Environ Geol (2006) 50: 101–121
DOI 10.1007/s00254-006-0190-8

ORIGINAL ARTICLE

Ming-Sheng Yeh
Yu-Pin Lin
Liang-Cheng Chang

Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms

Received: 19 September 2005
Accepted: 10 January 2006
Published online: 18 March 2006

M.-S. Yeh · L.-C. Chang
Department of Civil Engineering,
National Chiao Tung University,
Hsinchu 30010, Taiwan

Y.-P. Lin (✉)
Department of Bioenvironmental Systems
Engineering, National Taiwan University,
1 Section 4 Roosevelt Road,
Taipei 10617, Taiwan
E-mail: yplin@ntu.edu.tw
Tel.: +886-2-23686980
Fax: +886-2-23635854

Abstract The optimal selection of monitoring wells is a major task in designing an information-effective groundwater quality monitoring network which can provide sufficient and not redundant information of monitoring variables for delineating spatial distribution or variations of monitoring variables. This study develops a design approach for an optimal multivariate geostatistical groundwater quality network by proposing a network system to identify groundwater quality spatial variations by using factorial kriging with genetic algorithm. The proposed approach is applied in designing a groundwater quality monitoring network for nine variables (EC, TDS, Cl^-, Na, Ca, Mg, SO_4^{2-}, Mn and Fe) in the Pingtung Plain in Taiwan. The spatial structure results show that the variograms and cross-variograms of the nine variables can be modeled in two spatial structures: a Gaussian model with ranges 28.5 km and a spherical model with 40 km for short and long spatial scale variations, respectively. Moreover, the nine variables can be grouped into two major components for both short and long scales. The proposed optimal monitoring design model successfully obtains different optimal network systems for delineating spatial variations of the nine groundwater quality variables by using 20, 25 and 30 monitoring wells in both short scale (28.5 km) and long scale (40 km). Finally, the study confirms that the proposed model can design an optimal groundwater monitoring network that not only considers multiple groundwater quality variables but also monitors variations of monitoring variables at various spatial scales in the study area.

Keywords Groundwater quality · Monitoring network design · Factorial kriging · Optimization · Spatial Variation · Pingtung plain · Taiwan

Introduction

In environmental monitoring such as groundwater quality investigations, the collected data may harbor significant uncertainty, including complex or extremely complicated variations in the observed values of measurable characteristics of the investigated medium or pollution sources in time and space. Given the high cost and risks associated with such investigations, development of efficient procedures for designing and adjusting information-effective monitoring networks is an essential task for more accurately understanding the spatial distribution or variations of monitoring variables. Therefore, the information generated by such optimal monitoring networks should provide sufficient, but not redundant information to fully understand the spatial phenomena of monitoring variables or their variations. These networks can be used to characterize natural resources for the management of resources or to delineate polluted area and variation for remediation and risk assessment.

Geostatistics, a spatial statistical technique used in environmental monitoring, is widely applied to analyze and map distributions of concentrations and variations in space and time. Geostatistics uses variograms to characterize and quantify spatial variability, perform rational interpolation, and estimate the variance in the interpolated values. A variogram quantifies the commonly observed relationship between the values of data, pertaining to the samples, and the samples‘ proximity. Kriging, a geostatistical method, is a linear interpolation procedure that provides a best linear unbiased estimator (BLUE) for quantities that vary spatially. Recently, kriging has been widely used to analyze and map the spatial variability and distribution of investigated data in many fields. Multivariate geostatistical methods, such as factorial kriging, combine the advantages of geostatistical techniques and multivariate analysis, while incorporating spatial or temporal correlations and multivariate relationships to detect and map different sources of spatial variation on different scales (Lin 2002). Factorial kriging is a variant of kriging which aims at estimating and mapping the different sources of spatial variability identified on the experimental variogram (Goovaerts 1992 and 1998). Examples of factorial kriging studies include Goovaerts (1994), Goovaerts and Webster (1994), Dobermann and others (1995), Einax and Soldt (1998), Jiménez-Espinosa and Chica-Olmo (1999), Bocchi et al. (2000), Castrignano et al. (2000a, b), Batista and others (2001) and Lin (2002).

In monitoring network design studies, many researchers have considered geostatistical approaches to designing or adjusting environmental monitoring systems and quantifying the informational value of monitoring data and their variations, for example, Rouhani (1985), Rouhani and Hall (1988), Christakos and Olea (1988), Loaiciga (1989), Hudak and Loaiciga (1993), Benjemaa et al. (1994), Pesti et al. (1994) and Wang and Qi (1998). Recently, Brus et al. (1999) used a geostatistical sampling scheme to discuss sampling size and points for estimating the mean extractable phosphorus concentration of fields. Van Groenigen et al. (1999) extended spatial simulated annealing with the kriging method to optimize spatial sampling schemes for obtaining the minimal kriging estimation variance. Lark (2000) used fuzzy and kriging methods to define a sampling scheme for designing sampling grids from imprecise information of soil variability. Prakash and Singh (2000) applied kriging variance reduction to design a groundwater monitoring network, as well as locations of additional wells from predefined locations. Based on the variance reduction method, Lin and Rouhani (2001) have developed a multiple-point variance analysis (MPV), which utilizes both the multiple-point variance reduction analysis and the multiple-point variance increase analysis. This process expands on foregoing studies by providing automatic procedures for simultaneously identifying groups of sampling sites without any need for spatial discretization or sequential selection. The goal of MPV (Lin and Rouhani 2001) is to develop a framework for the optimal simultaneous selection of additional or redundant sampling locations. Lark (2002) used the maximum likelihood method to optimize and discuss the spatial sampling of soil for the estimation of variograms. Cameron and Hunter (2002) selected redundant groundwater monitoring wells that did not change the plume interpolation, the kriging estimation variance in the plume section, nor the averaging global kriging variance. Ferreyra et al. (2002) used the scaled variogram technique with spatial simulated annealing algorithms along with kriging methods to reduce the number of locations from a regular grid system to describe water content in an 8-ha study area. Passarella et al. (2003) used the cokriging estimation variance with the fuzzy method to assess the loss of information produced by the elimination of the selected well in a groundwater network. All of these approaches have only focused on one monitoring variable and its spatial distribution.

The genetic algorithms (GAs) are robust methods used to search for the optimum solution of a complex problem and can compute the near global optimal solutions. GAs have been widely used in solving optimization problems and have found applications in monitoring network design. Cieniawski et al. (1995) addressed the problem of how to select a system of monitoring wells with a GA and the method of optimization using GA which could consider the two objectives of (1) maximizing reliability and (2) minimizing the contaminated area at the time of first detection. Reed and others (2000) combined a fate-and-transport model, plume interpolation, and a GA to identify cost-effective sampling plans that accurately quantify the total mass of the dissolved contaminant. Al-Zahrani and Moied (2003) used a GA for optimizing monitoring stations for water quality in a water distribution network to select sampling locations which were representative of the whole network system.

Genetic algorithms in other hydrological and water resources management applications include McKinney and Lin (1994), Hsiao and Chang (2002), Chang and Hsiao (2002), Rogers and Dowla (1994), Wardlaw and Sharif (1999), Wang (1991) and Mohan (1997).

In fact, groundwater quality monitoring networks may not only consider multiple variables, but also delineate their major variations in space. Therefore, this study develops a multivariate geostatistical groundwater quality network design model to propose a network system to identify groundwater quality spatial variations by using factorial kriging with GAs. The proposed model can optimally design a groundwater monitoring network that not only considers multiple groundwater quality variables but also monitors their spatial variations at various spatial scales. The developed model also has been applied in a real groundwater quality monitoring case in Taiwan.

Materials and methods

Factorial kriging

Multivariate analysis provides techniques, such as principle component analysis (PCA) and factor analysis, for classifying the inter-relationship of measured variables. Multivariate geostatistical methods combine the advantages of geostatistical techniques and multivariate analysis while incorporating spatial or temporal correlations and multivariate relationships to detect and map different sources of spatial variation on different scales. Textbooks (Deutsch and Journel 1992; Wackernagel 1995; Goovaerts 1997) and papers (Goovaerts 1992; Wackernagel 1994) have further detailed multivariate geostatistical methods. Therefore, only a brief description of multivariate geostatistical methods is provided here.

Geostatistics provide a variogram of data within a statistical framework, including spatial and temporal covariance functions. As expected, these variogram models are termed spatial or temporal structures, and are defined in terms of the correlation between any two points separated either spatially or temporally. The variograms provide a means of quantifying the commonly observed relationship between the values of the samples and the samples' proximity (Lin et al. 2002).

The variogram $\gamma(h)$ of second-order stationary regionalized variables, $Z(x)$, is defined as

$$\gamma(h) = (1/2)\ \mathrm{Var}\ [Z(x) - Z(x+h)] \tag{1}$$

where h denotes the lag distance that separates pairs of points; Var represents the variance of the argument; $Z(x)$ is the value of the regionalized variable of interest at location x, and $Z(x+h)$ denotes the value at location $x+h$. An experimental variogram for the interval lag distance class h, $\gamma(h)$, is given by

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [Z(x_i + h) - Z(x_i)]^2 \tag{2}$$

where $n(h)$ represents the number of pairs separated by the lag distance, h. Similarly, the spatial correlations or cross-variograms ($\gamma_{\alpha\beta}(h)$) between two variables can be defined as

$$\gamma_{\alpha\beta}(h) = \frac{1}{2} E\left[[Z_\alpha(x_i + h) - Z_\alpha(x_i)][Z_\beta(x_i + h) - Z_\beta(x_i)]\right] \tag{3}$$

where α β represent the different regionalized variables. The experimental cross-variogram $\gamma_{\alpha\beta}(h)$ can be written as:

$$\hat{\gamma}_{\alpha\beta}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [Z_\alpha(x_i + h) - Z_\alpha(x_i)] \times [Z_\beta(x_i + h) - Z_\beta(x_i)]. \tag{4}$$

Multivariate regionalization of a set of random functions can be represented with a spatial, multivariate linear model which allows easy manipulation of multivariate data (Wackernagel 1995). The nested direct and cross-variogram can thus be modeled as linear combinations:

$$\gamma_{\alpha\beta}(h) = \sum_{u=1}^{S} \gamma_{\alpha\beta}^u(h) = \sum_{u=1}^{S} b_{\alpha\beta}^u g^u(h) \tag{5}$$

where S is the number of the spatial scale, $b_{\alpha\beta}^u$ are coefficients, and $g^u(h)$ are elementary variogram functions for the spatial scale u.

A set of second-order stationary regionalized variables, $\{Z_i(x);\ i=1,\ldots, N\}$, can be decomposed into sets of spatial components, $\{Z_i^u(x);\ i=1, \ldots, N; u=1, \ldots, S\}$:

$$Z_i(x) = \sum_{u=1}^{S} Z_i^u(x) + m_i, \tag{6}$$

where i represents the different regionalized variables, N is the number of regionalized variables, u represents the different spatial scale, and S the number of spatial scales. m_i is $E[Z_i(x)]$. Then, the set of spatial components $Z_i^u(x)$ can be decomposed into sets of spatially uncorrelated factors (Goovaerts 1992; Rouhani and Wackernagel 1990; Wackernagel 1995),

$$Z_i^u(x) = \sum_{p=1}^{N} a_{ip}^u Y_p^u(x) \tag{7}$$

where $Y_p^u(x)$ are the regionalized factors in which p denotes different factors at a given spatial scale u. According to Eqs. 6 and 7

$$Z_i(x) = \sum_{u=1}^{S} \sum_{p=1}^{N} a_{ip}^u Y_p^u(x) + m_i. \tag{8}$$

At a given spatial scale u, each uncorrelated factor $Y_p^u(x)$ is assigned the same elementary variogram function, $g^u(h)$. Because each factor is uncorrelated

$$\frac{1}{2} E\left[\left\{Y_v^u(x) - Y_v^u(x+h)\right\}\left\{Y_{v'}^{u'}(x) - Y_{v'}^{u'}(x+h)\right\}\right] = \begin{cases} g^u(h) & \text{if } u = u' \text{ and } v = v' \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

According to Eqs. 8 and 9, the direct and cross-variograms, between two variables, $\gamma_{\alpha\beta}(h)$, can be represented by $g^u(h)$ and a^u_{ip}

$$\begin{aligned}\gamma_{\alpha\beta}(h) &= \frac{1}{2}E\left[[Z_\alpha(x_i) - Z_\alpha(x_i+h)][Z_\beta(x_i) - Z_\beta(x_i+h)]\right] \\ &= \sum_{u=1}^{S}\sum_{u'=1}^{S}\sum_{p=1}^{N}\sum_{p'=1}^{N} a^u_{\alpha p} a^{u'}_{\beta p'} \\ &\quad \times \frac{1}{2}E\left[\left\{Y^u_p(x) - Y^u_p(x+h)\right\}\left\{Y^{u'}_{p'}(x) - Y^{u'}_{p'}(x+h)\right\}\right] \\ &= \sum_{u=1}^{S}\sum_{p=1}^{N} a^u_{\alpha p} a^u_{\beta p} g^u(h). \end{aligned} \tag{10}$$

Then, according to Eqs. 5 and 10

$$\gamma_{\alpha\beta}(h) = \sum_{u=1}^{S} b^u_{\alpha\beta} g^u(h) = \sum_{u=1}^{S}\sum_{p=1}^{N} a^u_{\alpha p} a^u_{\beta p} g^u(h). \tag{11}$$

The matrix form of Eq. 11 can be written as

$$\Gamma(h) = \sum_{u=1}^{S} B^u g^u(h) = \sum_{u=1}^{S} A^u A^{u^{\mathrm{T}}} g^u(h). \tag{12}$$

Then,

$$B^u = A^u A^{u^{\mathrm{T}}} \tag{13}$$

where B^u is called the coregionalization matrix for a given spatial scale u, and B^u must be a positive semi-definite matrix. Matrix A is the transformation coefficient between regionalized factors, $Y^u_p(x)$, and spatial components, $Z^u_i(x)$. Based on the above nested model, PCA can be applied to analyze the $N \times N$ coregionalization matrix of the coefficients $b^u_{\alpha\beta}$ as a covariance matrix of N regionalized variables on a spatial scale that can be decomposed and written as (Wackernagel 1995)

$$B^u = A^u A^{u\mathrm{T}} = \left(\left(Q^u\sqrt{\Lambda^u}\right)\left(Q^u\sqrt{\Lambda^u}\right)^{\mathrm{T}}\right) \tag{14}$$

where Q^u is the matrix of eigenvectors for spatial scale u, $\Lambda^u$is the diagonal matrix of eigenvalues for spatial scale u, and the relative eigenvalues are λ_1, λ_2,...,λ_N. The variance explanation of B^u by Y^u_p (x), i.e., proportion, can be represented as $\lambda_p / \sum\lambda$.

Based on the above, when $g^u(h)$ and a^u_{ip} have been obtained, the cokriging estimator of the regionalized factors, $Y^u_p(x_0)$, at a given point x_o is

$$Y^{u*}_p(x_0) = \sum_{i=1}^{N}\sum_{\alpha=1}^{m} \lambda_{i\alpha} Z_i(x_\alpha) \tag{15}$$

where $Z_i(x_\alpha)$ is the observed value of the regionalized variable, Z_i, at the data point x_α; m is the number of observed value data of the regionalized variable, Z_i; N is the number of regionalized variables; and $\lambda_{i\alpha}$ is the estimation weight of the observed value of the regionalized variable, Z_i, at the point x_α.

The cokriging system can be solved as

$$\begin{cases} \sum_{j=1}^{N}\sum_{\beta=1}^{m} \lambda_{j\beta}\gamma_{ij}\left(x_\alpha - x_\beta\right) - \mu_i \\ \quad = a^u_{ip} g^u(x_\alpha - x_0) & \text{for } i = 1,\ldots,N;\ \alpha = 1,\ldots,m \\ \sum_{\beta=1}^{m} \lambda_{i\beta} = 0 & \text{for } i = 1,\ldots,N \end{cases} \tag{16}$$

where μ_i is the ith Lagrange multiplier, g^u $(x_\alpha - x_0)$ is the value taken by the uth elementary variogram function, g^u (h), between the α th observed point and x_0.

Genetic algorithms

The concept of GAs has been derived from Darwin's theory of natural selection, and was first proposed in 1975 by John Holland (1992). In the 1960s and 1970s, several evolutionary computing models were simultaneously developed. GAs are becoming the most popular innovative methods of computing due to their ability to solve complex problems, simple interface, and their ability to be hybridized with existing simulation models. GAs are inspired by the mechanism of natural section, in which stronger individuals are likely to survive in a competing environment. GAs are computing procedures embodying important mechanisms of the adaptive process in natural systems.

Genetic algorithms are heuristic programming methods capable of locating near global optimal solutions for complex problems (Goldberg 1989). The basic principle of the GA is to simulate biological evolution. This process has been successfully applied to many situations. A single GA cycle, known as a "generation", includes three genetic operators: reproduction, crossover, and mutation, and can be considered to consist of the following steps (Mitchell 1998).

1. Start with a randomly generated population of n chromosomes (candidate solutions to a problem).
2. Calculate the fitness of each chromosome in the population.
3. Repeat the following steps until n offsprings have been created.
   (a) Select a pair of parent chromosomes from the current population, the probability of selection being an increasing function of fitness.
   (b) With the crossover probability, cross over the pair at a randomly chosen point to form two offsprings.

(c) Mutate the two offsprings at each locus with the mutation probability, and place the resulting chromosomes in the new population.

4. Replace the current population with the new population.
5. Go to step 2 until the required number of generations. For detailed procedures of GAs, refer to Mitchell (1998). In this study, the simple GA combines factorial kriging and GAs to develop a multivariate geostatistical groundwater monitoring network design model.

Multivariate geostatistical groundwater quality monitoring network design model

Definition of optimal problem

The aim of the optimal model is to minimize the estimation variance of a single-factor or multi-factors composed of groundwater quality variables for the purpose of establishing a monitoring network to monitor spatial variations. Factorial kriging can solve a multi-variable problem by applying regionalized factors as representative variables of multi-scale geostatistical structures. PCA, one of the components of factorial kriging, can address the proportion of each regionalized factor of multi-scale geostatistical structures. In this study a groundwater quality monitoring network design approach is developed by considering total variances of the regionalized factors composed by monitoring variables at various spatial scales.

The objective function of the optimal problem is to minimize the total variances involved when estimating regionalized factors of a study region under cost constraints. Factorial kriging is employed to estimate variances of regionalized factors in a study region. The optimal model can be formulated as

objective function

$$\underset{I\subset\Omega,\, s\subset S,\, n\subset N_u}{\text{Min}} J(I) = \sum_{u\in s}\sum_{p\in n}\sum_{d\in D} \omega_{Y_p^u}\sigma^d_{Y_p^u}(I) \tag{17}$$

subject to

$$N_I \leq N_{\max} \tag{18}$$

where

I a subset of Ω and is a possible alternative network design

Ω an index set that defines all of the candidate well locations in the study region

S an index set of all the spatial scales

s a subset of S and represents the set of spatial scales considering the network design

N_u an index set that represented all of the regionalized factors of a given spatial scale u

n a subset of N_u and represents the set of regionalized factors considering the network design of a given spatial scale u

Y_p^u regionalized factor

$\omega_{Y_p^u}$ the weighting of Y_p^u

D the set of all grids in the study region domain

d an element of D

$\sigma^d_{Y_p^u}$ the variance of the estimation of Y_p^u at a given grid d

N_I the number of a possible alternative network design, I

$N_{\max}$ the maximum limited number of monitoring wells.

In Eq. 17, the objective function represents the total variances of estimating the regionalized factors, Y_p^u, which are chosen under the value of the proportion of Y_p^u of the concerned spatial scale, s ($s \subset S$). The weighting of $Y_p^u, \omega_{Y_p^u}$, can be assigned based on the proportion of each Y_p^u. In Eq. 18, the constraint represents the cost limit of the monitoring network.

The optimal problem defined by Eqs. 17 and 18 has three key characteristics which are different from traditional network design problems.

First, the objective function is to minimize the total variances of not the regional variables themselves, but the chosen regionalized factors.

Second, an optimal network can be designed considering only one regionalized factor for a specific spatial scale, or the optimal network can be designed considering several main regionalized factors for more than one spatial scale.

Third, the weighting of Y_p^u, $\omega_{Y_p^u}$, can be assigned objectively based on the proportion of each Y_p^u. This differs from the kriging and cokriging methods, which use the subjective weight of regional variables.

Solution procedure: Integration of factorial kriging and GAs

To solve the optimal problem defined by Eqs. 17 and 18, factorial kriging is combined with GA to develop a groundwater monitoring network design model that considers multi-variables (Fig. 1). The algorithm is a simple GA with factorial kriging embedded in the total variance of the chosen regionalized factors. The FACTOR2D (Pardo-Iguzquiza and Dowd 2002) Fortran program was modified for factorial kriging analysis in this study. The program has two main features in this study. First, the GA accommodates the discreteness of the search for alternative optimal well locations among the candidate well sites. Second, factorial kriging is used to calculate the total variances of the chosen latent factors associated with each network alternative (chromosomes). These features are clarified in the following steps of the developed optimal model.

Step 1: standardizing monitoring variables

Solving the dimension of differing groundwater quality requires standardizing the monitoring variables before multivariate analysis.

$$\tilde{z}_{\alpha i} = \frac{z_{\alpha i} - \mu_i^*}{s_i}, \quad (19)$$

where

$\tilde{z}_{\alpha i}$ is the standardized variable value of the ith groundwater quality item of the αth monitoring well;

$z_{\alpha i}$ are the data of the ith groundwater quality item of the αth monitoring well;

μ_i^* is the mean of the data of the ith groundwater quality item; and

s_i is the standard deviation of the data of the ith groundwater quality item.

After standardizing groundwater quality data, a successive analysis can be conducted using the standardized variables.

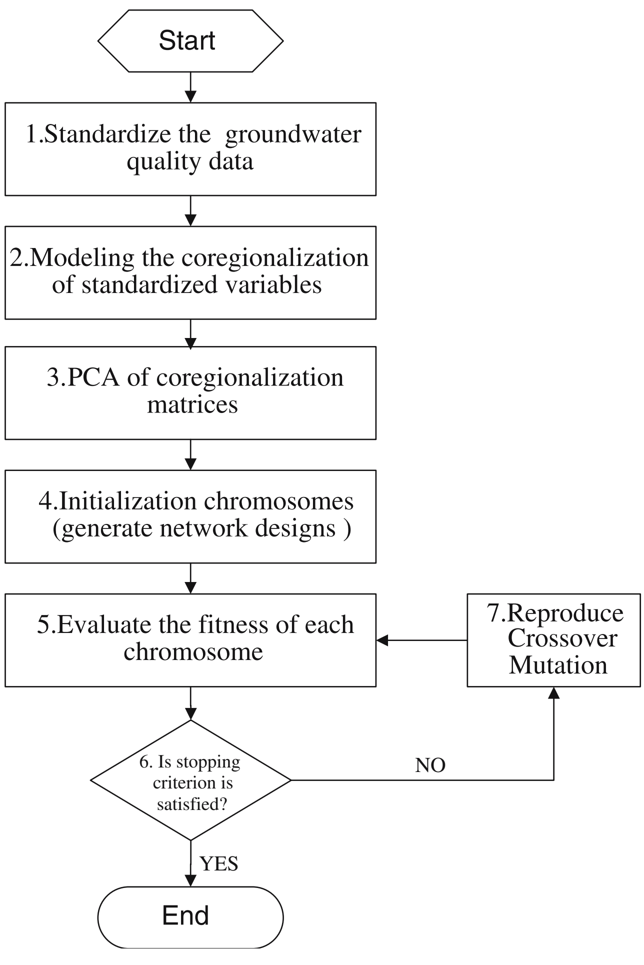

Fig. 1 Flow chart of the proposal model

Step 2: modeling the coregionalization of standardized variables

In this study, VARIOWIN2.2 (Pannatier 1996) is used to calculate and fit initial direct variograms and cross-variograms of the standardized variables. After calculations the experimental direct and cross-variograms of standardized variables are modeled as linear combinations of elementary variogram functions, $g^u(h)$, for each spatial scale u. Then, the variogram type and range of the elementary variogram functions $g^u(h)$ must be determined for each spatial scale u. Some studies (Goovaerts 1992; Pardo-Iguzquiza and Dowd 2002) offer more detail for modeling regionalization procedures. The procedures are simply described as follows.

1. All direct variograms and cross-variograms are estimated by using VARIOWIN 2.2 for the same number of lags and the same lag distances h.
2. The number and types of elementary variogram functions and their ranges are postulated.
3. The sills (coregionalization matrix) are fitted by repeat (1) and (2) to ensure the positive semi-definiteness of all coregionalization matrices.

Step 3: PCA of the coregionalization matrices

Based on the above, PCA can be applied to the coregionalization matrix of each spatial scale. In this study, the statistics software SPSS is employed in PCA and each coregionalization matrix is treated as a covariance matrix, making it possible to obtain the proportion and factor loading of each factor. The GA procedures are described in the subsequent paragraphs.

Step 4: initialization chromosomes

The network alternatives are encoded as chromosomes into the GA and randomly generate an initial population. The GA is widely known for using binary coding to represent a variable. This study uses a binary indicator to represent the status of a well installation at a candidate site. Accordingly, a chromosome, represented by a binary string, defines a network alternative. Each bit in a chromosome is associated with a candidate well, and the length of the chromosome equals the total number of candidate sites available for installation. If the value of a bit equals one, then a well will be installed at the associated candidate site; otherwise, the value of a bit is zero and no well will be installed at the associated candidate site. The selection of wells is binary, so the encoding and decoding of the chromosome are straightforward.

Step 5: evaluate the fitness of each chromosome

The objective function of the optimal problem is to minimize the total variances of latent factors of the grids for a study region under cost constraints. Therefore, the total variances can clearly represent the fitness of each chromosome. It should also be mentioned that the objective function of Eq. 17 can consider multi-spatial scales and multi-factors.

Some chromosomes may violate the maximum limited number of monitoring wells (Eq. 18). In this situation, the penalty function method is employed to modify the total variances $J(I)$ as fitness to avoid reducing the diversity of chromosomes. This is done because only a minority of the chromosomes can continue propagating if those chromosomes which violate the maximum limited number of monitoring wells are abandoned.

The modification of fitness by the penalty function method is performed as follows:

$$F(I) = \begin{cases} J(I) & \text{if } N_I = N_{\max} \\ J(I) \times (|N_I - N_{\max}|) & \text{if } N_I \neq N_{\max} \end{cases} \quad (20)$$

In Eq. 20, the modified fitness is equal to the total variances, $J(I)$, multiplied by a penalty factor $|N_I – N_{\max}|$ if the well number of the chromosome does not equal the maximum limited number of monitoring wells ($N_{\max}$). The fitness modified by the penalty function in Eq. 20 allows the chromosomes which violate the maximum limited number of monitoring wells to maintain a lower probability of reproduction instead of being abandoned.

Step 6: termination

The new population requires evaluating the total variances as in Step 5, which is employed to evaluate the stopping criterion. The stopping criterion is based on the change of either the value of the objective function or the optimized parameters. If the user-defined stopping criterion is satisfied or the maximum allowed number of generations is reached, the procedure terminates; otherwise, it performs Step 7 for another cycle. The success and performance of GAs depend on various parameters—population size, number of generations and the probabilities of crossover and mutation (Mckinney and Lin 1994). Goldberg (1989) has asserted that how well GAs perform depends on the choice of high-crossover and low-mutation probabilities and a moderate population size. Therefore, solutions obtained using a GA cannot be guaranteed to be optimal.

The stopping criterion requires two conditions be satisfied in the algorithm. The conditions are no further change of the value of the object function for 15 successive generations, and the population propagating for more than 50 generations.

Step 7: reproduce the best chromosomes, perform crossover and implement mutation

If the stopping criterion is not satisfied, one should reproduce the best strings, perform crossovers and implement mutations for a general new population, and then go back to Step 5. In this study, a uniform crossover using the tournament selection method is chosen; crossover probability (p_{cross}) equals 0.8, mutation probability (p_{mutat}) equals 0.1, and the population size equals 50.

Model application

The proposed model is utilized in designing the groundwater quality monitoring network for the second aquifer (Aquifer 2) of Pingtung Plain, Taiwan. The Pingtung Plain is located in southern Taiwan, and is the largest alluvial plain in the region. To the east lie the central mountains of Taiwan, to the north and west the low hills of the quaternary sediments, and to the south the Taiwan Strait. The area of the Pingtung Plain is about 1,140 km^2, approximately 60 km from north to south and 20 km from east to west (Fig. 2). The groundwater of the Pingtung Plain is an important water source in southern Taiwan. There are four major components of the aquifer system: Aquitard 1, Aquifer 2, Aquifer 3–1 and Aquifer 3–2 (Fig. 3).

In the Pingtung Plain, the intended monitoring program should produce information representative of the long-term water quality variations of the major aquifers. The current groundwater monitoring network established and operated by the Water Resources Agency has successfully provided valuable information on the major aquifers in the Pingtung Plain. The 34 existing monitoring wells system for the second aquifer is shown in Fig. 2.

Nine water quality variables, including EC, TDS, Cl^-, Na, Ca, Mg, SO_4^{2-}, Fe and Mn, have been selected as regionalized variables to assess the monitoring network design in the follow-up analysis procedures. The groundwater quality data used in this study were sampled in 2001 from a total of 34 wells in the regional monitoring network built by the Water Resources Agency. However, the cost of maintaining extensive monitoring of both the water level and the quality of groundwater is very expensive. Developing a cost-effective program for monitoring the quality of groundwater which involves

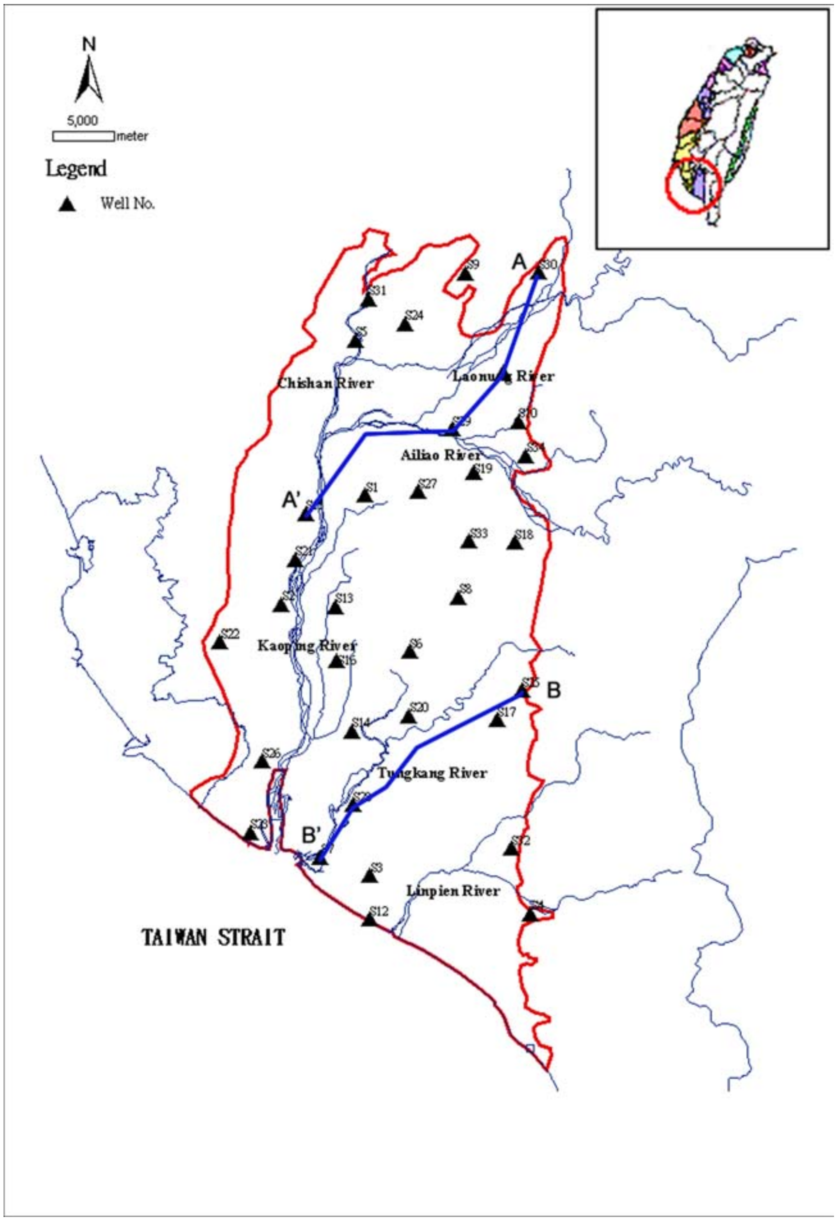

Fig. 2 Location of the groundwater monitoring wells in the second aquifer in Pingtung Plain

sampling from only a fraction of the existing monitoring wells is important. To produce a cost-effective monitoring system, the existing monitoring system has been re-evaluated and designed into 20-well, 25-well and 30-well monitoring systems using the proposed model.

Multivariate geostatistical analysis

This study calculates experimental direct variograms and cross-variograms for the standardized (zero mean and unit variance) EC, TDS, Cl^-, Na, Ca, Mg, SO_4^{2-}, Fe and Mn. A relatively consistent set of best-fit models was obtained to fit these variograms using VARIOWIN 2.2 (Pannatier 1996). The best-fit variogram models of these nine variables were specified as the sum of two structures by a Gaussian type model with an effective range of 28.5 km and a spherical type with an effective range of 40 km. The coregionalization matrix of spatial scales for 28.5 and 40 km are shown in Table. 1 and 2, respectively. After the PCA of the coregionalization matrix, the eigenvalues and the variance proportion of each factor are shown in Table 3. The factor loadings of the two spatial scales are shown in Table. 4 and 5, respectively.

In the 28.5 km scale, the first two factors explained 80.2% of the total variance for the nine variables as listed in Table 3. The first factor explained 69.1% of the

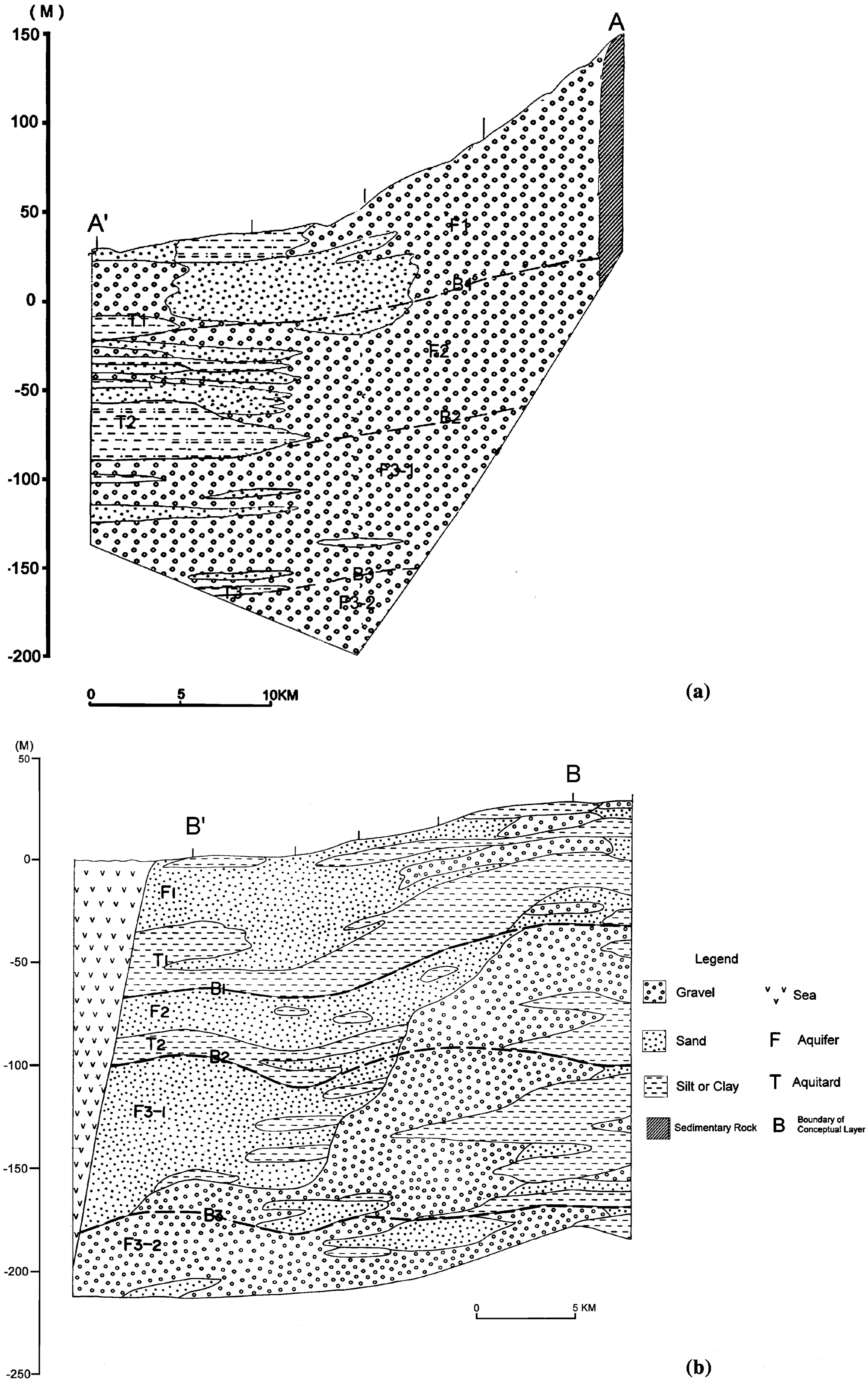

Fig. 3 Geological profile of Pingtung Plain **a** A–A′, **b** B–B′

Table 1 Coregionalization matrix of spatial scale 28.5 km for the standardized variables

	EC	TDS	Cl	Na	Ca	Mg	SO_4^{2-}	Fe	Mn
EC	0.4	–	–	–	–	–	–	–	–
TDS	0.36	0.4	–	–	–	–	–	–	–
Cl^-	0.36	0.36	0.39	–	–	–	–	–	–
Na	0.36	0.36	0.36	0.41	–	–	–	–	–
Ca	0.261	0.243	0.328	0.224	0.48	–	–	–	–
Mg	0.36	0.36	0.38	0.37	0.328	0.4	–	–	–
SO_4^{2-}	0.32	0.32	0.32	0.288	0.238	0.312	0.49	–	–
Fe	0.114	0.185	0.175	0.124	0.203	0.145	0.072	0.42	–
Mn	0.272	0.264	0.264	0.28	0.208	0.272	0.182	0.156	0.46

Table 2 Coregionalization matrix of spatial scale 40 km for the standardized variables

	EC	TDS	Cl	Na	Ca	Mg	SO_4^{2-}	Fe	Mn
EC	1.0	–	–	–	–	–	–	–	–
TDS	0.9	1.0	–	–	–	–	–	–	–
Cl^-	0.9	0.9	1.0	–	–	–	–	–	–
Na	0.9	0.9	0.9	1.0	–	–	–	–	–
Ca	0.9	0.9	0.8	0.8	0.97	–	–	–	–
Mg	0.9	0.9	0.9	0.9	0.8	1.0	–	–	–
SO_4^{2-}	0.8	0.8	0.8	0.8	0.7	0.8	1.0	–	–
Fe	0.6	0.5	0.5	0.4	0.7	0.5	0.4	1.0	–
Mn	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.6	1.0

Table 3 Eigenvalues of coregionalization matrix of spatial scale 28.5 and 40 km

Factor	28.5 km			40 km		
	Eigenvalues	Proportion (%)	Accumulated proportion (%)	Eigenvalues	Proportion (%)	Accumulated proportion (%)
Factor 1	2.66	69.1	69.1	7.211	80.394	80.394
Factor 2	0.427	11.091	80.191	0.796	8.874	89.267
Factor 3	0.298	7.734	87.925	0.27	3.016	92.283
Factor 4	0.217	5.637	93.562	0.242	2.694	94.977
Factor 5	0.155	4.021	97.583	0.167	1.86	96.837
Factor 6	0.04382	1.138	98.721	0.1	1.115	97.951
Factor 7	0.02849	0.74	99.461	0.1	1.115	99.066
Factor 8	0.01169	0.304	99.764	0.0658	0.734	99.8
Factor 9	0.00907	0.236	100	0.0179	0.2	100

Table 4 Factor loading of spatial scale 28.5 km

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
EC	0.934	−0.189	0.067	0.015	−0.102	0.250	0.120	0.029	−0.006
TDS	0.938	−0.059	0.072	0.230	−0.093	0.079	−0.202	−0.007	0.049
Cl^-	0.979	−0.005	−0.081	−0.017	−0.103	−0.025	0.017	−0.144	−0.054
Na	0.912	−0.168	0.194	0.077	−0.231	−0.172	0.095	0.014	0.071
Ca	0.740	0.322	−0.423	−0.409	0.016	0.011	−0.008	0.009	0.048
Mg	0.968	−0.059	−0.046	−0.093	−0.145	−0.086	−0.057	0.086	−0.097
SO_4^{2-}	0.767	−0.373	−0.284	0.202	0.387	−0.043	0.023	0.009	0.003
Fe	0.457	0.800	−0.029	0.381	0.052	−0.007	0.047	0.016	−0.011
Mn	0.710	0.160	0.570	−0.250	0.288	−0.006	−0.012	−0.005	−0.001

Table 5 Factor loading of spatial scale 40 km

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
EC	0.963	0.007	−0.003	−0.138	0.018	0.148	0.081	−0.150	−0.054
TDS	0.954	−0.103	−0.077	−0.129	−0.141	−0.148	−0.081	0.055	−0.076
Cl^-	0.941	−0.133	−0.035	−0.022	0.191	−0.182	0.155	−0.031	0.028
Na	0.932	−0.242	−0.109	−0.013	0.032	0.148	0.081	0.175	0.006
Ca	0.930	0.217	−0.073	−0.154	−0.225	0.000	0.000	−0.030	0.086
Mg	0.941	−0.133	−0.035	−0.022	0.191	0.033	−0.237	−0.031	0.028
SO_4^{2-}	0.852	−0.236	0.447	0.107	−0.081	0.000	0.000	0.006	0.008
Fe	0.625	0.760	0.127	−0.013	0.106	0.000	0.000	0.070	−0.018
Mn	0.884	0.113	−0.171	0.413	−0.074	0.000	0.000	−0.041	−0.010

total variance, and was highly positively correlated with EC, TDS, Cl^-, Na, Ca, Mg, SO_4^{2-} and Mn. The second factor explained 11.1% of the total variance, and was only highly positively correlated with Fe. In spatial scales of 40 km, the first factor explained 80.4% of the total variance, and had a highly positive loading on EC, TDS, Cl^-, Na, Ca, Mg, SO_4^{2-} and Mn. The second factor exhibited highly positive loading only on Fe. Figures 5a, and 8a show the spatial maps of the first two factors in 28.5 and 40 km scales. Figures 5a and 6a show that in both the spatial scales of 28.5 and 40 km, the sites with the high positive score of the first factor are almost located in coastal areas.

In both spatial scales of 28.5 and 40 km, the second factors have a positive correlation with Ca, Mn, and especially Fe. Since umber loam is the composition of the east gravel tableland of the Pingtung Plain, the aquifer should be abundant in iron oxide. Confined aquifers in the Pingtung Plain were created from alternating layers of permeable gravel and sand, and impermeable silts and clays that deposited in intermontane basins. The grain-size also becomes finer further towards the southwest. The components of the sediments include the mineral of $MgCO_3$ in a carbonate formation. Mn and Fe dissolve in the groundwater by dissolution and ion exchange; there is higher ion concentration of Fe around the aquitard. In this study, for both spatial scales of 28.5 and 40 km, the sites located further north have a lower score in the second factor. This phenomenon is primarily affected by the alluvium of the main river of Pingtung Plain, the Kaoping River.

Optimal multivariate geostatistical groundwater quality monitoring network

Based on the above multivariate geostatistical analysis, the existing 34 groundwater level monitoring wells in the second aquifer are treated as candidate wells for a groundwater quality monitoring network design in the optimization problem. The study area is divided into 1×1 km grids for calculating the total variances for the estimated regionalized factors.

Based on the PCA results, the nine groundwater quality variables in the second aquifer in the Pingtung Plain have 28.5 and 40 km spatial scales, respectively. The optimization of the monitoring network performed according to the following cases with different proposals.

Case 1: considering the first factors in the two spatial scales

In Case 1, the first factors in both 28.5 and 40 km spatial scales are considered, $Y_{p=1}^{u=1}$ and $Y_{p=1}^{u=2}$, simultaneously. Therefore, both the estimated variances of each first factor in 28.5 and 40 km scales should be minimized to obtain an accurate estimation of the factors by using the proposed optimal monitoring well system. To refer to the optimal problem defined by Eqs. 17 and 18, the problem definition of Case 1 is as follows:

Object function

$$\underset{I\subset\Omega}{\text{Min}} J(I) = \sum_{m\in M} \left\{ 0.691\sigma^m_{Y_1^1}(I) + 0.804\sigma^m_{Y_1^2}(I) \right\} \tag{21}$$

Subject to

$$N_I \leq N_{\max}. \tag{22}$$

According to the proportion of the total variance in Table 3, the weights of Y_1^1 and Y_1^2 are 0.691 and 0.804, respectively. The maximum number of monitoring wells is set to 20, 25 and 30, respectively.

Case 2: considering the first factor in the short spatial scale

In Case 2, only the first factor of the 28.5 km spatial scale is considered, $Y_{p=1}^{u=1}$. The estimated variance of the first factor in the 28.5 km scale should be minimized to obtain an accurate estimation of the factor by using the proposed optimal monitoring well system. In Case 2, the optimization problem can be defined as follows:

Object function

$$\underset{I\subset\Omega}{\mathrm{Min}} J(I) = \sum_{m\in M} \sigma^{m}_{Y^{1}_{1}}(I) \qquad (23)$$

Subject to

$$N_I \leq 25. \qquad (24)$$

The maximum number of monitoring wells is set to 25.

Case 3: considering multi-factors in the short spatial scale

In Case 3, two of the first factors in the 28.5 km spatial scales are considered, $Y^{u=1}_{p=1}$ and $Y^{u=1}_{p=2}$, simultaneously. The problem definition of Case 3 is as follows:

Object function

$$\underset{I\subset\Omega}{\mathrm{Min}} J(I) = \sum_{m\in M} \left\{0.691\sigma^{m}_{Y^{1}_{1}}(I) + 0.111\sigma^{m}_{Y^{1}_{2}}(I)\right\} \qquad (25)$$

Subject to

$$N_I \leq 25. \qquad (26)$$

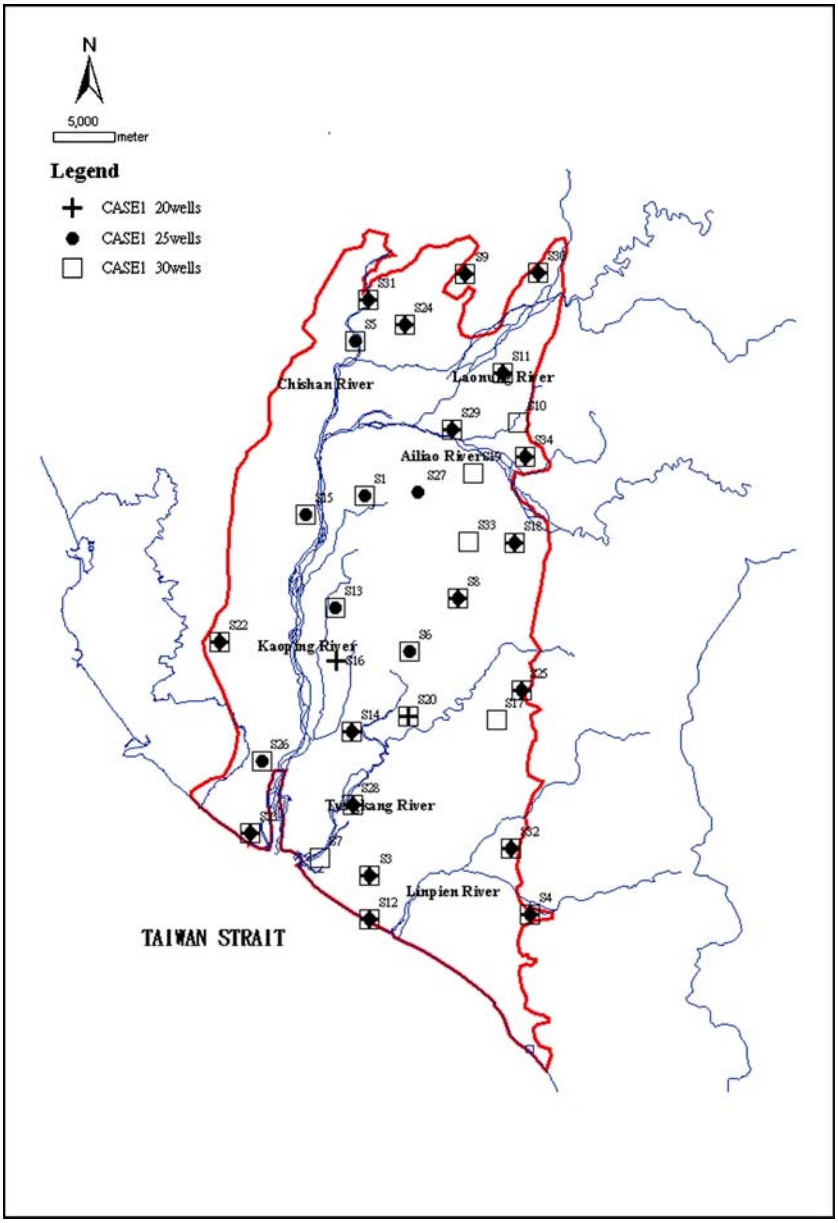

Fig. 4 Optimal network design of the 20, 25, 30 wells of case 1

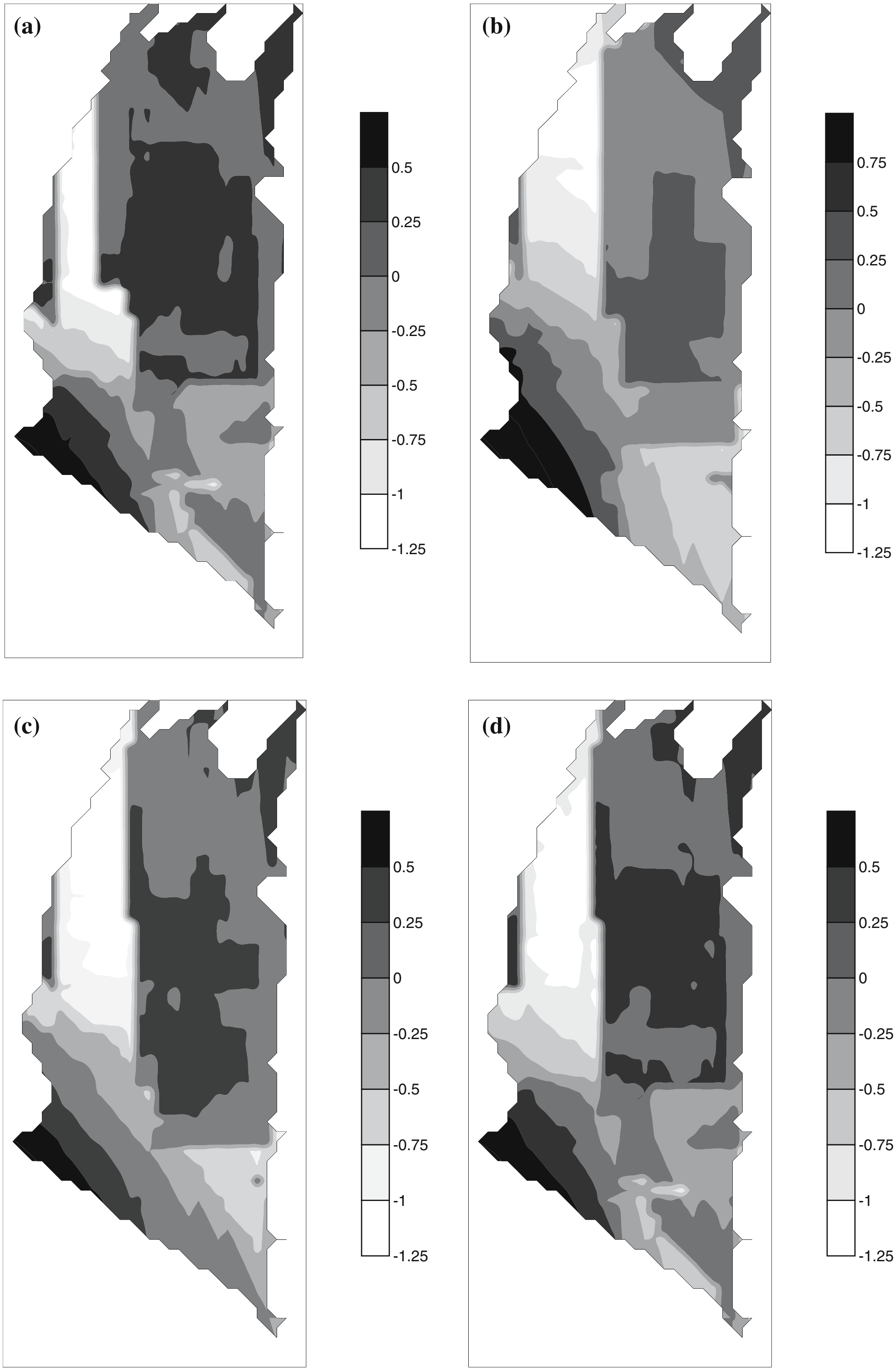

Fig. 5 Mappings of regionalized factors of factor 1 at the 28.5 km scale **a** by 34 well, **b** by Case 1, 20 well, **c** by Case 1, 25 well, **d** by Case 1, 30 well

According to the proportion column in Table 3, the weights of Y_1^1 and Y_2^1 are 0.691 and 0.111, respectively. The maximum number of monitoring wells is set to 25.

Considering the variations of monitoring variables in both short (28.5 km) and long (40 km) range scales, the sum of the estimation variances with weightings of both

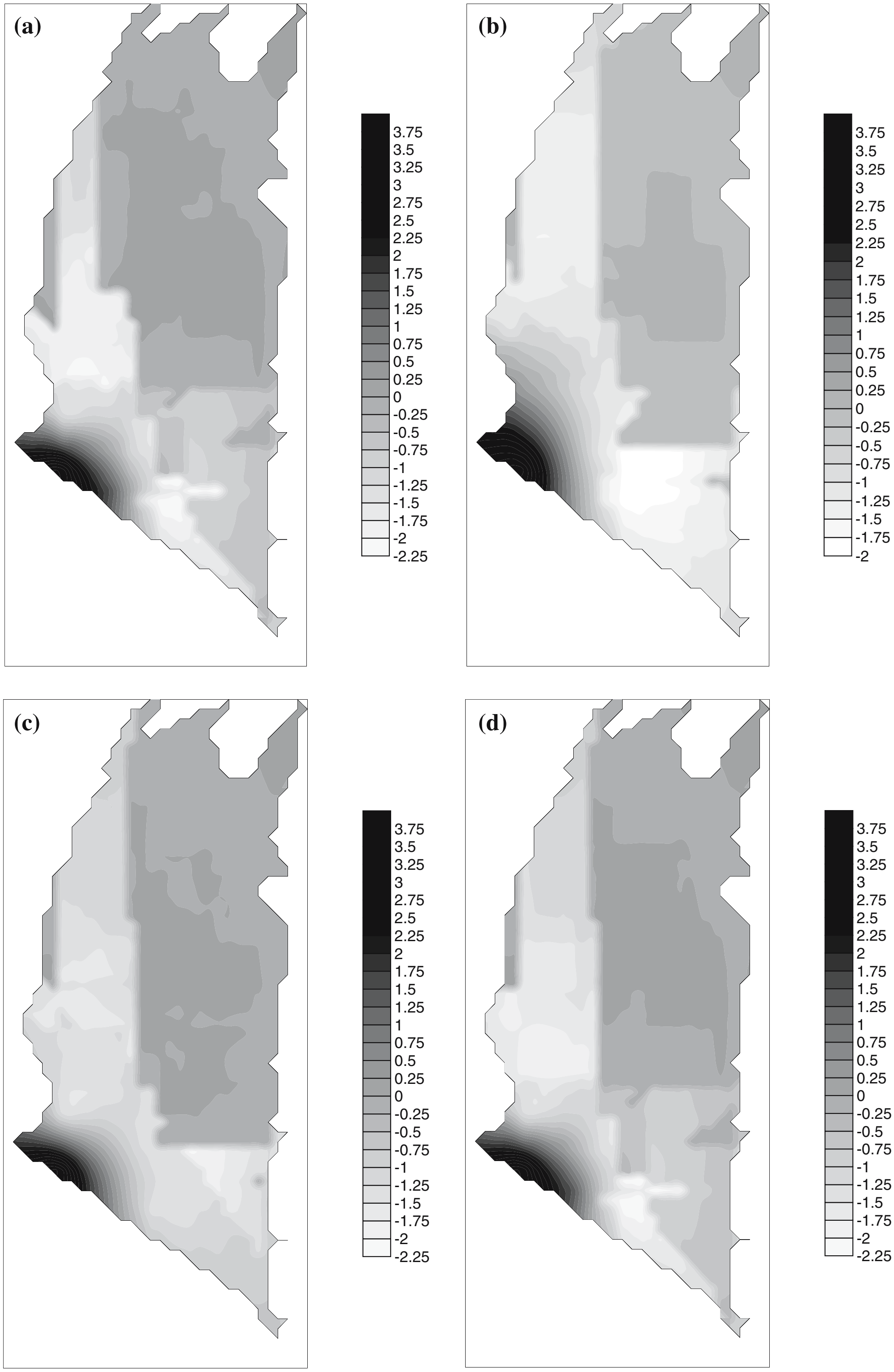

Fig. 6 Mappings of regionalized factors of factor 1 at the 40 km scale **a** by 34 well, **b** by Case 1, 20 well, **c** by Case 1, 25 well, **d** by Case 1, 30 well

first regionalized factors in the two scales is to be minimized using GA for selecting 20, 25 and 30 monitoring wells. The optimal 20, 25 and 30 selected monitoring wells are mapped in Fig. 4. These three monitoring systems focus on the spatial variations in both the local scale (28.5 km) and regional scale (40 km). In the 20-well monitoring system, most of the selected wells are located at the south and north parts of the study area,

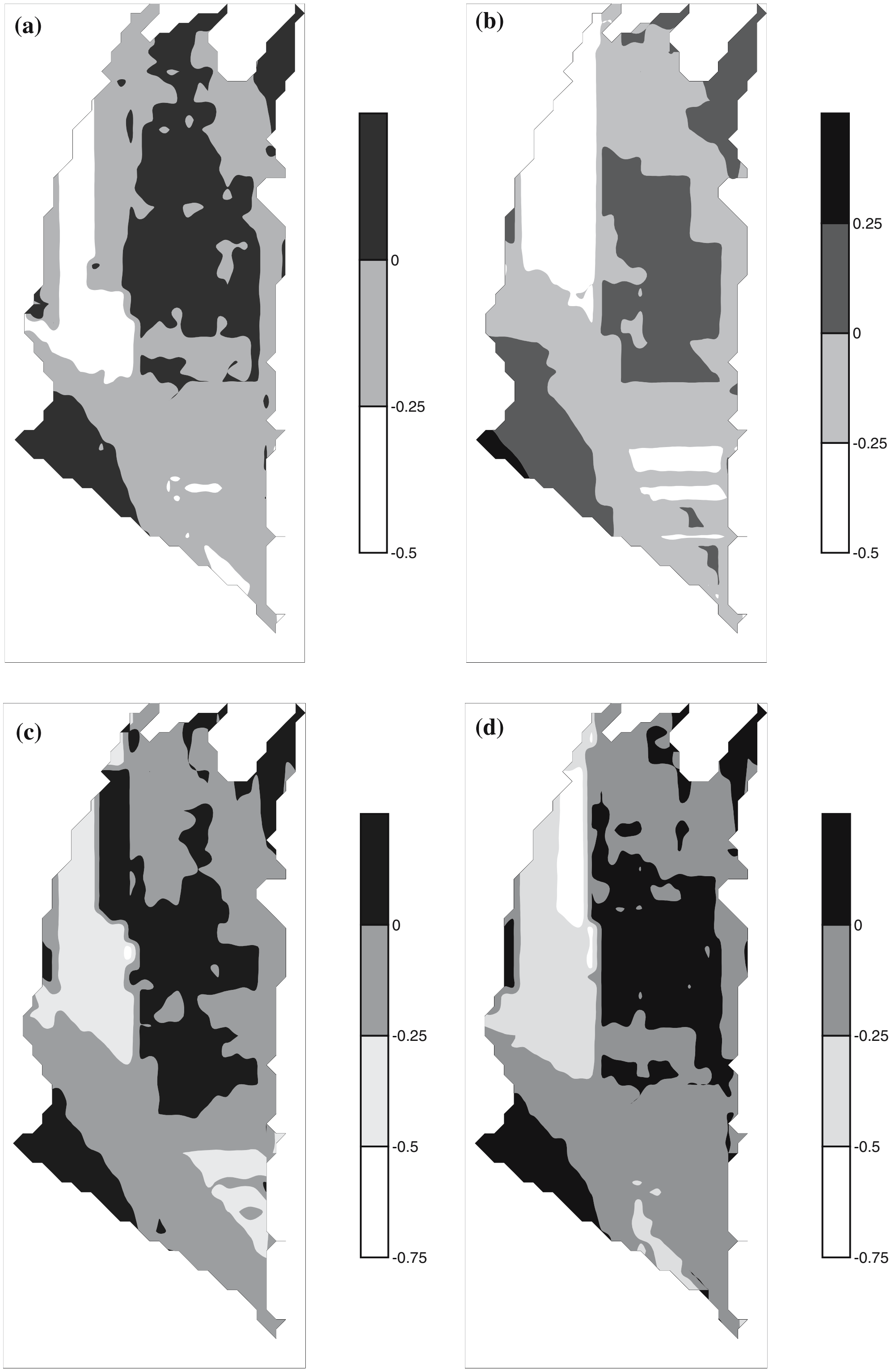

Fig. 7 Mappings of regionalized factors of factor 2 at the 28.5 km scale **a** by 34 well, **b** by Case 1, 20 well, **c** by Case 1, 25 well, **d** by Case 1, 30 well

and fewer wells are located at the eastern parts of the area because of improving estimations of the first factors in both 28.5 and 40 km scales (Figs. 5b, 6b). In the 25-well and 30-well monitoring systems the wells are distributed more uniformly over the study area (Fig. 4). There are 18 identical monitoring wells appearing in all

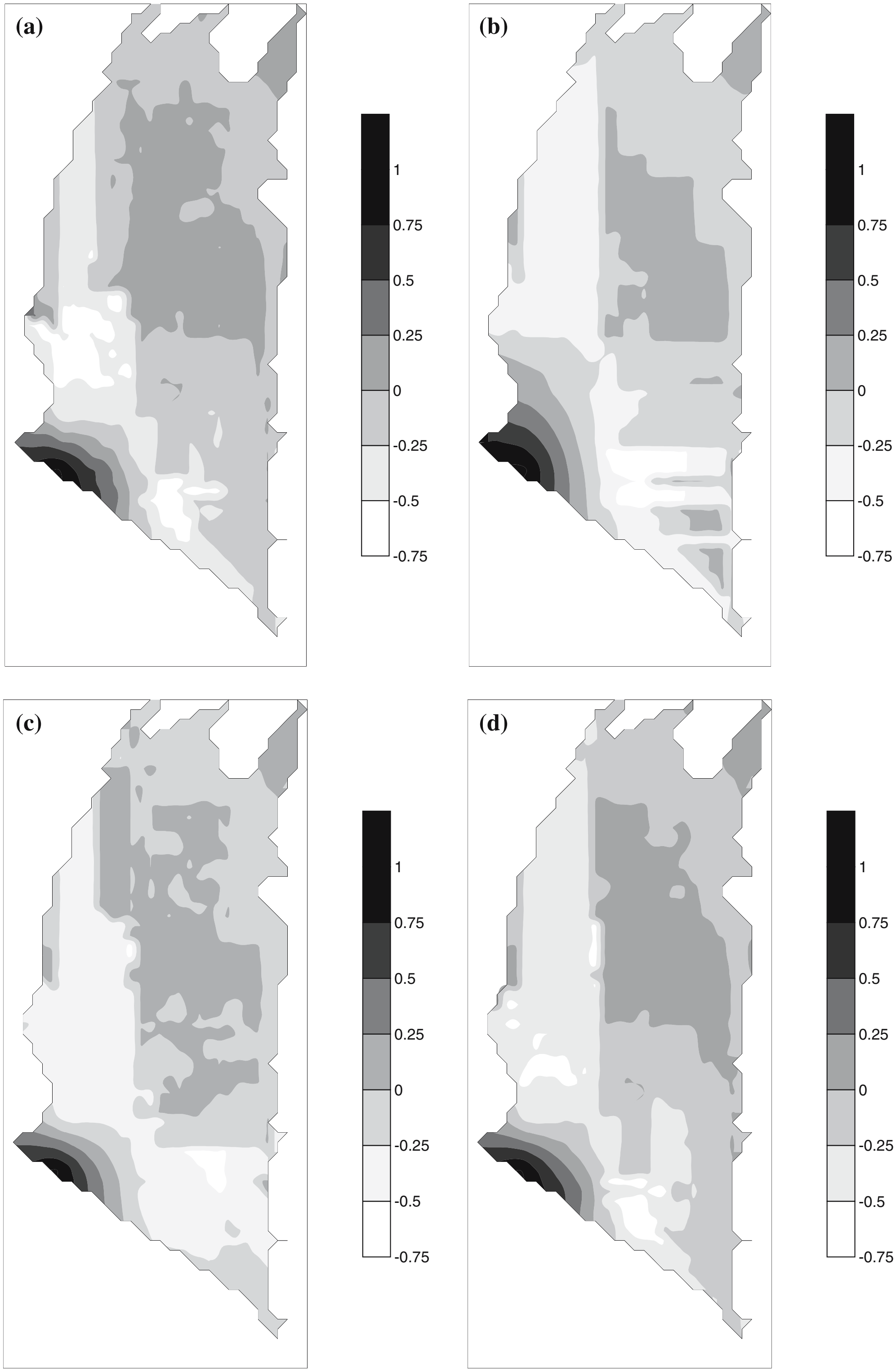

Fig. 8 Mappings of regionalized factors of factor 2 at the 40 km scale **a** by 34 well, **b** by Case 1, 20 well, **c** by Case 1, 25 well, **d** by Case 1, 30 well

20-well, 25-well and 30-well monitoring systems, and the same 24 wells appear in both the 25-well and 30-well systems (Fig. 4). These optimal selection results demonstrate that the 18 wells should be defined as minimum basic wells to delineate spatial variations in both short and long range scales.

Both the first factor scores in 28.5 and 40 km scales are mapped by using 34, 20, 25 and 30 wells with the

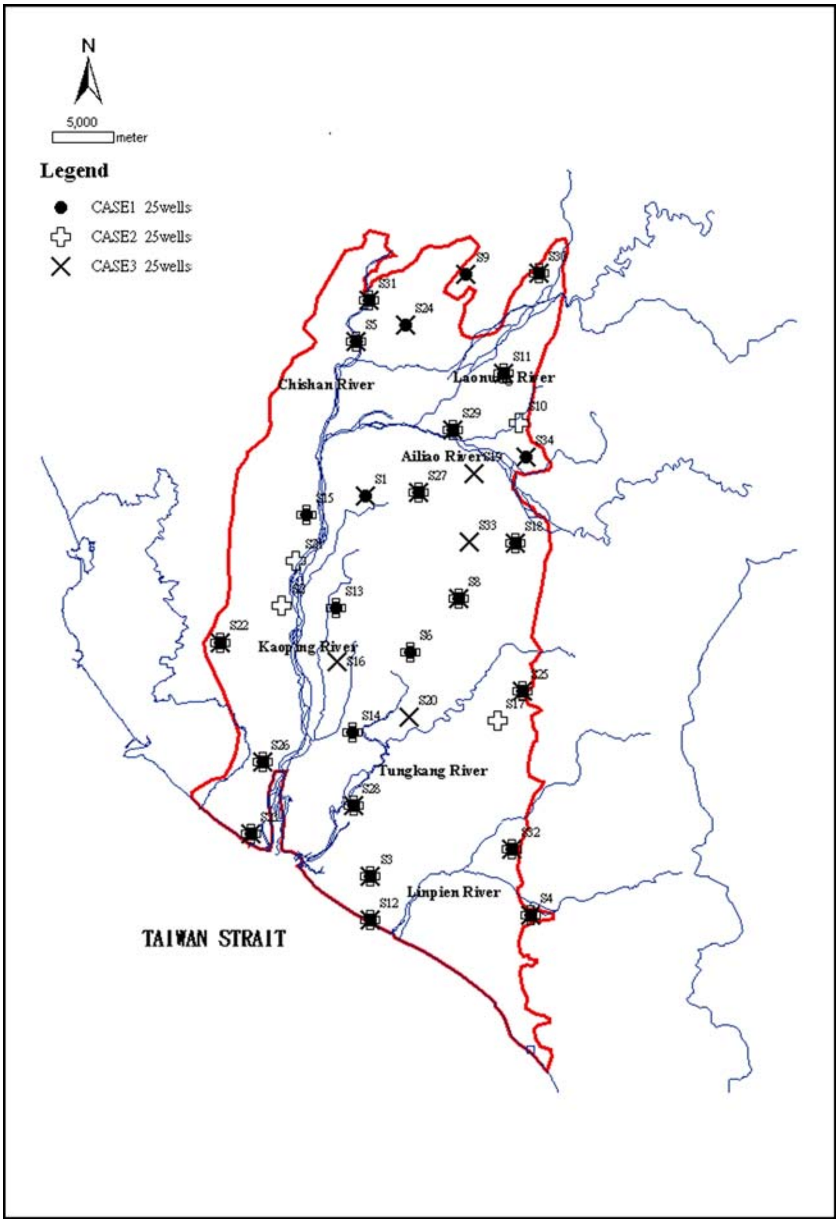

Fig. 9 Optimal network designs of 25 wells of Cases 1, 2 and 3

factor score (Figs. 5, 6, 7, 8). These factor score maps confirm that the more monitoring wells are installed, the more reliable factor score maps can be performed. To delineate spatial variations of the monitoring variables in the regional and local scales, the 30-well system is the best compared to the 25-well and 20-well systems. However, the proposed optimal monitoring design approach selected the monitoring systems which well capture spatial variations of monitoring variables in both scales (Figs. 5, 6, 7, 8).

After considering various purposes for the multi-factors in 28.5 and 40 km scales (Case 1), single factor in 28.5 km (Case 2) and multi-factor in 28.5 km (Case 3), there are 17 identical monitoring wells selected in 25-well systems for these three cases (Fig. 9). The 17 selected wells are likely to be homogeneously distributed in the study area except in the western part. These optimal well selection results imply that the 17 identical monitoring wells could be the baseline monitoring network system to provide information of the total spatial variations of monitoring variables for multi-purpose. The remaining eight monitoring wells are selected for monitoring spatial variations in various scales and purposes (Fig. 9).

Maps of factors mapped by 25-well systems in Cases 2 and 3 are shown in Figs. 10, 11, 12. Comparing Fig. 10 and Fig. 5a, c the 25-well system of Case 2 captures spatial variations of factor 1 in the 28.5 km scale slightly

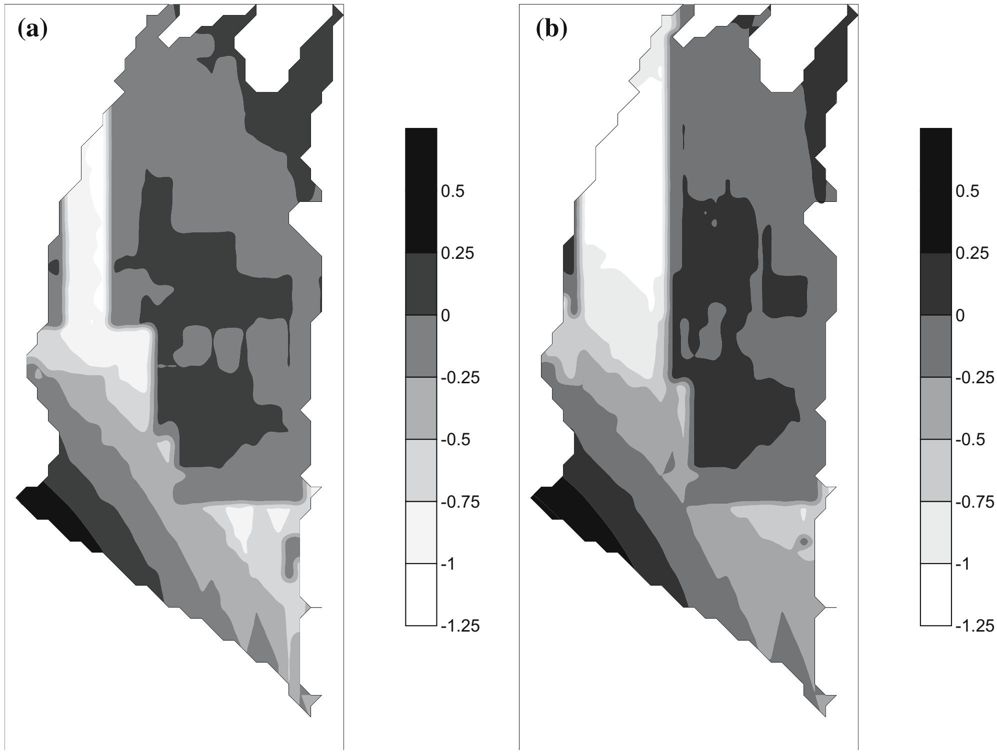

Fig. 10 Mappings of regionalized factors of factor 1 at the 28.5 km scale **a** by Case 2, 25 well, **b** by Case 3, 25 well

better than the systems of Cases 1 and 3. Figures 5a, c, 7a, c, 10 and 12 illustrate that the 25-well system of Case 3 captures the patterns of both factors 1 and 2 in the 28.5 km scale slightly better than the systems of Cases 1 and 2. However, the purposed approach obtains the monitoring system that does not only consider grouped monitoring variables but also delineate spatial variation of the grouped monitoring variables. Therefore, unlike

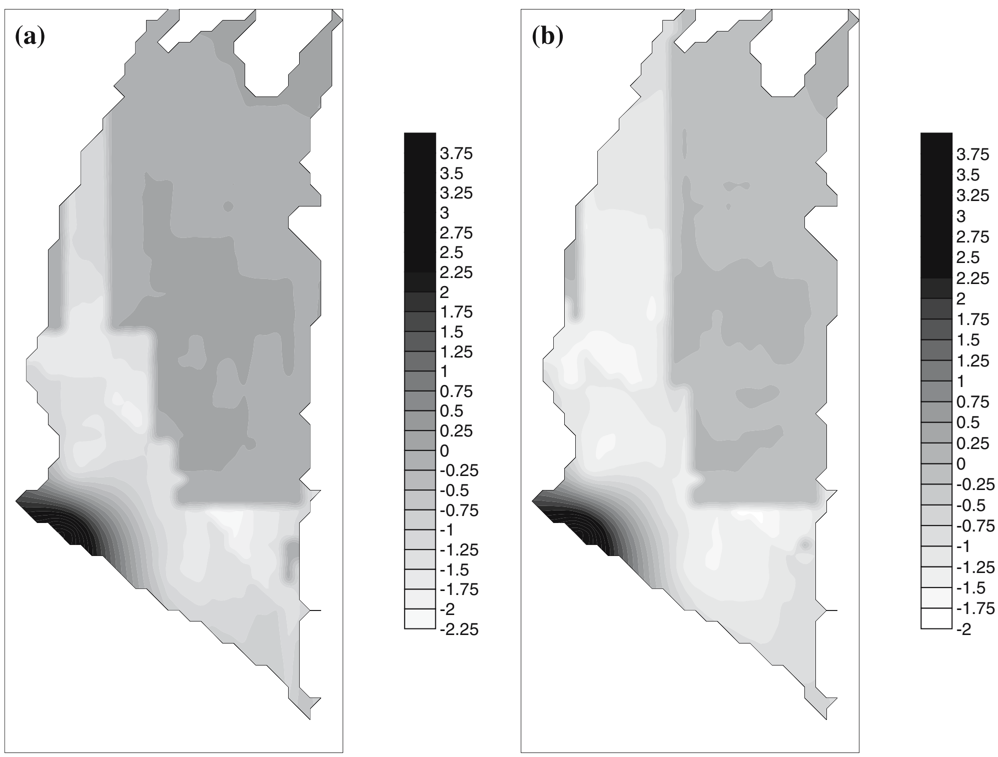

Fig. 11 Mappings of regionalized factors of factor 1 at the 40 km scale **a** by Case 2, 25 well, **b** by Case 3, 25 well

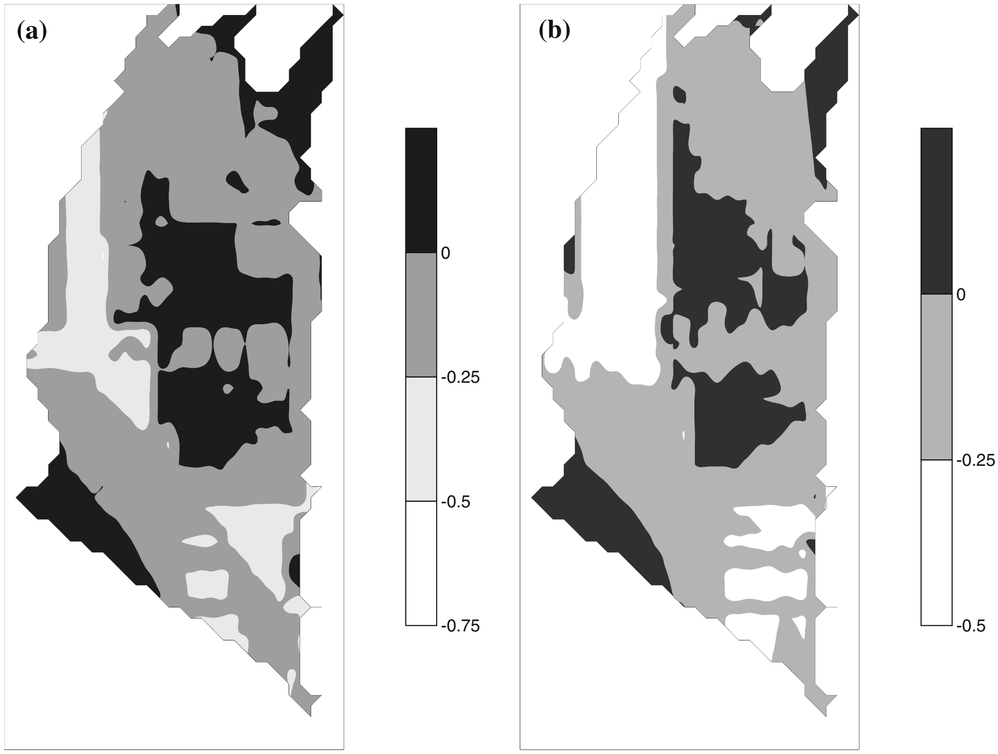

Fig. 12 Mappings of regionalized factors of factor 2 at the 28.5 km scale **a** by Case 2, 25 well, **b** by Case 3, 25 well

previous related studies different optimal systems are obtained by the purposed approach based on various factors to reach the monitoring purposes in different scales. For example, the monitoring well system based on the second factor is different to that based on the first factor or single monitoring variable. The monitoring well system based on the second factor will provide useful information of the monitoring variables included in the second factor for identifying the variations of the variables. The more the factors considered in the optimal monitoring system design, the more is the information provided by the designed monitoring system. Moreover, the purposed approach can obtain different monitoring systems that provide varied information of the monitoring variables in space. The maps and optimal results also confirmed that the proposed model can design an optimal groundwater monitoring network that not only considers various factors grouped by multiple groundwater quality variables but also monitors variations of monitoring variables at various spatial scales in the study area.

Conclusion

In the past groundwater quality monitoring design studies, a groundwater item was considered as a monitoring design variable for designing a system to monitor multiple items in real practice. This study develops a novel approach to design an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging with GAs. The proposed approach designs a monitoring system which not only considers multi-variables, but also monitors spatial variations of the variables in various scales. In the approach, a multivariate geostatistical analysis is used to decompose multiple variables into small sets of spatial factors in various spatial scales. Based on the multivariate geostatistical analysis the proposed optimization model minimizes the estimation variance of the spatial factor, needed to design a groundwater quality monitoring network considering one or multi-spatial scales in accordance with the different monitoring goals. GAs are suitable for use with factorial kriging to obtain optimal results. The designed monitoring system can be used to delineate spatial variation patterns and sources of multiple groundwater quality items. The proposed approach was also successfully applied in a real case in Taiwan to design optimal monitoring systems for various purposes in order to delineate spatial variations in various scales. In future studies, the developed model could also be modified and applied to design a monitoring system for multi-aquifer cases.

References

Al-Zahrani MA, Moied K (2003) Optimizing water quality monitoring stations using genetic algorithms. Arab J Sci Eng 28(1B):57–75

Batista AC, Sousa AJ, Batista MJ, Viegas L (2001) Factorial kriging with external drift: a cause study on the Penedono Region, Portugal. Appl Geochem 16(7–8):921–929

Benjemaa F, Mario MA, Loaiciga HA (1994) Multivariate geostatistical design of groundwater monitoring networks. J Water Resour Plan Manag ASCE 120(4):505–522

Bocchi S, Castrignano A, Fornaro F, Maggiore T (2000) Application of factorial kriging for mapping soil variation at field scale. Eur J Agron 13:295–308

Brus DJ, Spatjens LEEM, de Gruijter JJ (1999) A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. Goderma 89:129–148

Cameron K, Hunter P (2002) Using spatial models and kriging techniques to optimize long-term groundwater monitoring networks: a case study. Environmetrics 13(5–6):629–656

Castrignano A, Giugliarini L, Risaliti R, Martinelli N (2000a) Study of spatial relationships amon some soil physico-chemical properties of a field in central Italy using multivariate geostatistics. Geoderma 97:39–60

Castrignano A, Goovaerts P, Lulli L, Bragato G (2000b) A geostatistical approach to estimate probability of occurrence of Tuber melanosporum in relation to some soil properties. Geoderma 98:95–113

Chang LC, Hsiao CT (2002) Dynamic optimal groundwater remediation including fixed and operation costs. Ground Water 40(5):481–490

Christakos G, Olea R(1988) A multiple-objective optimal exploration strategy. Math Comput Model 11:413–418

Cieniawski SE, Eheart JW, Ranjithan S (1995) Using genetic algorithm to solve a multiobjective groundwater monitoring problem. Water Resour Res 31(2):399–409

Deutsch CV, Journel AG (1992) Geostatistical software library and user's guide. Oxford University Press, New York

Dobermann A, Goovaerts P, George T (1995) Sources of soil variation in an acid Ultisol of the Philippines. Geoderma 68:173–191

Einax JW, Soldt U (1998) Multivariate geostatistical analysis of soil contaminations. Fresen J Anal Chem 361:10–14

Ferreyra RA, Apezteguia HP, Rereno R, Jones JW (2002) Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing. Geoderma 110:265–289

Goldberg DE (1989) Genetic algorithm in search, optimization, and machine learning. Addison-Wesley, Reading

Goovaerts P (1992) Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. J Soil Sci 43:597–619

Goovaerts P (1994) Study of spatial relationships between two sets of variables using multivariate geostatistics. Geoderma 62:93–107

Goovaerts P (1997) Geostatistics for natural resource evaluation. Oxford University Press, Oxford

Goovaerts P (1998) Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Bio Fertil Soils 27:315–334

Goovaerts P, Webster R (1994) Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. Eur J Soil Sci 45(1):79–95

Holland JH (1992) Adaptation in Natural and Artificial Systems, 2nd edn. Mass Inst Technol, Cambridge

Hsiao CT, Chang LC (2002) Dynamic Optimal Groundwater Management with Inclusion of Fixed Costs. J Water Resource Plan Manag ASCE 128(1):57–65

Hudak PF, Loaiciga HA (1993) An optimization method for monitoring network design in multilayered groundwater flow systems. Water Resource Res 29(8):2835–2845

Jiménez-Espinosa R, Chica-Olmo M (1999) Application of geostatistics to identify gold-rich areas in the Finisterre-Fervenza region, NW Spain. Appl Geochem 14(1):133–145

Lark RM (2000) Design sampling grids from imprecise information on soil variability, and approach based on the fuzzy kriging variance. Geoderma 98:35–59

Lark RM (2002) Optimized spatial sampling of soil for estimation of the variogram by maximum likelihhod. Geoderma 105:49–80

Lin YP (2002) Multivariate geostatistical methods to identify and map spatial variations of soil heavy metals. Environ Geol 42:1–10

Lin YP, Rouhani S (2001) Multiple-point variance analysis for optimal adjustment of a monitoring network. Environ Monit Assess 69(3):239–266

Lin YP, Chang TK, Shih CW, Tseng CH (2002) Factorial and indicator kriging with geographic information system to delineate spatial variations and pollution sources of soil heavy metals. Environ Geol 42:900–909

Loaiciga HA (1989) An optimization approach for groundwater quality monitoring network design. Water Resour Res 25(8):1771–1782

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

Mitchell M (1998) An introduction to genetic algorithms. MIT Press, Cambridge

Mohan S (1997) Parameter estimation of nonlinear muskingum models using genetic algorithm. J Hydraul Eng 123(2):137–142

Pannatier Y (1996) Variowin: software for spatial data analysis in 2D. Springer, Berlin Heidelberg New York

Pardo-Iquzquiza E, Dowd PA (2002) FACTOR2D: a computer program for factorial cokriging. Comput Geosci 28:857–875

Passarella G, Vurro M, D'Agostino V, Barcelona MJ (2003) Cokriging optimization of monitoring network configuration based on fuzzy and non-fuzzy variogram evaluation. Environ Monit Assess 82:1–21

Pesti G, Kelly WE, Bogardi I (1994) Observation network design for selecting locations for water-supply wells. Environmetrics 5(2):91–110

Prakash MR, Sihgn VS (2000) Network design for groundwater monitoring—a case study. Environ Geol 39(6):628–632

Reed P, Minsker B, Valocchi AJ (2000) Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation. Water Resource Res 36 (12):3731–3741

Rogers LL, Dowla FU (1994) Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. Water Resource Res 30(2):457–481

Rouhani S (1985) Variance reduction analysis. Water Resource Res 21(6):837–846

Rouhani S, Hall TJ (1988) Geostatistical schemes for groundwater sampling. J Hydrol 103:85–102

Rouhani S, Wackernagel H, (1990) Multivariate geostatistical approach to space-time data analysis. Water Resource Res 26:585–591

Van Groenigen JW, Siderius W, Stein A (1999) Constrained optimal of soil sampling for minimization of the kriging variance. Geoderma 87:239–259

Wackernagel H (1994) Cokriging versus kriging in regionalized multivariate data analysis. Geoderma 62:83–92
Wackernagel H (1995) Multivariate geostatistics: an introduction with applications. Springer, Berlin Heidelberg New York
Wang QJ (1991) The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. Water Resource Res 27(9):2467–2471
Wang XJ, Qi F (1998) The effects of sampling design on spatial structure analysis of contaminated soil. Sci Total Environ 224:29–41
Wardlaw R, Sharif M(1999) Evaluation of genetic algorithms for optimal reservoir system operation. J Water Resource Plan Manag ASCE 125(1):25–33