Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2020.25456

192 (2020) 97–110
July

Water quality monitoring optimization using genetic algorithm, a case study: Mond River in Iran

Gholamreza Asadollahfardi*, Nima Heidarzadeh, Ali Sekhavati, Ali Rashtizadeh, Amirmohsen Samadi

Civil Engineering Department, Technical and Engineering Faculty, Kharazmi University, Tehran, 15719-14911, Iran, Tel. +98-912-1192-424; email: fardi@khu.ac.ir (G. Asadollahfardi), Tel. +989122279268; email: n.heidarzadeh@khu.ac.ir (N. Heidarzadeh), Tel. +989191577518; email: ali.sekhavati@alumni.ut.ac.ir (A. Sekhavati), Tel. +98-919-5269-960; email: rashtizadehali@gmail.com (A. Rashtizadeh), Tel. +98-936-8880-231; email: amirmohsen47@gmail.com (A. Samadi)

Received 27 August 2019; Accepted 14 January 2020

ABSTRACT

The economic management of surface and groundwater resources is a significant issue in the world. In today's industrial societies which increase the number of various pollutants is considerable, the existence of a monitoring network with appropriate station locations is essential for surface water. Because water quality monitoring of each station is expensive, the number of stations should be optimized, and also the collected data should be represented the water quality in surface water. First, we applied Strahler stream order to classify Mond Basin river and then used the genetic algorithm to optimize the number of existing stations in the river, which is located in Fars and Bushehr (Iran), using existing water quality data of 16 stations to reduce the number of monitoring stations. Based on Strahler ranking method, station 12 (Dahrom) with rank 4 had the maximum pollution index including 77.61 for the drinking usage and 82.95 for irrigation usage and the stations, which are located on rank 4 rivers had the maximum pollution index in various combinations of stations with 197.88 for irrigation and 232.49 for drinking uses. For irrigation, existing water quality monitoring stations can be reduced from 16 to 12 and for potable water, the number of stations can be reduced from 16 to 11.

Keywords: Mond Basin River; Genetic algorithm; Optimization of monitoring stations; Strahler number

1. Introduction

Water quality monitoring of surface water is a significant element of water resource management which determines temporal and spatial variations of physical, chemical, and biological properties of surface water. The selection of appropriate monitoring stations helps to achieve suitable locations for monitoring stations, which leads to a decline in uncertainty in the value of water quality data. Therefore, the study of the appropriate selection of water quality stations is vital.

Nowadays, water quality monitoring and water resources optimization have been carried out by various researchers which some of the water quality studies explained: Li et al. [1] indicated water usage of the riparian woodlands was low and similar in an arid area, and they proposed the long-term plant adaptation to the local weather and conditions of water obtainability. Yin et al. [2] indicated that temperature, pH, nitrate-nitrogen (NO_2–N), and total phosphorus are the most important abiotic factors affecting biological pollution distribution. Jafarabadi et al. [3] indicated the key sources of the contamination in the Persian Gulf were fuel and oil combustion mostly from offshore petroleum exploration and extraction, releasing of contamination from shipping activities. Yu et al. [4] described that plant water usage is the

* Corresponding author.

consequence of long-term adaptation to local climates and water obtainability.

The location of river water quality monitoring stations depends on monitoring objectives and the number of samples collected for each station as well as the funding available for monitoring [5] and [6]. The design of a water quality monitoring network is a repetitive-based approach so that an existing network must be reviewed again after certain periods based on changes in environmental requirements and objectives for water resources management. In recent years, regarding rising monitoring costs, approaches have been taken to reduce the number of stations and optimize their numbers [7]. Different mathematical methods are available for optimizing river quality monitoring stations, such as dynamic method, multivariate decision-making method, and genetic algorithm. We applied a genetic algorithm in this study and reviewed historical studies using this method.

Ritzel et al. [8] used a tournament selection method in a genetic algorithm to design a multiobjective groundwater quality monitoring system. Karpouzos et al. [9] used genetic algorithm research to achieve reliable water quality and could solve the inverse problem in hydrogeology. Icaga [10] carried out on the Gedi River with genetic algorithm in Turkey, optimized the number of quality monitoring stations by using other parameters such as population living around the river, the area under cultivation, biological data, and the impact of the point and non-point pollution sources of the river. Park et al. [11], using Geographic Information System (GIS) and integrating that with the genetic algorithm, designed a water quality monitoring network on the Nakdong River and found the exact location of the quality monitoring stations. Karamouz et al. [12] designed a water quality monitoring network on the Karun River in the southwest of Iran. They used an optimization model based on the genetic algorithm and a combination of the developed Keriging and analytic hierarchy process (AHP) methods. Lee et al. [13] proposed optimal water quality monitoring sites on the Logan and Albert Rivers Network (in the United States) using the new method of combining the cost function with the genetic algorithm. They indicated that using the similarity between the cost function and topographic features such as the number and length of the divided intervals of the rivers (distance between stations); it is possible to find the optimal positioning of the stations. Liyanage et al. [14] evaluated the number of the Kelani River monitoring stations in Sri Lanka using optimization methods such as multi-objective analysis and the genetic algorithm for selecting quality control stations in places.

Both accurate locations of water quality monitoring stations and choosing proper physical, chemical, and biological parameters are significant for water quality management in point of view of the quality of water usage and economic of monitoring station selection. Therefore, this study aims to optimize the number of existing water quality monitoring stations in the southern Fars province and Bushehr province, Mond River basin, (Iran) by using the genetic algorithm.

1.1. Study area

Mond basin rivers are located in the south of Fars province and Bushehr province with an area of 47,654 km^2, Fig. 1 [15].

This basin consists of the main Mond River; the most important feeders are Qareaghaj, Shour Jahrom, Firouzabad, and Shour Dahrom rivers. The main body of the Mond River is established by connecting Qareaghaj and Shour Dahrom rivers. Then, by connecting other branches to the main body of the Mond River, the mainstream discharges to the Persian Gulf [16].

The average annual rainfall in the long term period in the total basin area is about 307.5 mm. The total evaporation rate in all areas of the study zone increased, and it is maximum around low-lying beaches, but the average annual evaporation rate is variable, and it is between 2,131 and 3,975 mm [17] and [18].

In this basin, 30 water quality monitoring stations have been installed since 1966 but some of them are deactivated. Information on stations was collected from Iran Water Resources Management Company and Water Resources Company of Bushehr and Fars Provinces. Fig. 2 shows the hydrometric stations of the Mond River basin. We selected 16 active stations to study the water quality of rivers and the name and information of them listed in Table 1. The reasons for selecting these 16 stations are: (1) these stations are active until the end of the water year 2011–2012. (2) In these stations, the physical and the chemical water quality data were available, which overlap over 20 y (the water year 1991–1992 to 2011–2012). Thus, the study of quality and evaluation of the number of stations is based on a period of 20 y.

2. Materials and methods

We optimized some water quality monitoring stations by using a target function based on the pollution index, and the genetic algorithm optimization method. Similar to the work of Icaga [10], Cetinkaya and Harmancioglu [7,19], an index formed based on maximizing the values of the parameters. However, we used physical and chemical water quality parameters of each station to generate the pollution index. First, according to the existing physical and chemical water quality parameters, including total dissolved solids (TDS), sulfate (SO_4^{2-}), chloride (Cl^-), sodium cations (Na^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), acidity (pH), sodium absorption ratio, and electrical conductivity (EC) for drinking and irrigation usage, and based on standard guidelines 1053 (Iranian standard), and Food and Agriculture Organization, the effective pollutant parameters identified in different stations and was selected for forming pollution index. According to the mutual effect of sodium, calcium, and magnesium cations on the sodium absorption parameter and to avoid duplicate effects of parameters, only sodium absorption parameter was considered for irrigation usage. Also, according to direct relations between TDS and EC and to avoid duplicate effects of parameters, only EC was considered for irrigation usage.

2.1. Standardization and data harmonization

Before starting the calculation, the various inputs and parameters that were measured in different units, for measurement, comparison, and calculation should be standardized. Then, the elements of the transformed indices measured without a dimension [20]. Rolim da Paz et al. [21] standardized data and limited them to range 0.1–0.9 by using Eq. (1).

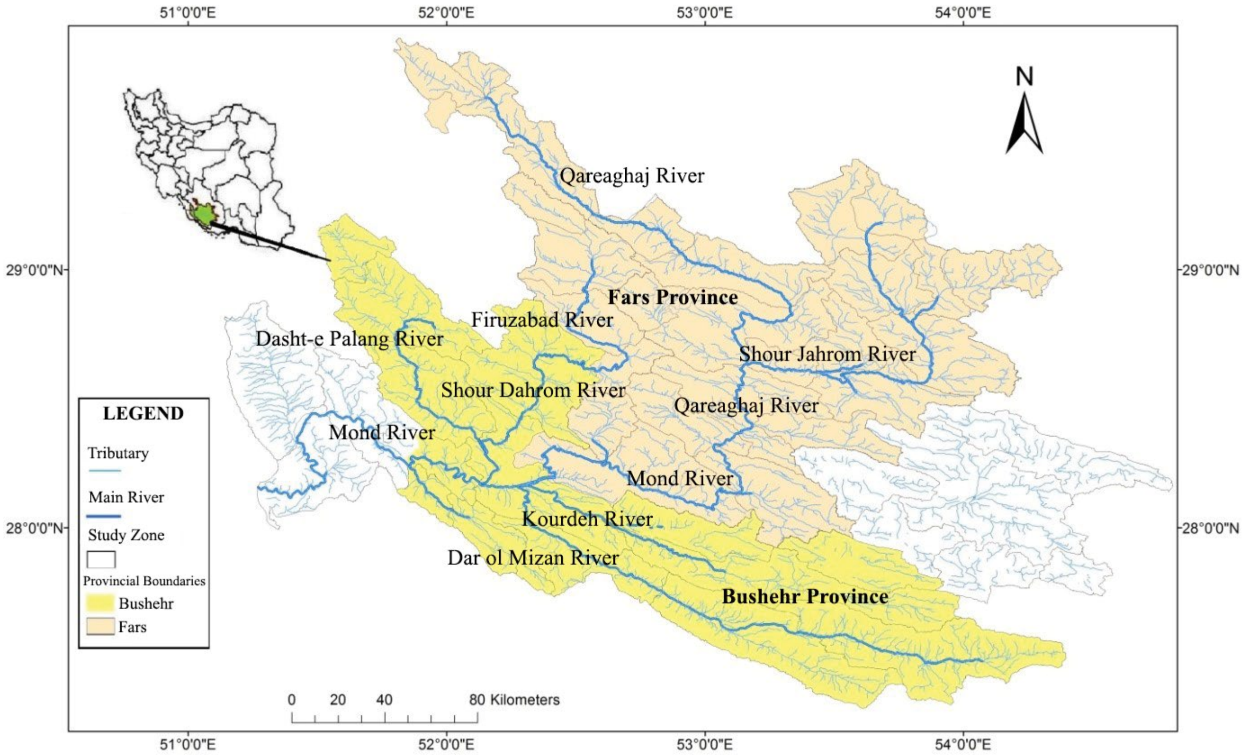

Fig. 1. Location of Mond waterway basin.

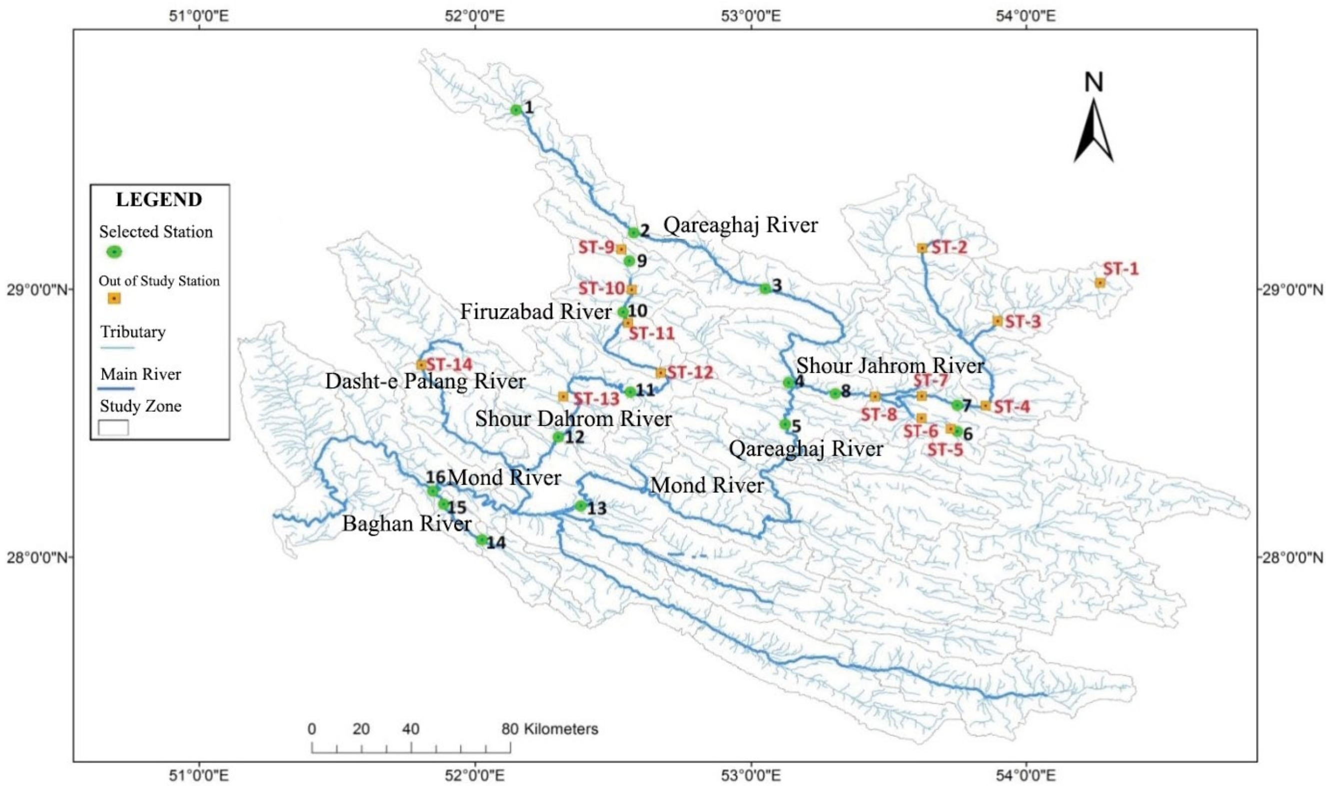

Fig. 2. Hydrometric stations of the waterway of the Mond Basin with Provincial separation.

Table 1
Name and location of selected stations in the Mond basin

No	Name of Station	Province	Main river	River	Latitude	Longitude	Height (m)
1	Khanzenian	Fars	Qareaghaj	Khatiri	29–40–00	52–09–00	1,940
2	Band Bahman	Fars	Qareaghaj	Qareaghaj	29–12–00	52–34–00	1,620
3	Aliabad Khofar	Fars	Qareaghaj	Qareaghaj	29–00–00	53–03–00	1,349
4	Borak	Fars	Qareaghaj	Simkan	28–38–00	53–08–00	876
5	Tang Karzin	Fars	Qareaghaj	Qareaghaj	28–29–59	53–07–44	760
6	Sarvo	Fars	Shour Jahrom	Shour	28–28–00	53–45–00	1,370
7	Baba Arab	Fars	Shour Jahrom	Shour Jahrom	28–34–00	53–45–00	1,095
8	Hokan	Fars	Shour Jahrom	Shour Jahrom	28–36–00	53–18–00	933
9	Hanifghan	Fars	Firuzabad	Hanifghan	29–06–16	52–33–45	1,585
10	Tongan	Fars	Firuzabad	Firuzabad	28–54–42	52–32–18	1,376
11	Dehroud	Fars	Firuzabad	Firuzabad	28–36–59	52–33–45	903
12	Dahrom	Fars	Firuzabad	Shour Jahrom	28–26–54	52–18–31	384
13	Dejgah	Fars	Mond	Mond	28–11–35	52–23–00	222
14	Galou Bardekan	Bushehr	Mond	Riz	27–54–00	52–14–00	540
15	Baghan	Bushehr	Mond	Baghan	28–14–00	51–22–00	83
16	Qantareh	Bushehr	Mond	Mond	28–15–00	51–51–00	70

We used this method to standardize data because of the effect of all parameters in the generation of a pollution index.

$$\bar{X} = 0.1 + 0.8 \times \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where $\bar{X}$ = standardized value, X = value of data, $X_{\min}$ = minimum value of data, and $X_{\max}$ = maximum value of data

2.1.1. Importance and weighting of parameters

The other innovation in this study is to find out the relative importance of selected parameters of drinking and irrigation water. We used the paired matrix, a method based on the AHP introduced in 1980 [22], and we sent a questionnaire to 15 experts. For parameters C_1, C_2, ……, C_n, a paired matrix was formed according to Eqs. (2) and (3), then they were asked to fill the matrix with either relative importance parameters or relative Preference parameters [23–25].

$$A = \left[a_{ij}\right], i,j = 1,2,\ldots,n = \begin{array}{c|c} \text{paired matrix} & C_1 \quad C_2 \quad \cdots \quad C_n \\ \begin{array}{c} C_1 \\ C_2 \\ \cdots \\ C_n \end{array} & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} A \end{array} \tag{2}$$

$$a_{ij} = \frac{1}{a_{ji}} \cdot \text{for } \ i = \text{then } a_{ij} = 1 \tag{3}$$

where a_{ij} is the preference of i_{th} element related to the j_{th} element, was sorted based on Table 2. When the paired matrix was formed according to Eqs. (4) and (5), we used a geometric mean to calculate the weight of each parameter, which is an approximate method with faster and lower error calculations [23] and [26].

Table 2
Verbal preference and relative importance of parameters relative to each other [23]

Value	Preference (verbal judgment)
9	Extremely preferred
7	Very strongly preferred
5	Strongly preferred
3	Moderately preferred
1	Equally preferred
2,4,6,8	Shows midway values (value 8 indicates the value of >7 and <9)

$$r_i = \left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}} = \begin{bmatrix} \sqrt[n]{a_{11} \times a_{12} \times \ldots \times a_{1n}} \\ \ldots \\ \ldots \\ \ldots \\ \sqrt[n]{a_{n1} \times a_{n2} \times \ldots \times a_{nn}} \end{bmatrix} \tag{4}$$

$$W_i = \frac{r_i}{\sum_{i=1}^{n} r_i} \tag{5}$$

By checking the rate of decision compatibility, we can trust the model easily. Inconsistency index (I.I.) by Eq. (6) and inconsistency ratio by Eq. (7) were calculated. Table 3

Table 3
Inconsistency Index of Random Matrix

Matrix dimension (n)	I.I.R.
1	0
2	0
3	0.58
4	0.9
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.45

indicates the value of the inconsistency index of random matrix (I.I.R.) for the paired matrix. The maximum value of the consistency rate is 0.1. If this value is >0.1, the decision-maker will revise his judgments. But if it is <0.1, it will be acceptable.

$$\text{I.I} = \frac{\lambda_{\max} - n}{n-1} \tag{6}$$

$$\text{I.R} = \frac{\text{I.I.}}{\text{I.I.R}} \tag{7}$$

where $\lambda_{\max}$ = Maximum Eigenvalue of matrix A; n = Dimension of paired matrix.

2.2. Estimate values of pollution index for each station

We used $SU_{j(i)l}$ as standard data, and l_{th} data being uniform in the i_{th} station that $j(i)$ is a counter index of the i_{th} station. Based on the importance of the parameter in comparison with each other and according to drinking and irrigation usage, we used the relative weight of parameters in Eq. (8) [7,19].

$$\text{TS}_{j(i)} = \sum_{l=1}^{l_N} W_l \times \text{SU}_{j(i)l} \tag{8}$$

where $TS_{j(i)}$ = Value of pollution index for drinking and irrigation usage; l_N = number of parameters in the station I; W_l = relative weight of parameter i.

First, we classified stations and involved each classification to decision steps. Then, the objective function was calculated for entering to optimization models.

The classification of stations is based on the ranking of the rivers in which stations are located. We used Strahler Number to the ranking of Mond Basin rivers and the classification of stations. This number is known as the rivers category. With this method, we ranked all the branches of the river. The river-level number at the center point represents the full extent of the river network in the upstream basin of that point. So that, at the end of waterways, the magnitude or rank of the final section of the waterways network is an equal sum of rank, previous waterways and contains total pollution of the network [27].

2.3. Estimation of objective function values for stations with various combinations

The estimation of objective function requires a pollution index of both irrigation and drinking water. To do so, we added subcategory subscript k to parameters [$SU_{j(i)l}$] and [$TS_{j(i)}$], to use in optimization models. Then, for each station in subcategory (k), the sum of standardized data [$SU_{j(i)kl}$], is represented in Eq. (9), according to the weight of data for drinking water and irrigation water with [$TS_{j(i)k}$] or value index of each station. The Value of each T_{RN} (number of the remaining stations that required) is necessary that we determined the number of stations was selected in subcategory k (R_k). Accordingly, the stations with the maximum sum of standardized data [$SU_{j(i)kl}$], were selected. For each station in each primary subcategory (k), the sum of normalized data [$SU_{j(i)kl}$], is shown by $TS_{j(i)kl}$.

$$\text{TS}_{j(i)k} = \sum_{l=1}^{l_N} W_l \times \text{SU}_{j(i)l} \tag{9}$$

where l_N = number of parameters in station i and subcategory k, W_l = relative weight of parameter l, $j(i)$ = parameter number of stations in primary subcategory k of stations.

Using Eq. (9), we obtain the total value of parameters in primary subcategory k and i_{th} station. In each primary subcategory k, there are different choices for stations depending on R_k and the value of $TS_{j(i)kl}$ which is different in each station.

$$\text{MTS}_{j(i)k} = \text{Max}\,\text{TS}_{j(i)k} \tag{10}$$

After determining T_{RN}, we choose R_K with maximum $MTS_{j(i)k}$.

$$\text{SMTS} = \text{Max}\sum_{k=1}^{N}\sum_{i=1}^{R_k}\text{MTS}_{j(i)k} \tag{11}$$

Limitation of problem parameters are as below:

$$\sum_{k=1}^{N} R_k = \text{TR}_N \qquad 0 \le R_k \le \text{TR}_N \tag{12}$$

$0 \le j(i) \le R_k$, $j(i) \ne j(h)$, $I \ne h$

The aim was to find out a combination of stations, which has maximum $MTS_{j(i)k}$ correspond to the clear value of $T_{RN.}$ Therefore, similar to Eq. (11), objective function will have two dimensions, and we used the genetic algorithm method to solve the objective function [6,7].

2.4. Using a genetic algorithm to select the monitoring station

In this method, each problem answer is like a chromosome, and each decision variable acts as genes within the chromosomes. Each chromosome includes N genes, and each N determines the number of stations that correspond to each subcategory of stations network rivers. In this study, the value of each gene is the maximum number of stations in each category or subcategory. Genetic algorithm method can be schematically illustrated in five stages, which is indicated in Fig. 3 [10,28,29].

First, we need to define strings of chromosomes as the population of the problem. As indicated in Table 4, there is a chromosome with n genes, according to our problem, each subcategory for the stations and then our decisions namely the stations fit into each gene.

Then, we have numeric strings as chromosome, which sum of these numbers represents the total number of remaining stations.

$$\text{Total number of required remained stations} = \sum_{j=1}^{N}\left(L_j\right) \tag{13}$$

On the other hand, each L_j represents the cumulative value (amount of index) of stations in each subcategory (Table 5).

Now, we check objective function which is SMTS. We maximized Eq. (14).

$$f(i) = \text{SMTS} = \text{Max}\sum_{k=1}^{N}\sum_{i=1}^{L_k}\text{MTS}_{j(i)k} \tag{14}$$

where i = No. i_{th} population, N = No. of subcategory, $\text{MTS}_{j(i)k}$ = value (amount of index).

After selecting some chromosomes as an initial population, we computed the fitness function. Fitness function (objective function) by entering $\text{MTS}_{j(i)k}$ or maximized value of different stations combinations for drinking water and irrigation water to calculate the sum of maximum values $\text{MTS}_{j(i)k}$ in each subcategory with MATLAB software and this result is called chromosome fitness [30] and [31]. Each chromosome, which had better fitness, as a new generation goes to the crossover stage for crossover [32]. We produced new chromosomes by crossover and mutation and then calculated their fitness until the maximum fitness was obtained by iteration and iterate for other generations. Max [$f(i)$] is the optimal answer to the problem [33,34].

If Fig. 4 is the exhibitor of the gene number or decision variable, we must determine the crossover point, and the number of crossover points can be $n_{\text{Var}} - 1$. We displaced parent's genes from the crossover point [33,35].

Several children, which can be obtained through crossover can be determined from Eq. (15) [34].

$$nc = 2 \times \text{round}\left(P_c \times \frac{\text{pop}_{\text{size}}}{2}\right) \tag{15}$$

where pop_{size} = n of population or initial chromosomes, p_c = crossover rate (%).

Round function in MATLAB R2015a software rounds the given number to the closest integer.

The number of chromosomes that will be mutated is determined by Eq. (16) [34].

$$nm = \text{round}\left(P_m \times \frac{\text{pop}_{\text{size}}}{2}\right) \tag{16}$$

Table 4
A chromosome and the placement of its genes

k_1	k_2	k_3	k_n
L_1	L_2	L_3	L_n

L_j = The remain stations in each subcategory, k_j = Subcategory of Stations in a network of rivers

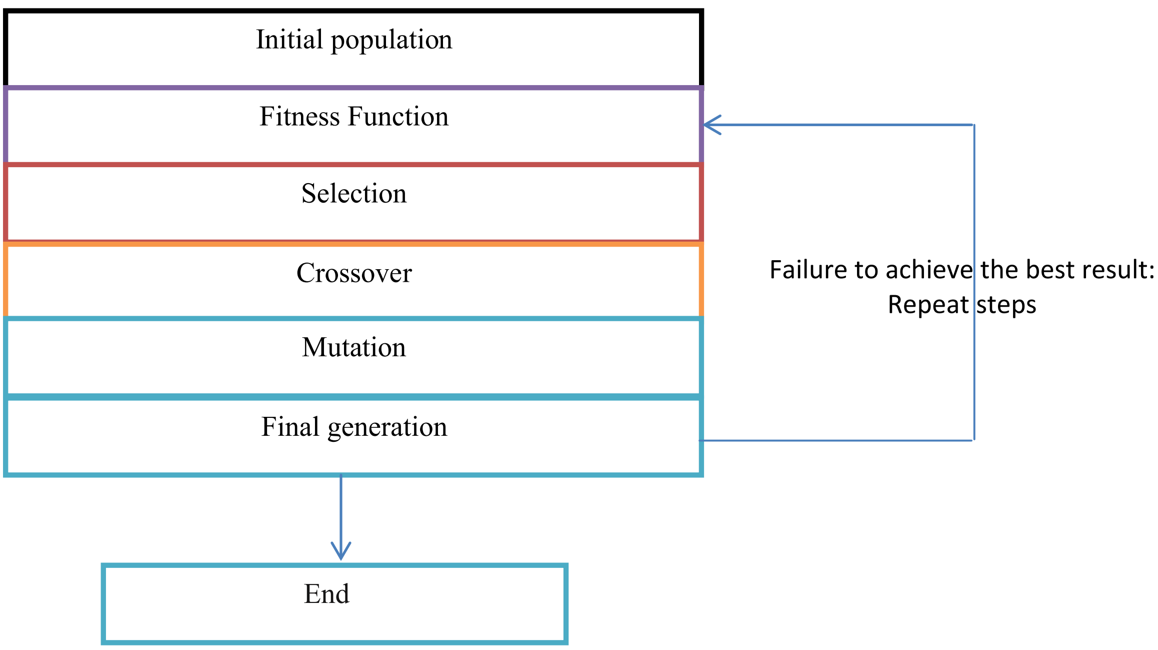

Fig. 3. Genetic algorithm steps.

Table 5
Value or amount of index for each stations combination in each subcategory

L_k	$K1$	$K2$	$K3$	$K4$	K_{N-1}	K_N
1	MT1.1	MT2.1	MT3.1	MT4.1	MTSN-1.1	MTSN.1
2	MT1.2	MT2.2	MT3.2	MT4.2	MTSN-1.2	MTSN.2
3	MT1.3	MT2.3	MT3.3	MT4.3	MTSN-1.3	MTSN.3
N–1	MT1.N–1	MT2.N–1	MT3.N–1	MT4.N–1	MTSN-1.N–1	MTSN.N–1
N	MT1.N	MT2.N	MT3.N–1	MT4.N	MTSN-1.N	MTSN.N

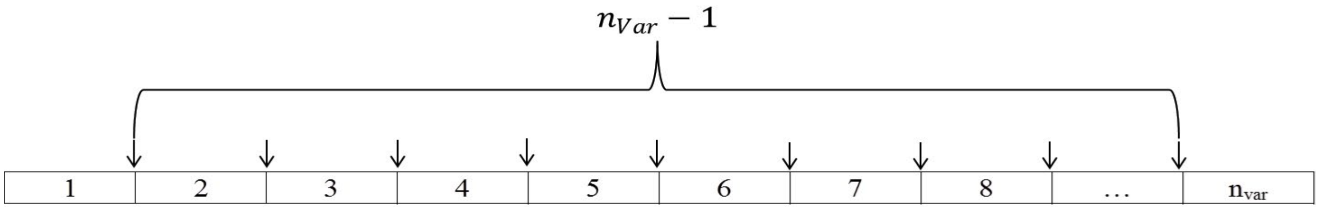

Fig. 4. Location of crossover point for crossover operation.

where P_m = mutation rate

These calculations would be iterated until the obtaining of the best population or optimum answer [31].

2.5. Some suggested stations for waterway network

The problem is to determine the station's combination with the more pollution index and the aim of optimizing is to realize the number of stations which finally would be remained in the system. After defining a maximization objective function (SMTS) for each T_{RN} in genetic algorithm parts, a maximization problem would be solved separately. For example, for each station category listed in four subcategories (k_1, k_2, k_3, and k_4), the most critical station's combination is selected according to an objective function for 4–16 (16 is the total number of stations examined) remaining stations. After finding the critical combination, we reduced the number of stations by the same methods of Cetinkaya and Harmancioglu [7]. In each category value of the objective function or value of the total index (SMTS) was estimated for any number of the remaining stations in the network (for example, for four categories, T_{RN} = 4–16). Then, the objective function difference values (DSMTS) were calculated according to Eq. (17) for two consecutive stations and by plotting DSMTS (vertical axis) vs. some stations (horizontal axis) the number of proposed stations was judged.

The number of suggestion stations is a number which after that number, the difference between DSMTS and variation of the objective function is very low and negligible. As a result, after that number of stations by adding a new station to the network, the changing of the objective function is negligible. The number of proposed stations is the minimum number required for the network [7].

$$\Delta \text{SMST} = \text{SMST}_{i+1} - \text{SMST}_i \quad (17)$$

where i = subscript of station number

3. Results

3.1. Classification of stations

Stations classification was carried out after rivers were ranked by maps and layers of Mond Basin through ARC GIS version 10.3(2014) software.

Fig. 5 presents the rank of rivers by using the Strahler and ARC GIS methods and selected stations on the network of rivers.

According to Table 6, stations are separated into five subcategories (k_1–k_5) based on the rank of rivers in which stations are located on them.

Based on Table 7, totaling three categories for the stations were generated and considered. Each category or subcategory should have at least one monitoring station. In the first category, since just one station is located on river rank 2, the stations which are located on river rank 2 and river rank 3 are contained in one subcategory ($K1$) together. The stations which are located on river rank 4,5, and 6, are contained in subcategory $K2$, $K3$, and $K4$ respectively. In the same way, for the second and third classification, subcategories and stations that are located on them are separated.

3.1.1. Calculation of objective function for various combinations of stations

To calculate the objective function of each station in each subcategory (k), the sum of standardized data [$SU_{j(i)kl}$] was determined according to the weight of data for drinking and irrigation usage through $TS_{j(i)k}$ or index value of each station. Table 8 indicates the index of pollution for drinking and irrigation usage.

To calculate $MTS_{j(i)k}$ or maximum value of various combinations in subcategories, according to the value of $TS_{j(i)k}$ in Table 7 (Station classification) and Table 8 were determined. Table 9 indicates the values of $MTS_{j(i)k}$ for 1st classification (subcategories $K1$, $K2$, $K3$, and $K4$).

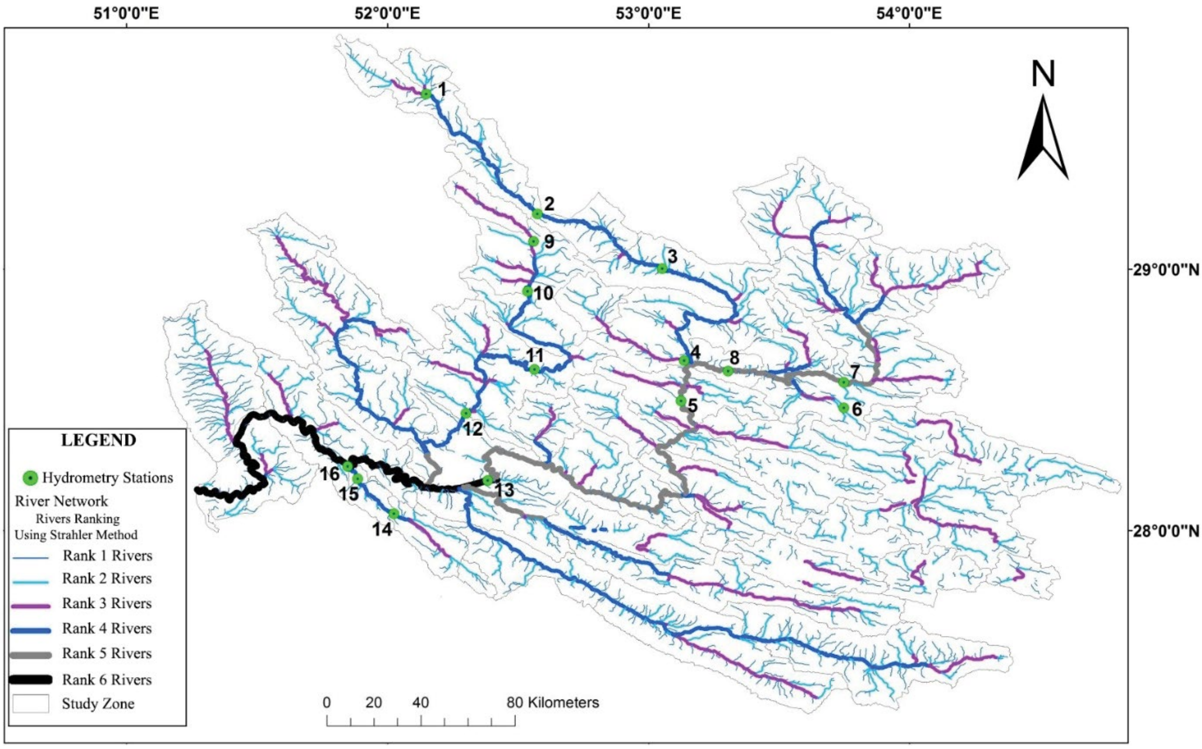

Fig. 5. Ranking rivers of Mond Basin by Strahler method.

Table 6
Stations classification according to the rank of rivers

Subcategory	Description	Number	Station	No. of stations
*K*1	Stations are located on rank 2 rivers	1	Sarvo	6
*K*2	Stations are located on rank 3 rivers	1	Khanzenian	1
		2	Borak	4
		3	Hanifghan	9
*K*3	Stations are located on rank 4 rivers	1	Band Bahman	2
		2	Aliabad Khofar	3
		3	Tongan	10
		4	Dehroud	11
		5	Dahrom	12
		6	Galou Bardekan	14
		7	Baghan	15
*K*4	Stations are located on rank 5 rivers	1	Tang Karzin	5
		2	Baba Arab	7
		3	Hokan	8
*K*5	Stations are located on rank 4 rivers	1	Dejgah	13
		2	Qantareh	16

Then, after calculating the $MTS_{j(i)k}$ in each subcategory, through determining T_{RN} (number of stations which will remain in the system), our choice is R_k (number of selected stations in each subcategory), which has a maximum sum value of $MTS_{j(i)k}$ (or SMTS). To obtain the value of an objective function or total value of stations, we used the genetic algorithm method.

3.2. Choose sampling stations through genetic algorithm

Calculations were performed by using the value of $TS_{j(i)K}$ or value index of each station (Table 8) for different usage and category of each station (Table 7). In solving problems by genetic algorithm method, we selected station combinations in each subcategory "*k*" (each subcategory calls a Gene) until

Table 7
Stations classification according to rivers with various ranks

Third classification			Second classification			First classification			No. of station	Stations
Subcategory	Description	R_k	Subcategory	Description	R_k	Subcategory	Description	R_k		
$K1$	Stations are located on rank 2,3,4 rivers	1«R1«11	$K1$	Stations are located on rank 2,3 rivers	1«R1«4	$K1$	Stations are located on rank 2,3 rivers	1«R1«4	6	Sarvo
									1	Khanzenian
									4	Borak
									9	Hanifghan
			$K2$	Stations are located on rank 4 rivers		$K2$	Stations are located on rank 4 rivers	1«R2«7	2	Band Bahman
									3	Aliabad Khofar
									10	Tongan
					1«R2«7				11	Dehroud
									12	Dahrom
									14	Galou Bardekan
									15	Baghan
$K2$	Stations are located on rank 5,6 rivers	1«R2«5	$K3$	Stations are located on rank 5,6 rivers	1«R3«5	$K3$	Stations are located on rank 5 rivers	1«R3«3	5	Tang Karzin
									7	Baba Arab
									8	Hokan
						$K4$	Stations are located on rank 6 rivers	1«R4«2	13	Dejgah
									16	Qantareh

the last subcategory, so that sum of $MTS_{j(i)k}$ or maximum value of different stations combinations or that objective function (SMTS) or Total final value (fitness target) for k_{th} step, was determined according to drinking and irrigation usage.

3.2.1. The operational process of choosing stations by using genetic algorithm method and MATLAB software for first stations classification [four subcategories (K)]

To begin operational by using the genetic algorithm, the values of $MTS_{j(i)k}$ or the maximum value of different station combinations in the first category is needed. By rewriting the value $MTS_{j(i)k}$ from Table 9, we began choosing the process by the genetic algorithm method (Table 10).

This algorithm was written by MATLAB software (version R2015a, 2015). Tables 11 and 12 indicate the results of coding for T_{RN} = 4–16 [we had four subcategories because at least one remaining station in each subcategory (k) must be selected].

Table 12 presents the final results of the remaining combinations in the network for drinking usage. By using the value of $MTS_{j(i)k}$ (Table 10) and the same method, select stations for irrigation usage for the first category.

Table 8
Values of $TS_{j(i)k}$ or value index of each station for different usage

No. station	$TS_{j(i)k}$ (Drinking usage)	$TS_{j(i)k}$ (Irrigation usage)
1	26.86	16.42
2	21.16	16.39
3	20.42	15.42
4	28.57	20.71
5	21.91	20.18
6	23.41	19.81
7	41.86	30.45
8	39.88	32.42
9	18.27	13.70
10	19.03	14.31
11	27.32	20.58
12	77.61	82.95
13	48.27	35.52
14	30.18	21.92
15	36.76	26.33
16	71.72	65.77

3.2.2. Some suggested stations for Mond Basin Network

The values of the objective function (SMTS) or total value index, based on T_{RN} are plotted in Figs. 6a and b for irrigation and drinking usage. The objective of our study was to determine the minimum number of a combination of stations in the basin, which remind from the total number of stations. For this purpose, first, the incremental trend of indicators presents with the increasing number of stations in the network and then, the number of stations is judged according to the difference of indicators. As the figures show if T_{RN} value increase, drinking and irrigation indicators value or target function value (SMTS) increase too.

Now, we plot Figs. 7a and b, which indicates the difference for every two consecutive stations (ΔSMTS = $SMST_{i+1}$– $SMST_i$) for irrigation and drinking usage in each three station classifications. Based on Fig. 7a for irrigation usage, after T_{RN} = 8 the behaviors are similar and trends converge. This convergence will continue till the end. Therefore, after this number of stations, we can make decisions about the number of remaining stations and as it is clear, the maximum difference index value belongs to a combination of 11 and 12 stations. So that after 12 stations, with adding any

Table 9
Values of $MTS_{j(i)k}$ or maximum value of various combinations for different usage in first subcategories

Subcategory		$r_N(N, a)a$: $0 \rightarrow R_k$ k = 1:N = 4	Irrigation usage		Drinking usage	
			Stations combinations	$MTS_{j(i)k}$	Stations combinations	$MTS_{j(i)k}$
R_1; k = 1; P_2 = 4	1	r (1,1)	4	20.71	4	28.57
	2	r (1,2)	6–4	40.52	4–1	55.43
	3	r (1,3)	6–4–1	56.93	6–4–1	78.84
	4	r (1,4)	9–6–4–1	70.63	9–6–4–1	97.11
R_2; k = 2; P_2 = 7	1	r (2,1)	12	82.95	12	77.61
	2	r (2,2)	15–12	109.28	15–12	114.37
	3	r (2,3)	15–14–12	131.19	15–14–12	144.56
	4	r (2,4)	15–14–12–11	151.77	15–14–12–11	171.87
	5	r (2,5)	15–14–12–11–2	168.16	15–14–12–11–2	193.04
	6	r (2,6)	15–14–12–11–3–2	183.57	15–14–12–11–3–2	213.46
	7	r (2,7)	15–14–12–11–10–3–2	197.88	15–14–12–11–10–3–2	232.49
R_3; k = 3; P_3 = 3	1	r (3,1)	8	32.42	7	41.86
	2	r (3,2)	8–7	62.87	8–7	81.74
	3	r (3,3)	8–7–5	83.05	8–7–5	103.65
R_4; k = 4; P_4 = 2	1	r (4,1)	16	65.77	16	71.72
	2	r (4,2)	16–13	101.29	16–13	119.99

Table 10
Values of $MTS_{j(i)k}$ or maximum value of various combinations for different usage

R_k		Irrigation usage		Drinking usage	
		Stations combinations	$MTS_{j(i)k}$	Stations combinations	$MTS_{j(i)k}$
K_1	1	4	20.71	4	28.57
	2	6–4	40.52	4–1	55.43
	3	6–4–1	56.93	6–4–1	78.84
	4	9–6–4–1	70.63	9–6–4–1	97.11
K_2	1	12	82.95	12	77.61
	2	15–12	109.28	15–12	114.37
	3	15–14–12	131.19	15–14–12	144.56
	4	15–14–12–11	151.77	15–14–12–11	171.87
	5	15–14–12–11–2	168.16	15–14–12–11–2	193.04
	6	15–14–12–11–3–2	183.57	15–14–12–11–3–2	213.46
	7	15–14–12–11–10–3–2	197.88	15–14–12–11–10–3–2	232.49
K_3	1	8	32.42	7	41.86
	2	8–7	62.87	8–7	81.74
	3	8–7–5	83.05	8–7-5	103.65
K_4	1	16	65.77	16	71.72
	2	16–13	101.29	16–13	119.99

Table 11
Stations combinations based on irrigation index and the number of selected stations and values of the objective function (fitness) for first category (four subcategories for each station)

T_{RN}	Stations combinations	Fitness
4	16–12–8–4	201.85
5	16–13–12–8–4	237.36
6	16–13–12–8–7–4	267.82
7	16–15–13–12–8–7–4	294.15
8	16–15–14–13–12–8–7–4	316.06
9	16–15–14–13–12–11–8–7–4	336.64
10	16–15–14–13–12–11–8–7–5-4	356.82
11	16–15–14–13–12–11–8–7–6–5–4	376.63
12	16–15–14–13–12–11–8–7–6–5–4–1	393.05
13	16–15–14–13–12–11–8–7–6–5–4–2–1	409.44
14	16–15–14–13–12–11–8–7–6–5–4–3–2–1	424.85
15	16–15–14–13–12–11–10–8–7–6–5–4–3–2–1	439.16
16	16–15–14–13–12–11–10–9–8–7–6–5–4–3–2–1	452.86

Table 12
Stations combinations based on Drinking index and number of selected stations and values of the objective function (fitness) for first category (four subcategories for each station)

T_{RN}	Stations combinations	Fitness
4	16–12–7–4	219.75
5	16–13–12–7–4	268.02
6	16–13–12–8–7–4	307.90
7	16–15–13–12–8–7–4	344.66
8	16–15–14–13–12–8–7–4	374.85
9	16–15–14–13–12–11–8–7–4	402.17
10	16–15–14–13–12–11–8–7–4–1	429.03
11	16–15–14–13–12–11–8–7–6–4–1	452.44
12	16–15–14–13–12–11–8–7–6–5–4–1	474.35
13	16–15–14–13–12–11–8–7–6–5–4–2–1	495.51
14	16–15–14–13–12–11–8–7–6–5–4–3–2–1	515.94
15	16–15–14–13–12–11–10–8–7–6–5–4–3–2–1	534.97
16	16–15–14–13–12–11–10–9–8–7–6–5–4–3–2–1	553.23

station, the added amount to final objective function (SMTS) is ignorable. Then, the suggested optimal number of stations (minimum number of stations) in-network for irrigation usage is 12.

Based on Fig. 7b for drinking usage, after T_{RN} = 8 the behaviors are similar and trends converge which this convergence will continue until the end. Therefore, after this number of stations, we can make decisions about the number of remaining stations and as it clear, after this number of stations, the maximum difference index value belongs to a combination of 10 and 11 stations. So that after 11 stations, by adding any station, the amount of objective function did not increase considerably. Then the suggested optimal number of stations (minimum number of stations) in networks for the drinking usage is 11.

4. Discussion

Comparison our study with Icaga [10], who studied in the Gediz River in Turkey using the genetic algorithm method, indicates that some of the water quality monitoring stations can be optimized by using other parameters such as population living around the river, the area under cultivation, biological data, and effect of point source and nonpoint source pollution of rivers. However, we optimized some water

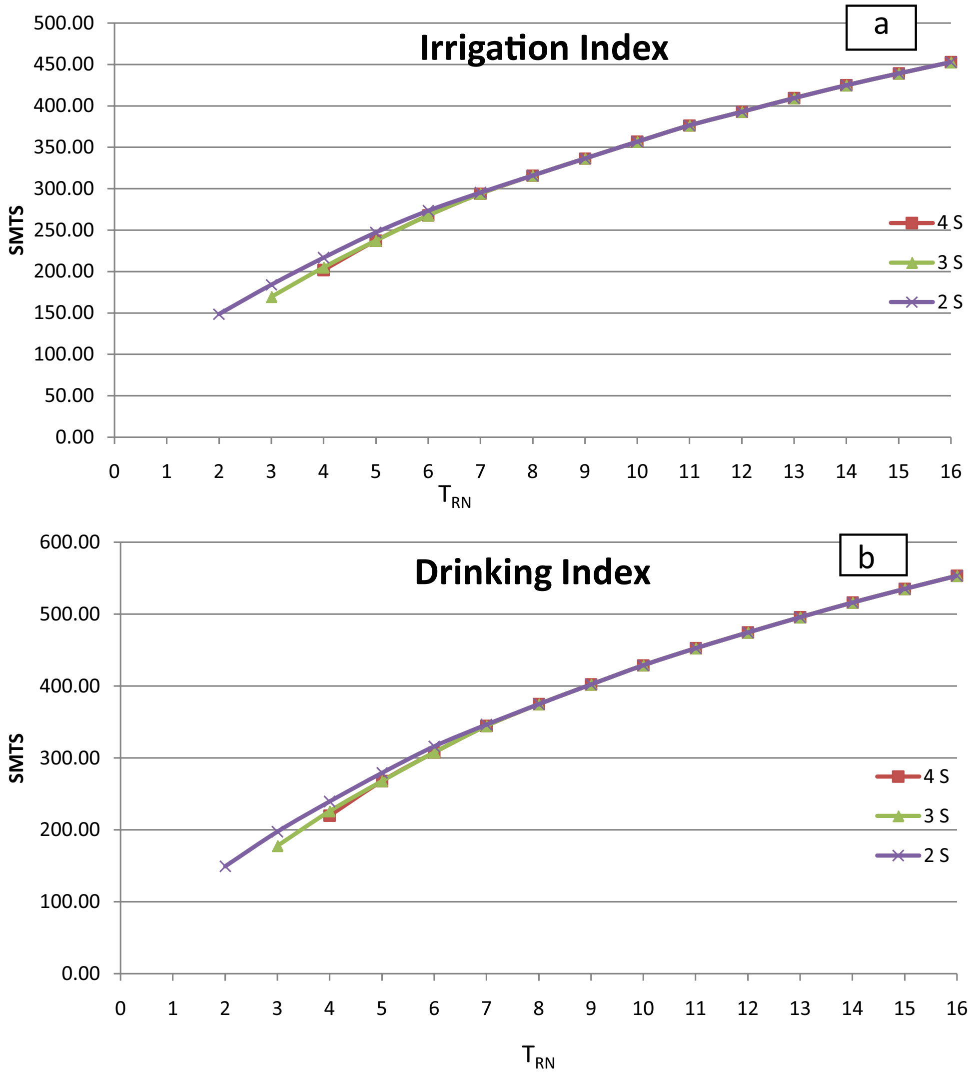

Fig. 6. Incremental irrigation index (a) and drinking index (b) value based on the incremental number of the remaining stations in the network for various classifications.

quality monitoring stations for drinking and irrigation usage applying AHP, Strahler stream ordering and GIS methods to find pollution index and then applied GA.

Comparison with Cetinkaya and Harmancioglu [7], who investigated the Gediz River in Turkey, which reduced the number of sampling stations by Dynamic Programming Approach, demonstrated using simple division of the basin with Sharp and Sanders methods for classification of stations. Therefore, in river systems based on waterways, first, found the center of gravity of basin and generated several subcategories through the division of basin according to the center of gravity of the basin. But in our study, the classification of stations was performed according to the location of stations on various rank rivers that we computed pollution index of the 16 monitoring stations using the AHP and Strahler stream ordering method. Also, in a study on the Gediz River, the other parameters were used such as the impact of population centers, drained areas, a period of use of stations and biological parameters, and heavy metals.

Results and the methods, which were used in our study, were compared with [5] and [6]. That both studies were related to the Sefid-Rud River network in the north of Iran using a genetic algorithm. Our method of generating pollution index for making objective functions is different from their method. We used average long-term physical and chemical water quality data (20 y) to generating pollution index which is a reliable criterion for judgment. However, in studying on the Sefid-Rud river, 2 y period data and four samples (with 6-month interval) were used to optimize several water quality stations. Also, we applied a weighting method based on the Paired Weighting method which is part of the AHP. However, in mentioned studies, independent weighting for goals was not used and instead of that, the weight of selected parameters in previous studies was used. Also, a method was presented to find the number of optimum stations, but previous studies determined a fixed number of stations to introducing a suitable number of the remaining stations in the basin.

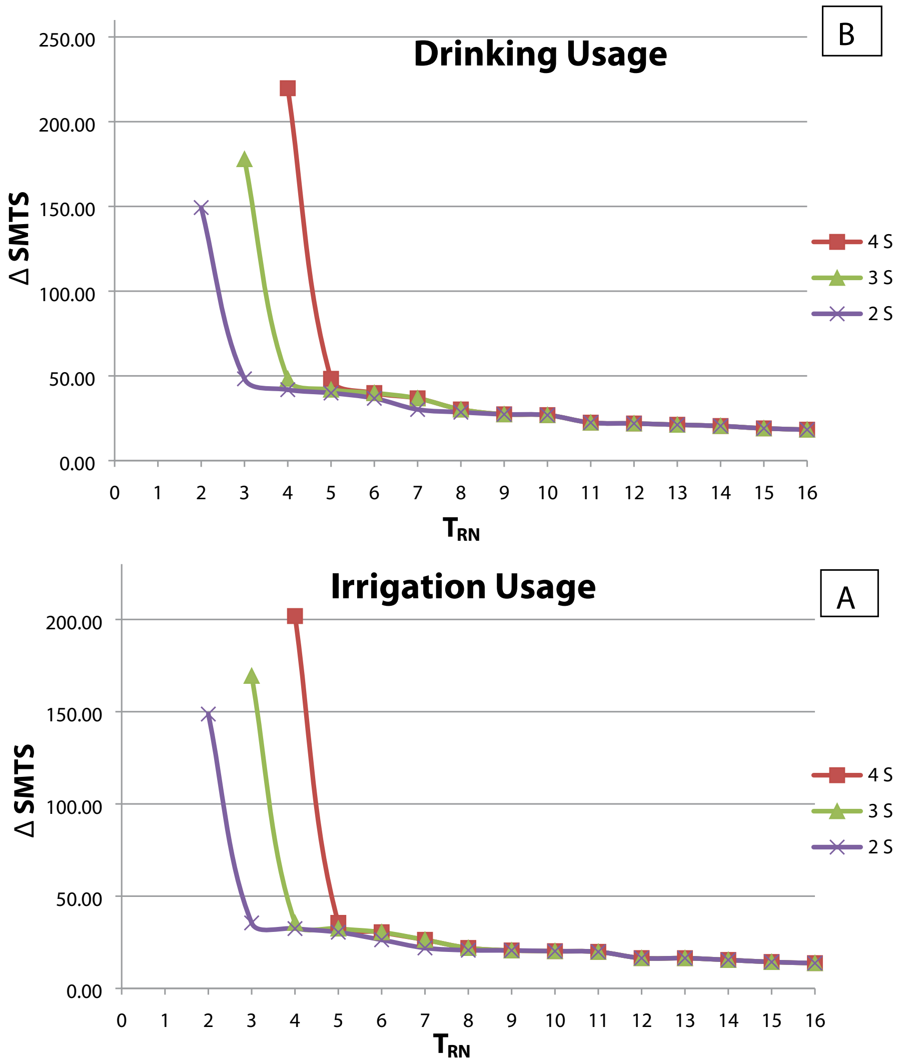

Fig. 7. Difference irrigation index (a) and drinking index (b) value for each two remained consecutive stations in the network for various classifications.

5. Conclusions

We achieved the following key results of the evaluation of available samples of the Mond River basin stations and using optimization method of genetic algorithm:

- Based on the Strahler stream order ranking method, station 12 with rank 4 has the maximum pollution index, 77.61 in terms of drinking usage and 82.95 in terms of irrigation usage.
- The Maximum pollution in various combinations of stations is related to the stations located on the rivers with rank 4. The pollution index of irrigation water is 197.88 and the pollution index of drinking water is 232.49.
- Reduction in the existing water quality monitoring stations from 16 to 12, in terms of irrigation usage, which means cutting redundant findings for maintaining 20% of the stations.
- Reduction in the existing water quality monitoring stations from 16 to 11, in terms of drinking usage, which means cutting redundant findings for maintaining 31% of the stations.
- By considering the combined effect of drinking and irrigation approaches, at least 12 stations will remain in-network such as station 16 (Qantareh), 15 (Baghan), 14 (Galou Bardekan), 13 (Dejgah), 12 (Dahrom), 11 (Dehroud), 8 (Hokan), 7 (Baba Arab), 6 (Sarvo), 5 (Tang Karzin), 4 (Borak), and 1 (Khanzenian).

References

[1] J. Li, C. Zhang, J. Lin, J. Yin, J. Xu, Y. Chen, Evaluating the bioavailability of heavy metals in natural-zeolite-amended aquatic sediments using thin-film diffusive gradients, Aquacult. Fish., 3 (2018) 122–128.

[2] L. Yin, Y. Ji, Y. Zhang, L. Chong, L. Chen, Rotifer community structure and its response to environmental factors in the Backshore Wetland of Expo Garden, Shanghai, Aquacult. Fish., 3 (2018) 90–97.

[3] A.R. Jafarabadi, A.R. Bakhtiari, M. Aliabadian, A.S. Toosi, Spatial distribution and composition of aliphatic hydrocarbons,

polycyclic aromatic hydrocarbons, and hopanes in superficial sediments of the coral reefs of the Persian Gulf, Iran, Environ. Pollut., 224 (2017) 195–223.
[4] T. Yu, Q. Feng, J. Si, X. Zhang, H. Xi, C. Zhao, Comparable water use of two contrasting riparian forests in the lower Heihe River basin, Northwest China, J. For. Res., 29 (2018) 1215–1224.
[5] G. Asadollahfardi, N. Heidarzadeh, A. Mosalli, A. Sekhavati, Optimization of water quality monitoring stations using genetic algorithm, a case study, Sefid-Rud River, Iran, Adv. Environ. Res., 7 (2018) 87–107.
[6] G. Asadollahfardi, M. Asadi, M. Nasrinasrabadi, A. Faraji, Dynamic programming approach (DPA) for assessment of water quality network: a case study of the Sefīd-Rūd River, Water Pract. Technol., 9 (2014) 135–149.
[7] C.P. Cetinkaya, N.B. Harmancioglu, Assessment of water quality sampling sites by a dynamic programming approach, J. Hydrol. Eng., 17 (2011) 305–317.
[8] B.J. Ritzel, J.W. Eheart, S. Ranjithan, Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, Water Resour. Res., 30 (1994) 1589–1603.
[9] D. Karpouzos, F. Delay, K. Katsifarakis, G. De Marsily, A multi-population genetic algorithm to solve the inverse problem in hydrogeology, Water Resour. Res., 37 (2001) 2291–2302.
[10] Y. Icaga, Genetic algorithm usage in water quality monitoring networks optimization in Gediz (Turkey) river basin, Environ. Monit. Assess., 108 (2005) 261–277.
[11] S.-Y. Park, J.H. Choi, S. Wang, S.S. Park, Design of a water quality monitoring network in a large river system using the genetic algorithm, Ecol. Modell., 199 (2006) 289–297.
[12] M. Karamouz, R. Kerachian, M. Akhbari, B. Hafez, Design of river water quality monitoring networks: a case study, Environ. Model. Assess., 14 (2009) 705.
[13] C. Lee, K. Paik, D.G. Yoo, J.H. Kim, Efficient method for the optimal placing of water quality monitoring stations for an ungauged basin, J. Environ. Manage., 132 (2014) 24–31.
[14] C.P. Liyanage, A. Marasinghe, K. Yamada, Comparison of optimized selection methods of sampling sites network for water quality monitoring in a river, Int. J. Affect. Eng., 15 (2016) 195–204.
[15] S. Shayan, G. Zare, M. Yamani, M. Sharifinia, M. Soltanpour, Analyses of the land morpho dynamic effect on the formation and expanding of Mond delta (Busher province), J. Nat. Geogr., 8 (2015) 39–56.
[16] S. Shyan G. Zare, M. Yamani, M. Sharifina, M. Soltanpour, Analysis of statistical changes of discharge and sediment catchment and its application in environmental planning, J. Appl. Geomorphol., 1 (2013) 37–50, (in Farsi).
[17] M.R. Sarvati, M. Mougholi, R. Shafei, A. Kasraiyan, The influence of Konar-Siyah salt dome on water and soil resources of Konar-Siyah plain (Firouzabad Fars province), J. Phys. Geogr., 5 (2012) 77–88.
[18] A. Darvishzadeh, Geology of Iran, Neda Publication, Tehran, 1991, pp. 1–901.
[19] N.B. Harmancioglu, C.P. Cetinkaya, P. Geerders, Transfer of Information Among Water Quality Monitoring Sites: Assessment by an Optimization Method, 18th International Conference Informatics for Environmental Protection, Geneva, Switzerland, 2004.
[20] M.J. Asgharpour, Multi-Criteria Decision Making, Vol. 120, Tehran University Publishing, Tehran, 2008.
[21] A. Rolim da Paz, C. Uvo, J. Bravo, W. Collischonn, H.R. da Rocha, Chapter 16 – Seasonal Precipitation Forecast Based on Artificial Neural Networks, H.A. do Prado, A.J.B. Luiz, H.C. Filho, Eds., Computational Methods for Agricultural Research: Advances and Applications, 2011, pp. 326–354.
[22] Y. Wind, T.L. Saaty, Marketing applications of the analytic hierarchy process, Manage. Sci., 26 (1980) 641–658.
[23] H. Ghodsipour, Analytical Hierarchy Process, Amirkabir University Pub, Tehran, 2008.
[24] M.E. Banihabib, M.H. Shabestari, Decision models for the ranking of agricultural water demand management strategies in an Arid region, Irrig. Drain., 66 (2017) 773–783.
[25] M. Chitsazan, G.R. Rahmani, M. Zaresefat, Hydrochemical assessment of ground resoureces for drinking by analytical hierarchy process (AHP) and fuzzy logic in GIS (Case study: Aghili Plain in northeast Khuzestan), J. Water Soil Conserv., 20 (2013) 131–146.
[26] T.L. Saaty, L.G. Vargas, Inconsistency and rank preservation, J. Math. Psychol., 28 (1984) 205–214.
[27] M. Ghodrati, Learning the Application of HEC-GeoHMS Model in ArcGIS 10, Simaye Danesh Book Pub, Tehran, 2014.
[28] M. Melani, An Introduction to Genetic Algorithms, A Bradford Book MIT Press, USA, 1998.
[29] A. Osyczka, Evolutionary Algorithms for Single and Multi-criteria Design Optimization, Physica-Verlag, Heidelberg, 2002.
[30] M. Affenzeller, S. Wagner, S. Winkler, A. Beham, genetic algorithms and Genetic Programming: Modern Concepts and Practical Applications, Chapman and Hall/CRC, London, 2009.
[31] P.A. Diaz-Gomez, D.F. Hougen, Initial Population for genetic algorithms: A Metric Approach, Proceedings of the 2007 International Conference on Genetic and Evolutionary Methods (GEM 2007), Las Vegas Nevada, USA, 2007, pp. 43–49.
[32] M. Karamouz, A. Ahmadi, A. Falahi, System Engineering, Amirkabir University of Technology, Tehran, 2006.
[33] S. Simonovic, Managing Water Resources: Methods and Tools for a Systems Approach. Earthscan, London and UNESCO Publishing, Paris, 2008.
[34] K. Sabaghalvani, genetic algorithm Based Discrete Optimum Design of Geometrically Non-Linear Plane and Space Trusses, Master of Engineering (Hons.) thesis, Faculty of Engineering, University of Wollongong, Australia, 2002, https://ro.uow.edu.au/theses/2535.
[35] D. Goldberg, genetic algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Boston, MA, 1989.