Chemometric Analysis of Surface Water Quality Data: Case Study of the Gorganrud River Basin, Iran

Roohollah Noori · Abdulreza Karbassi · Amir Khakpour · Mohammadreza Shahbazbegian · Hassan Mohammadi Khalf Badam · Mohammadreza Vesali-Naseh

Abstract Canonical correlation analysis (CCA), principal component analysis (PCA), and principal factor analysis (PFA) have been adopted to provide ease of understanding: interpretation of a large complex data set in the Gorganrud River monitoring networks, evaluation of the temporal and spatial variations of water quality, and finally identification of monitoring stations and parameters which are most important in assessing annual variations of water quality in the river. In accomplishing the research, 11 surface water quality data related to both of physical and chemical parameters have been collected from seven monitoring stations from 1996 to 2002. In general, our results from CCA method indicated strong relationship between physical and chemical parameters in the Gorganrud River. In addition, analyzing data through the PCA and PFA techniques revealed that all monitoring stations are important in explaining the annual variation of data set. From the point of view of the degree of importance of parameters contributing to water quality variations, further investigations by running two scenarios (rotated factor correlation coefficient value equal to 0.95 and 0.90 for the first and second scenarios, respectively) showed that the important parameters in one season may not be important for another season. For example, unlike in summer, water temperature, total suspended solids, total phosphorous, and nitrate parameters were important, electrical conductivity, and turbidity parameters had been realized as important parameters in spring through the first scenario.

1 Introduction

Surface water pollution by chemical, physical, and biological contaminants all over the world can be considered as a worldwide problem [1, 2]. Anthropogenic inputs such as municipal, industrial, and agricultural wastewater discharge and natural processes, i.e., weathering and soil erosion, are major factors determining the quality of the water resources. Many studies have been done on anthropogenic contamination of ecosystems [3, 4]. However, due to spatial and temporal variations in water quality, which are often difficult to interpret, a monitoring program providing a representative and reliable estimation of the quality of surface waters is necessary [5]. Literature demonstrated that chemometric data analysis methods such as canonical correlation analysis (CCA), principal component analysis (PCA), and principal factor analysis (PFA) are suitable techniques to achieve the goals. Through the CCA approach, Larson et al. [6] analyzed an 11-year-long measurement time series of waves and profiles from Duck North Carolina in order to determine covariability between waves and profile response. Gangopadhyay et al. [7] applied the PCA and PFA techniques to identify importance of monitoring wells predicting dynamic variations related to potentiometric head at a location in Bangkok, Thailand. Simeonov et al. [8] using PCA, clustering analysis (CA), and principal component regression interpreted a large and

R. Noori (✉) · A. Khakpour · M. Shahbazbegian · H. M. K. Badam
Department of Civil Engineering, Faculty of Engineering,
Bushehr Branch, Islamic Azad University,
Bushehr, Iran
e-mail: roohollahnoori@gmail.com

A. Karbassi · M. Vesali-Naseh
Department of Environmental Engineering,
Graduate Faculty of Environment, University of Tehran,
Tehran, Iran

complex data matrix of surface water parameters in northern Greece. However, water quality data set from alluvial region in northern India have been analyzed by means of PCA, discriminate analysis, and partial least squares in order to investigate the three parts: first compositional differences between surface and groundwater samples; second spatial variations in groundwater composition; and finally influence of natural and anthropogenic factors [9]. Sarbu and Pop [10] illustrated a data set concerning the water quality in the Danube River through a robust fuzzy PCA algorithm. Ouyang [11] adopted PCA and PFA to identify important water quality parameters in 22 stations located at the main stem of the lower St. Johns River in Florida, USA. Results revealed that total organic carbon, dissolved organic carbon, total nitrogen, dissolved nitrate and nitrite, orthophosphate, alkalinity, salinity, magnesium, and calcium were the most important parameters in assessing variations of water quality in the river. Through the PCA and geographical information system approaches, Terrado et al. [12] analyzed the main contamination sources of heavy metals, organic compounds, and other physicochemical parameters in Ebro River surface waters. Furthermore, they evaluated their temporal and spatial distributions. Noori et al. [13] applied PCA and PFA techniques for selecting the monitoring stations in assessing annual variations of river water quality. They selected eight monitoring stations, located at the Karoon River in Iran. Finally, authors suggested that PCA and PFA techniques were useful tools for identifying the importance of surface water quality monitoring stations. Sherestha and Kazama [14] applied CA, PCA, PFA, and discriminant analysis techniques to evaluate temporal and spatial variations and interpret a large complex water quality data set of the Fuji River basin. Liu et al. [15] applied CCA to investigate relationship between personal exposure to ten volatile organic compounds and biochemical liver tests. Noori et al. [16] proposed a multivariate statistical method, i.e., canonical correlation analysis for investigating the relationship between physical and chemical parameters of the Karoon River.

However, it is pointed out that previous studies try to identify annual variations of the water quality parameters in the water quality monitoring networks. It is clear that water quality parameters are affected by arid, semi-arid, and wet conditions; thus they can be different in each season of the year. So it is an important task to investigate the seasonal variations of the water quality in the monitoring networks. Hence, the research aims are to analyze seasonal variations of 11 physio-chemical parameters recorded in seven surface water quality monitoring stations for 7 years in the Gorganrud River, Iran, by means of PCA and PFA techniques. In addition, investigation of the relationship between physical and chemical parameters in the Gorganrud River is carried out using CCA method.

2 Materials and Methods

2.1 Case Study and Data

The Gorganrud River basin (between 54°00′ to 56°07′E and 36°36′ to 37°47′N) is located in Golestan province, northern part of Iran (Fig. 1). Gorganrud River originates from the Alborz Mountains and after passing from the residential, agricultural, and industrial areas flows down to the Caspian Sea. It has a catchment area of 10,200 km^2 and average annual rainfall of 500 mm. The main stream length of the Gorganrud River catchment is 350 km. The Increasing water withdrawal that leads to enhance wastewater discharge to the river endangered the aquatic life of this ecosystem. As a result, there is an increasing trend gap between current water quality and standard water quality. Agricultural and agro-industrial return flows, domestic wastewater of the cities–rural area, and industrial wastewaters are known as the main pollution sources of the surface and groundwater resources in the Gorganrud River basins.

In this study, 11 physio-chemical parameters related to seven monitoring stations are used for chemometric analysis (Table 1).

2.2 CCA Method

In some sets of multivariate data, the variables are divided naturally into two groups (i.e., response data and predictor variable). A canonical correlation analysis can then be used to investigate relationships between the two groups. As an exploratory tool, it is used as a data reduction method. The goal of CCA is to construct two new sets of canonical variates $U=\alpha X$ and $V=\beta Y$ that are linear combinations of the original variables such that the simple correlation between U and V is maximal, subject to the restriction that each canonical variate U and V has unit variance (to ensure uniqueness, except for sign) and is uncorrelated with other constructed variates within the set [17]. Assume that the $(p + q) \times (p + q)$ correlation matrix between the variables $X_1, X_2, \ldots, X_p$ and $Y_1, Y_2, \ldots, Y_q$ takes the following form when it is calculated from the sample for which the variables are recorded:

$$
\begin{array}{cc}
 & \begin{array}{cc} X_1\ X_2\ \ldots\ X_p & \quad Y_1\ Y_2\ \ldots\ Y_q \end{array} \\
\begin{array}{c} X_1 \\ X_2 \\ . \\ X_p \\ \\ Y_1 \\ Y_2 \\ . \\ Y_q \end{array} &
\left[\begin{array}{c|c}
\begin{array}{c} p \times p\ \text{matrix} \\ A \end{array} & \begin{array}{c} p \times q\ \text{matrix} \\ C \end{array} \\
\hline
\begin{array}{c} q \times p\ \text{matrix} \\ C' \end{array} & \begin{array}{c} q \times q\ \text{matrix} \\ B \end{array}
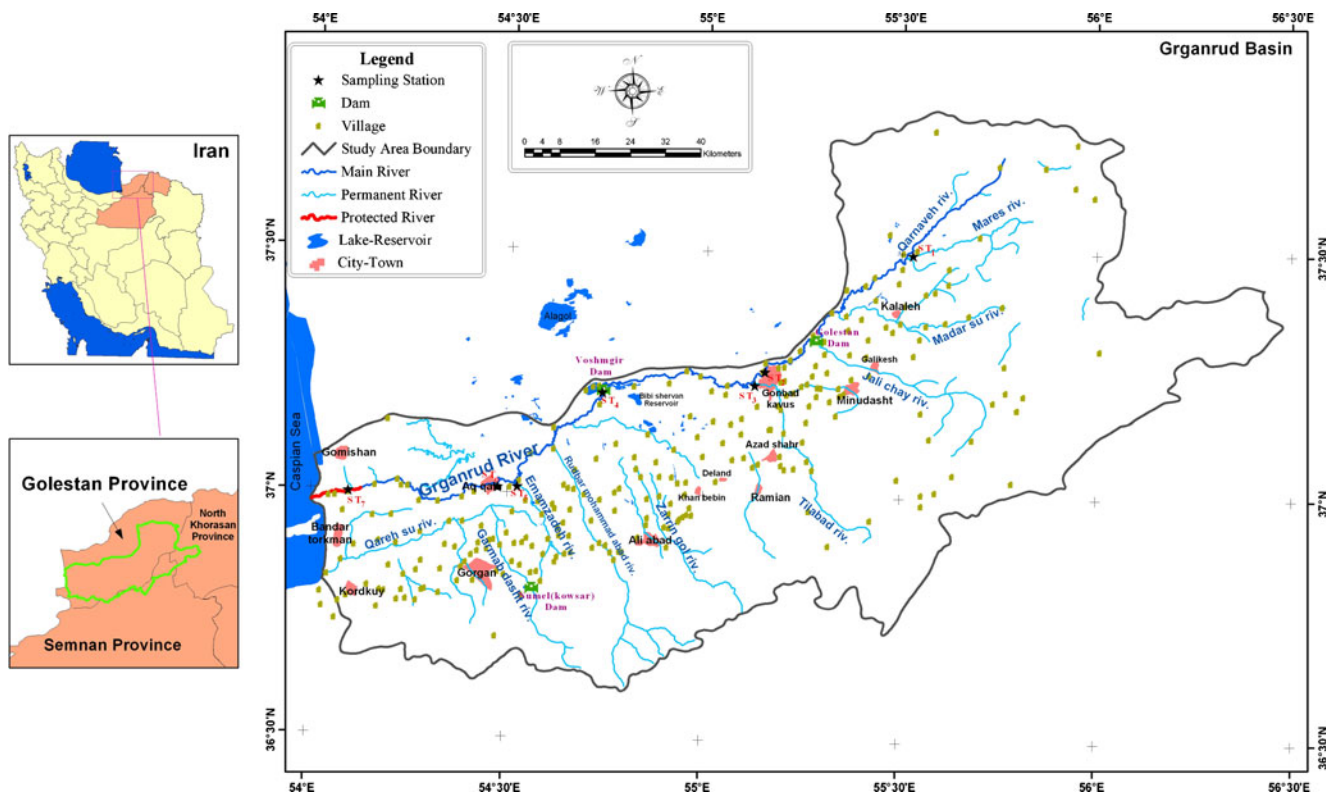\end{array} \right]
\end{array}
$$

Fig. 1 Territorial layout of the Gorganrud River basin and the location of the river sampling sites

From this matrix, a $q \times q$ matrix $\mathbf{B}^{-1}\mathbf{C}'\mathbf{A}^{-1}\mathbf{C}$ can be calculated, and the eigenvalue problem can be considered as [18]:

$$\left(\mathbf{B}^{-1}\mathbf{C}'\mathbf{A}^{-1}\mathbf{C} - \lambda\mathbf{I}\right)\mathbf{b} = 0 \qquad (2)$$

It turns out that the eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ are then the squares of the correlations between the canonical

variates. The r subscript is smaller than p and q. The corresponding eigenvectors $\mathbf{b}_1$, $\mathbf{b}_2$, ..., $\mathbf{b}_r$ give the coefficients of the Y variables for canonical variates. The coefficients of linear combination of X variables (U_i) and the ith canonical variate for the X variables are given by the elements of the $\mathbf{a}_i$ vector [19].

$$\mathbf{a}_i = \mathbf{A}^{-1}\mathbf{C}\mathbf{b}_i \qquad (3)$$

In these calculations, it is assumed that the original X and Y variables are in a standardized form with a mean of zero and standard deviation of unity. The coefficients of the canonical variates are for these standardized X and Y variables.

2.3 PCA and PFA

PCA and PFA are multivariate statistical methods which can be used for reducing complexity of input variables when there are large volumes of information and it is intended to have a better interpretation of variables [20, 21]. In mathematical terms, PCA and PFA involve the following five major steps: (1) start by coding the variables X_1, X_2,..., X_p to have zero means and unit variance; (2) calculate the

Table 1 Water quality parameters for surface water of the Gorganrud River basin

Parameter	Abbreviation	Unit
Water temperature	T	°C
Electrical conductance	EC	μmhos/cm
Dissolved oxygen	DO	mg/l
5-Day biochemical oxygen demand	BOD_5	mg/l
Chemical oxygen demand	COD	mg/l
pH	pH	pH units
Nitrate	NO_3^-	mg/l
Total phosphorus	TP	mg/l
Turbidity	Turb	NTU
Total dissolved solids	TDS	mg/l
Total suspended solids	TSS	mg/l

Table 2 Descriptive statistics of created PCs

PCs	Eigen value	Variance proportion	Cumulative variance proportion
PC1	3.40	48.59	48.59
PC2	2.20	31.35	79.94
PC3	1.36	19.49	99.43
PC4	0.04	0.57	100.00
PC5	2E-05	2.3E-04	100.00
PC6	3E-06	4.9E-05	100.00
PC7	3E-08	3.6E-07	100.00

correlation matrix R; (3) find the eigenvalues $\lambda_1, \lambda_2,\ldots, \lambda_p$ and the corresponding eigenvectors $a_1, a_2,\ldots, a_p$ by solving Eq. 4:

$$|R - I\lambda| = 0. \qquad (4)$$

(4) Discard any components that only account for a small proportion of the variation in datasets and (5) develop the factor loading matrix and perform a Varimax rotation on the factor loading matrix to infer the principal parameters [22]. Details for mastering the art of PCA and PFA are published elsewhere [23–25].

3 Results and Discussion

3.1 Relationship Between Physical and Chemical Parameters

According to Table 1 there are five variables in the response data set, i.e., physical parameters including T, DO, TDS, Turb, TSS, and six variables in the predictor set, i.e., chemical parameters including BOD_5, COD, NO_3^-, TP, EC, and pH. CCA results indicated that correlation coefficient for canonical variates 1, 2, and 3 were 0.94, 0.86, and 0.72, respectively. Correlation coefficients for fourth and fifth canonical variates were 0.38 and 0.45,

and then they were neglected in conclusion. Among the first three canonical variates, only the first and second canonical correlation was statistically significant (p value< 0.05). Therefore, there is no real evidence of any relationships between the physical and chemical variables based on canonical variate 3. It is pointed out that the first and second canonical variates represent the most variations in the response and predictor data set. Thus based on correlation coefficients of the first and second canonical variates, it can be concluded that a strong relationship between physical and chemical parameters exists in the Gorganrud River.

3.2 Identification of Important Monitoring Stations

Early correlation symmetrical matrix R is formed with dimensions 7×7 (equivalent to the number of input variables or stations) for PCA application. From solving Eq. 4, seven eigenvalues are obtained. Then for each of the eigenvalue, seven eigenvectors are calculated. Finally, using obtained eigenvectors, seven principal components (PCs) are computed. The characteristics of the PCs are presented in Table 2.

In this table, eigenvalues, variance proportion, and cumulative variance proportion are shown. Clearly, the first three components accounted approximately 48.59%, 31.35%, and 19.49% of the total variance in the data sets, respectively. These three components together

Table 3 Eigenvectors obtained through PFA application

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7
ST_1	0.242	0.065	0.968	0.006	−3E-05	−6E-05	−2E-06
ST_2	0.262	0.950	0.169	−0.006	0.001	−1E-04	−3E-05
ST_3	0.644	0.763	0.013	0.046	−0.002	−1E-04	−1E-04
ST_4	0.780	0.620	−0.040	−0.071	0.001	0.002	6E-07
ST_5	0.822	0.358	0.433	−0.090	0.002	0.000	1E-04
ST_6	0.905	0.292	0.307	−0.036	−0.002	−8E-04	−6E-05
ST_7	0.869	0.326	0.338	0.153	0.000	4E-05	2E-05

Table 4 Descriptive statistics of created PCs for spring and summer seasons

PCs	Spring			Summer		
	Eigenvalue	Variance proportion	Cumulative variance proportion	Eigenvalue	Variance proportion	Cumulative variance proportion
PC1	4.53	41.18	41.18	5.50	50.04	50.04
PC2	3.29	29.92	71.09	2.73	24.85	74.89
PC3	1.28	11.66	82.75	1.32	12.02	86.92
PC4	1.05	9.53	92.29	1.23	11.19	98.10
PC5	0.55	4.98	97.26	0.11	0.97	99.08
PC6	0.30	2.74	100.00	0.10	0.93	100.00
PC7	3E-16	3E-15	100.00	3E-15	3E-15	100.00
PC8	2E-16	2E-15	100.00	2E-16	2E-15	100.00
PC9	1E-16	1E-15	100.00	7E-17	6E-16	100.00
PC10	5E-17	5E-16	100.00	6E-17	5E-16	100.00
PC11	2E-17	2E-16	100.00	5E-18	5E-17	100.00

accounted for about 99.43% of the total variance and the rest only accounted for about 0.57%. Therefore, our discussions will focus only on the three components calculated as:

$$PC1 = 0.242ST_1 + 0.262ST_2 + 0.644ST_3 + 0.780ST_4 + 0.822ST_5 + 0.905ST_6 + 0.869ST_7$$
$$PC2 = 0.065ST_1 + 0.950ST_2 + 0.763ST_3 + 0.620ST_4 + 0.358ST_5 + 0.292ST_6 + 0.326ST_7 \qquad (5)$$
$$PC3 = 0.968ST_1 + 0.169ST_2 + 0.103ST_3 - 0.040ST_4 + 0.433ST_5 + 0.307ST_6 + 0.338ST_7$$

where ST_i is the monitoring station, the subscripts denote the station numbers, and the coefficients are the eigenvectors. PC1 (Eq. 5) indicated that there are difference between ST_1 and ST_2 coefficients and other coefficients. So the two coefficients have little effects on PC1 leading to realize that these stations are less important in monitoring water quality variations. In addition, based on the results of PC2, ST has

lowest absolute loading (eigenvector) values and a similar trend could be obtained for PC3. However, any conclusion based upon the PC1, PC2, and PC3 would be inappropriate since they only accounted for 48.59%, 31.35%, and 19.49% of the total variance, respectively. For determining the important water quality stations, a PFA technique should be established. In the PFA technique, similar to PCA, the

Table 5 Descriptive statistics of created PCs for autumn and winter seasons

PCs	Autumn			Winter		
	Eigenvalue	Variance proportion	Cumulative variance proportion	Eigenvalue	Variance proportion	Cumulative variance proportion
PC1	6.12	55.66	55.66	4.57	41.55	41.55
PC2	3.09	28.09	83.75	3.13	28.43	69.97
PC3	1.05	9.54	93.29	2.18	19.83	89.80
PC4	0.51	4.61	97.90	0.75	6.78	96.58
PC5	0.19	1.72	99.62	0.32	2.91	99.48
PC6	0.04	0.38	100.00	0.06	0.52	100.00
PC7	3E-15	2E-15	100.00	3E-16	3E-15	100.00
PC8	1E-16	1E-15	100.00	1E-16	1E-15	100.00
PC9	2E-17	2E-16	100.00	3E-18	3E-17	100.00
PC10	2E-18	2E-17	100.00	1E-18	9E-18	100.00
PC11	1E-18	9E-18	100.00	5E-19	5E-18	100.00

Table 6 PCs correlation coefficients for spring and summer seasons

Variables	Spring						Summer					
	PC1	PC2	PC3	PC4	PC5	PC6	PC1	PC2	PC3	PC4	PC5	PC6
T	0.979	0.077	−0.082	0.116	0.029	−0.123	0.563	0.681	0.050	0.448	−0.038	0.116
pH	0.213	0.840	−0.363	0.314	−0.065	−0.124	0.885	−0.401	−0.031	0.213	0.086	0.042
EC	0.400	0.120	0.805	0.381	−0.013	0.182	0.877	0.341	−0.240	−0.189	−0.043	0.136
Turb	−0.651	0.620	0.094	0.349	0.178	0.169	−0.787	0.610	−0.021	0.039	−0.052	−0.061
TDS	0.720	0.537	0.041	−0.414	0.105	0.092	0.774	0.602	0.166	0.000	0.040	−0.100
TSS	−0.958	0.262	0.055	−0.038	0.012	0.091	−0.732	0.659	−0.049	−0.145	−0.059	−0.059
DO	0.394	−0.414	−0.637	0.346	0.242	0.297	−0.509	0.710	−0.429	0.085	0.201	0.077
BOD$_5$	0.734	0.541	0.164	0.193	0.286	−0.153	−0.857	−0.074	0.381	0.269	−0.109	0.176
COD	−0.206	0.782	−0.059	−0.509	0.270	0.102	0.819	0.392	0.087	0.380	−0.118	−0.095
TP	−0.721	0.587	−0.158	0.282	−0.079	−0.160	−0.462	−0.290	−0.064	0.830	0.069	−0.070
NO$_3^-$	−0.499	−0.635	0.149	0.047	0.528	−0.211	0.040	0.277	0.943	−0.120	0.133	0.011

number of factors is equal to the number of variables. Table 3 shows the eigenvectors, which assess the coefficients for formation of factors. In the research, the correlation coefficient considered significant is the one that is greater than 0.75 (or >75%). The main reason of selecting the conservative criterion is that the study area (Gorganrud River basin) is large and the river system is highly nonlinear and dynamic. In addition, some researchers [11, 16] proposed approximately similar value which is used in this research. The stations with less rotated factor correlation coefficients than mentioned value are not considered principal stations. Table 3 indicated that all monitoring stations have coefficient values which are greater than 0.75. Therefore, to explain the annual variation of the data set, all water quality monitoring stations are considered important and thereby their location in the river system could be suitable.

3.3 Data Analysis Based on Seasonal Water Quality Parameters

Eleven variables related to water quality parameters have been used for each season. So there are four seasonal correlation symmetrical matrixes for spring, summer, autumn, and winter seasons. Similar to previous section, after solving Eq. 4 for correlation matrixes, the characteristics of 11 PCs for each season is calculated (Tables 4 and 5). In this section, according to PCA results, PCs with eigenvalues higher than 1 are selected, as a result, only four PCs for spring and summer and three for autumn and winter are allocated. The PCs indicated 92.29%, 98.10%, 93.29%, and 89.80% of total variance proportion of input variables in spring, summer, autumn, and winter seasons, respectively. In addition, eigenvectors are obtained through PCA application

Table 7 PCs correlation coefficients for autumn and winter seasons

| Variables	Autumn						Winter					
	PC1	PC2	PC3	PC4	PC5	PC6	PC1	PC2	PC3	PC4	PC5	PC6
T	−0.847	−0.241	0.287	0.367	−0.040	0.074	−0.177	0.640	0.662	0.338	0.049	−0.070
pH	−0.781	0.529	0.047	−0.283	0.160	0.047	−0.037	0.979	−0.156	0.091	0.081	−0.043
EC	0.787	0.040	0.584	−0.001	0.183	0.057	0.929	−0.084	0.258	0.195	−0.152	−0.046
Turb	0.907	−0.290	0.304	−0.023	−0.019	0.009	0.897	0.304	0.111	−0.300	0.034	0.024
TDS	0.328	0.869	0.178	0.318	−0.054	0.034	0.904	−0.027	0.407	0.050	0.058	−0.100
TSS	0.936	−0.334	0.039	−0.096	−0.033	−0.039	0.603	0.127	−0.539	0.496	−0.280	0.069
DO	−0.847	0.509	−0.093	−0.085	−0.074	0.036	−0.262	−0.800	0.463	0.252	0.111	0.041
BOD$_5$	0.667	0.693	−0.076	−0.257	0.032	0.055	0.653	−0.608	−0.166	0.269	0.321	0.017
COD	0.596	0.527	−0.460	0.328	0.214	−0.044	0.708	0.641	−0.130	−0.044	0.238	0.112
TP	−0.577	−0.766	−0.109	0.026	0.260	0.023	−0.686	0.462	0.493	0.248	0.039	0.098
NO$_3^-$	−0.719	0.443	0.510	−0.021	0.070	−0.146	−0.472	0.096	−0.825	0.218	0.177	−0.093
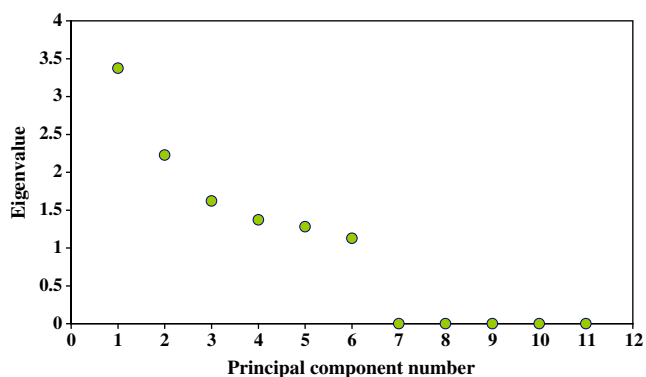
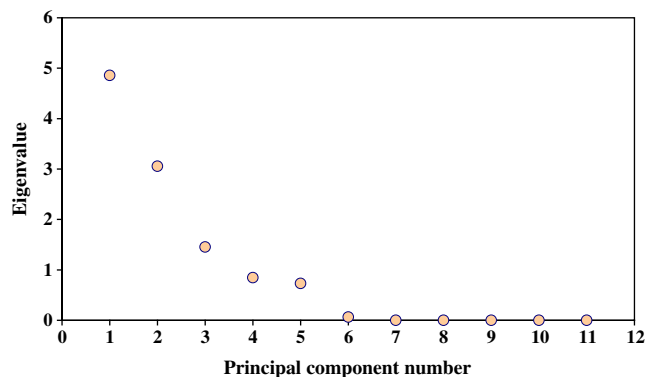**Fig. 2** Eigenvalues of principal factors in spring season

Fig. 4 Eigenvalues of principal factors in autumn season

(Tables 6 and 7) for each season. It should be pointed out that for retaining the PCs, a criterion equal to 10^{-6} is used. It resulted to six PCs for each season (Tables 6 and 7). In these tables, most effective variables to form the PCs are shown by bold font. Table 6 shows that T and TSS as two water quality parameters that have the highest absolute loading (eigenvector) values for the first component (PC1) in spring season. However, important parameters based on PC1 for summer are pH, EC, BOD_5, and COD. Furthermore, the important parameters for autumn and winter seasons are presented by bold font in Table 7. Similar to the previous section, any conclusion based on PC1 in all seasons would be inappropriate since they only accounted for 41.18%, 50.04%, 55.66%, and 41.55% of the total variance in spring, summer, autumn, and winter seasons, respectively. For example, in order to select the important parameter in spring season, although T is the most important parameter in formation of PC1, it has the lowest effect on formation of PC2 (0.077). Also, in the winter, although EC is the most important parameter based on PC1, it is one of the few parameters which affected PC2. Many details are available in Tables 6 and 7.

3.4 Extraction of Important Seasonal Water Quality Parameters

As demonstrated in the previous section, the PCA is not proper technique for extracting the important seasonal water quality parameters and it should be carried out by means of the PFA technique. Thus, using PFA method, results of the eigenvalues for each season are plotted in Figs. 2, 3, 4, and 5. Also, Tables 8 and 9 contain the eigenvectors or rotated factor correlation coefficients for each season. Similar to previous section, a criterion as 10^{-6} is used to retain the principal factors. Furthermore, an absolute rotated factor correlation coefficient value equal to 0.95 (or >95%) is considered for selecting the important parameter contributing to seasonal variations of the water quality of Gorganrud River. It is pointed out, if the value of this criterion is selected close to 1, the numbers of less importance stations or parameters increase. Therefore, due to negative impact of ignored stations is more than ignored parameters, the value of 0.95 was considered for choosing the principal seasonal water quality parameters. Besides, another scenario is run by the value of 0.90 for selecting the principal seasonal water quality parameters.
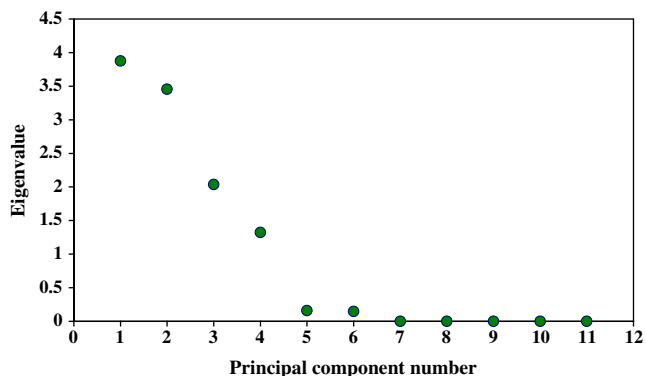

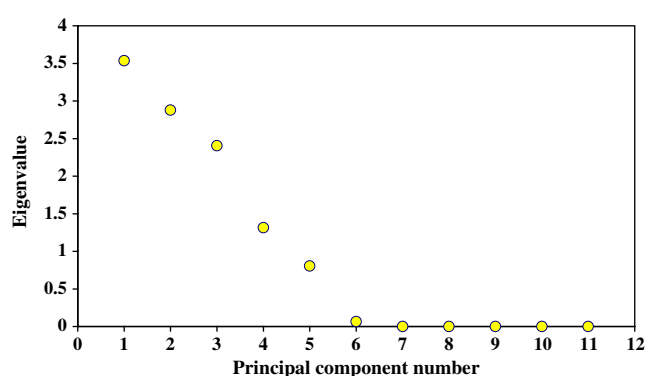**Fig. 3** Eigenvalues of principal factors in summer season

Fig. 5 Eigenvalues of principal factors in winter season

Table 8 Rotated factor (*F*.) correlation coefficients for spring and summer seasons

Variables	Spring						Summer					
	F. 1	*F*. 2	*F*. 3	*F*. 4	*F*. 5	*F*. 6	*F*. 1	*F*. 2	*F*. 3	*F*. 4	*F*. 5	*F*. 6
T	−0.573	0.697	0.316	−0.016	0.173	0.235	0.075	0.986	−0.013	0.059	0.076	0.109
pH	0.471	0.636	0.576	0.124	−0.132	0.091	−0.893	0.396	−0.087	−0.143	−0.082	0.102
EC	−0.079	0.230	0.058	−0.091	0.953	−0.145	−0.279	0.668	−0.649	−0.205	0.019	0.113
Turb	0.970	−0.036	−0.012	0.178	0.153	−0.050	0.967	−0.091	0.210	0.092	0.039	−0.051
TDS	−0.376	0.490	0.433	0.635	0.143	−0.088	−0.089	0.854	−0.434	0.218	−0.165	0.001
TSS	0.787	−0.486	−0.178	0.135	−0.155	−0.266	0.987	−0.113	0.013	0.092	0.030	−0.057
DO	−0.259	0.116	−0.017	−0.196	−0.160	0.925	0.902	0.137	0.065	−0.260	−0.063	0.304
BOD$_5$	−0.102	0.866	0.210	0.256	0.358	0.039	0.420	−0.439	0.633	0.346	0.329	−0.047
COD	0.383	0.101	0.158	0.854	−0.182	−0.238	−0.291	0.944	−0.092	0.033	−0.062	−0.101
TP	0.939	0.000	0.105	−0.042	−0.243	−0.214	0.028	−0.082	0.979	−0.177	−0.025	0.036
NO$_3^-$	0.049	−0.245	−0.946	−0.185	−0.085	0.031	0.041	0.141	−0.066	0.987	0.012	−0.017

According to Tables 8 and 9, for rotated factor correlation coefficient value equal to 0.95 (first scenario), the important parameters in contributing to water quality variations for one season may not be important for another season. The numbers of important variables in spring, summer, autumn, and winter seasons are 2, 5, 3, and 2 parameters, respectively. In contrast with other seasons, summer, and autumn seasons have the more important parameters because in these seasons, the Gorganrud River has the least amount of flow leading to deteriorate water quality of the river. However, water temperature parameter is one of the most important parameters in summer and winter seasons because in these seasons it affects water quality more than the other seasons. Furthermore, Table 8 denotes that in summer, TP and NO$_3^-$ are included in the important parameters. In the Gorganrud River basin, the most volumes of phosphate and nitrate fertilizers are commonly used in summer. In addition, in summer, activity of aquatic plants is very high. So the mentioned reasons cause TP and NO$_3^-$ to have more variations. Generally, the important parameters in the spring season are electrical conductivity and turbidity, while important parameters for summer season are water temperature, turbidity, total phosphorous, nitrate, and total suspended solids. However, the main parameters attributed to autumn are pH, total dissolved solids, and dissolved oxygen; and that attributed to winter are water temperature and dissolved oxygen.

In the second scenario, the rotated factor correlation coefficient value is selected to be 0.90. In this scenario, the numbers of important variables in spring, summer, autumn, and winter seasons achieved as five, seven, six, and six parameters, respectively. It concluded that the summer has the more important parameters. Generally, the important parameters in the spring season are electrical conductivity,

Table 9 Rotated factor (*F*.) correlation coefficients for autumn and winter seasons

Variables	Autumn						Winter					
	F. 1	*F*. 2	*F*. 3	*F*. 4	*F*. 5	*F*. 6	*F*. 1	*F*. 2	*F*. 3	*F*. 4	*F*. 5	*F*. 6
T	0.455	−0.379	0.779	−0.208	−0.019	−0.015	0.197	0.143	0.963	−0.076	−0.042	0.070
pH	0.986	0.010	−0.036	−0.130	0.094	−0.013	−0.173	0.836	0.500	0.103	−0.082	0.060
EC	−0.624	0.409	−0.104	−0.094	0.651	0.008	0.871	0.042	−0.132	0.425	0.195	0.064
Turb	−0.906	0.195	−0.187	−0.088	0.314	0.004	0.784	0.568	−0.226	0.047	0.089	−0.040
TDS	0.123	0.952	0.051	0.227	0.158	0.008	0.932	0.103	−0.065	0.156	0.289	0.095
TSS	−0.921	0.104	−0.350	0.024	0.133	0.012	0.106	0.306	−0.233	0.908	0.128	0.004
DO	0.964	0.057	0.131	−0.099	−0.200	−0.021	0.052	−0.952	0.059	−0.158	0.245	−0.052
BOD$_5$	−0.102	0.743	−0.616	0.170	0.157	−0.068	0.338	−0.250	−0.507	0.316	0.683	0.005
COD	−0.183	0.565	−0.260	0.759	−0.066	−0.006	0.418	0.843	0.018	0.183	0.265	−0.103
TP	0.098	−0.946	0.304	−0.006	−0.030	−0.035	−0.283	−0.074	0.903	−0.240	−0.175	−0.104
NO$_3^-$	0.787	0.140	0.343	−0.318	0.294	0.239	−0.918	0.212	−0.154	0.223	0.144	0.132

turbidity, dissolved oxygen, total phosphorous, and nitrate while important parameters for summer season are water temperature, turbidity, total phosphorous, nitrate, total suspended solids, dissolved oxygen, and chemical oxygen demand. However, the main parameters attributed to autumn are pH, turbidity, total dissolved solids, total suspended solids, dissolved oxygen, and total phosphorous; and those attributed to winter are water temperature, total dissolved solids, total suspended solids, dissolved oxygen, total phosphorous, and nitrate.

4 Conclusions

In this research, water quality of the Gorganrud River basin from 1996 to 2002 is evaluated. To achieve this goal, canonical correlation analysis, principal component analysis, and principal factor analyses are used. The following conclusions are drawn in the study through:

a. Generally, multivariate statistical techniques such as CCA, PCA, and PFA were effective tool for environmental quality assessment of the Gorganrud River.
b. CCA results indicated strong relationship between physical and chemical parameters in the Gorganrud River.
c. Results from the PFA technique showed that all water quality monitoring stations are considered important in explaining the annual variance of the data set, and thereby the location of them in the river system could be suitable.
d. In the first scenario (rotated factor correlation coefficient value equal to 0.95) the important parameters in the spring season were EC and Turb, while important parameters for summer season were T, Turb, TP, NO_3^-, and TSS. However, the main parameters attributed to autumn were pH, TDS, and DO; and that attributed to winter were T and DO.
e. In the second scenario (rotated factor correlation coefficient value equal to 0.90) the important parameters in the spring season were EC, Turb, DO, TP, and NO_3^- while important parameters for summer season were T, Turb, TP, NO_3^-, TSS, DO, and COD. However, the main parameters attributed to autumn are pH, Turb, TDS, TSS, DO, and TP; and that attributed to winter were T, TDS, TSS, DO, TP, and NO_3^-.
f. Generally, important parameters in contributing to water quality variations in the first and second scenario for one season may not be important for another season.
g. The presented methodology in this study can be a good tool for authorities in order to program the monitoring stations and water quality parameters.

References

1. Noori, R., Karbassi, A., Farokhnia, A., & Dehghani, M. (2009). Predicting the longitudinal dispersion coefficient using support vector machine and adaptive neuro-fuzzy inference system techniques. *Environmental Engineering Science, 26*(10), 1503–1510.
2. Noori, R., Karbassi, A. R., Mehdizadeh, H., Vesali-Naseh, M., & Sabahi, M. S. (2011). A framework development for predicting the longitudinal dispersion coefficient in natural streams using artificial neural network. *Environmental Progress & Sustainable Energy, 30*(3), 439–449.
3. Szymanowska, A., Samecka-Cymerman, A., & Kempers, A. J. (1999). Heavy metals in three lakes in West Poland. *Ecotoxicology and Environmental Safety, 43*(1), 21–29.
4. Issa, Y. M., Elewa, A. A., Rizk, M. S., & Hassouna, A. F. A. (1996). Distribution of some heavy metals in Qaroun lake and river Nile, Egypt, Menofiya. *Journal of Agricultural Research, 21*(5), 733–746.
5. Dixon, W., & Chiswell, B. (1996). Review of aquatic monitoring program design. *Water Research, 30*(9), 1935–1948.
6. Larson, M., Capobianco, M., & Hanson, H. (1999). Relationship between beach profiles and waves at Duck, North Carolina, determined by canonical correlation analysis. *Journal of Marine Geology, 163*(1–4), 275–288.
7. Gangopadhyay, S., Gupta, A. D., & Nachabe, M. H. (2001). Evaluation of ground water monitoring network by principal component analysis. *Ground Water, 39*(2), 181–191.
8. Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., et al. (2003). Assessment of the surface water quality in Northern Greece. *Water Research, 37*(17), 4119–4124.
9. Singh, K. P., Malik, A., Singh, V. K., Mohan, D., & Sinha, S. (2005). Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India. *Analytica Chimica Acta, 550*(1–2), 82–91.
10. Sarbu, C., & Pop, H. F. (2005). Principal component analysis versus fuzzy principal component analysis, a case study: The quality of Danube water (1985–1996). *Talanta, 65*(5), 1220–1225.
11. Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research, 39*(12), 2621–2635.
12. Terrado, M., Barcelo, D., & Tauler, R. (2006). Identification and distribution of contamination sources in the Ebro river basin by chemometrics modelling coupled to geographical information systems. *Talanta, 70*(4), 691–704.
13. Noori, R., Kerachian, R., Khodadadi, A., & Shakibayinia, A. (2007). Assessment of importance of water quality monitoring stations using principal component and factor analyses: A case study of the Karoon River. *Journal of Water & Wastewater, 63*(3), 60–69 (Persian).
14. Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software, 22*(4), 464–475.
15. Liu, J., Drane, W., Liu, W., & Wu, T. (2009). Examination of the relationships between environmental exposures to volatile organic compounds and biochemical liver tests: Application of canonical correlation analysis. *Journal of Environmental Research, 109*(2), 193–199.
16. Noori, R., Sabahi, M. S., Karbassi, A. R., Baghvand, A., & Taati-Zadeh, H. (2010). Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination, 260*(1–3), 129–136.

17. Facchinelli, A., Sacchi, E., & Mallen, L. (2001). Multivariate statistical and GIS-based approach to identify heavy metals sources in soils. *Environmental Pollution, 114*(3), 313–324.

18. Hotelling, H. (1936). Relation between two sets of variates. *Biometrica, 28*(3/4), 321–329.

19. Noori, R., Abdoli, M. A., Ameri, A., & Jalili-Ghazizade, M. (2009). Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of Mashhad. *Environmental Progress & Sustainable Energy, 28*(2), 249–258.

20. Noori, R., Karbassi, A. R., & Sabahi, M. S. (2010). Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste predicting. *Journal of Environmental Management, 91*(3), 767–771.

21. Noori, R., Abdoli, M. A., Jalili-Ghazizade, M., & Samifard, R. (2009). Comparison of ANN and PCA based multivariate linear regression applied to predict the weekly municipal solid waste generation in Tehran. *Iranian Journal of Public Health, 38*(1), 74–84.

22. Manly, B. F. J. (1986). *Multivariate statistical methods: A primer*. London: Chapman & Hall.

23. Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. London: Allyn and Bacon.

24. Noori, R., Ashrafi, K., & Ajdarpour, A. (2008). Comparison of ANN and PCA based multivariate linear regression applied to predict the daily average concentration of CO: A case study of Tehran. *Journal of the Earth and Space Physics, 34*(1), 135–152.

25. Noori, R., Khakpour, A., Omidvar, B., & Farokhnia, A. (2010). Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Systems with Applications, 37*(8), 5856–5862.