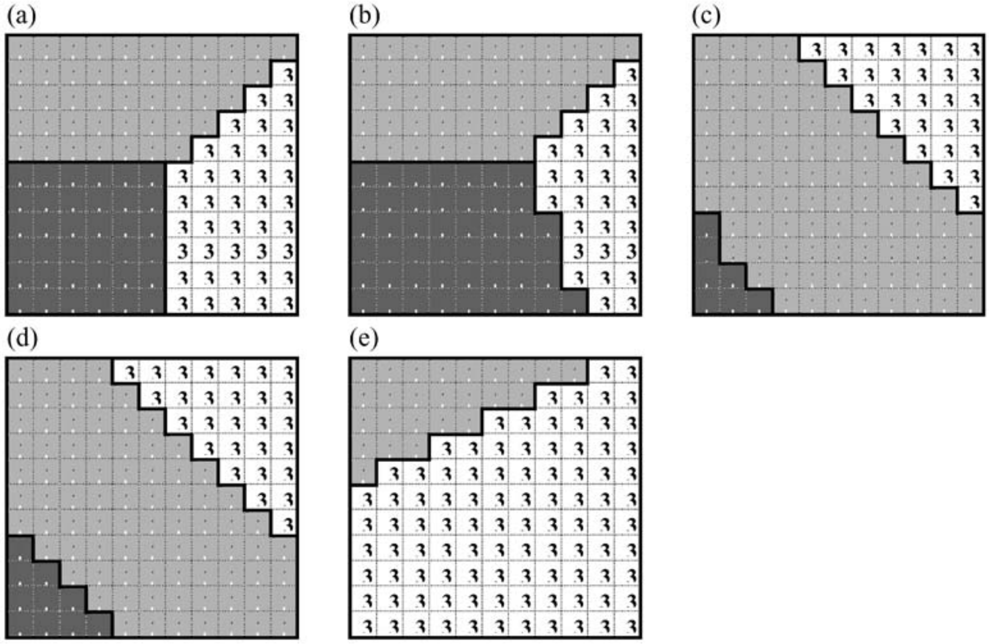


Fig. 4 Five initial zonations, case A

DOI 10.1007/s10040-004-0344-2

Table 3 Simulation results based on five initial zonations and different lengths of the tabu list for case A

	A1	A2a	A2b	A3a	A3b	A4	A5
Total iterations	400	400	400	500	500	500	500
Length of tabu list	12	12	24	12	24	12	12
Mean square error	0	0.059	0	0.022	0.054	0	0.022
Cycling	No	Yes	No	Yes	No	No	No

i=1, 2, and 3. Five initial zonations are given as Fig. 4, and results are summarized in Table 3.

Initially, the assigned length of the tabu list is 12. Two (case A1 and A4) out of five initial zonations are directed to the optimal one. Figure 5 shows simulation results of the hydraulic heads of case A1, in which the simulated hydraulic heads of five wells coincided with the observed ones. Cases A2a and A3a fall into cycling, and thus the length of the tabu list is increased to 24, which is equal to the period of cycling. Case A2b indicates that the initial zonation 2 with the longer tabu list can result in finding the global optimum. Initial zonations 3 and 5 fail to be directed to optimal zonation in exploring very limited solution space—12,000 (500×24) out of about 1.2 billion (11^6×3^6) solutions. If more iterations with the proper length of the tabu list are allowed, it can be expected that the global optimal solution will be found.

Case B: Known Number of Patterns Only

In this section, average pumping rates are assumed to be unknown and added as decision variables. Two average pumping rates were considered for wet and dry seasons in a year. Thus, decision variables are (μ_{xi}, μ_{yi}, σ_{xi}, σ_{yi}, w_{1i}, $w_{2\mathrm{i}}$) and i=1, 2, and 3. The size of the solution space will increase from 1.2 billion to about 20 thousand billion (11^6×3^6×5^6).

Table 4 lists the results based on five initial zonations as in Fig. 4, and randomly given initial pumping rates. The results indicate that both initial zonations 1 and 3 can be directed to find the global optimal solution, but the other three zonations are not. There is no cycling observed for all five cases. Thus, the reason that they can not find the optimal solution may be due to the limited size of the explored space. The number of explored solutions for each case is 14,400 [400×2×3×(4+2)] solutions, which is far smaller than the whole solution space.

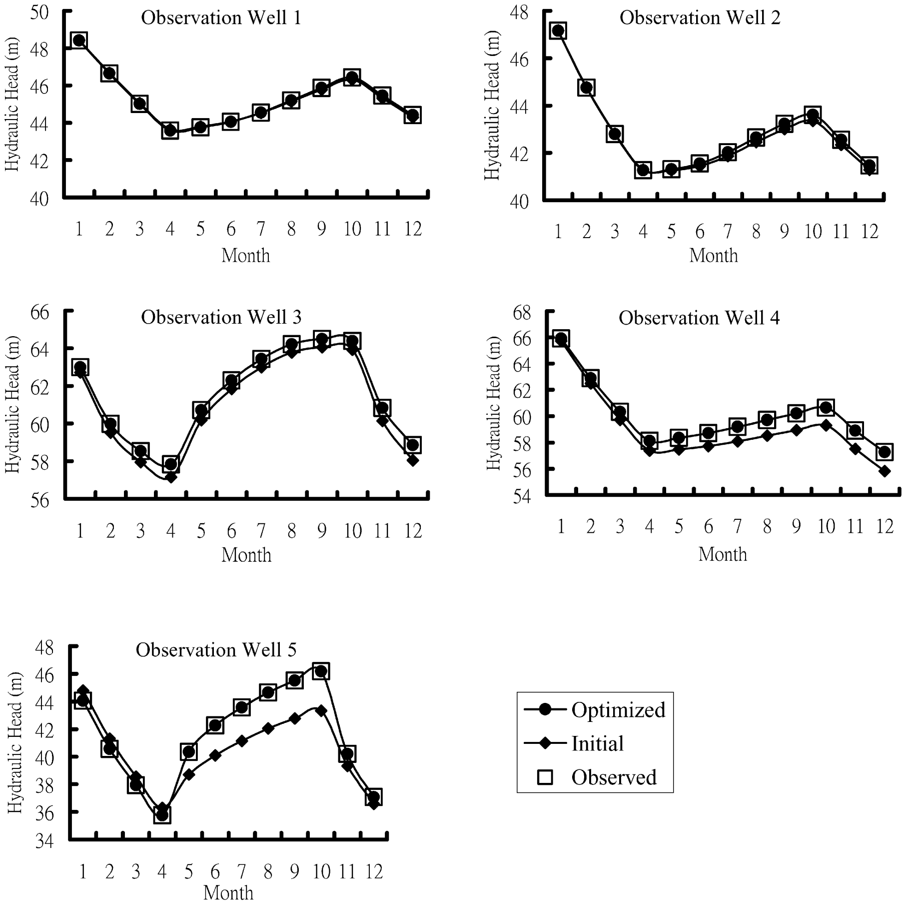


Fig. 5 Observed and simulated heads of five observation wells based on the first initial and optimized zonations in case A1, respectively. Hydraulic head values are in meters above an arbitrary datum

DOI 10.1007/s10040-004-0344-2

Table 4 Simulation results based on five initial zonations, case B

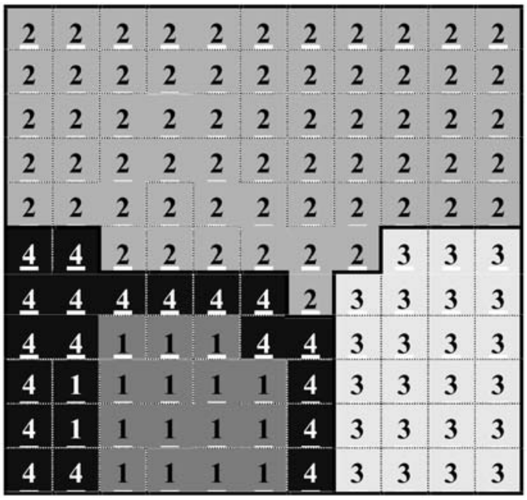
	B1	B2	B3	B4	B5
Total iterations	400	400	400	400	400
Length of tabu list	18	18	18	18	18
Mean square error	0	0.17	0	0.21	0.44
Cycling	No	No	No	No	No

Case C: Number of Patterns, Zonation, and Pumping Rates are Unknown

In most of the problems, the number of patterns (*N*), the zonation, and associated average pumping rates are unknown. The procedure given in the section, "Problem Formulation and Search Procedure", is applied to solve such problems. If the optimal number of patterns is N^* and given that $N<N^*$, the global optimal solution can not be found. The previous two sections have shown that if $N=N^*$, the global optimal solution can be obtained. If the given number of patterns *N* is equal to N^*+M (M>0), the spatial distribution is still dominated by the N^* patterns. The rest of the *M* patterns will not result in new zones or with different pumping rates due to one or more of following conditions:

1. The *M* membership functions will be pushed out of the study area or their values are totally smaller than the other functions within the study site
2. There are more than N^* zones, but only N^* pumping rates will be assigned, i.e., there will be some zones having the same pumping rates

The following simulation results provide evidence for the above conclusions. The simulated problem has three zones. The proposed method is further tested as to whether the simulated zonation can be identified when two or four membership functions are given. Each is tested with three different initial zonations. Cases C1–C3 in Table 5 are the results of giving two membership functions. The best values of the objective function for the three cases are only close to 2.5, and zonation consisting of two zones is different from the simulated one. Cases C4–C6 in Table 5 show the results of giving four member functions. The first initial zonation is successfully directed to the simulated zonation, but the other two failed. The possible reasons for failure are that there are not enough iterations and the length of the tabu list is too short. Figure 6 shows the optimal zonation of the average pumping rate of case C4; because zones 1 and 4 have the same average pumping rates, they represent only one zonation pattern.



2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
4	4	2	2	2	2	2	2	3	3	3
4	4	4	4	4	4	2	3	3	3	3
4	4	1	1	1	4	4	3	3	3	3
4	1	1	1	1	1	4	3	3	3	3
4	1	1	1	1	1	4	3	3	3	3
4	4	1	1	1	1	4	3	3	3	3

Fig. 6 Optimal zonation of case C4

Conclusions

A procedure integrating pattern classification and tabu search is proposed to identify the spatial pattern of groundwater pumping rates. The procedure can be applied for optimizing (1) zonation only with predetermined pumping rates and number of zones; (2) pumping rates and zonation with predetermined number of zones; and (3) all pumping rates, number of zones, and zonation. The results indicate that the procedure successfully identifies the spatial pattern of groundwater pumping. It is worth noting that some initial zonations may not be directed to the optimal solution due to limited searching iteration or improper length of the tabu list. More than one initial zonation is recommended for increasing diversification of the search. The procedure is successfully applied to a simulated problem based on real conditions, and will be applied to the Choushui Creek Alluvial in Taiwan in a future study. However, more uncertainties can be expected in a field study due to the sparsity of hydrogeologic information.

The probability density functions (PDF) are used for determining spatial patterns. During the optimization process, it was noticed that small changes of parameter values of PDF may not significantly influence spatial patterns and thus limit the search efficiency. Many neighbor solutions, which have similar values of decision variables, result in the same zonation and thus cause slow improvement of the value of objective function. The value of the objective function is improved quickly in early iterations and then changes very slowly in subsequent iterations. Intensification and diversification are two impor-

Table 5 Simulation results for case C based on three different initial zonations. Cases C1–C3 given two membership functions and C4–C6 given four membership functions

	C1	C2	C3	C4	C5	C6
Total iterations	400	400	400	400	400	400
Length of tabu list	24	24	24	32	32	32
Mean square error	2.53	2.53	2.53	0	0.47	0.14
Cycling	Yes	No	Yes	No	No	No

DOI 10.1007/s10040-004-0344-2

tant strategies in searching for the optimal solution. The changes in values of the PDF parameters can be larger to increase diversification at an early stage and gradually reduced for intensification.

The procedure can be applied to different problems of pattern classification such as the spatial distribution of values of hydraulic conductivity and area weighting of rainfall stations. On the other hand, different heuristic algorithms such as simulated annealing and genetic algorithm, can be applied to replace tabu search. However, the question of which algorithms work more efficiently was not tested in this study. Besides, when the proposed method is applied to a real field, an objective function for minimizing structure errors may need to take the uncertainty in relation to field conditions into account (Sun et al. 1998 ; Tsai et al. 2003).

References

Černy V (1985) Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. J Optimiz Theory Appl 45:41–51

Dougherty DE (1991) Optimal groundwater management 1. Simul Anneal 27(10):2493–2508

Kirkpatrick S, Gelatt CD, Jr., Vecchi MP (1984) Optimization by simulated annealing. Science 220:671–680

Glover F (1989) Tabu search, part I. ORSA J Compu 1:190–206

Glover F (1993) A user's guide to tabu search. Ann Oper Res 41: 3–28

Glover F, Laguna M (1997) Tabu search. Kluwer, Boston, Massachusetts

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, Michigan

McKinney DC, Lin MD (1994) Genetic algorithm solution of groundwater management models. Water Resour Res 30(6): 1897–1906

McDonald MG, Harbaugh AW (1998) A modular three-dimensional finite-difference ground-water flow model. Techniques of Water Resources Investigations, 06-A1, U.S. Geological Survey, Reston, Virginia

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21(6):1087–1092

Pan L, Wu LS (1998) A hybrid global optimization method for inverse estimation of hydraulic parameters: annealing-simplex method. Water Resour Res 34(9):2261–2269

Schürmann, J (1996) Pattern classification: a unified view of statistical and neural approaches. Wiley, New York

Sun NZ, Yang SL, Yeh WWG (1998) A proposed stepwise regression method for model structure identification. Water Resour Res 34(10):2561–2572

Tsai FTC, Su NZ, Yeh WWG (2003) A combinatorial optimization scheme for parameter structure identification in groundwater modeling. Ground Water 41(2)156–169

Tung CP, Chou CA (2002) Application of tabu search to groundwater parameter zonation. J Amer Water Res Assoc 38(4): 1115–1126

Yang SL, Yeh WWG (1998) Formulation and evaluation of alternatives for mitigating groundwater overdraft in Taiwan's coastal area. Water Resources Bureau, MOEA, Taipei

Yeh WWG (1975) Aquifer parameter identification. J Hydraul Div Amer Soc Civil Eng 101(HY9):1197–1209

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

Zheng C, Wang P (1999) An integrated global and local optimization approach for remediation system design. Water Resour Res 35(1):137–148

DOI 10.1007/s10040-004-0344-2