Environ Monit Assess (2010) 161:247–257
DOI 10.1007/s10661-008-0742-3

Optimal redesign of groundwater quality monitoring networks: a case study

Fariborz Masoumi · Reza Kerachian

Received: 29 July 2008 / Accepted: 23 December 2008 / Published online: 6 February 2009

Abstract Assessment and redesign of water quality monitoring networks is an important task in water quality management. This paper presents a new methodology for optimal redesign of groundwater quality monitoring networks. The measure of transinformation in discrete entropy theory and the transinformation–distance (*T*–*D*) curves are used to quantify the efficiency of sampling locations and sampling frequencies in a monitoring network. The existing uncertainties in the *T*–*D* curves are taken in to account using the fuzzy set theory. The C-means clustering method is also used to classify the study area to some homogenous zones. The fuzzy *T*–*D* curve of the zones is then used in a multi-objective hybrid genetic algorithm-based optimization model. The proposed methodology is utilized for optimal redesign of monitoring network of the Tehran aquifer in the Tehran metropolitan area, Iran.

F. Masoumi · R. Kerachian (✉)
Center of Excellence for Infrastructure Engineering and Management, Faculty of Civil Engineering, University of Tehran, Tehran, Iran
e-mail: kerachian@ut.ac.ir

F. Masoumi
e-mail: fariborz_mg@yahoo.com

Keywords Groundwater quality · Monitoring · C-means clustering method · Discrete entropy theory · Transinformation · Fuzzy sets theory · Hybrid genetic algorithms

Introduction

The groundwater quality monitoring networks should be revised from time to time in concern with the changing data objectives and needs. The maim objective of a monitoring network is to gather information efficiently for various purposes. This basic goal can be achieved by undertaking (1) an investigation of the information required, (2) development of design strategies for an efficient collection of sufficient information, and (3) optimization of the monitoring network design according to a fixed cost. Design of water quality monitoring networks is still a controversial issue because there are difficulties in the selection of temporal and spatial sampling frequencies, the variables to be monitored, the sampling duration, and the objectives of sampling (Harmancioglu et al. 1999).

Essentially, available design methodologies serve to assess the efficiency of existing water quality monitoring networks. Although, each method focuses on the problem from a different perspective, using different criteria, there is still the question of how one relates such criteria in the assessment process to the value of data. Within this respect, one of the most promising methods for network assessment purposes is based on the entropy concept of information theory, which has been used to evaluate not only water quality but also other hydrometric networks.

Harmancioglu and Yevjevich (1987) used the entropy theory to assess the goodness of information transfer by regression using the monthly water quality data of a polluted river basin. Husain (1989) presented a simple entropy-based concept to estimate regional hydrologic uncertainties at both gauged and ungauged stations in a river basin. Yang and Burn (1994) showed that entropy theory can be efficiently used for evaluating the information transmission characteristics of a monitoring network.

Ozkul et al. (2000) presented a method using entropy theory for assessing existing water quality monitoring networks. Discrete and analytical entropy theories were utilized by Mogheir et al. (2004) to characterize the spatial variability of groundwater quality. The analytical entropy theory was also used by Salark and Sorman (2006) for evaluating existing monitoring stations in rivers. The existing salinity monitoring network in Tehran aquifer was assessed by Masoumi and Kerachian (2008a). They proposed a crisp optimization model for proposing sampling stations for water quality variable electrical conductivity (EC). A fuzzy version of this optimization model has also been presented by Masoumi and Kerachian (2008b). Karamouz et al. (2009) utilized the measure of transinformation in entropy theory for selecting the best monitoring stations from a set of potential monitoring sites along a river.

In this paper, a new version of the optimization models proposed by Masoumi and Kerachian (2008a, b) is proposed so that it can consider several water quality variables. This multi-objective optimization model can be utilized for optimally updating the location of monitoring wells and the sampling frequency of water quality indicators. The optimization model uses a fuzzy transinformation–distance (T–D) curve, which is obtained using discrete entropy theory. An entropy-based methodology, which uses the transinformation–time lags (T–T) curves, is also presented for updating sampling frequencies. The proposed mythology is used for redesigning the groundwater monitoring system of the Tehran Aquifer, Tehran, Iran.

Entropy theory

Entropy theory has been fully described in literature (e.g., Ozkul et al. 2000 and Mogheir et al. 2004). In entropy theory, the measures of information include the marginal entropy, joint entropy, conditional entropy, and transinformation. The measure of transinformation between two random variables x and y is interpreted as the information content of x that is contained in y. Transinformation ($T(x, y)$) in the discrete form can be expressed as follows:

$$T(x, y) = -\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p(x_i, y_j) \ln\left[\frac{p(x_i, y_j)}{p(x_i)\, p(y_j)}\right] \tag{1}$$

where x and y are two discrete variables with values x_i, $i = 1,2,3,...,n$; y_i, $j = 1,2,3,...,m$. $p(x_i)$ and $p(y_i)$ are the discrete probabilities of occurrence of x_i and y_j; and $p(x_i, y_j)$ is the joint probability of x_i and y_j.

In discrete entropy theory, the contingency tables record the frequency for the values that fall into each possible combination of two categories. To construct a two-dimensional contingency table, the following steps are involved (Mogheir and Singh 2003):

- Consider v categories (class intervals) for random variable x and u categories (class intervals) for random variable y. The number of class intervals can be calculated as follows:

  $$\mathrm{NCI} = 1 + 1.33\,\mathrm{Ln}\,(n) \tag{2}$$

  where NCI is the number of class intervals, and n is the size of the time series of the variables.

Table 1 A two-dimensional contingency table (Mogheir and Singh 2003)

		y					
		1	2	3	...	u	Total
x	1	f_{11}	f_{12}	f_{13}	...	f_{1u}	$f_{1.}$
	2	f_{21}	f_{22}	...	...	f_{2u}	$f_{2.}$
	3	f_{31}	...	...	...	f_{3u}	$f_{3.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	v	f_{v1}	f_{v2}	f_{v2}	...	f_{vu}	$f_{v.}$
	Total	$f_{.1}$	$f_{.2}$	$f_{.3}$	...	$f_{.u}$	f_x or f_y

- The joint frequencies (cell densities) for (i, j) is denoted by f_{ij}, where the first subscript is related to the row, and the second one is related to the column. The cell density is the count of measurements with the corresponding u class interval of variable y and v class interval of variable x.
- The marginal frequencies are denoted by f_i and f_j, which are the summation of the cell densities for each category of variables x and y, respectively (see Table 1 for more details).
- $P(x_i, y_j)$ is calculated by dividing the cell density by total number of the data recorded in one monitoring well.

In this paper, the transinformation is calculated for each pair of the existing monitoring wells, and the values of the transinformations are plotted against distance between the monitoring wells (Fig. 1). An exponential T–D curve is usually fitted to these points because the initial transinformation decreases and reaches a minimum value of $T_{\min}$ at a distance equal to the range (d^*). d^* can be considered as the optimal distance between two adjacent monitoring wells.

Model formulation

To incorporate the existing uncertainty in the transinformation–distance (T–D) curve, the fuzzy set theory is used and for each distance; a triangular membership function is considered for the corresponding transinformation. As shown in Fig. 1, upper and lower bounds are considered for the T–D curve to set the membership functions.

The objective functions of the proposed optimization model are as follows:

1. Maximizing the satisfaction level (f) in the fuzzy membership functions
2. Minimizing the redundant information in the system
3. Maximizing the coverage of the monitoring system

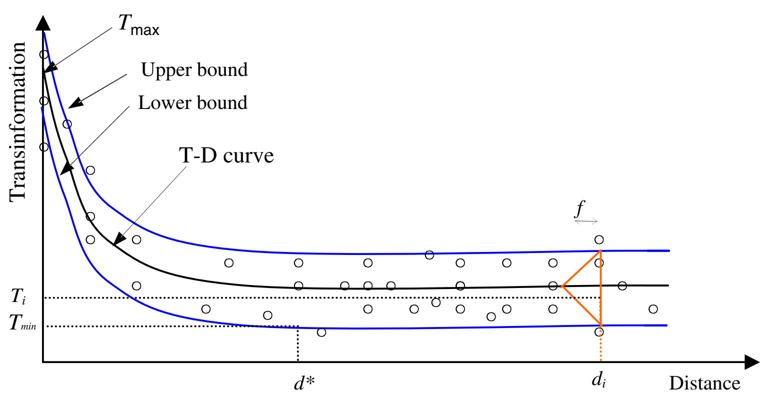


Fig. 1 A typical transinformation–distance curve (adopted from Masoumi and Kerachian 2008b)

The second and third objective functions are quantified using the variables $\bar{L}_2$ and $\bar{L}_3$. The objective functions are standardized between 0 and 1 and weighed to provide a single objective function:

$$\text{Maximize } Z = w_1 f + w_2 \bar{L}_2 + w_3 \bar{L}_3 \tag{3}$$

$$\bar{L}_2 = \sum_{k=1}^{K} \alpha_k \sum_{z=1}^{Z} \sum_{i=1}^{n_z} \frac{L_{2,i,k,z}}{n_z} \tag{4}$$

$$\bar{L}_3 = \sum_{k=1}^{K} \alpha_k \sum_{z=1}^{Z} \sum_{i=1}^{n_z} \frac{L_{3,i,k,z}}{n_z} \tag{5}$$

$$L_{2,i,z,k} = \begin{cases} X_{i,z} \times \dfrac{T_{\max,z,k} - T_{i,z,k}}{T_{\max,z,k} - T_{\min,z,k}} & \text{if } d_{i,z,k} < d^*_{i,z,k} \\ X_{i,z} & \text{if } d_{i,z,k} \geq d^*_{i,z,k} \end{cases} \quad \forall i\ , z, k \tag{6}$$

$$L_{3,i,z,k} = \begin{cases} X_{i,z} \times \dfrac{d_{i,z,k} - d^*_{z,k}}{d_{\max,k,z} - d^*_{z,k}} & \text{if } d_{i,z,k} > d^*_{i,z,k} \\ X_{i,z} & \text{if } d_{i,z,k} \leq d^*_{i,z,k} \end{cases} \quad \forall i, z, k \tag{7}$$

$$X_{i,z} \in [0,\ 1] \tag{8}$$

$$\sum_{z=1}^{Z} \sum_{i=1}^{I_z} X_{i,z} = n \tag{9}$$

$$T_{i,z,k} = g\left(d_{i,z,k},\ f\right) \qquad \forall i, z, k \tag{10}$$

$$\sum_{k=1}^{K} \alpha_k = 1 \tag{11}$$

$$\sum_{i=1}^{I} w_i = 1 \tag{12}$$

where

w_i	relative importance weight of objective function i, $i = 1,2,3$
K	Total number of water quality variables
Z	Total number of zones
a_k	Relative importance weight of water quality variable k, $k = 1, 2$
n_z	Total number of required monitoring wells in zone z
n	Total number of required monitoring wells in the study area
$d_{i,z,k}$	Distance between potential well i and its nearest neighboring potential well in zone z for water quality variable k
$T_{i,z,k}$	The transinformation value corresponding to distance d_i for potential well i in zone z for water quality variable k
$T_{\max,z,k}$	Maximum value of the transinformation on the T–D curve in zone z for water quality variable k
$T_{\min,z,k}$	Minimum value of the transinformation on the T–D curve in zone z for water quality variable k
$d^*_{z,k}$	Minimum distance corresponding to $T_{\min}$ in zone z for water quality variable k (see Fig. 1)
f	The satisfaction level in the fuzzy membership functions (see Fig. 1)
$d_{\max,k,z}$	Maximum distance between two neighboring potential wells in zone z for water quality variable k
$L_{2,i,z,k}$	Value of the second objective function, which is less than 1 when d_i is less than d^* in zone z for water quality variable k
$L_{3,i,z,k}$	Value of the third objective function, which is less than 1 when d_i is greater than d^* in zone z for water quality variable k
$X_{i,z}$	An integer variable with the values of 0 or 1. Its value is one when a monitoring well is located in the potential well i in zone z
$g(d_{i,z,k},\ f)$	A function which is presented by the fuzzy T–D curve

Flowchart of the proposed methodology is presented in Fig. 2. Different components of this flowchart are described in following sections. As shown in Fig. 2, entropy theory can also be used for updating the existing sampling frequencies of water quality variables. Transinformation–time lags (T–T) curves can be calculated for water quality variables at some indicator monitoring wells. Variations of T–T curve of a water quality variable can be used for selecting the best sampling frequency. In a typical T–T curve for a water quality variable, when transinformation decreases and reaches a minimum value of $T_{\min}$ at a time lag equal to the range t^*. t^* can be considered as the optimal sampling frequency for the water quality variable.

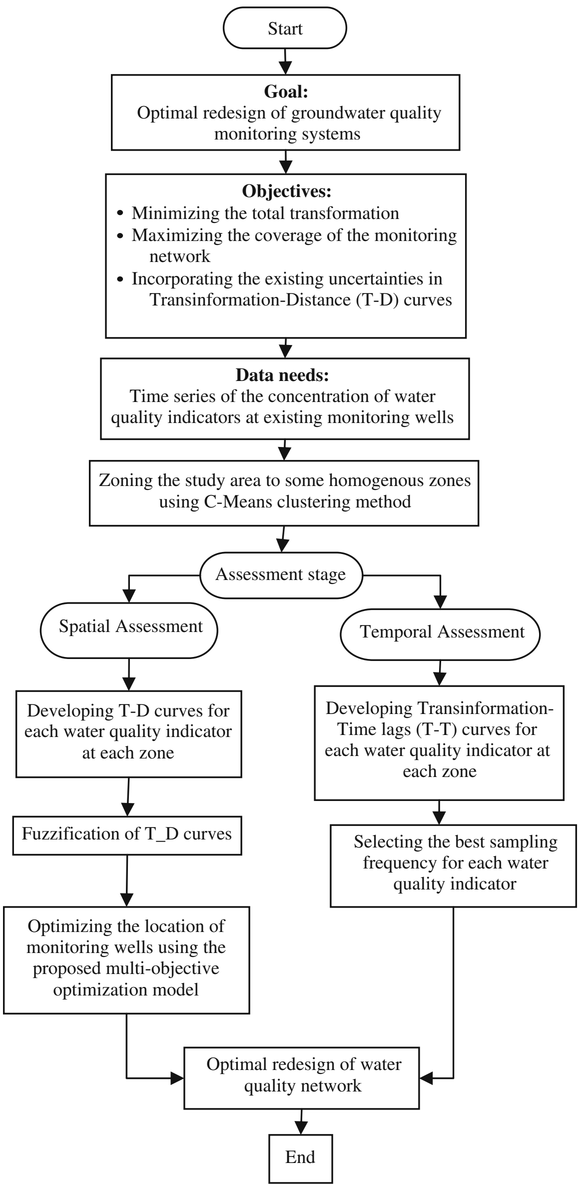


Fig. 2 Flowchart of the proposed methodology for optimal redesign of groundwater quality monitoring systems

Hybrid genetic algorithms

In this study, a hybrid genetic algorithm is used to solve the proposed optimization problem. Genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, the process that drives biological evolution. The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals randomly from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.

One can apply the genetic algorithm to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, non-differentiable, stochastic, or highly nonlinear. More detailed information about GAs can be found in Gen and Cheng (2000).

In the hybrid genetic algorithm (HGA), a pattern search algorithm is run after the genetic algorithm terminates in order to improve the value of the fitness function. The hybrid function uses the final point from the genetic algorithm as its initial point.

Pattern search algorithms compute a sequence of points that get closer and closer to the optimal point. At each step, the algorithm searches a set of points, called a mesh, around the current point, the point computed at the previous step of the algorithm. If the pattern search algorithm finds a point in the mesh that improves the objective function at the current point, the new point becomes the current point at the next step of the algorithm.

C-means clustering method

Clustering is the process of grouping a set of objects into classes of similar objects. One of the most popular clustering techniques is C-means, to the extent that it is hardly possible to find a clustering software that would not offer C-means as a method of classification. The C-means algorithm minimizes the sum of squared distance of all points included in a cluster space to the center of the cluster. The distance between the center of cluster i and kth data sample (d_{kj}) can be calculated as follows:

$$d_{ik} = \left[\sum_{j=1}^{m} \left(x_{kj} - v_{ij} \right)^2 \right]^{0.5} \qquad \forall i, k \tag{13}$$

where x_{kj} is the jth coordinate of kth data sample ($j = 1, 2, \ldots, m$) and v_{ij} is the ith coordinate of the center of cluster i. In this paper, the C-means method is used to cluster the study area to some homogenous zones considering the groundwater quality and the spatial location of potential monitoring wells.

Case study

The proposed methodology is utilized for optimal redesign of the groundwater quality monitoring system in Tehran Aquifer, Iran. More than eight million people are living in the Tehran City, and about 60% of domestic water consumption in this region returns to the Tehran aquifer via traditional absorption wells. Total domestic water demand in Tehran metropolitan area is about one billion cubic meters per year. This water demand is supplied from Tehran aquifer and Karaj, Lar, and Latyan Dams. The share of groundwater in water supply to Tehran is raised up to 60% during drought conditions (Karamouz et al. 2004, 2007).

The Tehran aquifer is mainly recharged by inflows at the boundaries, precipitation, local rivers, and return flows from domestic, industrial, and agricultural water users. The discharge from the aquifer is through water extraction from wells, springs, and qanats as well as groundwater outflow and evapotranspiration. Wastewater disposal in Tehran is carried out through more than three million absorption wells, which are often 15–20 m deep. The use of absorption wells has caused groundwater pollution and a significant rise of the water table in the southern part of Tehran (Bazargan-Lari et al. 2009). Comparison between the average groundwater table in April 1985 and April 2007 shows that there is about a 10 m increase in water table level in the central and the Southern parts of the aquifer that is mainly due to aquifer recharge by domestic wastewater and return flows (Masoumi and Kerachian 2008a, b). According to the existing groundwater quality data, some water quality variables such as total dissolved solids, nitrate, and caliform bacteria are violating the standards. In this study, the existing groundwater quality monitoring systems is assessed and redesigned considering the water quality variables of EC and sodium absorption ratio (SAR). The available data about the other water quality variables is limited so that they

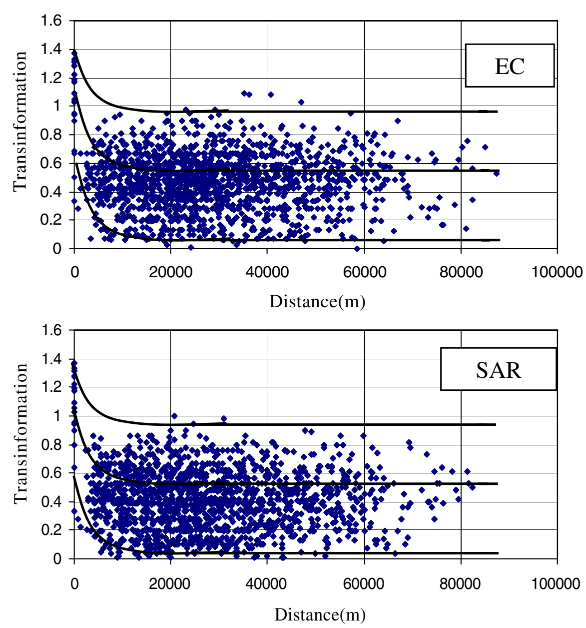


Fig. 3 Variation of the transinformation versus distance in the study area for the water quality variables EC and SAR

cannot be considered for evaluating the effectiveness of the proposed methodology.

Results and discussion

As the Tehran aquifer supplies the water demands of several agricultural zones in the southern part of Tehran, EC and SAR are considered as water quality variables. The available water quality data are used for developing the contingency tables and the T–D curves.

Figure 3 shows the variation of transinformation versus distance in the study area. To find the equation of the T–D curve and its upper and lower bounds, three exponential curves are fitted to the data presented in Fig. 3. To improve the accuracy of the T–D curves, the existing monitoring wells are clustered to three classes (zones), considering the criteria of the location of monitoring wells, total number of upstream wells, the average, and standard deviation of the concentrations of EC and SAR (Fig. 4).

Figure 5 presents the variations of transinformation versus distance in each zone. Tables 2 and 3 present the equations of T–D curves and their upper and lower bounds for the study area and each zone. These tables also present the optimal distance between monitoring wells (d*).

As shown in Fig. 4, the existing groundwater quality monitoring system does not completely cover the Tehran aquifer, and the distances between some of the wells are much more or less than the optimum distance d*.

As mentioned before, a HGA is used to solve the proposed optimization model. In the HGA, each gene, which shows the value of the decision variables X_i (i = 1,..., m), has one bit. Therefore, for m potential monitoring wells, each chromosome has m genes. In order to find a robust solution, the probability of mutation and crossover were calculated through a trial and error process as 0.001 and 0.8, respectively. For a population size of 80, the HGA-based optimization model converges to the optimal or near optimal solution after 120 generations. The total run-time of

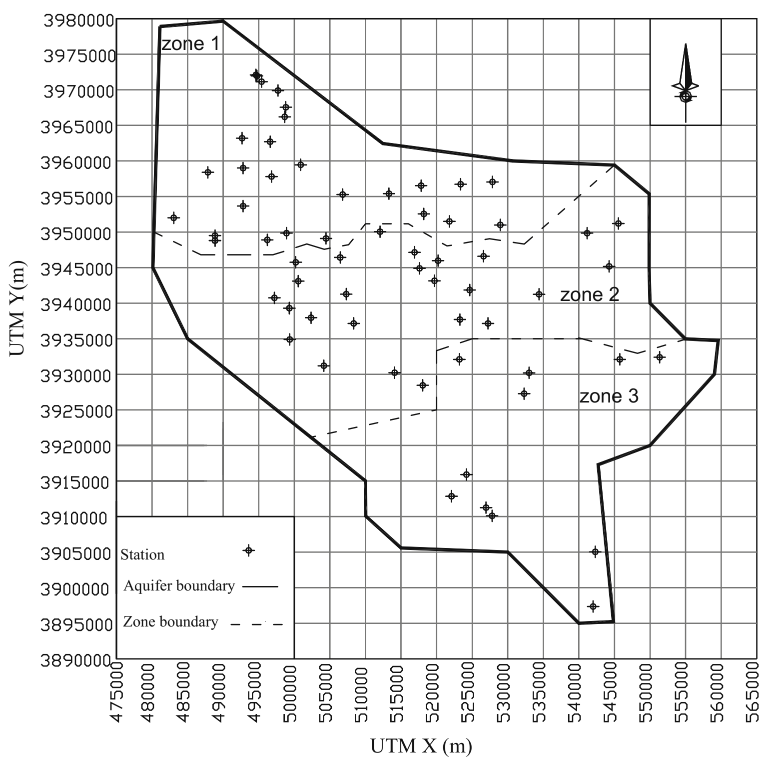


Fig. 4 Clustering of monitoring wells in the Tehran Aquifer using the C-means method

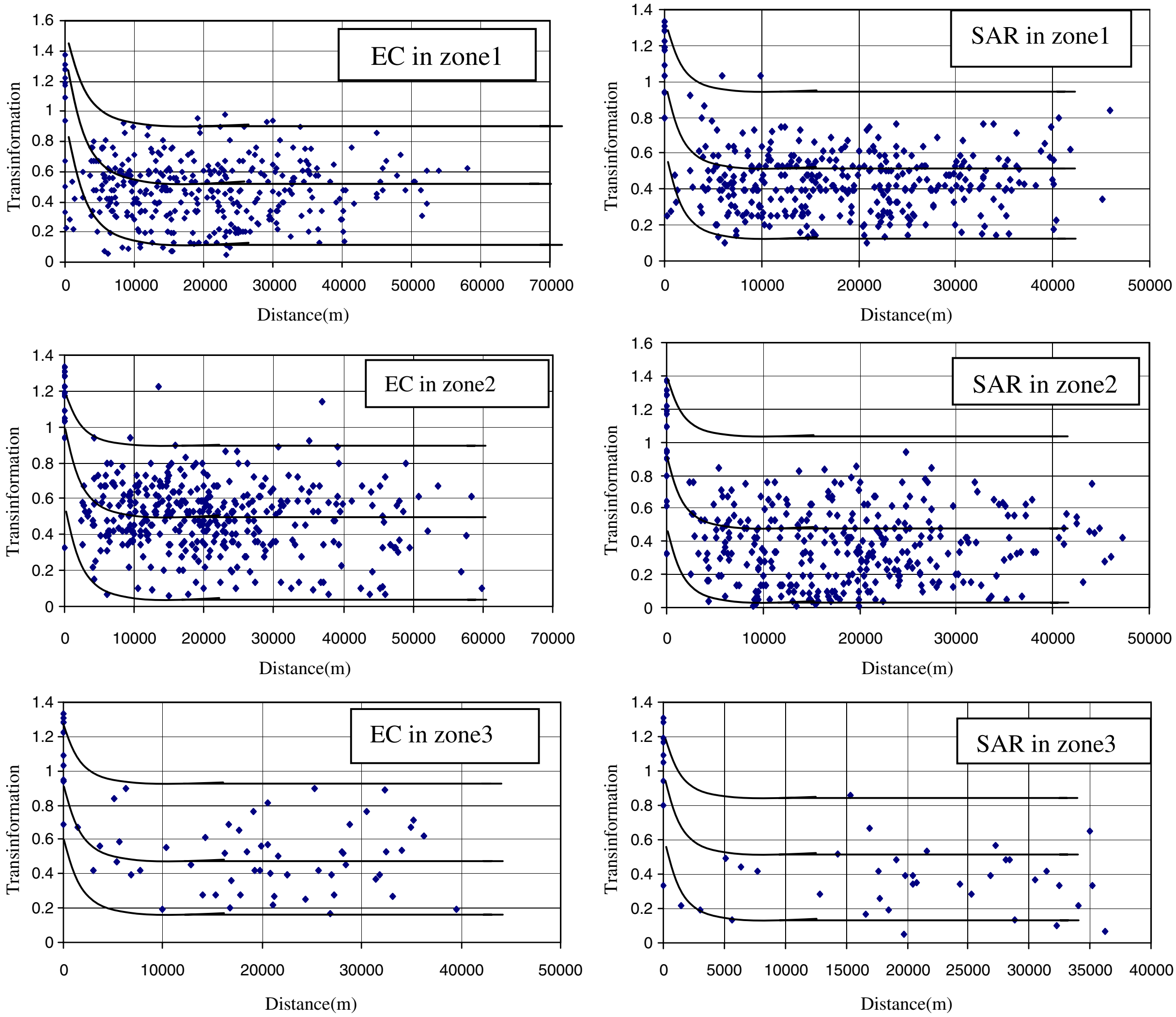


Fig. 5 Variation of transinformation versus distance in different zones in Tehran aquifer

the optimization model is less than 120 min in the MATLAB environment using a Pentium 4, 3.2 GHz computer. Figure 6 shows the optimal location of the monitoring wells when the total number of monitoring wells is equal to 62 and the clustering method is utilized. Figure 7 also presents the location of the monitoring wells when the clustering method is not used. Comparing

Table 2 The equations of T–D curves for EC and SAR variables in Tehran Aquifer (d is distance between two monitoring wells (meter))

Water quality variable	Curve	Curve equation	$d^*(m)$
EC	Upper	$T = 0.5 \times (\exp(-0.001 \times d)) + 0.79$	1,000
	Lower	$T = 0.55 \times (\exp(-0.002 \times d)) + 0.05$	
	Middle	$T = 0.92 \times (\exp(-0.0015 \times d)) + 0.4$	
SAR	Upper	$T = 0.7 \times (\exp(-0.001 \times d)) + 0.65$	8,000
	Lower	$T = 0.7 \times (\exp(-0.002 \times d)) + 0.05$	
	Middle	$T = 1.05 \times (\exp(-0.0015 \times d)) + 0.25$	

Table 3 The equations of T–D curves for different zones in Tehran Aquifer (d is distance between two monitoring wells (meter))

Zone	Water quality variable	Curve	Curve equation	$d^*(m)$
1	EC	Upper	$T = 0.6 \times (\exp(-0.001 \times d)) + 0.8$	10,000
		Lower	$T = 0.9 \times (\exp(-0.002 \times d)) + 0.1$	
		Middle	$T = 0.8 \times (\exp(-0.0015 \times d)) + 0.4$	
	SAR	Upper	$T = 0.55 \times (\exp(-0.001 \times d)) + 0.8$	6,500
		Lower	$T = 0.6 \times (\exp(-0.002 \times d)) + 0.12$	
		Middle	$T = 0.8 \times (\exp(-0.0015 \times d)) + 0.4$	
2	EC	Upper	$T = 0.5 \times (\exp(-0.001 \times d)) + 0.9$	7,000
		Lower	$T = 0.75 \times (\exp(-0.002 \times d)) + 0.03$	
		Middle	$T = 0.98 \times (\exp(-0.0015 \times d)) + 0.4$	
	SAR	Upper	$T = 0.38 \times (\exp(-0.001 \times d)) + 0.9$	6,000
		Lower	$T = 0.62 \times (\exp(-0.002 \times d)) + 0.03$	
		Middle	$T = 0.88 \times (\exp(-0.0015 \times d)) + 0.42$	
3	EC	Upper	$T = 0.53 \times (\exp(-0.001 \times d)) + 0.95$	8,000
		Lower	$T = 0.67 \times (\exp(-0.002 \times d)) + 0.17$	
		Middle	$T = 0.15 \times (\exp(-0.0015 \times d)) + 0.40$	
	SAR	Upper	$T = 0.5 \times (\exp(-0.001 \times d)) + 0.9$	10,000
		Lower	$T = 0.7 \times (\exp(-0.002 \times d)) + 0.18$	
		Middle	$T = 1.03 \times (\exp(-0.0015 \times d)) + 0.27$	

Figs. 4 and 6 shows that 30 new monitoring wells should be added to the monitoring system, and 30 existing monitoring wells should be omitted.

As mentioned before, entropy theory can also be used for updating sampling frequencies. The existing sampling interval in Tehran Aquifer is 6 months. The T–T curves are developed for water quality variables in each zone (Fig. 8). As shown in Fig. 8, by increasing the sampling interval from 6 months to 1 year, the transinformation is significantly reduced. However, this increase

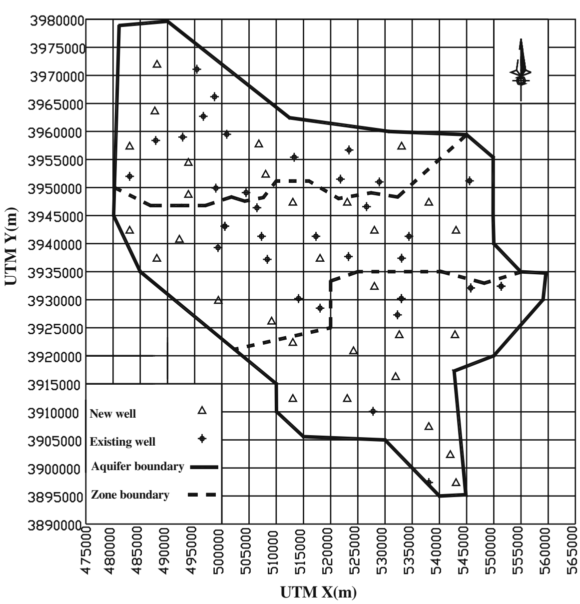


Fig. 6 The location of the selected existing and potential monitoring wells based on the result of the optimization model with zoning

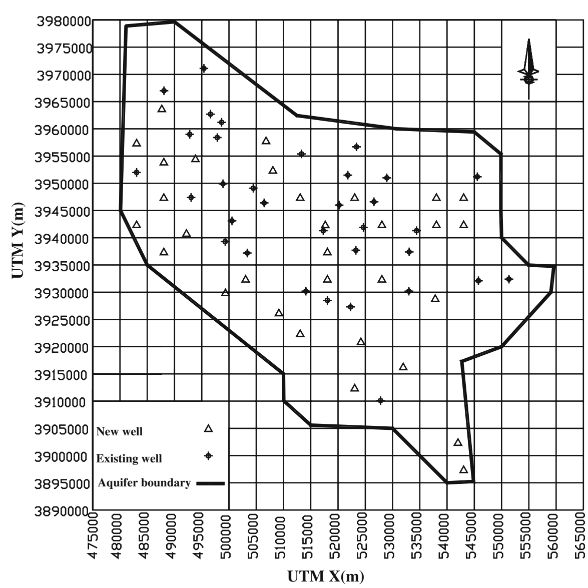
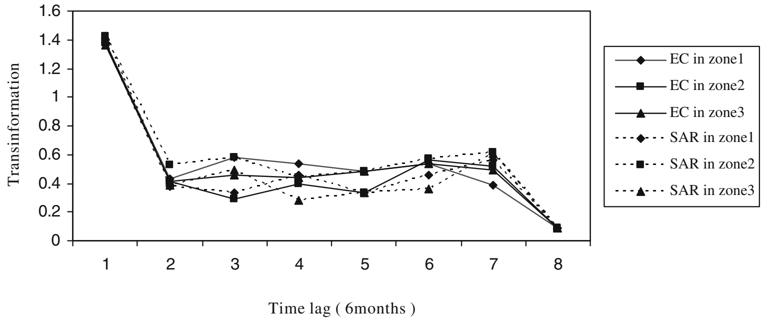


Fig. 7 The optimal location of the monitoring wells based on the result of the optimization model without zoning

Fig. 8 The transinformation–time lags (T–T) curves for water quality indicators in different zones



in sampling interval is not recommended due to significant seasonal variation of the concentration of water quality variables in the study area.

Summary and conclusions

Entropy method is basically a network assessment procedure that focuses on the variability of water quality in time and space. Its basic advantages are that it (1) provides a quantitative measure of the information content of a sampling site and of an observed time series; (2) assesses, again in quantitative terms, transfer of information in space; and (3) can be used to assess jointly several features of a network.

In this paper, a new optimization model was developed for optimal redesign of groundwater quality monitoring systems. The measure of transinformation in discrete entropy theory was used to find the optimal distance between the monitoring wells. The existing uncertainty in the transinformation–distance (T–D) curve was also incorporated using the fuzzy set theory. The fuzzy T–D curve was then used in a HGA-based optimization model. To improve the accuracy of the T–D curves, the C-means clustering method was also used to cluster the study area to some homogenous zones, and different T–D curves were calculated for different zones. The sampling frequency of water quality variables can also be updated using transinformation–time lags (T–T) curves.

The proposed methodology was applied to groundwater resources in the southern part of Tehran, Iran. The results show the efficiency of the model for the optimal redesign of the groundwater monitoring systems. In future studies, this methodology can be extended so that the sampling sites and sampling frequencies are jointly optimized using an integrated optimization model. The proposed approach can also be used for designing an integrated groundwater monitoring network that covers all objectives related to groundwater quality and quantity.

References

Bazargan-Lari, M. R., Kerachian, R., & Mansoori, A. (2009). A conflict-resolution nodel for the conjunctive use of surface and groundwater resources that considers water-quality issues: A case study. *Environmental Management*, Springer. doi:10.1007/s00267-008-9191-6.

Gen, M., & Cheng, R. (2000). *Genetic algorithm and engineering optimization*, (512 pp.). New York: Wiley.

Harmancioglu, N. B., & Yevjevich, V. (1987). Transfer of hydrologic information among river points. *Journal of Hydrology* (*Amsterdam*), *91*, 103–118. doi:10.1016/0022-1694(87)90131-4.

Harmancioglu, N. B., Fistikoglu, O., Ozkul, S. D., Singh, V. P., & Alpaslan, N. (1999). *Water quality monitoring network design*. Boston: Kluwer.

Husain, T. (1989). Hydrologic uncertainty measure and network design. *Water Resources Bulletin, 25*, 527–534.

Karamouz, M., Kerachian, R., & Zahraie, B. (2004). Monthly water resources and irrigation planning: Case study of conjunctive use of surface and groundwater resources. *Journal of Irrigation and Drainage Engineering, 130*(5), 391–402. doi:10.1061/(ASCE)0733-9437(2004)130:5(391).

Karamouz, M., Tabari, M. M. R., & Kerachian, R. (2007). Application of genetic algorithms and artificial neural networks in conjunctive use of surface and groundwater resources. *Water International. IWRA*, *32*(1), 163–176.

Karamouz, M., Nokhandan A. K., Kerachian, R., & Maksimovic, C. (2009). Design of on-line river water quality monitoring systems using the entropy theory: A case study. *Environmental Monitoring and Assessment*, Springer. doi:10.1007/s10661-008-0418-z.

Masoumi, F., & Kerachian, R. (2008a). Assessment of the groundwater salinity monitoring network of the Tehran Region: Application of the discrete entropy theory. *Water Science and Technology. IWA, 58*(4), 765–771. doi:10.2166/wst.2008.674.

Masoumi, F., & Kerachian, R. (2008b). Optimal groundwater monitoring network design using the entropy theory. *Water and Wastewater, 65*, 2–12 (In Persian).

Mogheir, Y., & Singh, V. P. (2003). Specification of information needs for groundwater management planning in developing country. *Groundwater Hydrology, Balema Publisher, Tokyo, 2*, 3–20.

Mogheir, Y., Lima, J. L. M. P., & Singh, V. P. (2004). Characterizing the special variability of groundwater quality using the entropy theory. *Hydrological Processes, 18*, 2165–2179. doi:10.1002/hyp.1465.

Ozkul, S., Harmancioglu, N. B., & Singh, V. P. (2000). Entropy-based assessment of water quality monitoring networks. *Journal of Hydrologic Engineering, 5*(1), 90–100. doi:10.1061/(ASCE)1084-0699(2000)5:1(90).

Salark, N., & Sorman, A. U. (2006). Evaluation and selection of streamflow network stations using entropy methods. *Turkish Journal of Engineering and Environmental Sciences, 30*, 91–100.

Yang, Y., & Burn, D. (1994). An entropy approach to data collection network design. *Journal of Hydrology (Amsterdam), 157*, 307–324. doi:10.1016/0022-1694(94)90111-2.