Journal of Hydroscience and Hydraulic Engineering
Vol. 18, No.2 November, 2000, 1-14

COMPARISON OF GENETIC ALGORITHMS WITH OTHER METHODS FOR GROUNDWATER MONITORING NETWORK PLANNING BASED ON GEOSTATISTICS

By

Hui-chen Su, Liang-Cheng Chang
National Chiao Tung University, Hsinchu, TAIWAN, R.O.C.

SYNOPSIS

Groundwater monitoring network planning largely concerns itself with constructing a flexible and efficient monitoring network. While attempting to solve a related problem, this study presents a novel procedure that combines genetic algorithms with geostatistical theory. The proposed method is compared to other methods that also integrate geostatistical theory with other optimization schemes, including the sequential design method (SDM), branch and bound method (BBM) and non-linear programming method (NPM). Those methods are implemented and applied to a simplified field case. The finding indicates that the BBM method is computationally infeasible for practical applications even if it can obtain the global optimal solution in principle. Genetic algorithms (GAs) are characterized by their ability to obtain a set of near optimal solutions instead of a single solution. In general, the total variations among the network defined by these methods do not significantly differ. For a network design problem, the SDM method provides a computationally efficient solution for a preliminary study. On the other hand, the multiple choices given by the GAs provide decision makers with flexibility to consider factors that geostatistics can not.

INTRODUCTION

Regardless of the hydrogeological problem, understanding groundwater behavior is a relevant task. Estimating the number of samples and obtaining them are always problematic owing to the high cost of hydrogeological investigation. To resolve this problem, a sufficient number of monitoring wells should be constructed. However, how much information is sufficient? The different planning methods were discussed in the last century, most were based on Kriging (12) method owing to the following characteristics of this method: (a) The estimated variance can be obtained when the value is interpolated by Kriging; and (b) The statistical spatial dependency between the sampling sites is characterized by the variogram function when Kriging is implemented. Kriging method is therefore an effective means of monitoring network design. This method was initially applied to determine the optimal locations of sampling sites for estimating fluoride concentrations in groundwater (3). Rouhani (15) developed a sequential algorithm, variance reduction analysis (VRA), to identify the new sampling locations that could enhance the maximum estimated accuracy (information gain) for the entire analyzed field. VRA selects a point from regular nodes to provide maximum information gain sequentially.

Theoretically, the solution of VRA is not a purely optimal network distribution. The following questions arise: can any algorithm locate the optimal location of sampling sites? Previous studies have utilized the branch and bound method (1, 5), non-linear programming method (10, 9), and Genetic algorithms (2, 4) to identify the optimal locations in the monitoring network. Differences among the above three techniques can be summarized as follows; **(a) Difference in the search space.** Non-linear programming (NPM) is completely continuous in the search domain, while the other two search techniques are based on a set of finite candidate points; **(b) Difference in the meaning of optimization.** Theoretically, branch and bound method (BBM) can obtain the global optimal solution, while the other techniques can not necessarily do so; and **(c) Difference in the number of solutions.** While the genetic algorithms (GAs) can provide users with multiple solutions, the other two can only offer a single solution.

For the above three methodologies, NPM and BBM are traditional optimal methods, GA is not. The underlying concept of genetic algorithm is inspired by natural processes of the selection of individuals and the evolution of species as well as reproduction mechanisms and the genetic transmission of characteristics (11). In addition, genetic algorithms generally function as searching procedures capable of optimizing functions based on a limited sample of function values. In sum, genetic algorithms are promising alternatives to conventional optimal methodologies in searching for optimal groundwater monitoring sites.

This study integrates geostatistics and different optimal methodologies (NPM, BBM and GAs) to perform the monitoring network design. This incorporation is able to consider the structure of hydrogeological parameters on space and the optimal search on mathematics. Besides these three algorithms, the forth monitoring planning algorithm based on VRA is termed herein as the sequential design method (SDM). Co-Kriging is selected as the common interpolation method in the four algorithms. All of the algorithms are applied to a simplified field case in Ping-Tung Plain, southern part of Taiwan. Moreover, the computational reliability of the algorithms and issues related to application are discussed.

GEOSTATISTICAL INTERPOLATION METHOD

Co-Kriging is adopted in this study to compute the estimation variance. Therefore, features of Co-Kriging are highlighted as follows.

Structural Analysis

As widely considered, hydrogeological parameters are regionalized variables with spatial statistical structures (7). In the following, we present the structural aspect of a regionalized variable by a theoretically mathematical form of semi-variogram:

$$\gamma_h = \frac{1}{2}\mathrm{E}[Z(x+h) - Z(x)]^2 \quad (1)$$

In geostatistical techniques on interpolation, Kriging, semivariogram is a basic and important term.

Co-Kriging

Co-Kriging is a type of Kriging that interpolates the value using the data of interrelated variables. The estimator is expressed as a linear combination of all the available data (12, 13):

$$U^*(X_o) = \sum_{i=1}^{N} \lambda_i^U U_i + \sum_{k=1}^{M} \lambda_k^V V_k \quad (2)$$

where $U^*(X_o)$ = the estimator of variables U to be estimated at location X_0; λ_i^U, λ_k^V = co-Kriging coefficients which present the weight to the variables U and V by the observation point of both variables; N = the total observation number of primary variables U; and M = total observation number of secondary variable V. Then, by applying the criteria of unbiased condition and minimum variance, the co-Kriging coefficients are determined and the co-Kriging system can be derived (13). The variance of the co-Kriging estimation is

$$\sigma^2 = \sum_{i=1}^{N} \lambda_i^U \overline{\gamma_{io}} + \sum_{k=1}^{M} \lambda_k^V \overline{\gamma_{ko}} + \mu_1 - \overline{\gamma_{oo}} \text{ ; subject to } \sum_{i=1}^{N} \lambda_i^U = 1 \quad \& \quad \sum_{k=1}^{M} \lambda_k^V = 0 \quad (3)$$

where σ^2 = the variance of co-Kriging estimation; $\overline{\gamma_{io}}$ = the auto-variogram of variable (13) U at location X_i and X_o; $\overline{\gamma_{ko}}$ = the cross variogram of variables V and U at location X_k and X_o; $\overline{\gamma_{oo}}$ = nugget effect; and μ_1 = the Langrange multipliers.

By assuming that multiple points ($[X_\beta, \beta = 1,2\cdots,Q]$) are to be estimated, the total estimation variance can be written as follows:

$$\sigma_T^2 = \sum_{\beta=1}^{Q}\left[\sum_{i=1}^{N}\lambda_i^U\overline{\gamma_{i\beta}} + \sum_{k=1}^{M}\lambda_k^V\overline{\gamma_{k\beta}} + \mu_1 - \overline{\gamma_{oo}}\right] \quad (4)$$

where X_β = the locations to be estimated.

Variation of Estimation Variance

For the co-Kriging method, the variance reduction caused by adding new monitoring wells into the network can be expressed as

$$VR_{*0} = V_0(N) - V_0(N+L) \quad (5)$$

In Eqs. 5, the network is assumed to have N monitoring wells before adding new wells and VR_{*0} is the variance reduction at any location X_{*i} (i=1,2,…,L) after adding the new monitoring wells. In contrast to the variance reduction based on the universal Kriging as in Rouhani (15), Eqs. 5 can not be rewritten as a compact and computationally efficient form owing to the complexity in computing the variance as shown by Eqs. 3. Therefore, the total estimation variance of $V_0(N)$ and $V_0(N+L)$ must be evaluated independently using the Eqs. 4, thus increasing the computational load significantly. For the optimal algorithms adopted herein, the SDM algorithm adds a new well at each iteration, and L equals 1. However, multiple wells are added into or removed from the network at each iteration for the BBM and GAs algorithms. As for the NPM algorithm, the variation of variance continuously varies with its decision variables, the locations of the new wells (10).

OPTIMAL ALGORITHMS

In this section, we introduce the optimal schemes for optimal groundwater monitoring network design. All of the algorithms search for the optimal network with minimum variance of total estimation or maximum reduction of estimation variance.

Genetic Algorithms (GAs)

Genetic algorithms (GAs) are computational paradigms inspired by the mechanics of natural evolution, including survival of the fittest, reproduction, and mutation. Developed by Holland in the 1960s, GAs allow computers to derive solutions from difficult problems, such as function optimization and artificial intelligence. In this study, we apply GAs and co-Kriging to optimal groundwater network design. The procedures are summarized as follows.

[Step 0] Express the objective function as

$$\max VR_{*0}(S), \quad S \subset \lfloor X_\eta \rfloor \quad num(S) = L \quad (6)$$

where S = the set of solution; $[X_\eta]$ = the set of candidate sites and will be further defined in section "study area"; L = the number of new wells that the decision maker plans to construct; and $VR_{*0}(S)$ = the value of the total reduced variance because of the new additional sites in set S.

[Step 1] Generate an initial set of chromosomes (network designs):

Herein, each chromosome refers to a possible network design or network alternative. The chromosomes are represented by binary strings. Consider the following example:

For a network design, a chromosome that selects the first, third and fifth sites from a candidate set with totally 10 sites is encoded as a binary string, [1 0 1 0 1 0 0 0 0 0]. In this study, the algorithm randomly generates 100 chromosomes initially and the number of chromosomes remains constant for all the generations.

[Step 2] Evaluate the fitness for all the chromosomes.

The measure of fitness for each chromosome (network design) consists of two parts: the variance reduction for the design network and the penalty related to the total number of wells that the network design will increase. In GAs, the number of wells for each chromosome may be greater or less than the required number of well numbers. Therefore, a penalty factor is defined as a situation in which the proposed number of wells is greater than the required well numbers. The fitness value for each chromosome is then formulated as (7)

$$F = \Delta V \times \left(\frac{1}{c}\right)^2 \text{ ; where c = the number of generated sites} \quad (7)$$

In Eqs. 7, c equals the number of wells for the chromosome when they are greater than L. On the other hand, c equals 1 when the number of wells is less than the specified. Where ΔV = the total variance reduction for the (chromosome) network design.

[Step 3] Verify the terminating conditions.

This step examines the terminating conditions to determine whether or not the algorithm is to be continued. Herein, we apply the simple criteria to stop the calculation after fifty generations.

[Step 4] Select (reproduction) the chromosomes:

The roulette wheel method is applied to select the chromosomes for crossover in this study. In the roulette wheel reproduction, each chromosome has the probability p_i being selected. This operation is simulating the natural selection. The more high fitness chromosomes have the more probability to survive. All the selected chromosomes form a pool of chromosomes that are ready to crossover.

[Step 5] Perform crossover of the chromosomes.

Randomly select two chromosomes from the pool and, then, perform the crossover to generate new offspring. In this study, one-point crossover is utilized.

[Step 6] Conduct mutation of the chromosomes.

The mutation restores the lost or unexplored genetic material into the population to prevent the GAs from prematurely converging to a local minimum. A mutation probability allows random mutations to be made to individual genes. Chen (4) supposed a simplified case, and three mutation probabilities (0.1,0.05,0.01) are tested and implemented five times for each. 0.1 is the most flexible probability among the above three values. Therefore, mutation probability 0.1 is finally adopted in this work. After completing this procedure, a new population of chromosomes is generated. The algorithm then goes back to step 2.

Sequential Design Method (SDM)

Although the basic procedure of SDM resembles the VRA method developed by Rouhani (15), co-Kriging is adopted herein instead of universal Kriging. The procedure of SDM is summarized as follows.

[Step 0] Define $\max VR_{*0}$, VR_{*0} as in Eqs. 5.

[Step 1] Define a set of points $[X_\beta, \beta = 1,2\cdots,Q]$ to evaluate the total estimation variance

[Step 2] Examine the set of available data points $[X_k, k = 1,2\cdots,N]$, which is the basis for interpolation.

[Step 3] Define the set of candidate locations $[X_\eta, \eta = 1,2\cdots,p]$ from which the new well locations are selected.

[Step 4] Select a point X_* from the candidate set $[X_\eta]$, then, evaluate the new total estimation variance V_η due to the addition of X_* by Eqs. 3. Repeat the same procedure for all the candidate points. The well location that causes the new network to have minimum total variance is the optimal well location, X_{*m}.

[Step 5] Move point X_{*m} out of the candidate set $[X_\eta]$ and into the observation data set $[X_k]$. Therefore, the number of wells in the candidate set is reduced to $p-1$ and the number of observation points in $[X_k]$ increases by one for the next evaluation.

[Step 6] Repeat the procedure 4 and 5 until the required number of monitoring wells is selected.

Branch and Bound Method (BBM)

While performing the search along pre-constructed tree branches, the branch and bound method is a simultaneous search as Ben-Jemaa (1) proposed. This algorithm is formed as a sequential procedure in which a well is conceptually deleted (dropped) at a time while moving along the tree branches. On the other hand, each node in the tree represents a possible network design. Fig. 1 illustrates a search tree of this algorithm in which three wells from a total of five candidate sites. Based on the evaluated estimation variance at each node, the search either proceeds forward along the same branch to delete a new well or leaves the current branch and moves one step back to another branch. The operation of BBM is described as follows.

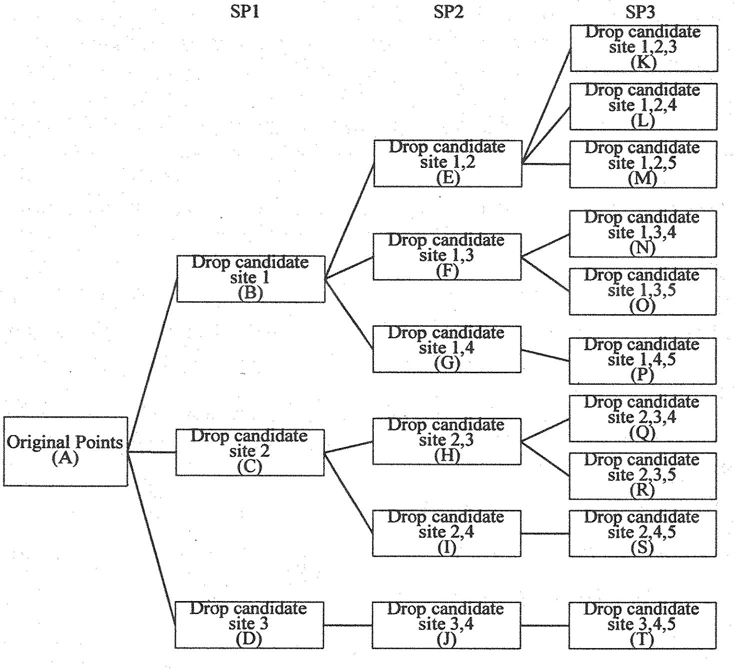

Fig. 1 The diagram of branch and bound method

[Step 0] Define the objective function in this algorithm as $\min \sigma_T^2$, σ_T^2 which is described as Eqs. 4.

[Step 0] Utilize the total estimation variance by the SDM method as the initial 'Bound' value.

[Step 1] Evaluate the total variance of the nodes starting from the root.

[Step 2] Describe this step according to Fig. 1 owing to its complexity. The following conditions should be considered.
By assuming that the process is visiting the node E, for a situation in which the total estimation variance due to the node E is less than the initial 'Bound' from step 1, the algorithm proceeds to the branches of the node E which has not been checked, e.g. node K. Consider a situation in which the node K is the lowest level

of the tree and its total estimation variance V_{nb} is less than the 'Bound'. Under this circumstance, the algorithm replaces the 'Bound' by V_{nb} as the new 'Bound'. Notably, nothing changes if the total variance of the node K is greater than the 'Bound'. The process then continues to the other branches of node E, e.g. node L.

- Assume that the total variance of the node E exceeds or equal the 'Bound', thus making it impossible to obtain a smaller total variance when going down to its branches. Therefore, all the branches starting from node E can be omitted. The process then goes directly to the other nodes at the same level of node E.
- After calculating the branches of the node E, the algorithm proceeds to the other node at the same level as node E, for example node F. The process repeats itself.
- The final 'Bound' value is the global optimal total estimation variance after visiting all the branches. Meanwhile, the set of associated sites is the global optimal monitoring network.

In principle, the BBM method can determine the global optimal solution for the network design problem. Nevertheless, computational loading of this algorithm is too heavy for a practical problem.

Non-linear Programming Method (NPM)

If the well locations of the design network are treated as continuous decision variables, the problem is obviously a non-linear programming one and the objective function is $\max VR_{*0}$. This study adopts general non-linear programming algorithms, the steep descent method, to search for the optimal locations in space. Steepest descent is an algorithm that takes the optimal move at each iteration. The implementation is straightforward and is summarized as follows:

[Step 0] Define the objective function as $\max VR_{*0}$, in which VR_{*0} was defined as Eqs. 5; The objective function also is a function of coordinates (10). The initial total estimation variance is evaluated from the set of observation networks by using Eqs. 4.

[Step 1] Choose a initial design of well locations $X^{k=0}=\left(X_1^0,X_2^0,\cdots,X_n^0\right)$

[Step 2] Evaluate the gradient d^k (via central-difference approximation) of objective function on the designing well locations. The vector shows the direction when searching the optimal locations.

[Step 3] Select a search step S^k by a one-dimensional search method and, then, update the well locations by $X^{k+1}=X^k+S^k d^k$ and the value of the objective function, $f\left(X^k+S^k d^k\right)$.

[Step 4] Define an acceptable stopping criterion ε. If the ratio of the variance reduction is less than or equal to ε, $\frac{\Delta f(X)}{f(X)}\le\varepsilon$, then stop the computation. Otherwise, repeat procedures 2 and 3.

APPLICATIONS AND COMPARISONS

Study Area

This study applies the above algorithms to a simplified groundwater monitoring network design problem in Ping-Tung Plain, one of nine groundwater regions in Taiwan. Located in the southern part of Taiwan, Ping-Tung Plain is about 1200km^2. Groundwater table (W) and hydraulic conductivity (K) are the parameters in this monitoring network analysis. The former one is the parameter that is of primary concern, while the latter one is the auxiliary parameter. Three sets of sites must be defined before the optimal schemes are employed:

1. Set of existed data sites $[X_k]$: the set of existed monitoring wells. Herein, the initial set of existed data sites contains fifteen installed observation wells.
2. Set of candidate sites $[X_\eta]$: the set of possible locations to construct new monitoring wells in the analyzed field. Herein, the set contains forty candidate

sites.

3. Set of basis nodes $[X_\beta]$: Herein, the estimation variances for a design network are evaluated at a set of nodes distributed in space. The total variance is the summation of all the estimation variance at the nodes. In practice, the nodes on uniform grids are adopted for this purpose. Cumulatively, thirty eight nodes are defined uniformly in our study area to calculate the total estimation variance. Fig. 2 presents the distribution of the above three sets of sites for the Ping-Tung Plain. Co-Kriging is the primary means of evaluating the estimation variance in all the algorithms. Therefore, the structure analysis of geostatistics, the semivariogram, must be completed before the algorithms can be proceeded.

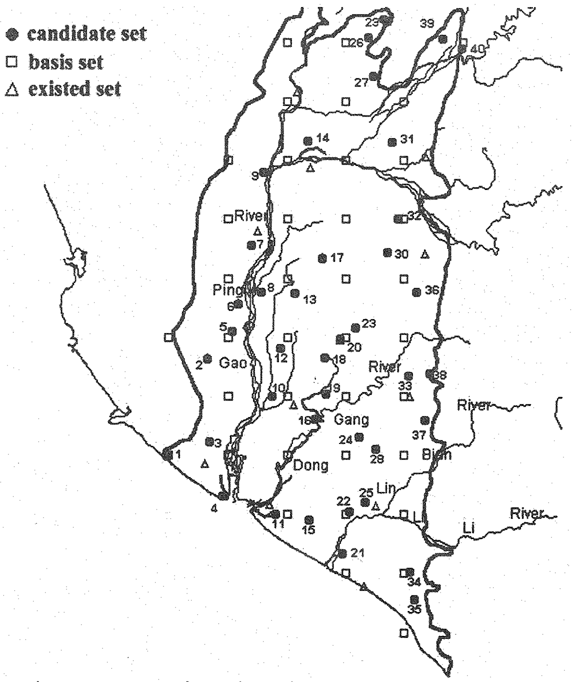

Fig. 2 The distribution of the data sets

Geostatistical Structure Analysis

By analyzing the data of Ping-Tung plain using the GeoEAS which was developed by EPA, the semivariogram and cross-semeivariogram for the groundwater table and hydraulic conductivity can be summarized as follows:

1. Semivariogram of water table (W): Exponential model, sill=1.7, range=30km

$$\gamma_W = 1.7\left[1 - e^{\frac{-3h}{30}}\right]$$; where h = the distance between any two points

2. Semivariogram of natural log of hydraulic conductivity (ln(K)): Exponential model, sill=0.7, range=20km

$$\gamma_{\ln(K)} = 0.7\left[1 - e^{\frac{-3h}{20}}\right]$$

3. Cross-semivariogram $\gamma_{W\ln(K)}(h)$ was taken by the following formula (13):

$$\begin{aligned}\gamma_{W\ln(K)}(h) &= \frac{1}{2}\left[\gamma^{+}_{W\ln(K)}(h)-\gamma_{W}(h)-\gamma_{\ln(K)}(h)\right]\\ &= \frac{1}{2}\left[0.85\left(1-e^{\frac{-3h}{25}}\right)-0.7\left(1-e^{\frac{-3h}{20}}\right)-1.7\left(1-e^{\frac{-3h}{30}}\right)\right]\end{aligned}$$

where $\gamma^{+}_{W\ln(K)}(h)$ = the summation of semivariogram of water table and natural log of hydraulic conductivity.

Models Application

The four monitoring design algorithms are developed using MATLAB. This study initially examines the programs by a simple case. The simple case is defined as follows; the numbers for the set of existed data sites $[X_k]$, the set of candidate sites $[X_\eta]$ and the set of basis nodes $[X_\beta]$ are 8, 7 and 21, respectively, as illustrated in Fig. 3. The objective is to add three more wells to the existing network. Fig. 4 summarizes the solutions for all the algorithms, and the solutions of SDM, BBM and GAs are the same for this simplified case. The solution of NPM is somewhat different from the other methods. The optimal locations for the NPM can be anywhere within the study area and not restricted to the candidate set. Therefore, two sites of NPM are not in the study area owing to the lack of a constraint for the boundary in the NPM system. Since the solution of BBM is theoretically global optimal, the SDM and GAs could thereby provide the optimal solution in this simplified case. This finding also demonstrates that the program is ready to solve a practical problem.

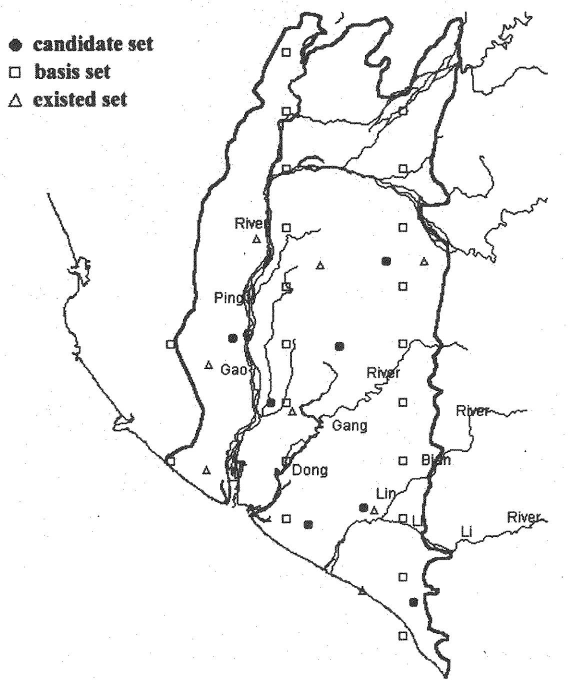

Fig. 3 The distribution of the data sets for simplified case

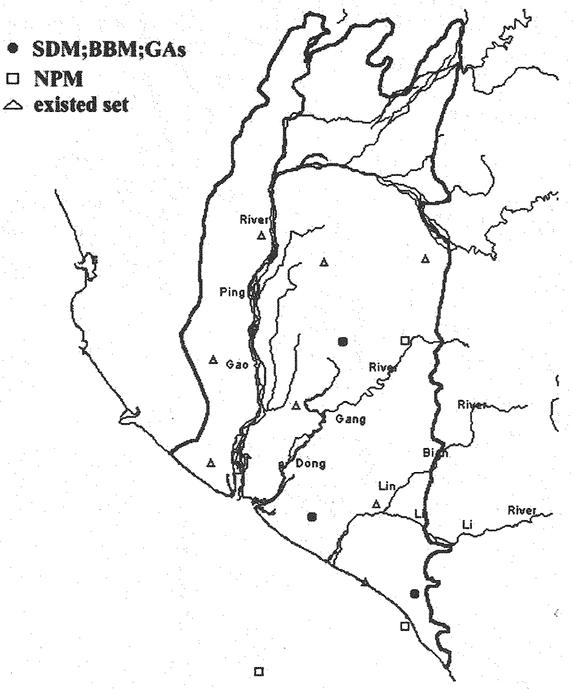

Fig. 4 The distribution of solutions from each optimal algorithm (simplified case)

The problem for the field case has been described in section "Study Area". The objective is to add five wells to the existing monitoring network; the final network contains a total of twenty monitoring wells. Table 1 and Fig. 5 summarize the results of all the different algorithms.

Method	SDM		NPM		BBM	GAs	
Coordinate (km)	X	Y	X	Y		X	Y
Site 1	212.903	2532.350	212.995	2532.996	—	212.903	2532.350
Site 2	207.600	2479.200	207.013	2478.991	—	207.600	2479.200
Site 3	203.306	2533.460	200.994	2532.993	—	203.306	2533.460
Site 4	182.900	2490.930	182.998	2491.011	—	182.900	2490.930
Site 5	206.360	2515.100	206.993	2515.005	—	206.360	2515.100
Initial TEV	59.3571						
TEV after the above 5 sites added	48.8479		46.4772		—	48.8479	

TEV: total estimation variance; [m^2]

Table 1 Solutions of the different methods

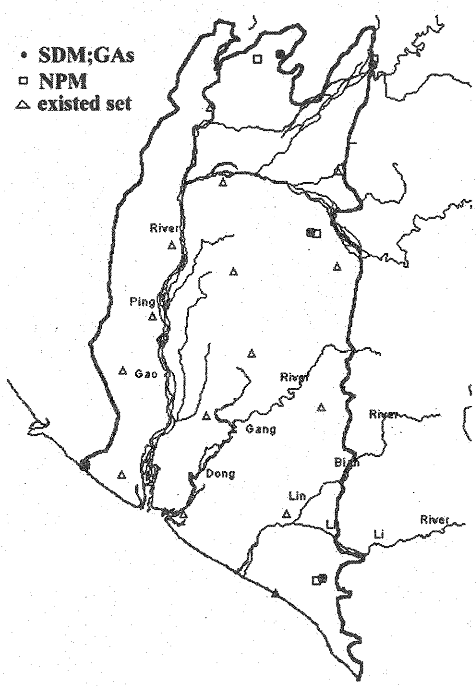

Fig. 5 The distribution of solutions from each optimal algorithm

Based on above results, the following points are worth mentioning:

1. The computational load for BBM algorithm does not grow linearly with candidate points. For the both cases, the number of branches is $C_3^7=35$ while three new monitoring locations are picked up from seven candidate points; in addition, the number of branches sharply becomes $C_5^{40}=658008$ while 5 new monitoring locations are selected from forty candidate points. However, depending on the value of initial 'BOUND', the tree search does not necessarily visit all the branches. Nevertheless, the number of total bracnches is still a fair index to estimate the computational load. For this simplified field case, more than one month was spent separately on three different computers (PC AMD-K6, PC P-II MMX and workstation Ultra-5); the problem was still not solved. Therefore, the BBM is not a feasible algorithm for a practical problem owing to its high comptational demand. By the way, the other three methods are completed in one day.
2. An initial solution must be defined to start the NPM computation. In this study, the solution of SDM is used as the initial solution. Although the solution of SDM is not necessarily an optimal one in principle, its solution is very close to the optimal solution by experience; and it also does well for the optimal solutions in many cases. Therefore, the solution of SDM is a good starting point for the NPM methods.
3. As mentioned earlier, GAs can provide users with multiple solutions simultaneously. For our problem, as indicated in Fig. 6, the optimal fitness for each generation is almost the same after twenty three generations. This finding implies that chromosomes have the fitness value near the optimal fitness value, 10.509, after twenty three generations. After collecting all the chromosomes with the fitness value near the optimal one, there are a total of fifty one chromosomes. In addition, each chromosome represents a network design, implying that there are fifty one near optimal network designs for this problem. Table 2 lists the solutions. For all candidate sites, points 40, 1, 34, 32, 27, and 26 are selected with a high frequency; points 2, 3, 4, 6, 7, 8, 9, 10, 11, 14, 15, 19, 20, 21 and 23 are never selected. Table 3 displays the frequency for each candidate site. From table 3, the candidate site with the higher frequency demonstrates that the gene is

more suitable for GAs. Therefore, it is always selected to reproduce. This table could also be a reference for the decision maker during the construction of new wells.

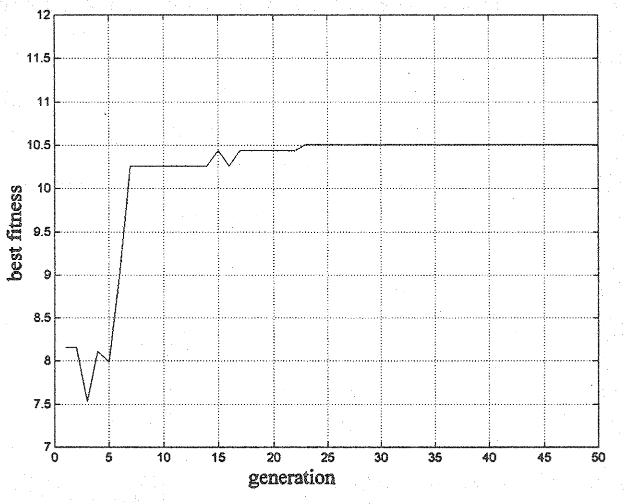

Fig. 6 Best fitness of each generation for GAs

Solution (Selected Sites*)	Total estimation variance	Solution (Selected Sites*)	Total estimation variance	Solution (Selected Sites*)	Total estimation variance
1,26,32,34,40	48.8479	1,29,32,34,40	49.5796	1,24,34,39,40	50.6162
1,27,32,34,40	48.9201	1,26,33,34,40	49.9291	1,22,26,38,40	50.7000
1,22,26,34,40	49.0289	22,26,32,34,40	49.3527	1,26,32,37,40	50.4437
1,16,26,34,40	49.6892	1,27,33,34,40	49.9985	16,22,27,34,40	50.2999
1,24,32,34,40	49.4071	1,26,32,36,40	49.8682	1,29,32,36,40	50.5993
1,24,26,34,40	49.0505	1,27,32,35,40	49.2357	18,22,26,34,40	50.5493
1,26,31,34,40	49.3415	1,31,32,34,40	49.6464	1,31,33,34,40	50.6540
1,27,30,34,40	49.6438	12,26,32,34,40	49.6186	1,22,27,36,40	50.1133
1,24,26,32,40	49.5738	1,24,31,34,40	49.7754	1,27,33,36,40	51.1237
1,13,27,34,40	49.5796	25,27,32,35,40	50.6055	1,25,33,34,40	51.3321
1,22,27,34,40	49.0984	1,26,27,32,40	50.0563	1,29,32,38,40	51.2499
1,24,27,35,40	49.4355	1,22,30,34,40	50.1121	1,30,32,36,40	51.4386
1,24,27,34,40	49.1198	1,25,27,32,40	50.4911	1,13,27,33,34	51.9558
1,22,32,34,40	49.3857	1,28,31,32,40	50.4687	1,30,32,38,40	51.9200
1,24,27,32,40	49.6482	1,27,34,39,40	50.2811	1,17,33,34,40	51.6981
1,25,27,34,40	49.9632	1,30,32,34,40	50.2362	5,25,27,34,40	51.0006
1,26,27,34,40	49.5304	1,25,32,34,40	50.2505	1,22,33,36,40	51.5941

*: Selected Sites are from candidate set

Table 2 The total estimation variance of GAs' solutions

No. of candidate point	1	2	3	4	5	6	7	8	9	10
Frequency	45	0	0	0	1	0	0	0	0	0
No. of candidate point	11	12	13	14	15	16	17	18	19	20
Frequency	0	1	2	0	0	2	1	1	0	0
No. of candidate point	21	22	23	24	25	26	27	28	29	30
Frequency	0	10	0	8	6	15	20	1	3	5
No. of candidate point	31	32	33	34	35	36	37	38	39	40
Frequency	5	23	8	33	3	6	1	3	2	50

Table 3 The frequency of each candidate site for the solutions of GAs

4. If the last (50^{th}) generation of GAs is considered, then six different solutions ([1,26,32,34,40], [1,24,27,34,40], [1,27,32,34,40], [1,26,27,34,40], [1,26,33,34,40], [1,27,30,34,40]) have nearly the same optimal fitness value, 10.509. Furthermore, for the above six solutions, sites 1, 34 and 40 are chosen all the time.
5. As indicated in Eqs. 7, the fitness of GAs contains a penalty factor to restrict the number of wells to be added. Therefore, the total estimation variance for their represented network designs may not be the same value when the chromosomes have the same fitness value. Hence, the total estimation variance must be re-calculated alone to evaluate the efficiency of the network design represented by each chromosome. The actual values of total estimation variance for the previous fifty one different chromosomes are re-evaluated by co-Kriging as summarized in Table 2. According to this table, the difference between the minimum total variance and the maximum one is 2.7462, i.e. only 5.6% on average, which is insignificant for hydrogeological estimation in practice. This finding confirms that all fifty one solutions can be accepted as the monitoring network designs. Our results also suggest that the network design can be flexible. In fact, the geostatistics can not consider factors such as the geological condition, land availability and other administrative constraints. The multiple choices given by the GAs can provide decision-makers with flexibility in considering these factors. On the other hand, the single solution provided by the other methods places unnecessary restrictions on the decision-makers.

CONCLUSION

This study presents a novel algorithm to solve a problem related to groundwater monitoring network design. The proposed algorithm is also compared with other methods. All the algorithms are all based on the co-Kriging theorem and the differences of these methods are based on the optimization schemes. The proposed algorithms are tested and applied to a field case. Based on the results in this study, we conclude the following:

1. For a small sized test problems all the algorithms can find the optimal solutions and the solutions for SDM, BBM and GAs method are all the same. The solution of BBM method is a global optimal one, indicating that the SDM and GAs can find the optimal solution as well. Nevertheless, for the simplified field case problem of a moderate size, the BBM method fails to solve the problem within a reasonable computational time, implying that it may be inappropriate for practical applications.
2. The decision variables for NPM method are continuous in space. Nevertheless, for a practical regional network project, a future study should determine whether or not the available sites to construct the monitoring wells should be implemented before the network is developed. Therefore, the continuous nature of the NPM method may not be beneficial for practical applications. Results in this study indicate that, when the density of the candidate sites is high, the solution of the NPM method is close to solutions of other methods.
3. The monitoring wells should be well maintained when located at schools and national regions in Taiwan. Though the candidate sites are distributed as a mesh, the most troublesome points (such as: private land, river, etc.) can be eliminated before the optimal algorithms are implemented. As we know, the geostatistics cannot consider the special geological situations or external factors. For the four algorithms what we use, SDM, NPM and BBM methods can only offer a single solution, GAs can provide multiple near optimal solutions. The multiple choices given by the GAs can provide decision-makers with the flexibility to consider the other real factors (such as: the geological situations) which are not easily defined in the monitoring design algorithms. On the other hand, the single solution provided by other methods might place unnecessary restrictions on decision-makers.
4. The solutions of the SDM method are closer to the solutions of other three methods even though the total estimation variance is the highest among the four algorithms. Therefore, the SDM method is the practical algorithm for a network design problem

if the external factors (geological situation, land use, etc.) are not concerned.

ACKNOWLEDGMENT

The authors would like to thank the Water Resources Bureau and the Taiwan Provincial Government Water Resource Department of the Republic of China for financially supporting this research.

REFERENCES

1. Ben-Jemaa, F., M.A. Marino, H.A. Loaiciga: Multivariate Geostatistical Design of Ground-Water Monitoring Networks, Journal of Water Resource Planning and Management, Vol.120, No.4: pp.505-522, 1994.
2. Brian J. Wagner: Sampling design methods for groundwater modeling under uncertainty, Water Resources Research, Vol.31, No.10, pp.2581-2591, 1995.
3. Carrera, J., E. d Usunoff, and F. Szidarovszky: A Method for Optimal Observation Network Design for Groundwater Management, Journal of Hydrology, Vol.73, pp.147-163, 1984.
4. Chen, H.-C..: Application of Genetic Algorithms to the Groundwater Monitoring Network Planning, Master Dissertation, Institute of Civil Eng., National Chiao Tung University, Taiwan, R.O.C, 1998.
5. Chiu, Y.-F.: Application of Co-Kriging Method to Groundwater Monitoring Network Planning, Master Dissertation, Institute of Civil Eng., National Chiao Tung University, Taiwan, R.O.C, 1997.
6. Delhomme, J.P.: Spatial Variability and Uncertainty in Groundwater Flow Parameters: A Geostatistical Approach, Water Resources Research, Vol.15, No.2, pp.269-280, 1979.
7. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.
8. Journel, A.G. and J.C. Huijbregts: Mining Geostatistics, Academic Press, New York, 1978.
9. Lin, C.-I.: Application of Optimal Multiple-Point Variance Analysis to the Groundwater Monitoring Network Planning, Master Dissertation, Institute of Civil Eng., National Chiao Tung University, Taiwan, R.O.C., 1996.
10. Lin, Y.P.: Multiple-Point Variance Analysis for Groundwater Monitoring Network Design, Ph.D. Thesis, 1995.
11. Lis, J.: Genetic Algorithm with the Dynamic Probability of Mutation in the Classification Problem, Pattern Recognition Letters, Vol.16, pp.1311-1320, 1995.
12. Matheron G.: The Theory of Regionalized Variables and its Applications, Paris School of Mines, Cah. Cent. Morphologic Math., Fontainnebleau, France, 1971.
13. Myers, D.E.: Matrix Formulation of Co-Kriging , Math. Geol., Vol.14, No.3, pp.249-257, 1982.
14. Reklaitis, G.V., A. Ravindran, and K.M. Ragsdell: Engineering Optimization - Methods and Applications, 1983.
15. Rouhani, S: Variance Reduction Analysis, Water Resources Research, Vol.21, No.6, pp.837-846, 1985.
16. Sophocleous, M., J.E. Paschetto and R.A. Olea: Ground-Water Network Design for Northwest Kansas, Using the Theory of Regionalized Variables, Ground Water, Vol.20, No.1, pp.48-58, 1982.

APPENDIX - NOTATION

The following symbols are used in this paper:

c = the number of sites what GAs generate;

F = fitness in Gas;

h = the distance between any two points;

K = hydraulic conductivity;

L	= the number of constructed wells what decision makers plan;
S^k	= the search step in Steepest descent algorithm
W	= groundwater table;
VR_{*0}	= total reduction estimation variance;
X_k	= location of the existed(observed) data sites;
X_η	= location of the candidate sites;
X_β	= location of the basis nodes;
$\bar{\gamma}_{ij}$	= estimated cross variogram of variable U and V at location X_i and X_j;
$\bar{\gamma}_{oo}$	= nugget effect;
λ	= co-Kriging coefficients;
μ	= Langrange multiplier; and
σ_T^2	= total estimation variance.

(Received December 20, 1999 ; revised July 26, 2000)