



Redesigning and monitoring groundwater quality and quantity networks by using the entropy theory

Mohammad Nazeri Tahroudi · Abbas Khashei Siuki ·
Yousef Ramezani

Received: 15 February 2018 / Accepted: 5 March 2019 / Published online: 27 March 2019
© Springer Nature Switzerland AG 2019

Abstract This study aimed at redesigning and monitoring the groundwater network of Naqadeh plain in the southwest of Lake Urmia to examine the number and position of optimal wells for the salinity information transfer (EC) and survey of groundwater level at aquifer. In this regard, groundwater level data (35 wells) and electrical conductivity values (24 wells) were used during a 10-year period (2002–2012). In the first stage, simulation was conducted using the multivariate regression method and quantitative and qualitative values and the interaction of wells was observed. In the next stage, number of different classes was considered for clustering quantitative and qualitative values. The results of studying different classes of data clustering showed that the 12-class cluster had more accurate results based on the root mean square error and coefficient of determination. The root mean square error was improved by about 40, 21, and 15%, respectively, compared to the 3, 5, and 9-class clusters. Finally, by choosing proper cluster of data, entropy indicators were investigated for quantitative and qualitative values at the aquifer level. The results of entropy indices at the aquifer showed that there was a severe shortage of information in terms of salinity in the Northwest of the aquifer, which necessitates drilling a new well in this area to accurately monitor the EC values. However, since more than 90% of the

basin area is in surplus and approximately surplus conditions in terms of transferring information, the studied area has a good dispersion for qualitative monitoring. Information transfer index for the quantitative groundwater network monitoring showed that piezometers near Lake Urmia were faced with a lack of information, which according to piezometers ranking, is ranked last in terms of value of maintaining or keeping the network. Eastern areas of aquifer are also faced with shortage of piezometers accounting for about 3% of the total area. The results of survey of surplus wells in the aquifer showed that nine and six surplus wells are in the aquifer for the qualitative and quantitative network, respectively. There were also wells in which information transfer was not well done and their information could not be assured. Finally, based on the conditions, a new arrangement of wells and a new optimal network were proposed.

Keywords Information transfer · Quality monitoring · Entropy theory · Groundwater network · Lake Urmia

Introduction

Designing groundwater quality monitoring systems has always been considered as one of the most complex issues in the field of water resources and environment. Since several issues such as spatial and temporal sequences of samples as well as choosing quality standard variables must be considered. These systems are designed to gather quality and quantity information, while

M. Nazeri Tahroudi · A. Khashei Siuki (✉) · Y. Ramezani
Department of Water Engineering, Faculty of Agriculture,
University of Birjand, Birjand, Iran
e-mail: abbaskhashei@birjand.ac.ir
e-mail: abbaskhashei@yahoo.com

designing them requires this basic information. Therefore, designing these systems is conducted in an iterative process (Van Luin and Ottens 1997). Given the importance of information and communication, evaluating monitoring networks and designing an optimum network have always been in the spotlight. For this reason, evaluating the existing network in order to design an optimum network seems essential and first, the initial condition should be considered, which means the demand for an optimum network should be investigated as the first step and then the optimized network must be provided. For example, if the numbers of wells in an area are few and are far apart, it is not necessary to provide sophisticated methods for an optimized network. When the groundwater is polluted, in many cases, eliminating the pollution lasts decades or longer and water takes back its original quality late. The reason for this event is the long retention time due to the slow movement of water through ground and the low degradation rate of pollutants. The main objective is achieving a good status for all groundwater bodies and ensuring that these resources will not be destroyed in future. A good status is the status when the groundwater combination is in a condition in which first, the pollutants, salinity or other interferences (as the change in the electric conductivity) are not observed in the groundwater zone. Second, it does not exceed the applicable quality standards for that water body. Third, no major reduction in the ecological or chemical quality of such water bodies are seen or no major damage is exposed to terrestrial ecosystems which all depend on the groundwater zone directly (Quevauviller 2009). Adequate and sufficient information is essential for planning and managing groundwater aquifers. Monitoring is in a close bond with the management of groundwater, whereat the results of monitoring can change the management of aquifer. The data collected from a groundwater monitoring network may reflect a lack or redundancy of aquifer information. The entropy theory has been used in various fields of evaluating and designing monitoring system network. For example, in the water quality monitoring network, researches such as Wu and Zidek (1992) and Harmancioglu et al. (1998) and for rain monitoring stations some researches including Krstanovic and Singh (1992) can be pointed out. But the studies of Jaynes (1957) and Shannon (1948) presented a new area of research for using entropy in a wide range of science and technology. Shannon (1948) developed vast studies on the entropy theory in various

engineering fields including the assessment of economic time series and ecological issues and developed a lot of unknown concepts on this theory. Chapman (1986) explained entropy as a quantity for measuring the uncertainty of hydrological data and the performance of hydrological models. He also examined the effect of different clusters on the entropy indices, and concluded that the entropy shows varying amounts by changing the number of classes. Harmancioglu and Alpaslan (1992) used the entropy theory for designing a water quality monitoring network. They developed standards spatial, temporal, and combinations of spatial-temporal indices based on the entropy theory. The results of their research indicated the great capability of entropy in designing quality monitoring networks. Ozkul et al. (2000) presented an approach for assessing river water quality monitoring systems by using the continuity entropy theory. Mogheir and Singh (2003) and Mogheir et al. (2004) showed that among four different types of entropy (i.e., the marginal, joint, conditional, and mutual (transinformation) entropies), the transinformation entropy is the best and the most appropriate method for assessing groundwater quality monitoring systems. They also suggested a method for evaluating water quality monitoring systems by means of the contour maps of the boundary entropy. Şarлак and Şorman (2006) evaluated and selected the wells of a hydrometric network using the entropy theory. They examined the effect of normal, lognormal, and gamma distributions on the ranking results of the wells and concluded that the distribution considered for the discharge data in the continuity entropy theory is important and let to various amounts of well ranks. Chen et al. (2008) used Kriging method for interpolating the data of monthly precipitation of 13 rain wells which recorded 87 months of rainfall in Shimen area in Taiwan in a network of 7.5×7.5 km². The resulting number of networks in the region was 17 where a well was placed in the center of each network and the well was the basis for further studies. Based on the precipitation statistics, they used the information transfer entropy between wells to prioritize new wells. Finally, according to 95% of the transinformation entropy between wells, six wells were localized for the area. Chadalavada et al. (2011) also used the uncertainty entropy along with the aforementioned methods and optimized the network of their interest. Zhu et al. (2015) also examined groundwater quality using entropy in China. Keum et al. (2017) evaluated application of entropy methods

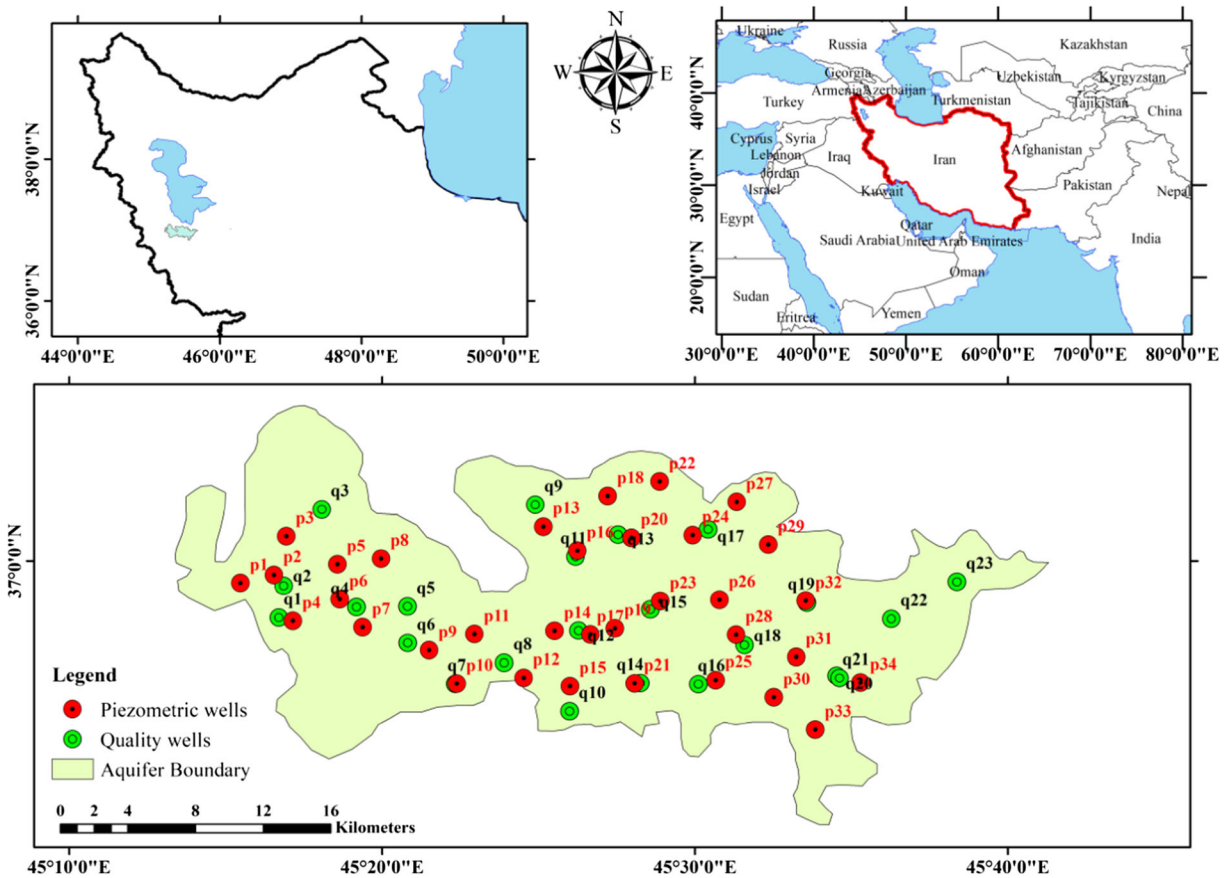


Fig. 1 Location of studied wells in Naqadeh plain

regarding precipitation, streamflow and water level, water quality, soil moisture and groundwater networks. They presented a summary of results of various researchers and argued that entropy theory is well suited for designing a monitoring network. However, more studies are required to provide design standards, and guidelines are also needed to implement them.

The quantitative and qualitative assessment and monitoring of groundwater is always an important challenge with specific problems. In fact, it is hard to understand what is happening under the earth surface. Water quality and quantity monitoring programs can ensure the quality and quantity of water resources for different uses. It is not possible to determine exact location of the surplus wells or deficiency of wells by entropy theory, but combining discrete and marginal entropy theories can make this possible. In the studies on entropy theory, the current network has been studied more and design and presentation of a new network has not been considered. Moreover, in

the present study, the quantitative and qualitative network of groundwater has been studied to provide a regular monitoring network of groundwater in the region. The studied area has undergone quantitative and qualitative changes in groundwater due to the frequent droughts in the Lake Urmia basin. By examining the results of this study, we can decide on the current monitoring network and determine the valuable and value less wells in the region. In fact, rating of wells regarding information transfer can be useful in making decision on optimizing monitoring of groundwater network by reducing or increasing the wells. Accordingly, this study aimed at providing a quantitative and qualitative monitoring of the groundwater network to investigate number of redundant and required wells in Nagadeh plain aquifer to monitor the salinity and groundwater level. Moreover, in this study, using the entropy theory, an appropriate monitoring network in the region has been proposed for salinity and groundwater level.

Table 1 Statistical properties of the studied data in the case study

Variable	Minimum	Maximum	Mean	Std. deviation	Variable	Minimum	Maximum	Mean	Std. deviation
q1	397.22	647.04	481.35	73.86	p6	1.92	3.49	2.81	0.58
q2	412.92	926.09	641.12	148.61	p7	2.64	3.42	3.12	0.28
q3	571.97	798.65	627.93	67.53	p8	3.07	4.29	3.77	0.39
q4	342.73	655.26	487.19	90.52	p9	2.78	3.65	3.23	0.26
q5	488.89	1365.82	649.64	261.49	p10	2.60	8.30	4.42	1.89
q6	385.24	606.66	477.96	77.75	p11	2.90	7.64	5.18	1.76
q7	493.67	954.56	664.20	146.44	p12	4.33	5.06	4.60	0.20
q8	425.63	694.34	532.94	87.80	p13	1.87	3.32	2.62	0.51
q9	1861.30	20,967.30	5239.87	5613.93	p14	1.68	2.33	1.89	0.20
q10	253.67	678.43	507.84	124.56	p15	3.11	4.12	3.58	0.26
q11	1007.56	2057.45	1742.83	300.35	p16	2.57	5.90	4.22	1.20
q12	455.50	1281.21	875.63	234.88	p17	1.72	2.91	2.24	0.38
q13	520.74	1265.32	976.32	245.78	p18	1.20	2.46	1.60	0.40
q14	148.54	2227.27	569.28	617.99	p19	1.98	3.14	2.52	0.30
q15	479.00	984.15	680.18	173.85	p20	2.83	4.73	3.77	0.53
q16	657.27	2224.21	1469.55	483.58	p21	2.31	3.68	2.78	0.43
q17	690.30	1245.99	1014.88	176.06	p22	1.07	2.86	1.73	0.65
q18	384.72	1391.63	900.41	269.26	p23	1.87	2.92	2.36	0.32
q19	445.71	2291.57	965.95	572.90	p24	2.62	4.35	3.37	0.53
q20	580.96	1615.85	1150.46	359.18	p25	4.95	8.72	6.68	1.09
q21	512.64	2387.36	965.30	580.75	p26	3.27	4.89	4.08	0.58
q22	826.98	3828.74	1987.83	829.31	p27	2.70	4.59	3.41	0.56
q23	2967.69	5099.07	4188.87	700.46	p28	4.41	5.20	4.81	0.23
q24	1381.80	2895.06	2143.22	642.18	p29	2.48	4.58	3.35	0.77
p1	2.27	3.60	2.85	0.45	p30	4.07	6.99	5.13	0.99
p2	2.79	3.50	3.06	0.20	p31	3.04	9.17	5.47	2.26
p3	9.09	13.31	10.39	1.38	p32	2.60	4.02	3.16	0.41
p4	16.75	18.15	17.44	0.47	p33	1.61	2.78	2.18	0.36
p5	2.09	4.07	2.60	0.56	p34	4.10	11.03	6.78	2.68

q quality well, p piezometric well

Materials and methods

Study area

Iran, with an area of over 16,480,000 km², is situated in the northern hemisphere and southwest of Asia. Almost all parts of Iran have four seasons. In general, a year can be divided into two warm and cold seasons. Iran, with an average annual precipitation of 62.1–344.8 mm, is located between two meridians of eastern 44° and 64° and two orbits of northern 40° and 25°. Approximately 94.8% of the country is either arid or semi-arid with low atmospheric precipitation and high Evapotranspiration (Khalili et al. 2016). In this paper, the study area is

Naqadeh plain located in the northwest of Iran, in the north of Azerbaijan Province, and in the southwest of Lake Urmia. The area of it is about 481 km² and its

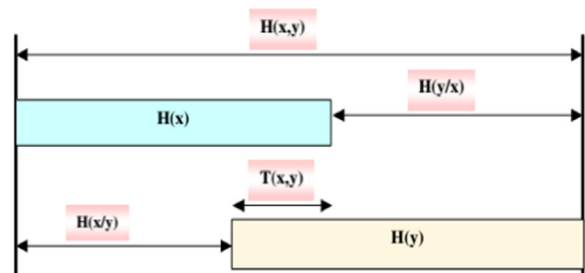


Fig. 2 Schematic plan of the coefficients of entropy theory (Jessop 1995)

Table 2 Classification values of the information transfer index (ITI)

Regional importance	Index value
Severe shortage	0–0.2
Shortage	0.2–0.4
Average	0.4–0.6
Approximately surplus	0.6–0.8
Surplus	> 0.8

perimeter is 172 km. In this study, the values of EC measured in 24 wells are employed to monitor the water quality and water level and 35 wells are utilized to monitor the quantity of groundwater in the statistical years between 2002 and 2012. The study area is shown in Fig. 1 and the statistical properties of the data under groundwater quality and quantity monitoring analysis are presented in Table 1.

Geological situation of the study area

Morphologically, across the region, there are ridges with an average height of 1500 to 2500 m above sea level. The permeability of formations in the region with the same hydrological conditions can be divided into three groups:

A: Permeable group

Includes Permian limestone, Cretaceous, and Miocene, which are featured in the eastern and southeastern

regions of the study. Due to gaps, dissolution cavities, and other karstic agents, they have high permeability and play an important role in aquifer feeding.

B: Low permeation group

It includes Cambrian, Permian, and Miocene rocks. These formations are visible in the south-east of the study area. Due to the presence of alluvial sediments containing clay with silt and clay with fine sand, they have little permeability.

C: Impermeable group

These sediments include Cambrian, Cretaceous, and Miocene marl, and have been observed in the north-east and south of the study area. Miocene marls have salty minerals, are outside the study area, and cause groundwater contamination.

Entropy theory

Basically, entropy means disorder, the higher the randomness (disorder) in a system, the more entropy. According to the definition of entropy provided by Shannon, for two discrete variables of x and y , where $x_i, i = 1, 2, 3, \dots, n$ and $y_j, j = 1, 2, 3, \dots, m$, and are in the same likelihood space, each one has a discrete probability of $p(x_i)$, a joint probability of $p(x_i, y_j)$, and a conditional probability of $p(x_i | y_j)$ for x_i provided that y_j occurs.

Table 3 Results of evaluating the accuracy and error values of multivariate regression in estimating the data of well

Variable	RMSE	N-S	MSE	R	Variable	RMSE	N-S	MSE	r
q1	2.47	0.9988	6.12	1.0000	q13	3.24	0.9998	10.49	1.0000
q2	2.22	0.9998	4.95	1.0000	q14	31.26	0.9974	977.10	0.9988
q3	3.70	0.9965	13.73	1.0000	q15	9.33	0.9970	87.02	0.9990
q4	4.96	0.9968	24.64	1.0000	q16	3.30	1.0000	10.88	1.0000
q5	34.80	0.9799	1211.24	0.9998	q17	30.99	0.9625	960.14	0.9883
q6	1.03	0.9998	1.05	1.0000	q18	1.20	1.0000	1.44	1.0000
q7	0.34	1.0000	0.11	1.0000	q19	10.15	0.9997	103.09	1.0000
q8	2.23	0.9993	4.95	1.0000	q20	24.52	0.9948	601.08	0.9976
q9	48.83	0.9999	2384.59	1.0000	q21	6.59	0.9999	43.48	1.0000
q10	0.45	1.0000	0.20	1.0000	q22	23.54	0.9991	554.28	1.0000
q11	15.54	0.9971	241.60	0.9988	q23	35.75	0.9968	1278.31	0.9985
q12	88.87	0.8425	7897.77	0.9908	q24	21.86	0.9987	477.91	1.0000

Marginal entropy

$$E(I(x)) = H(x) = - \sum_{i=1}^{\infty} p(x_i) \ln p(x_i) \tag{1}$$

where $E(I(x))$ indicates the mathematical expectation of data. In fact, by definition, the average of data (the average of $I(x)$) is used to measure the uncertainty. Moreover, in some books $H(x)$ is mentioned as a measurement of the data uncertainty. However, it can be true since the uncertainty is an indicative of the potential information, therefore, for a random variable like x , the marginal entropy, $H(x)$, can be defined as a potential information of the variable.

Joint entropy: indicates the data that exit in both x and y .

$$H(x,y) = - \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p(x_i, y_j) \ln p(x_i, y_j) \tag{2}$$

Conditional entropy: for random variables of x and y , the conditional entropy represents the data from x which does not exist in y (Mogheir and Singh 2003).

$$H(x,y) = - \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p(x_i|y_j) \ln p(x_i|y_j) \tag{3}$$

Mutual (transinformation) information entropy: is interpreted as a reduction in the uncertainty of x respect to the random variable y . It can also be defined as the information from x that exists in y (Lubbe 1996).

$$T(x,y) = - \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p(x_i, y_j) \ln \left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right) \tag{4}$$

Where $p(x)$ is the occurrence probability of x , $p(x,y)$ is the joint probability of x and y , and $p(x|y)$ is the probability of x provided that y occurs. Note that $T(x,y) = T(y,x)$ (Jessop 1995). The information transfer entropy can also be calculated in the following equivalent ways:

$$T(x,y) = H(x) - H(x|y) \tag{5}$$

$$T(x,y) = H(x) + H(y) - H(x,y) \tag{6}$$

$$T(y,x) = H(y) - H(y|x) \tag{7}$$

$$T(y,x) = H(y) + H(x) - H(x,y) \tag{8}$$

For understanding the relationship between information transfer and the entropy coefficients better, refer to Fig. 2.

Information transfer can also be expressed by using a normal indicator transfer index which is introduced by ITI, which indicates the standard information which is transferred from one place to another. Clustering the information transfer is presented in Table 2

$$ITI = \frac{T(x,y)}{H(x,y)} \tag{9}$$

Three indices of $R(i)$, $S(i)$, and $N(i)$ can be introduced as follows and as an entropy fractional converter of x by $R(x,y)$ symbol, so that if y is known, it is also a reduction of uncertainty of x , and in fact the information received by x is of y as well, which is defined as the discrete entropy:

$$R(x,y) = \frac{T(x,y)}{H(x)} \tag{10}$$

Which can be used as a reduction of the uncertainty of x if y is known, in other words, it is the amount of information received by well x from well y . The information that is sent from x to y is defined by relationship 11:

$$S(x,y) = \frac{T(x,y)}{H(y)} \tag{11}$$

Aforementioned equations express the relationship between x and y variables. By using Eqs. 10 and 11, the same reasoning can be applied to each of the wells. Moreover, the received and transferred information of the i th station is defined as follows:

$$R(i) = R(x(i), \hat{x}(i)) \tag{12}$$

$x(i)$ represents the data of the i th well and $\hat{x}(i)$ is based on the following linear relationship:

$$\hat{x}(i) = a(i) + \sum_{j=1}^{I-1} y_j(i) \times b_j(i) \tag{13}$$

$y(i)$ is the data matrix of all the other stations and $a(i)$ and $b(i)$ are the regression parameters between the i th station and all the other stations that fit a linear equation. Larger values of $R(i)$ and $S(i)$ respectively mean receiving and sending more information between the i th station and the other stations in the network. In other words, it means establishing a better communication between the base station and the other ones. Thus, the greater amounts of $R(i)$ and $S(i)$ for a station mean higher values of the station, and maintaining and preserving them is

Table 4 Results of investigating the correlation between the quantity data of aquifer

Correl	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	p12	p13	p14	p15	p16	p17	p18
p1	1.00																	
p2	0.81	1.00																
p3	0.62	0.30	1.00															
p4	0.91	0.75	0.67	1.00														
p5	0.84	0.84	0.13	0.65	1.00													
p6	0.79	0.67	0.39	0.89	0.68	1.00												
p7	0.84	0.69	0.60	0.87	0.66	0.86	1.00											
p8	0.89	0.66	0.52	0.89	0.76	0.93	0.86	1.00										
p9	0.50	0.70	-0.04	0.61	0.63	0.82	0.56	0.68	1.00									
p10	0.88	0.56	0.89	0.81	0.53	0.54	0.71	0.71	0.12	1.00								
p11	0.63	0.81	0.53	0.64	0.50	0.46	0.61	0.47	0.49	0.59	1.00							
p12	0.77	0.94	0.16	0.70	0.86	0.62	0.57	0.63	0.70	0.48	0.72	1.00						
p13	0.82	0.86	0.62	0.83	0.67	0.68	0.78	0.72	0.60	0.75	0.94	0.79	1.00					
p14	0.37	0.70	-0.19	0.33	0.62	0.30	0.19	0.20	0.55	0.11	0.57	0.80	0.55	1.00				
p15	0.72	0.91	0.13	0.72	0.82	0.76	0.62	0.72	0.89	0.38	0.69	0.93	0.78	0.70	1.00			
p16	0.76	0.89	0.56	0.74	0.63	0.56	0.67	0.58	0.53	0.68	0.96	0.83	0.96	0.62	0.78	1.00		
p17	0.25	0.64	-0.21	0.25	0.48	0.39	0.29	0.25	0.73	-0.09	0.57	0.67	0.51	0.72	0.76	0.60	1.00	
p18	0.54	0.44	-0.12	-0.03	0.08	-0.16	-0.10	-0.26	0.21	-0.12	0.62	0.47	0.38	0.65	0.40	0.57	0.70	1.00
p19	0.27	0.25	-0.48	-0.19	0.00	0.01	-0.09	-0.20	0.50	-0.54	0.30	0.25	0.11	0.45	0.39	0.23	0.83	0.73
p20	0.39	0.52	-0.28	0.47	0.62	0.69	0.42	0.54	0.85	-0.01	0.25	0.64	0.40	0.68	0.74	0.33	0.65	0.12
p21	0.15	0.60	-0.11	0.19	0.33	0.29	0.29	0.13	0.64	-0.09	0.71	0.55	0.57	0.66	0.64	0.64	0.91	0.78
p22	0.82	0.79	0.66	0.81	0.62	0.57	0.67	0.63	0.41	0.82	0.87	0.77	0.94	0.60	0.66	0.92	0.36	0.36
p23	0.76	0.84	0.24	0.84	0.77	0.93	0.75	0.84	0.92	0.45	0.61	0.81	0.76	0.58	0.92	0.69	0.60	0.13
p24	0.33	0.73	0.03	0.34	0.43	0.37	0.46	0.28	0.62	0.06	0.69	0.62	0.58	0.42	0.70	0.66	0.76	0.65
p25	0.49	0.55	-0.19	0.21	0.25	0.29	0.15	0.08	0.67	-0.16	0.62	0.59	0.49	0.75	0.67	0.59	0.91	0.82
p26	0.79	0.83	0.69	0.84	0.54	0.59	0.70	0.59	0.43	0.79	0.90	0.75	0.93	0.53	0.67	0.95	0.40	0.43
p27	0.43	0.79	0.30	0.51	0.37	0.39	0.43	0.27	0.55	0.31	0.87	0.69	0.75	0.61	0.70	0.86	0.73	0.78
p28	0.06	0.48	-0.26	0.14	0.25	0.23	0.12	0.12	0.66	-0.24	0.51	0.56	0.39	0.54	0.67	0.49	0.87	0.74
p29	0.48	0.82	0.13	0.56	0.51	0.51	0.43	0.36	0.70	0.25	0.80	0.82	0.75	0.83	0.82	0.84	0.84	0.75
p30	0.55	0.84	0.09	0.59	0.65	0.58	0.49	0.47	0.77	0.29	0.80	0.87	0.80	0.90	0.87	0.83	0.82	0.64
p31	0.43	0.56	0.68	0.44	0.15	0.11	0.37	0.15	0.05	0.59	0.86	0.44	0.73	0.29	0.32	0.81	0.25	0.62
p32	0.61	0.44	-0.06	0.16	0.08	0.17	0.11	-0.03	0.50	-0.11	0.65	0.45	0.49	0.68	0.50	0.59	0.81	0.85
p33	0.56	0.87	0.15	0.60	0.62	0.64	0.59	0.53	0.83	0.26	0.81	0.82	0.79	0.68	0.91	0.83	0.89	0.62
p34	0.52	0.61	0.73	0.52	0.24	0.20	0.46	0.26	0.10	0.67	0.89	0.48	0.79	0.31	0.38	0.86	0.29	0.57
p35	0.57	0.45	0.87	0.54	0.19	0.16	0.45	0.31	-0.10	0.82	0.74	0.35	0.71	0.11	0.21	0.73	-0.04	0.30

Correl
p1
p2
p3

Table 5 Observation matrix of the 3-class cluster of q1 data

Estimation values of well1					
Classification		405–490	490–575	575–660	Sum
Observation values of well1	405–490	6	0	0	6
	490–575	1	2	0	3
	575–660	0	0	1	1
Sum		7	2	1	10

recommended, but the $N(i)$ index, which is called the net exchange information, is defined as follows:

$$N(i) = S(i) - R(i) \tag{14}$$

Since the value for each of the stations is measured by $N(i)$, this index is important. $N(i)$ represents the total net information of each well and the station with the lowest value of $N(i)$ has the lowest rank and importance in the monitoring network (Markus et al. 2003).

One of the factors contributing to the value of marginal entropy is the interval (size and number). When performing frequency analysis, it is clear that when the intervals are smaller, the number of classes (NCI) is more (Zhou 1996). However, there is no general rule for choosing NCI. In 1996, Zhou proposed using the following formula to avoid such problems when calculating the number of classes:

$$NCI = 1 + 1.33 \ln(n) \tag{15}$$

$$NCI < 5 \log n \tag{16}$$

In the above equation, n is the number of observations and NCI is the number of intervals (classes) in the time series of the variable of interest. However, in many cases, an arbitrary number of classes are chosen, in

which the classes are usually less than 5 and no more than 20 classes (the number of classes is usually between 5 and 20). If the number of classes is less than 5, the data may lose their information. On the other hand, increasing the number of classes over 20 will make the calculations longer and more time consuming. In any case, this method is quite prone to subjectivity and is an experimental approach.

T-model

To study the transmission of information at different distances, T-model is used as follows (Mogheir et al. 2006):

$$T(d) = (T_0 - T_{\text{Min}})e^{(-Kd)} + T_{\text{Min}} \tag{17}$$

where T_0 is the initial value of information transfer. K is the decay rate of information transfer. T_{min} is the minimum of information transfer and d is the distance between the wells. The percentage of the redundant wells can be calculated by relationship 18.

$$\%NRI = \frac{T(d)^{\text{NET}}}{T_0 - T_{\text{Min}}} \times 100 \tag{18}$$

where $T(d)^{\text{NET}}$ equals the network information transfer which is based on Eq. 19 (Mogheir et al. 2006):

Table 6 Statistical matrix of the 3-class cluster of q1 data

Estimation values of well1					
Classification		405–490	490–575	575–660	Sum
Observation values of well 1	405–490	0.6	0	0	0.6
	490–575	0.1	0.2	0	0.3
	575–660	0	0	0.1	0.1
Sum		0.7	0.2	0.1	1

Table 7 Results of evaluating the information transfer (T) in different classifications

Classifications	MSE	RMSE	R^2
Quality data			
3 class	7.02	2.65	0.06
5 class	4.13	2.03	0.50
9 class	3.57	7.89	0.41
12 class	2.57	1.60	0.61
quantity data			
3 class	0.79	0.89	0.19
5 class	0.35	0.59	0.26
9 class	0.90	0.95	0.09
12 class	0.11	0.34	0.58

$$T(d)^{NET} = T(d) - T_{Min} \tag{19}$$

The size of the optimized network monitoring can be calculated using Eq. 20 (Mogheir et al. 2006):

$$a = 0.354 L \tag{20}$$

where a is equal to the network side and L is the maximum distance between the wells.

Results and discussion

In this study, for monitoring the groundwater quality of Naqadeh plain in order to measure the salinity values in the area optimally, annual data of EC obtained from 24 wells in the aquifer of Naqadeh plain were used between 2002 and 2012. Also, the data of groundwater level in

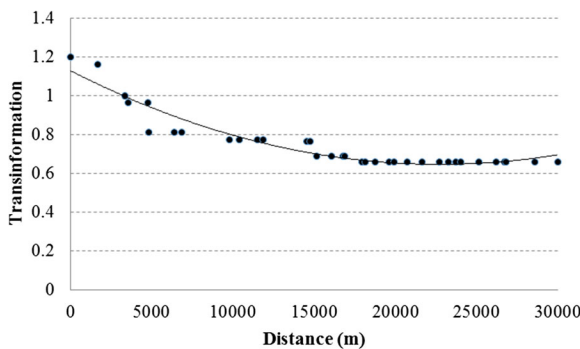


Fig. 3 Transinformation model for the water level variable in Naqadeh aquifer

35 piezometric wells were used to monitor the quantity groundwater network of the study area in the same period. The first step of the present study was applying the modified Mann-Kendall test to examine the trend changes of the data. The results indicated a significant trend in EC values of wells number q7, q5, q17, and q19 whereat using differentiating method, the trend in the data was removed. In most parametric tests, there are many basic assumptions, hence, if the assumptions are not met, the results of the test will not be valid.

Among these assumptions, the most important and common one is the normality assumption of data. Recent researches have shown that the probability distribution function of many quality and quantity variables of the water resources systems do not follow a normal distribution. The discrete Entropy is an approach to revise this shortcoming in applying entropy to water related issues (Mogheir and Singh 2003). In this respect, the values of electric conductivity and the water level in all wells are assessed.

After preparing the studied data using Eq. 13, new data were generated for each well by using the data from the other wells. In other words, rather than using data directly from a well, by creating a multivariate regression among the other wells, the predictions were made for the well of interest. Multivariate regression is an extension of the linear regression that creates the common linear models by taking more than one independent variable as well as one specific case into account. EC values of each well were estimated using multivariate regression and the values of the other 23 wells. The results of investigating the accuracy of multivariate regression method are presented in Table 3. The results of investigating the accuracy of the multiple regression showed that according to Nash-Sutcliffe (N-S) statistic, the performance of the model is acceptable for estimating the EC values of each well. By employing the map of elevations, the directions of hydraulic slope as well as the direction of hydraulic conductivity of the groundwater in the study area were investigated, where it was revealed that the direction of the groundwater flow is toward Lake Urmia. So p1 was assigned as the basic well for the quantity groundwater network monitoring of the area and the values of information transfer were calculated based on it. The results of the correlation between piezometric wells are presented in Table 4. Eventually, the data were analyzed and divided into different classes. For classifying the data, the categories of 3

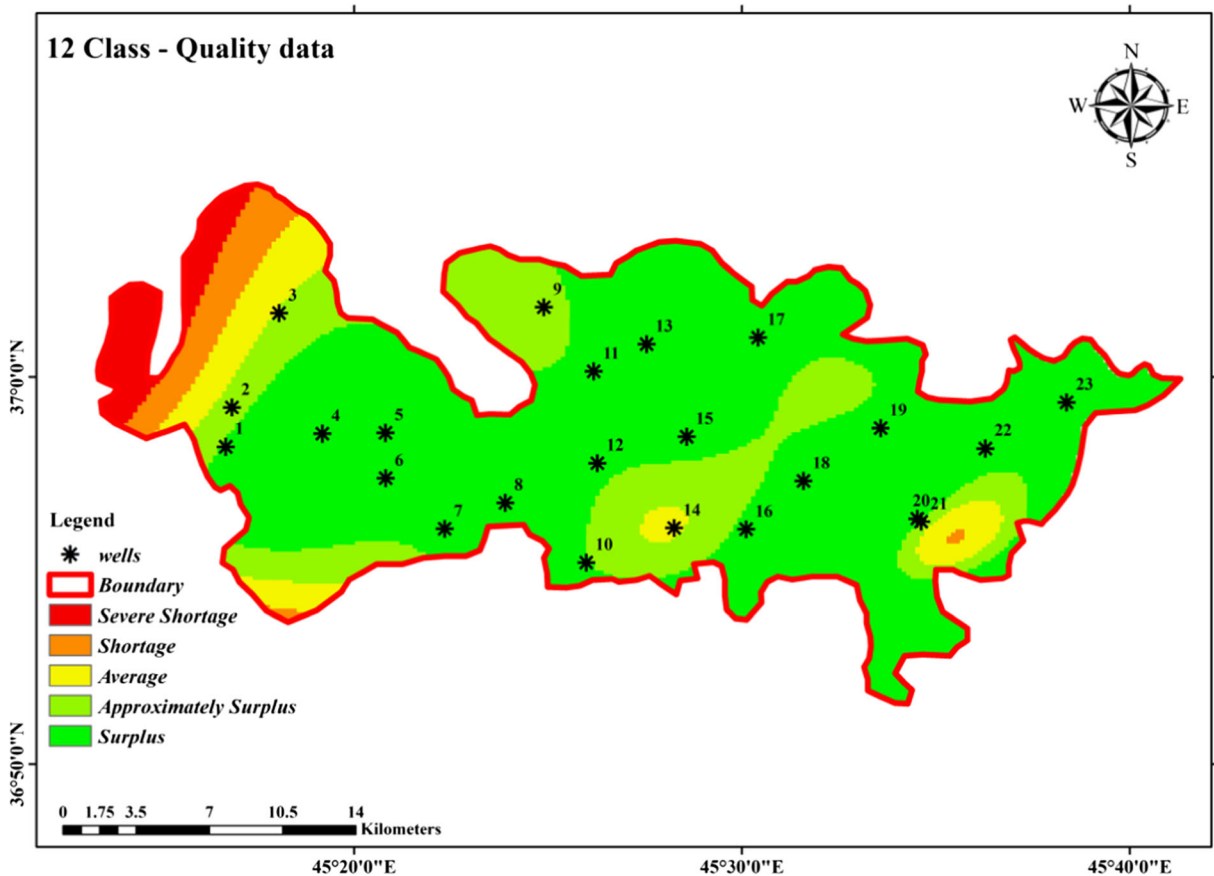


Fig. 4 Results of zoning the entropy value index in a 12-class cluster to monitor the groundwater quality network

classes, 5 classes, 9 classes, and 12 classes (using Eq. 16) were studied. In fact, the reason of selecting 9 and 12 classes is investigating the accuracy of the network monitoring. By specifying the number of investigation classes, the observation and likelihood matrices were produced. The results of calculating the possibility and observation matrices of the 3-class cluster related to well q1 are presented in Tables 5 and 6.

By calculating the frequency and probability matrices, the values of the transfer entropy (T) were calculated for all classes. The statistic of T-model was calculated for all of the classes using Eq. 17. The results of T-model showed that the 12-class cluster is more accurate than the other classes in estimating the data. In this method, first the relationship 17 was calculated for the values of EC (quality) and water table (quantity) of the present wells and then using them, the values of information transfer of each well (T) were estimated based on the distance between them. The results of investigating the errors raised from examining the classes are presented in

Table 7. Figure 3 also shows the information transfer based on the distances between piezometric wells in investigating the 12-class cluster of quantity values (water table).

The results of investigating the accuracy of clustering information transfer indicated that the 12-class cluster is of the lowest error rate and the highest correlation compared to the other classes. By selecting the 12-class cluster, the standard values of entropy were determined and the results obtained from the information transfer index (ITI) were estimated in monitoring the groundwater quantity and quality networks in the study area. The results of monitoring the quality and quantity networks of Naqadeh plain are presented in Figs. 4 and 5, respectively.

As Fig. 4 illustrates the results of zoning the information transmission index (ITI) at the regional level showed that in some parts of the southeast of basin and near q22 and q21 wells, the optimal status of the number of wells is medium and even a shortage of well

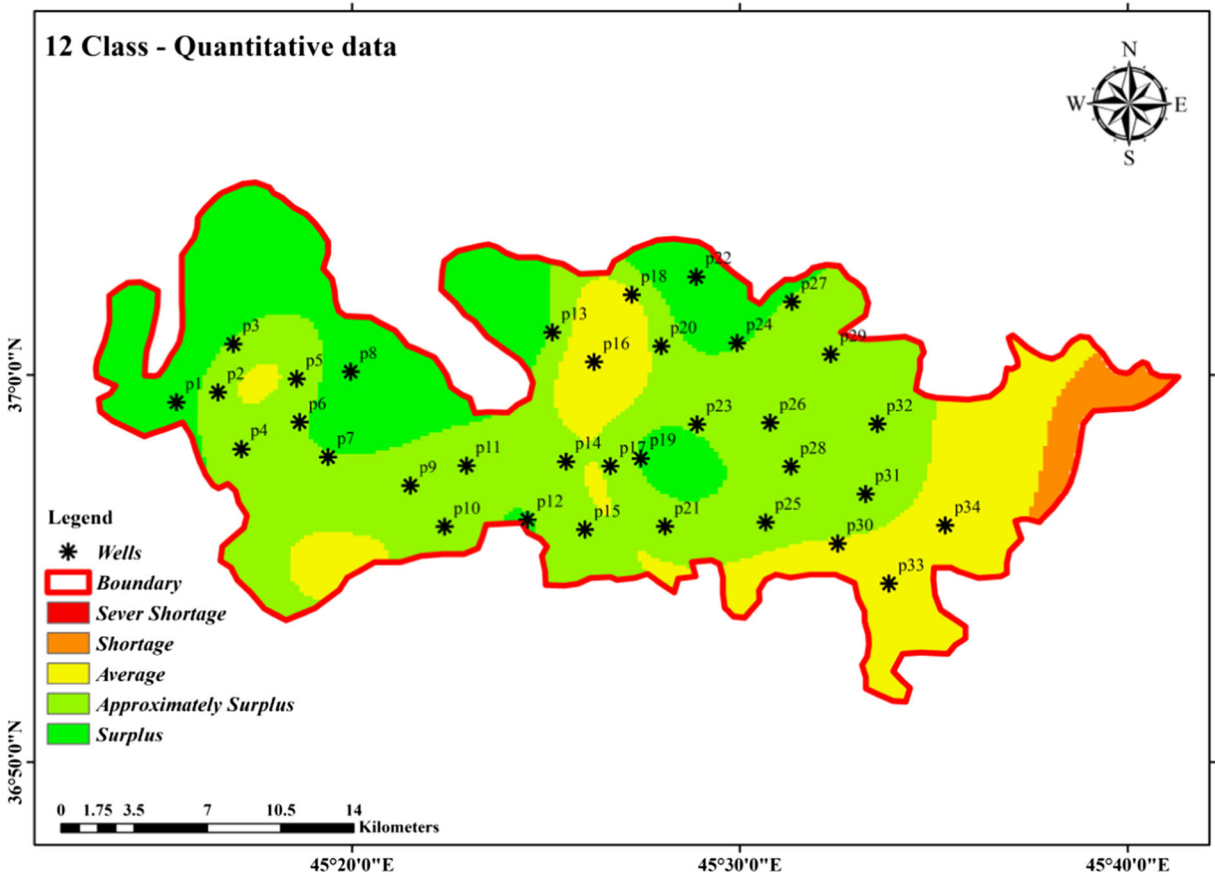


Fig. 5 Results of zoning the entropy values index for monitoring the groundwater quantity network

is suspected. This represents the poor information transfer in q22 and q21 wells and also shows that the relationship between the two wells is weak. But the shortage of wells in this area is not acute and there is no urgent need to dig new wells. In the northwest parts of basin,

there is also a severe shortage of well. Severe shortage of well in this area represents the insufficiency of the number of wells in the area for monitoring quality properties. In other words, in these areas, enough information on the salinity of groundwater (the areas of the

Table 8 Coverage of ITI index in groundwater quality monitoring

The regional importance	Quantitative/qualitative	Percent of surface	Area (km ²)
Severe shortage	EC	3.85	18.50
	Water level	0.00	0.00
Shortage	EC	3.54	17.00
	Water level	2.79	13.45
Average	EC	6.35	30.50
	Water level	22.23	106.9
Approximately surplus	EC	20.16	97.00
	Water level	48.62	233.85
Surplus	EC	66.10	318.00
	Water level	26.36	126.80

Table 9 Ranking the studied wells based on $N(i)$ index

Variable	Rank	Variable	Rank
q1	6	p6	17
q2	21	p7	11
q3	22	p8	3
q4	5	p9	2
q5	11	p10	34
q6	4	p11	27
q7	19	p12	26
q8	7	p13	6
q9	24	p14	10
q10	18	p15	25
q11	23	p16	24
q12	8	p17	29
q13	17	p18	12
q14	20	p19	32
q15	16	p20	5
q16	9	p21	23
q17	15	p22	28
q18	14	p23	7
q19	2	p24	9
q20	13	p25	4
q21	12	p26	22
q22	10	p27	21
q23	3	p28	8
q24	1	p29	15
p1	1	p30	14
p2	13	p31	31
p3	18	p32	20
p4	16	p33	33
p5	19	p34	30

northwest border of the aquifer) cannot be obtained with this scattering of wells. Most of the study area (66%) is in the redundant cluster regarding the number of wells. Considering the current scattering of wells, a relatively good quality monitoring of the groundwater salinity is possible. Based on NRI (number of redundant index), it was determined that about 38% of the present wells are redundant in the groundwater quality network of Naqadeh aquifer. The results of studying the groundwater quality monitoring showed that 9 wells are redundant in the aquifer under study, but the spatial scattering of wells is not uniform.

The results of surveying the values of information transfer index showed that there is no problem with the

groundwater quantity monitoring network in the study area, and by means of the piezometers in the area, an appropriate monitoring is possible on the groundwater quantity values of the aquifer. The information transfer is weak merely in the eastern part of the aquifer which does not cover a large area. If this issue is solved, the quantity monitoring of the groundwater aquifer will be conducted fully and by using its information, the quantity management of aquifer will be carried out better. Since digging and constructing new piezometers are costly, it is suggested to utilize the available deep and semi-deep wells in the network to monitor and assess the groundwater quality in eastern areas of the aquifer. Based on NRI, for the quantity data of groundwater, it was also found that about 17% of the wells are redundant in the groundwater quantity network of Naqadeh aquifer. On this basis, in the groundwater monitoring network of the study area, 6 wells are redundant.

The areas covered by each of the clusters of ITI are presented in Table 8. According to Figs. 4 and 5 and Table 8, it can be seen that more than 90% of the study area has a good quality and quantity monitoring on the groundwater resources considering the distribution of salinity and groundwater levels. It seems that there is a well shortage in less than 10% of the study area, where in terms of salinity and quantity variables much of the shortage is located in the northwest and east of the aquifer, respectively.

It is suggested to remove the wells with low ranking from the monitoring network in order to reduce or eliminate redundant wells. Because a well with a higher $N(i)$ is of higher value and this high rank indicates that the aquifer status is presented better by the well and the information of it can be easily attributed to the aquifer. That is why ranking wells based on the net information (N_i) is essential. The values of net information index (N_i) are calculated for the quality and quantity monitoring networks and are presented in Table 9.

The results of ranking the quality wells in the region showed that q19, q23, and q24 wells are holding a 1st to 3rd rank which means that these wells receive and transfer more information than the other ones. In other words, these wells have more preserving values than the other wells and the data of them can be collected and used with a complete confidence. These wells are located in the lower areas of the aquifer. On the other hand, q9, q3, and q11 wells obtained the lowest values among the other wells. q3 well is in the farthest distance from the exit of basin and is within the critical region of ITI. q19, q23, and

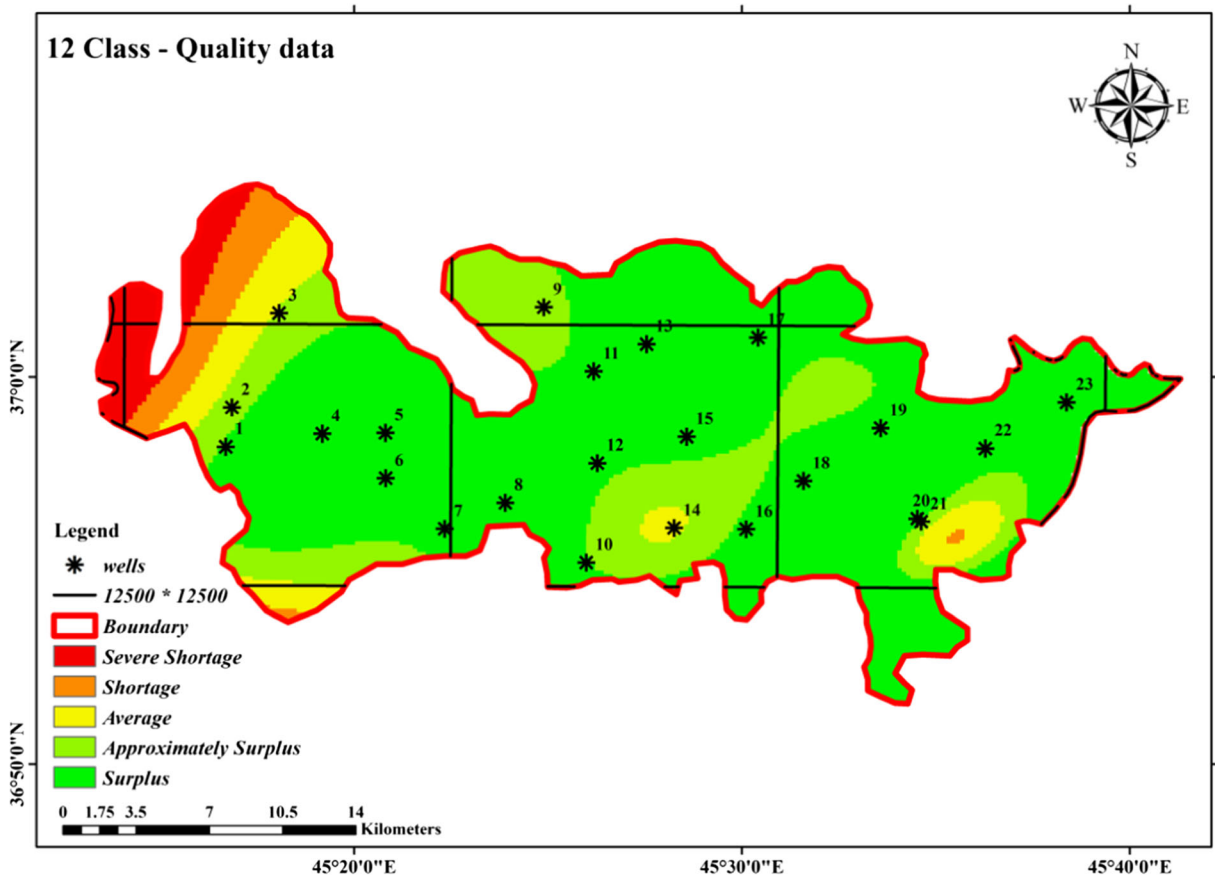


Fig. 6 Results of gridding the studied aquifer for monitoring the groundwater quantity network

q24 wells generate the best information and q9, q3, and q11 wells produce the least information about the groundwater quality monitoring network. In the similar way, p1, p8, and p9 piezometers showed the most and the best information and p10, p33, and p19 piezometers demonstrated the least information on the groundwater quality network of aquifer. By means of the 12-class cluster and Eqs. 17 to 19, the numbers of redundant wells were calculated in the region. Considering the distance between the wells, the number of the redundant quality wells in the aquifer equaled 9 and the number of the redundant piezometric wells in the network was 6. Due to the irregular distribution of the wells within the aquifer and also because of finding places with decreasing or increasing information, the gridding of the study area was employed. Using Eq. 20, the optimal size of network was selected for gridding the study area in order to monitor the quantity and quality values. According to the maximum distance between the wells, the results showed that a network with a size of $12,500 \times 12,500$ m

and a network with a size of $10,000 \times 10,000$ m are suitable for the quality and the quantity values of the study area, respectively. The results of gridding the study area are presented in Figs. 6 and 7.

Taking into account the redundancy of 9 wells in the study area along with considering the area of the grids, it seems that 15 wells could reflect a good quality monitoring of the study area. According to the established networks, 5 wells per network, which are distributed regularly, can provide enough information on the salinity of aquifer. If the goal of monitoring is optimizing and improving the groundwater quality network, it is possible either to dig new wells based on the results or to remove the wells with less information transfer from the system. Besides, given the redundancy of six piezometers in the quantity groundwater monitoring network of the aquifer, almost seven piezometers in each of the $10,000 \times 10,000$ m cells can provide the best quantity monitoring in the aquifer. For networks with redundant wells, the wells that are

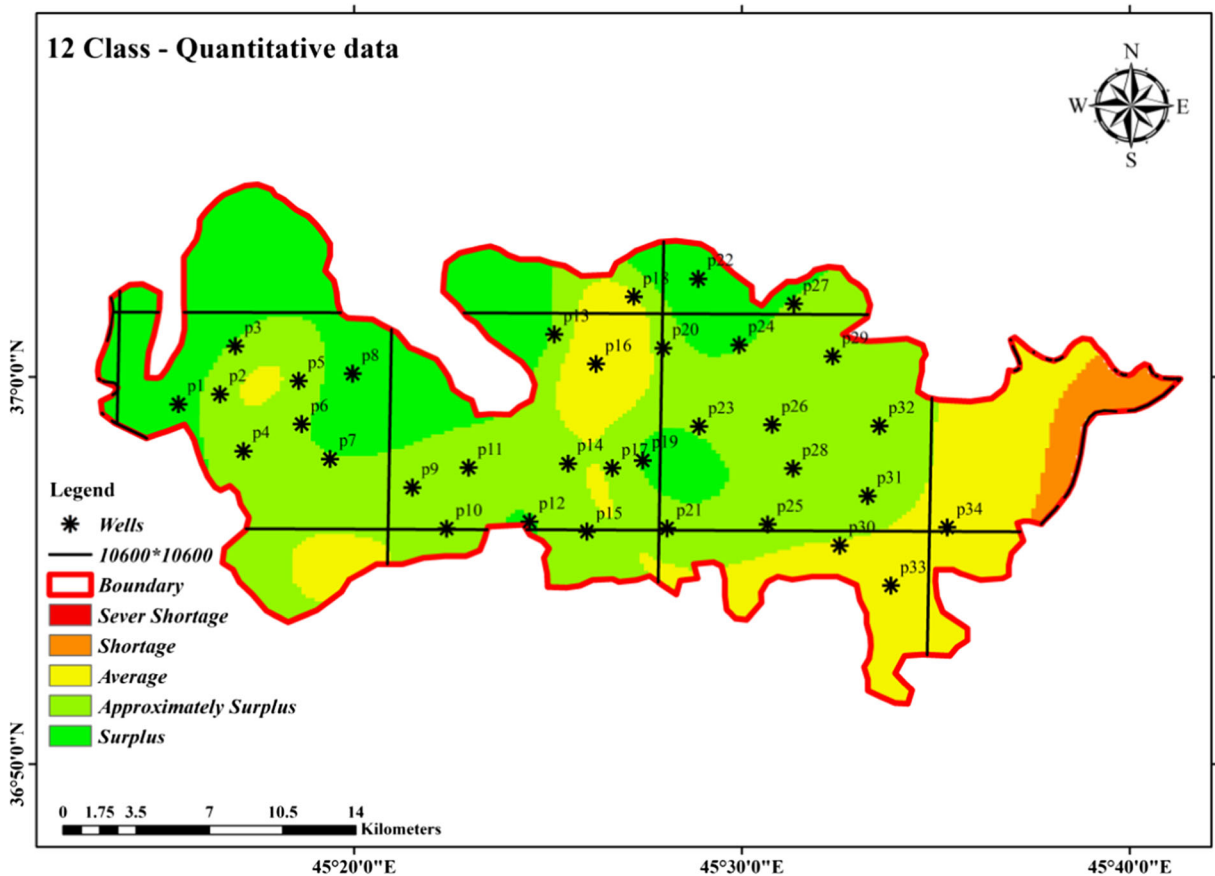


Fig. 7 Results of gridding the studied aquifer for monitoring the groundwater quantity network

not holding high ranks can be removed and for networks with insufficient number of wells, the recommendation of Mogheir et al. (2006) can be applied, which says to utilize abandoned wells or drinking water wells. According to the available networks, on average a piezometric well is required to monitor the groundwater quantity in each region with an area of 14.28 km² (or a circle with a radius of 2.13 km). However, in the groundwater quantity network in the northwest parts of the studied plain, three wells in the eastern part of the basin and two wells in the southwest parts of the basin are required to monitor the network more accurately.

Conclusion

This study conducted a quantitative and qualitative monitoring network of the groundwater of Naqadeh plain located south of Lake Urmia using entropy theory.

Regarding the quantitative (groundwater) and qualitative (salinity) networks, the following results were obtained:

- By developing the T-model and examining different classes of data clustering, the gap was eliminated in this study. In previous studies, a method has been proposed for the number of different classes of data clusters which was not true for each region.
- The results of integration of entropy indices and estimation of information transfer showed that the northwest of the aquifer has high shortage of information meaning that, in these areas, there was a need to construct new wells to better monitor groundwater quality in terms of salinity.
- Approximately 66% of the aquifer area was surplus in terms of quality information transfer of wells and their number. On the whole, the existing network maintains a significant statistical relationship between wells, but constructing two wells in the

northwest of aquifer will make this connection 100%. One of the most important goals of entropy theory is introduction of monitoring networks with 100% coverage.

- Regarding the quantitative groundwater monitoring network in the studied area, there were good data transfer capacities for the network, and the current network can provide good information on the groundwater in the plain. However, monitoring the quantitative network of groundwater has been associated with a reduction in information transfer in eastern boundaries of the basin, indicating a lack of three wells in these areas. However, due to the location and reduction of Lake Urmia water level and its retention, there is a possibility of a sharp decrease in the water level in these areas. Moreover, due to the quality and salinity of Lake Urmia water, it is likely that information transfer from the wells of these areas will be ceased in the future. Entropy indicators such as $R(i)$ and $S(i)$ show this very clearly.
- Another use of the entropy index in aquifer management is related to net exchange information. The results of examining net exchange information index ($N(i)$) showed that wells close to Lake Urmia had more value than other wells and had achieved a higher ranking in terms of groundwater quality monitoring in the studied area. These include q19, q23, and q24. Moreover, in the groundwater quality monitoring of the aquifer, the q9, q3, and q11 wells received the lowest value and their future activity requires a serious revision and information of these areas are less reliable.
- The ranking results of the piezometers in the quantitative monitoring of groundwater network showed that piezometers near the Lake Urmia had lower values and, in fact, their information should be considered carefully for managing aquifer. On the other hand, central wells of the aquifer are of great importance and have higher values. High value wells are considered valuable in terms of providing information and, if possible, measurements of groundwater levels and EC parameters of these wells should be carried out within a shorter period of time.
- The provision of regular monitoring networks is another capability of entropy theory which has been less studied. The results showed that in the studied area, the best network for monitoring the quality and quantity of groundwater were $12,500 \times 12,500$ and

$10,000 \times 10,000$ m, respectively. On the other hand, entropy theory can be used to design a new network which can be effective in designing an efficient and strategic network based on information.

Acknowledgements The authors would like to thank West Azerbaijan Regional Water Authority for providing the data.

References

- Chadalavada, S., Datta, B., & Naidu, R. (2011). Uncertainty based optimal monitoring network design for a chlorinated hydrocarbon contaminated site. *Environmental Monitoring and Assessment*, 173(1), 929–940. <https://doi.org/10.1007/s10661-010-1435-2>.
- Chapman, T. G. (1986). Entropy as a measure of hydrologic data uncertainty and model performance. *Journal of Hydrology*, 85(1–2), 111–126. [https://doi.org/10.1016/0022-1694\(86\)90079-X](https://doi.org/10.1016/0022-1694(86)90079-X).
- Chen, Y. C., Wei, C., & Yeh, H. C. (2008). Rainfall network design using kriging and entropy. *Hydrological Processes*, 22(3), 340–346. <https://doi.org/10.1002/hyp.6292>.
- Harmancioglu, N. B., & Alpaslan, N. (1992). Water quality monitoring network design: A problem of multiobjective decision making. *Journal of the American Water Resources Association*, 28(1), 179–192. <https://doi.org/10.1111/j.1752-1688.1992.tb03163.x>.
- Harmancioglu, N. B., Ozkul, S. D., & Alpaslan, M. N. (1998). Water quality monitoring and network design. *Environmental Data Management*, 27, 61–106. <https://doi.org/10.1007/978-94-015-9056-34>.
- Jaynes E. T. 1957. Information theory and statistical mechanics, *I. Phys. Rev*, 106, 620–630.
- Jessop, A. (1995). *Informed assessments, an introduction to information, entropy and statistics*. New York: Ellis Horwood.
- Keum, J., Kornelsen, K., Leach, J., & Coulibaly, P. (2017). Entropy applications to water monitoring network design: a review. *Entropy*, 19(11), 1–21. <https://doi.org/10.3390/e19110613>.
- Khalili, K., Tahoudi, M. N., Mirabbasi, R., & Ahmadi, F. (2016). Investigation of spatial and temporal variability of precipitation in Iran over the last half century. *Stochastic Environmental Research and Risk Assessment*, 30(4): 1205–1221.
- Krstanovic, P. F., & Singh, V. P. (1992). Evaluation of rainfall networks using entropy: I. Theoretical development. *Water Resources Management*, 6(4), 279–293. <https://doi.org/10.1007/BF00872281>.
- Lubbe, C. (1996). *Information theory*. Cambridge: Cambridge University Press.
- Markus, M., Knapp, H. V., & Tasker, G. D. (2003). Entropy and generalized least square methods in assessment of the regional value of stream gages. *Journal of Hydrology*, 283(1), 107–121. [https://doi.org/10.1016/S0022-1694\(03\)00244-0](https://doi.org/10.1016/S0022-1694(03)00244-0).

- 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., De Lima, J. L. M. P., & Singh, V. P. (2004). Characterizing the spatial variability of groundwater quality using the entropy theory: II. Case study from Gaza strip. *Hydrological Processes*, 18(13), 2579–2590.
- Mogheir, Y., Singh, V. P., & de Lima, J. L. M. P. (2006). Spatial assessment and redesign of a groundwater quality monitoring network using entropy theory, Gaza Strip, Palestine. *Hydrogeology Journal*, 14(5), 700–712. <https://doi.org/10.1007/s10040-005-0464-3>.
- 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. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:1\(90\)#sthash.eBtsb7ac.dpuf](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:1(90)#sthash.eBtsb7ac.dpuf).
- Quevauviller, P. (2009). *Groundwater monitoring*. Hoboken: John Wiley & Sons.
- Şarlak, N., & Şorman, A. (2006). Evaluation and selection of streamflow network stations using entropy methods. *Turkish Journal of Engineering and Environmental Sciences*, 30(2), 91–100.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423. <https://doi.org/10.1145/584091.584093>.
- Van Luin, A. B., & Ottens, J. J. (1997). Conclusions and recommendations of the international workshop monitoring tailor-made II–information strategies in water management. *European Water Pollution Control*, 4(7), 53–55.
- Wu, S., & Zidek, J. V. (1992). An entropy-based analysis of data from selected NADP/NTN network sites for 1983–1986. *Atmospheric Environment, Part A: General Topics*, 26(11), 2089–2103. [https://doi.org/10.1016/0960-1686\(92\)90093-Z](https://doi.org/10.1016/0960-1686(92)90093-Z).
- Zhou, Y. (1996). *Spatial data generation program (COVRAN)*. The Netherlands: Delft.
- Zhu, Q., Shen, L., Liu, P., Zhao, Y., Yang, Y., Huang, D., & Yang, J. (2015). Evolution of the water resources system based on synergetic and entropy theory. *Polish Journal of Environmental Studies*, 24(6), 2727–2738. <https://doi.org/10.15244/pjoes/59236>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.