ISSN 0097-8078, Water Resources, 2023, Vol. 50, No. 1, pp. 152–167. © Pleiades Publishing, Ltd., 2023.



# Development of a Water Quality Index Using Sparse Principal Component Analysis for the Tigris River in Iraq

**Safaa H. Ali[a], Tyler Cook[b], Salam H. Ewaid[c], and Sanjeewa Gamagedara[d, *]**

[a] *Department of Chemistry and Physiology, College of Veterinary Medicine, University of Thi-Qar, Al-Shatrah, Thi-Qar, 64007 Iraq*

[b] *Department of Mathematics & Statistics, University of Central Oklahoma, Edmond, OK 73034 United States*

[c] *Al-Shatrah Technical Institute, Southern Technical University, Al-Shatrah, Thi-Qar 64007 Iraq*

[d] *Department of Chemistry, University of Central Oklahoma, Edmond, OK 73034 United States*

**e-mail: sgamagedara@uco.edu*

Received January 11, 2022; revised April 12, 2022; accepted July 14, 2022

**Abstract**—Freshwater levels in the Tigris River significantly reduced during the last two decades due to global warming and geopolitics issues around Iraq. Thus, continuous and regular assessment for water resources became critically essential and this study was designed to evaluate the water quality of Tigris River and to develop a novel Water Quality Index (WQI). The raw water (untreated) and drinking water (treated) samples were collected from twelve stations. Twenty parameters were assessed for each sample based on the standard methods including physical properties of water such as total dissolved solids, suspended solids, temperature, turbidity, PH, color, conductivity and also, chemical species such as $F^-$, $Cl^-$, $Na^+$, $K^+$, $Ca^{+2}$, $Mg^{+2}$, $Fe^{+3}$, $Al^{+3}$, $NH_3$, $SO_4^{-2}$, $PO_4^{-3}$, $SiO_2$, $NO_2^-$, $NO_3^-$. Using the above data a novel WQI was created using sparse principal component analysis. This sparse principal component WQI successfully identified a small subset of important variables that contribute to water quality.

**Keywords:** water quality index, sparse principal component, water quality parameters, tigris river, ion chromatography

**DOI:** 10.1134/S0097807823010037

## INTRODUCTION

Water is an essential natural bio-resource for all life forms [34]. About 97% of the earth's water is saline water in the oceans and 3% is fresh water contained in the poles (in the form of ice), groundwater, lakes, and rivers. Nearly, 70% of the world's fresh water is frozen in glaciers, permanent snow cover, ice, and permafrost [24]. The other Thirty percent of all freshwater is ground, most of it in deep, hard-to-reach aquifers. Lakes and rivers together contain just a little more than 0.25% of all freshwater; lakes contain most of it [10, 19]. There is a fewer number of permanent rivers with freshwater across the world which has about 0.01% of all water present on the earth [1]. Two of the major rivers in the world are located in Iraq namely Tigris and Euphrates. This project is focused on the Tigris river and it is the main sole source of supplying freshwater in Baghdad City, Iraq. Tigris River provides drinking water to 100% of the Baghdad city population. The water flow of the Tigris River has been declining for the last twenty years and is expected to a further decrease in the future. Thus, will lead to critical changes to the quality of water.

The water quality of the Tigris River is influenced by a range of natural variables that lead to change the physicochemical characteristics such as anthropogenic factors [20, 26], hydrological conditions, topography, and lithology, climate [12], precipitation inputs [20, 26], catchment area [12], tectonic and edaphic factors [20], erosion, weathering of crustal materials and bedrock geology [26]. But the fundamental factor is the rate of water flow from Turkey where the river source. Also, it is well known that water resources that pass through large cities such as the Tigris River in Baghdad are more likely to receive a large variety and quantity of pollutants because of the intensive and large-scale human and industrial activities [37]. As a result, large and significant impacts on the environment and human health due to leaking through the soil or directly discharging of these pollutants into water [27]. Most of the developing countries suffer from water pollution. About 80% of water pollution in these countries is a result of domestic waste and adequate sanitation infrastructures, which often results in the death of about two million infants annually [21, 38]. Thus, determining water quality quantitatively has been become a great concern [32] and a health priority for societies and governments [38].

The purpose of this study was to quantitatively evaluate the water quality status of the Tigris river for raw water (untreated) and drinking water (treated) in Baghdad City and to develop a novel Water Quality Index (**WQI**). A number of techniques currently exist for creating water quality indices. Originally, much of the work focusing on the development of water quality indices relied on expert opinion in selecting and weighting relevant parameters [2]. One of the earliest and most influential of these methods was Horton's index [2, 18, 36]. This WQI was based on the selection of 10 commonly measured variables believed to be important in water quality. Horton, however, did not elect to use any toxic chemicals in his WQI which demonstrates the potentially undesirable influence individual researchers can have on an index based on expert opinion. Numerous other WQI models with similar designs have been established by researchers throughout the world, and many of these methods allow for parameter selection based on locally available variables [36]. Some examples include the National Sanitation Foundation WQI, the Scottish Research Development Department WQI, the Canadian Council of Ministers of the Environment (**CCME**) WQI, and the Malaysian WQI [36]. In order to provide an integral, complex assessment of water quality in Russia, the Russian Federal Service for Hydrometeorology and Environmental Monitoring (RosHydroMet) implemented a water quality assessment method, and a comprehensive evaluation of water quality based on the data was first discussed in Nikanorov and Yemelyanova [28]. This method examines hydrodynamic, hydrobiological, and toxicological parameters that are flexible in allowing for the inclusion of important local parameters and can assess changes over time. The method proposed by Nikanorov and coworkers entails applying a set of 18 evaluation criteria to analyze and describe the state of the water body in question from various angles [28]. This comprehensive method enables users to unambiguously estimate the degree of water contamination using scalar values for a wide range of pollutants and water quality parameters. It also classifies the investigated water according to the degree of contamination and prepares the analytical information in an easy-to-understand format for regulatory agencies [28]. More recently, much attention has been given to creating water quality indices using more advanced statistical techniques for the analysis of multivariate data [6, 15]. Indices based on more advanced statistical methods have the benefit of relying on fewer assumptions rooted in expert opinion. For example, researchers have proposed methods based on elements from fuzzy logic [22, 23, 30], non-parametric probability distributions [29], and artificial neural networks [17]. In addition, several authors have explored the use of principal component analysis for constructing water quality indices [11, 31]. Here, we propose a new water quality index based on the popular machine learning technique sparse principal component analysis [40]. This method is an extension of traditional principal component analysis that aids in variable selection. A comprehensive overview detailing the development of methods for analyzing water quality can be found in Abbasi and Abbasi [2], and extensive reviews of recent advances can be found in Gupta and Gupta [16] and Uddin et al. [36].

## EXPERIMENTAL

### *Study Areas*

This study was carried out in Baghdad which is the capital city of Iraq with an area of about 1554 km$^2$. It is situated in the center of Iraq on latitude 34°–38° north and longitude 46°–43° east 600 m above the sea. The topography is relatively flat and it is the most densely populated city in Iraq [8] with a population of approximately 4 million per capita, according to the last residential census issued by the Ministry of Planning in 2015. This population increase has resulted in a growing demand for potable water, and thus, raises concerns about increasing potable water production with low specification control [3]. The city is characterized by two climatic seasons, namely the rainy seasons and the dry seasons. The rainy season extends from December to May while the dry seasons run from June to November.

This study was designed to carry out a quantitative assessment of some environmental toxic elements in the water of the Tigris River within Baghdad city, as a surface water resource. The data used in this study for all parameters are monthly averages collected from ten fixed stations in Baghdad city. These stations are supplied drinking water about 24 h per day for about 8 million residents of Baghdad city. The water supply system of Baghdad city is using these ten main pumping stations to supply the system, from the Tigris River at Baghdad city.

*Sampling*

Samples were taken from along the banks of the sampling stations between 2014–2016. Samples were collected in High-density PVC bottles (1 L capacity) which had been thoroughly washed and filled with deionized water and then taken to the sampling site. Each bottle was washed with deionized water before the next sample and then rinsed several times with the water to be collected and filled up to the brim and immediately sealed to avoid exposure to air [9] and the temperature was taken [33]. Samples were collected from surface water in the sampled location in 0.5 m depth [13]. The physiochemical water quality parameters were analyzed according to the International analytical standard methods as described by Greenberg and Clesceri [14]. All types of equipment were duly calibrated with standards and samples were analyzed in replicates.

*Determination Physico-Chemical Parameters*

The parameters such as temperature (**Temp**; °C), water pH, electrical conductivity (**EC**; μS/cm), total dissolved salts (**TS**, mg/L), suspended solids (**SS**, mg/L), and turbidity (**Tur**; NTU), were measured in-situ during sampling. Water pH was measured using a pH meter (HANNA, HI 9125) and EC, TS using a calibrated conductivity meter (HANNA, Conductivity meter). Turbidity measurements were conducted using a portable turbidity meter (LaMotte 2020E) [34]. Total alkalinity (**TA**) and water-soluble anions fluoride (mg/L $F^-$), chloride (mg/L $Cl^-$), nitrate (mg/L $NO_3^-$), nitrite (mg/L $NO_2^-$), sulphate (mg/L $SO_4^{-2}$), silica (mg/L $SiO_2$), phosphate (mg/L $PO_4^{-3}$), ammonia (mg/L $NH_3$), were determined using Ion Chromatography (IC) (Dionex™ ICS 2000). The IC analytical column and guard column are Dionex™ IonPac™ AS11-HC IC Columns (Thermo Fischer Scientific Inc). The cationic water-soluble constituents ($Na^+$, $K^+$, $Ca^{+2}$, $Mg^{+2}$, $Fe^{+3}$, $Al^{+3}$ in mg/L) were analyzed with a Dionex™ DX 500 system using a Dionex™ IonPac™ CS 12A analytical column. (Thermo Fischer Scientific Inc).

*Water Quality Index (WQI)*

A novel WQI was developed using sparse principal component analysis (**SPCA**). SPCA is an adaptation of a traditional principal component analysis (**PCA**) designed to ease the difficulty of interpreting the resulting loadings. With PCA, each independent variable will have a loading value on each principal component. This can make interpretation of the principal components difficult. Many researchers choose to apply a rotation to the PCA results in order to help identify the influential variables within each principal component. However, these rotation methods do not completely solve the variable selection problem. SPCA addresses this issue by creating principal components with sparse loadings. An SPCA will produce loadings that are exactly equal to zero for a subset of the independent variables [40]. This aids in the interpretation of the resulting principal components and can help perform variable selection. This can be particularly useful in cases where there are a large number of variables and it is of interest to identify a small subset of important variables contributing to water quality.

The SPCA-based WQI presented here is a variation on the WQI proposed in Fathy et al [11]. Here, the WQI is defined as:

$$\mathrm{WQI} = \sum_{i=1}^{k} \frac{\lambda_i}{\lambda} \mathrm{PC}_i,$$

where $k$ is the number of chosen sparse principal components, $\lambda_i$ is the eigenvalue for the $i$th chosen sparse principal component, $\lambda$ is the sum of all the eigenvalues, and $PC_i$ is the $i$th chosen sparse principal component score from the SPCA. Many approaches exist for deciding on the number of principal components to use. Popular choices include Kaiser's rule and inspecting scree plots. Here, we selected the first two principal components in order to increase the interpretability of the results.

The proposed WQI has several advantages over similar existing methods. First, as a primarily statistical method, it relies less on subjective expert opinion. Next, the aforementioned sparsity of SPCA produces a smaller set of important variables related to water quality. This makes the proposed WQI more accessible and understandable. As described in Abbasi and Abbasi [2], limiting the number of features to avoid unnecessary complexity was one of the key characteristics Horton argued for when developing his early WQI. Moreover, the proposed WQI is very flexible and can be adapted to handle any variables based on local availability or importance. Finally, Uddin et al.

[36] describe the structure of a WQI model based on a breakdown of four components: parameter selection, sub-indexing, parameter weighting, and aggregation. The unique mathematical features of SPCA result in a WQI that implements new approaches to both parameter weighting and aggregation since variable loadings of zero are not seen in other WQI methods that use traditional PCA.

## RESULTS AND DISCUSSION

Among all evaluated water quality parameters, the pH was used to assess the acidity and alkalinity since most aquatic species have a restricted pH range of 6–8. The average annual pH range for this study was ranged from 7.4 to 8.0. The alkalinity is typical for Iraqi rivers due to the natural existence of carbonates and bicarbonates. The water temperature of the Tigris river ranges from 7–38°C during the year of study due to the seasonal variations and slight differences among the stations observed during the same month. Turbidity measures cloudiness or haziness due to the dissolved solid and it decreases the light infiltration it affects photosynthesis and aquatic life. Also, high solid levels can increase the water temperature. Domestic sewage, agricultural waste, soil erosion, fertilizers are major sources of nitrates and phosphates. The world health organization **(WHO)** quality standard for nitrate and phosphate is ≤10 mg/L and 0.2 mg/L respectively. All the nitrates and phosphate levels are within the water quality standards except the phosphate level in station 10.

The main purpose of the WQI is to convert complex water quality data into a simple, understandable, and useable form for the public. The WQI provides a single number that represents the overall water quality at a certain location and time based on multiple water quality parameters. To begin, summary statistics of the quantitative variables were calculated for each of the twelve stations under study. These results are presented in Table 1 for untreated water and in Table 2 for treated water. Values of SS for Station 4 and values of aluminum for Station 10 and Station 12 were unable to be recorded for the untreated water. After treatment, values of $NO_3^-$ and $NH_3$ were not able to be measured for Station 4 from March through September.

The proposed WQI was then calculated for each station by month. This allowed for the comparison of water quality amongst the different stations over time. The sparse principal components were calculated in the R statistical software using the sparse Eigen package [7] after scaling the data. For untreated water, the resulting loadings for the first two sparse principal components can be found in Table 3. It is apparent that the SPCA procedure produced principal components that contain a small number of variables with nonzero loadings. In January, for example, the first sparse principal component had five variables with nonzero loadings and the second sparse principal component had three. For this month, the first principal component is a contrast between turbidity and $Cl^-$, EC, $SO_4^{-2}$, and TS. The maximum number of nonzero loadings for any principal component during any month was seven. In addition, there is some monthly change in the composition of the subset of variables that have nonzero loadings on the first two principal components. $NO_2^-$, for example, only appears in either of the principal components in November and December. The WQI was then recalculated using measurements from the treated water. Suspended solids had to be removed prior to this analysis because all of the recorded values were equal to zero and the lack of variability made SPCA impossible. The loadings corresponding to the WQI from treated water are reported in Table 4. These values again show the sparsity of significant variables used to calculate water treatment. There is an important difference relative to the untreated values because many of the specific variables selected each month have changed. Also, the loadings change for the variables that are represented both before and after treatment. Comparing untreated and treated results can give insight into the efficacy of the water treatment as one can identify which variables remain important in calculating the WQI.

A plot of the first two sparse principal component scores calculated from untreated water at each station separated by month can be seen in Fig. 1. The plot exhibits significant clustering for several of the months. This could suggest that the resulting WQI values will display a seasonal variation. The large range of values taken by the principal component scores from month-to-month reflects the changing groups of variables that have nonzero loadings. A second plot of the principal component scores calculated from water measured after treatment is in Fig. 2. Again, monthly clustering can be seen. However, the overall structure is quite different when compared to values from treated water. This is expected since the groups of variables with nonzero loadings changes after treatment.

**Table 1.** Summary statistics for each water quality parameter calculated by the stations for untreated water. Values reported are the mean (top) and standard deviation (bottom)

| Station | Temp | Tur | TA | Hard | $Ca^{+2}$ | $Cl^-$ | $Mg^{+2}$ | pH | EC | $SO_4^{-2}$ | TS | SS | $Fe^{+3}$ | $F^-$ | $Al^{+3}$ | $NO_2^-$ | $NO_3^-$ | $NH_3$ | $SiO_2$ | $PO_4^{-3}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 23.000 | 136.667 | 126.125 | 280.000 | 65.875 | 46.958 | 27.458 | 7.888 | 710.417 | 117.500 | 461.250 | 70.417 | 7.496 | 0.089 | 0.010 | 0.004 | 0.881 | 0.012 | 3.929 | 0.040 |
| | 5.543 | 219.217 | 6.806 | 26.853 | 5.753 | 10.105 | 3.230 | 0.053 | 90.440 | 42.131 | 61.980 | 50.484 | 22.840 | 0.020 | 0.000 | 0.003 | 0.249 | 0.004 | 0.408 | 0.004 |
| 2 | 21.667 | 100.042 | 147.500 | 301.833 | 72.250 | 67.500 | 29.667 | 8.038 | 835.00 | 204.208 | 559.042 | 147.083 | 19.078 | 0.129 | 0.010 | 0.006 | 0.940 | 0.072 | 2.100 | 0.010 |
| | 6.534 | 110.069 | 7.954 | 31.721 | 6.051 | 14.487 | 4.589 | 0.041 | 122.918 | 24.324 | 82.270 | 180.549 | 58.714 | 0.012 | 0.000 | 0.002 | 0.412 | 0.103 | 0.884 | 0.000 |
| 3 | 20.667 | 80.792 | 146.833 | 310.958 | 71.792 | 77.167 | 32.250 | 7.895 | 833.250 | 205.417 | 576.750 | 116.125 | 13.734 | 0.127 | 0.010 | 0.009 | 0.536 | 0.112 | 5.642 | 0.010 |
| | 5.698 | 92.405 | 11.054 | 34.706 | 5.533 | 14.656 | 4.565 | 0.134 | 133.832 | 39.373 | 92.189 | 128.130 | 41.670 | 0.028 | 0.000 | 0.002 | 0.071 | 0.021 | 0.844 | 0.000 |
| 4 | 22.292 | 72.375 | 152.333 | 325.417 | 79.167 | 74.000 | 31.125 | 7.468 | 860.333 | 182.875 | 586.583 | – | 0.186 | 0.062 | 0.010 | 0.008 | 0.360 | 0.200 | 3.275 | 0.013 |
| | 4.952 | 42.357 | 10.599 | 46.975 | 12.886 | 18.474 | 4.647 | 2.352 | 147.172 | 54.046 | 103.348 | – | 0.154 | 0.036 | 0.000 | 0.004 | – | 0.000 | 1.274 | 0.012 |
| 5 | 21.750 | 118.583 | 156.667 | 315.667 | 75.750 | 66.583 | 31.167 | 7.960 | 854.167 | 195.375 | 572.625 | 99.958 | 13.672 | 0.160 | 0.010 | 0.006 | 0.938 | 0.118 | 3.825 | 0.022 |
| | 6.638 | 152.860 | 11.005 | 40.820 | 12.289 | 16.209 | 3.576 | 0.022 | 130.103 | 55.350 | 87.063 | 109.997 | 38.042 | 0.038 | 0.000 | 0.002 | 0.276 | 0.030 | 0.383 | 0.009 |
| 6 | 22.542 | 104.667 | 156.958 | 329.917 | 78.292 | 71.250 | 33.833 | 8.021 | 838.583 | 236.5042 | 625.417 | 159.542 | 21.929 | 0.165 | 0.015 | 0.004 | 1.041 | 0.019 | 4.892 | 0.058 |
| | 7.014 | 118.082 | 7.653 | 38.683 | 12.952 | 13.815 | 4.218 | 0.040 | 134.561 | 40.312 | 61.717 | 180.456 | 72.926 | 0.022 | 0.000 | 0.001 | 0.373 | 0.003 | 0.520 | 0.013 |
| 7 | 22.750 | 61.792 | 162.875 | 320.333 | 77.625 | 73.458 | 31.125 | 7.796 | 881.583 | 212.208 | 573.208 | 85.542 | 9.238 | 0.122 | 0.035 | 0.020 | 0.812 | 0.074 | 5.021 | 0.044 |
| | 5.101 | 46.828 | 12.782 | 41.359 | 11.144 | 16.033 | 4.401 | 0.040 | 125.522 | 50.203 | 81.750 | 79.423 | 27.653 | 0.048 | – | 0.009 | 0.369 | 0.051 | 0.563 | 0.016 |
| 8 | 21.875 | 91.958 | 160.333 | 314.167 | 80.250 | 70.500 | 30.250 | 7.754 | 847.083 | 245.208 | 558.292 | 196.167 | 36.011 | 0.102 | 0.029 | 0.004 | 0.982 | 0.056 | 3.208 | 0.091 |
| | 6.183 | 72.492 | 11.216 | 43.828 | 11.458 | 15.448 | 4.707 | 0.050 | 141.752 | 50.951 | 94.572 | 146.738 | 118.413 | 0.015 | 0.006 | 0.002 | 0.448 | 0.039 | 0.394 | 0.018 |
| 9 | 21.083 | 83.542 | 160.292 | 328.917 | 81.583 | 76.958 | 30.625 | 7.963 | 888.042 | 207.583 | 579.250 | 55.875 | 12.243 | 0.103 | 0.010 | 0.009 | 0.747 | 0.434 | 5.067 | 0.069 |
| | 5.575 | 106.129 | 12.237 | 43.856 | 10.927 | 17.730 | 5.091 | 0.196 | 158.990 | 46.288 | 104.060 | 34.155 | 37.405 | 0.017 | 0.000 | 0.004 | 0.254 | 0.144 | 0.703 | 0.023 |
| 10 | 22.708 | 110.000 | 158.875 | 310.375 | 76.958 | 70.375 | 28.250 | 7.803 | 883.542 | 187.833 | 592.000 | 137.708 | 10.700 | 0.095 | – | 0.012 | 0.768 | 0.250 | 4.942 | 0.114 |
| | 6.917 | 147.464 | 11.605 | 42.096 | 12.682 | 15.501 | 3.659 | 0.098 | 145.612 | 48.757 | 97.553 | 117.386 | 31.603 | 0.026 | – | 0.005 | 0.330 | 0.131 | 0.556 | 0.265 |
| 11 | 20.625 | 110.958 | 159.000 | 309.667 | 77.375 | 71.417 | 28.917 | 7.877 | 864.958 | 192.292 | 602.708 | 134.625 | 8.910 | 0.079 | 0.015 | 0.013 | 0.731 | 0.245 | 4.871 | 0.042 |
| | 6.789 | 137.248 | 11.107 | 41.205 | 11.863 | 14.582 | 3.417 | 0.087 | 221.202 | 44.356 | 105.819 | 120.443 | 26.643 | 0.035 | 0.010 | 0.004 | 0.275 | 0.114 | 0.796 | 0.023 |
| 12 | 23.208 | 44.708 | 142.500 | 311.958 | 78.083 | 73.083 | 28.583 | 7.947 | 852.333 | 207.083 | 570.958 | 79.708 | 7.967 | 0.178 | – | 0.010 | 0.889 | 0.420 | 5.158 | 0.041 |
| | 6.013 | 39.245 | 10.768 | 44.041 | 10.935 | 15.018 | 4.762 | 0.067 | 129.633 | 53.342 | 86.868 | 89.037 | 21.292 | 0.016 | – | 0.004 | 0.333 | 0.231 | 0.512 | 0.035 |

**Table 2.** Summary statistics for each water quality parameter calculated by the stations for treated water. Values reported are the mean (top) and standard deviation (bottom)

| Station | Temp | Tur | TA | Hard | $Ca^{+2}$ | $Cl^-$ | $Mg^{+2}$ | pH | EC | $SO_4^{-2}$ | TS | SS | $Fe^{+3}$ | $F^-$ | $Al^{+3}$ | $NO_2^-$ | $NO_3^-$ | $NH_3$ | $SiO_2$ | $PO_4^{-3}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 21.667 | 2.183 | 121.833 | 280.750 | 66.208 | 48.208 | 27.500 | 7.483 | 713.333 | 125.833 | 464.583 | 0.000 | 0.082 | 0.090 | 0.121 | 0.001 | 0.981 | 0.010 | 4.133 | 0.040 |
| | 5.754 | 0.729 | 6.877 | 24.849 | 5.163 | 10.192 | 3.233 | 0.103 | 88.839 | 44.202 | 60.919 | 0.000 | 0.017 | 0.010 | 0.032 | 0.000 | 0.240 | 0.000 | 0.434 | 0.004 |
| 2 | 21.292 | 3.021 | 138.708 | 301.417 | 71.667 | 66.958 | 29.750 | 7.646 | 832.708 | 206.500 | 559.125 | 0.000 | 0.142 | 0.096 | 0.132 | 0.002 | 0.973 | 0.010 | 1.925 | 0.010 |
| | 6.781 | 0.454 | 6.649 | 28.745 | 5.154 | 12.860 | 4.798 | 0.037 | 122.307 | 25.417 | 74.134 | 0.000 | 0.044 | 0.014 | 0.028 | 0.001 | 0.253 | 0.000 | 0.878 | 0.000 |
| 3 | 22.833 | 2.329 | 138.833 | 311.292 | 72.917 | 78.667 | 32.750 | 7.620 | 838.708 | 207.208 | 572.250 | 0.000 | 0.089 | 0.063 | 0.122 | 0.001 | 0.565 | 0.051 | 5.658 | 0.010 |
| | 6.834 | 0.308 | 12.289 | 38.238 | 6.663 | 17.615 | 4.454 | 0.158 | 148.879 | 42.763 | 97.088 | 0.000 | 0.015 | 0.031 | 0.031 | 0.000 | 0.073 | 0.005 | 0.874 | 0.000 |
| 4 | 22.333 | 3.300 | 142.000 | 330.583 | 79.875 | 78.500 | 32.208 | 7.753 | 871.667 | 195.208 | 594.167 | 0.000 | 0.076 | 0.040 | 0.083 | 0.003 | 0.635 | 0.136 | 3.725 | 0.010 |
| | 4.731 | 0.750 | 9.427 | 49.302 | 13.053 | 19.820 | 5.079 | 0.132 | 137.600 | 55.570 | 97.497 | 0.000 | 0.047 | 0.018 | 0.017 | 0.003 | 0.236 | 0.125 | 1.361 | 0.000 |
| 5 | 22.542 | 3.142 | 145.625 | 317.875 | 79.583 | 69.875 | 31.167 | 7.509 | 867.625 | 196.958 | 581.458 | 0.000 | 0.088 | 0.089 | 0.090 | 0.001 | 0.975 | 0.020 | 3.812 | 0.015 |
| | 6.827 | 1.322 | 11.376 | 39.260 | 11.291 | 18.530 | 3.762 | 0.056 | 139.453 | 56.084 | 93.348 | 0.000 | 0.040 | 0.039 | 0.021 | 0.000 | 0.299 | 0.000 | 0.431 | 0.006 |
| 6 | 22.625 | 2.742 | 153.042 | 334.250 | 78.875 | 71.375 | 33.917 | 7.725 | 842.708 | 239.833 | 624.292 | 0.000 | 0.128 | 0.150 | 0.068 | 0.001 | 0.983 | 0.010 | 4.154 | 0.045 |
| | 7.349 | 0.415 | 8.286 | 41.820 | 13.255 | 13.373 | 4.226 | 0.050 | 136.665 | 38.997 | 66.168 | 0.000 | 0.072 | 0.017 | 0.006 | 0.000 | 0.339 | 0.000 | 0.573 | 0.027 |
| 7 | 22.833 | 1.554 | 135.958 | 319.042 | 78.083 | 80.208 | 31.167 | 7.029 | 876.042 | 217.250 | 570.750 | 0.000 | 0.033 | 0.070 | 0.056 | 0.004 | 0.668 | 0.014 | 4.342 | 0.034 |
| | 5.232 | 0.556 | 11.091 | 45.946 | 12.360 | 16.530 | 4.668 | 0.058 | 135.377 | 53.797 | 88.948 | 0.000 | 0.027 | 0.026 | 0.019 | 0.002 | 0.265 | 0.009 | 0.757 | 0.026 |
| 8 | 21.833 | 3.696 | 152.375 | 317.417 | 82.500 | 72.708 | 30.250 | 7.525 | 851.833 | 245.750 | 562.208 | 0.000 | 0.055 | 0.095 | 0.086 | 0.001 | 1.032 | 0.017 | 3.146 | 0.060 |
| | 6.202 | 0.821 | 11.027 | 39.580 | 10.311 | 16.155 | 5.101 | 0.045 | 128.339 | 49.648 | 84.687 | 0.000 | 0.027 | 0.014 | 0.018 | 0.000 | 0.469 | 0.008 | 0.434 | 0.015 |
| 9 | 21.125 | 3.354 | 149.000 | 325.208 | 80.583 | 75.792 | 30.125 | 7.460 | 878.625 | 197.208 | 572.542 | 0.000 | 0.067 | 0.100 | 0.106 | 0.001 | 1.026 | 0.014 | 4.629 | 0.038 |
| | 5.909 | 0.436 | 11.743 | 41.292 | 12.041 | 16.977 | 4.986 | 0.183 | 161.008 | 41.020 | 104.020 | 0.000 | 0.033 | 0.017 | 0.025 | 0.000 | 0.335 | 0.010 | 0.693 | 0.016 |
| 10 | 22.750 | 2.121 | 146.708 | 308.083 | 77.625 | 72.292 | 28.250 | 7.350 | 894.625 | 196.125 | 599.500 | 0.000 | 0.100 | 0.026 | 0.042 | 0.002 | 1.066 | 0.095 | 4.729 | 0.032 |
| | 7.172 | 1.245 | 10.648 | 44.684 | 13.106 | 15.397 | 3334 | 0.098 | 148.798 | 49.453 | 99.762 | 0.000 | 0.085 | 0.009 | 0.016 | 0.001 | 0.352 | 0.121 | 0.507 | 0.017 |
| 11 | 21.625 | 3.696 | 146.333 | 312.667 | 78.792 | 72.208 | 28.083 | 7.340 | 900.667 | 198.167 | 606.375 | 0.000 | 0.127 | 0.028 | 0.071 | 0.002 | 0.899 | 0.092 | 4.592 | 0.029 |
| | 7.164 | 0.884 | 10.393 | 72.390 | 11.610 | 14.905 | 3.741 | 0.086 | 147.929 | 43.207 | 97.570 | 0.000 | 0.170 | 0.015 | 0.022 | 0.001 | 0.275 | 0.117 | 0.607 | 0.017 |
| 12 | 23.458 | 2.812 | 128.917 | 310.500 | 78.083 | 75.750 | 28.500 | 7.402 | 851.917 | 210.833 | 570.917 | 0.000 | 0.093 | 0.053 | 0.092 | 0.001 | 1.010 | 0.010 | 4.858 | 0.016 |
| | 6.144 | 0.492 | 10.475 | 43.953 | 10.615 | 15.796 | 4.880 | 0.140 | 133.748 | 53.158 | 89.548 | 0.000 | 0.044 | 0.014 | 0.020 | 0.001 | 0.347 | 0.000 | 0.527 | 0.010 |

**Table 3.** Untreated water monthly loadings for each variable from the first two sparse principal components. Each month is recorded in a pair of columns with the first column representing the first sparse principal component (PC1) and the second column showing values for the second sparse principal component (PC2)

| | Jan | | Feb | | Mar | | Apr | | May | | Jun | | Jul | | Aug | | Sep | | Oct | | Nov | | Dec | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 |
| Temp | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| Tur | 0.43 | – | – | 0.53 | – | – | – | – | – | – | – | –0.56 | – | – | –0.44 | – | – | – | – | 0.54 | – | – | – | – |
| TA | – | – | –0.51 | – | – | – | – | – | –0.41 | – | –0.41 | – | –0.56 | – | – | –0.67 | –0.47 | – | –0.36 | – | – | – | – | – |
| Hard | – | – | –0.51 | – | –0.26 | – | –0.39 | – | –0.28 | – | –0.40 | – | – | – | – | – | –0.50 | – | –0.34 | – | –0.35 | – | –0.44 | – |
| $Ca^{+2}$ | – | – | –0.62 | – | – | – | –0.41 | – | –0.38 | – | –0.41 | – | –0.48 | – | – | –0.39 | – | – | –0.47 | – | –0.44 | – | –0.41 | – |
| $Cl^-$ | –0.51 | – | – | – | –0.42 | – | –0.43 | – | –0.34 | – | –0.37 | – | – | – | – | –0.62 | –0.36 | – | –0.45 | – | –0.39 | – | –0.37 | – |
| $Mg^{+2}$ | – | – | –0.21 | – | – | – | –0.31 | – | – | – | – | – | – | – | – | – | –0.41 | – | – | – | – | –0.49 | – | – |
| pH | – | 0.63 | – | –0.13 | – | – | – | 0.34 | – | – | – | – | – | – | – | – | – | – | – | – | – | 0.63 | – | – |
| EC | –0.49 | – | – | – | –0.42 | – | –0.44 | – | –0.44 | – | –0.35 | – | –0.46 | – | – | – | –0.16 | – | –0.37 | – | –0.45 | – | –0.42 | – |
| $SO_4^{-2}$ | –0.30 | – | –0.21 | – | –0.51 | – | – | –0.50 | –0.33 | – | –0.25 | – | – | –0.39 | – | – | – | – | – | – | – | –0.40 | –0.39 | – |
| TS | –0.49 | – | – | – | –0.51 | – | –0.45 | – | –0.42 | – | –0.24 | – | –0.50 | – | – | – | –0.44 | – | –0.44 | – | –0.43 | – | –0.42 | – |
| SS | – | –0.39 | – | 0.65 | – | 0.71 | – | –0.27 | – | –0.59 | – | –0.46 | – | –0.22 | – | – | – | –0.23 | – | – | – | – | – | – |
| $Fe^{+3}$ | – | – | – | – | – | 0.71 | – | –0.40 | – | – | – | – | – | – | 0.53 | – | – | – | – | 0.46 | – | – | – | – |
| $F^-$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| $Al^{+3}$ | – | – | – | – | – | – | – | –0.55 | – | –0.56 | – | – | – | –0.56 | –0.49 | – | – | –0.71 | – | 0.70 | – | –0.39 | – | 0.62 |
| $NO_2^-$ | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 0.25 | – | –0.42 |
| $NO_3^-$ | – | – | – | – | – | – | – | – | – | –0.40 | – | – | – | –0.38 | – | – | – | –0.61 | – | – | – | – | – | – |
| $NH_3$ | – | 0.67 | – | – | – | – | – | – | – | – | – | 0.55 | – | – | – | – | – | 0.25 | – | – | –0.39 | – | – | – |
| $SiO_2$ | – | – | – | – | – | – | – | – | – | – | – | 0.41 | – | – | – | – | – | – | – | – | – | – | – | – |
| $PO_4^{-3}$ | – | – | – | – | – | – | – | –0.32 | – | –0.42 | – | – | – | –0.59 | –0.54 | – | – | – | – | – | – | – | – | 0.66 |

**Table 4.** Treated water monthly loadings for each variable from the first two sparse principal components. Each month is recorded in a pair of columns with the first column representing the first sparse principal component (PC1) and the second column showing values for the second sparse principal component (PC2)

| | Jan | | Feb | | Mar | | Apr | | May | | Jun | | Jul | | Aug | | Sep | | Oct | | Nov | | Dec | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 | PC1 | PC2 |
| Temp | – | 0.55 | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 0.65 |
| Tur | – | – | – | 0.82 | – | –0.47 | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| TA | – | – | – | – | – | – | – | – | –0.42 | – | –0.42 | – | –0.50 | – | 0.33 | –0.42 | –0.50 | – | – | – | – | – | – | –0.56 |
| Hard | –0.38 | – | – | – | 0.44 | – | –0.48 | – | –0.23 | – | –0.42 | – | –0.22 | 0.38 | – | – | –0.48 | – | –0.27 | –0.25 | –0.44 | – | –0.39 | – |
| $Ca^{+2}$ | –0.35 | – | – | – | 0.39 | – | – | – | –0.42 | – | –0.34 | – | –0.39 | – | 0.35 | – | – | – | –0.45 | – | –0.42 | – | –0.35 | – |
| $Cl^{-}$ | –0.44 | – | –0.54 | – | 0.47 | – | –0.52 | – | –0.26 | – | – | – | – | – | 0.44 | – | –0.39 | – | –0.41 | – | –0.48 | – | –0.43 | – |
| $Mg^{+2}$ | –0.41 | – | – | – | 0.13 | – | – | – | – | – | – | – | – | 0.57 | – | –0.50 | –0.33 | – | – | –0.59 | – | – | – | – |
| pH | – | 0.63 | – | – | – | –0.54 | – | – | – | – | – | – | – | – | – | –0.52 | – | –0.58 | – | –0.44 | – | 0.24 | – | –0.51 |
| EC | –0.40 | – | –0.54 | – | 0.41 | – | –0.55 | – | –0.46 | – | – | – | –0.41 | –0.20 | 0.37 | 0.22 | – | – | –0.53 | 0.12 | –0.43 | – | –0.44 | – |
| $SO_4^{-2}$ | –0.15 | – | –0.38 | – | 0.44 | – | – | – | –0.38 | – | –0.46 | – | – | – | 0.25 | – | –0.27 | – | – | – | – | – | –0.41 | – |
| TS | –0.44 | – | –0.53 | – | 0.21 | – | –0.44 | – | –0.42 | – | –0.41 | – | –0.47 | – | 0.47 | – | –0.42 | – | –0.52 | – | –0.46 | – | –0.44 | – |
| $Fe^{+3}$ | – | – | – | – | – | – | – | – | – | 0.67 | – | – | – | – | – | – | – | –0.50 | – | 0.44 | – | – | – | – |
| $F^{-}$ | – | –0.33 | – | – | – | –0.23 | – | –0.65 | – | –0.32 | – | – | – | 0.40 | – | – | – | –0.61 | – | –0.42 | – | –0.40 | – | – |
| $Al^{+3}$ | – | – | – | – | – | –0.33 | – | – | – | – | – | 0.47 | 0.39 | – | –0.40 | –0.15 | – | –0.22 | – | – | – | – | – | – |
| $NO_2^{-}$ | – | – | – | – | – | 0.56 | – | – | – | – | – | –0.62 | – | – | – | – | – | – | – | – | – | – | – | – |
| $NO_3^{-}$ | – | –0.54 | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| $NH_3$ | – | 0.55 | – | – | – | – | – | 0.76 | – | 0.68 | – | –0.63 | – | –0.57 | – | –0.48 | – | – | – | – | – | 0.68 | – | – |
| $SiO_2$ | – | – | – | –0.57 | – | – | – | – | – | – | – | – | – | –1 | – | – | – | – | – | – | – | 0.58 | – | – |
| $PO_4^{-3}$ | – | – | – | – | – | – | – | – | – | – | –0.40 | – | – | – | – | – | – | – | – | – | – | – | – | – |

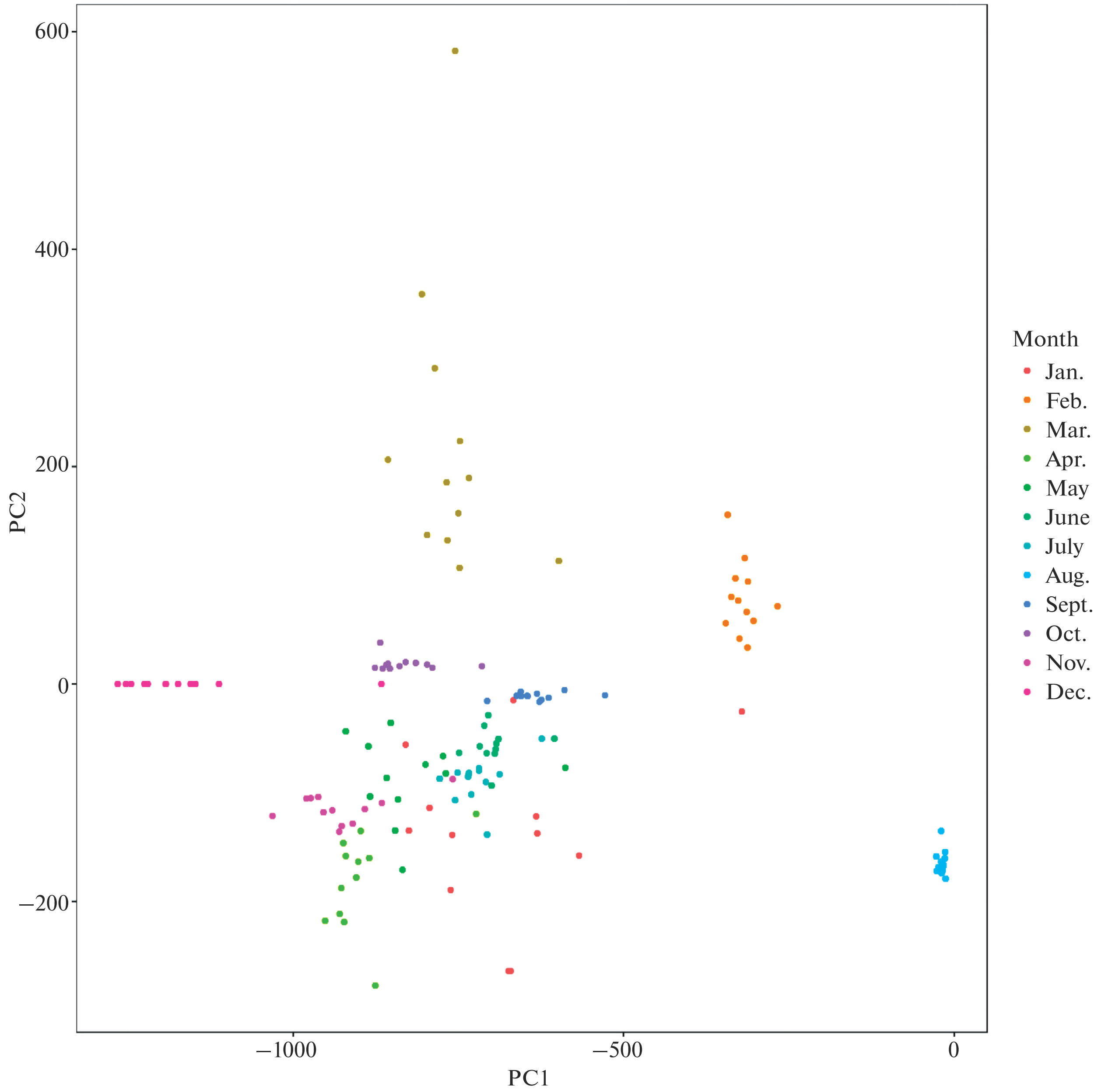


**Fig. 1.** The plot of the first two sparse principal component scores for each station by month for untreated water. The horizontal axis contains scores from the first sparse principal components while the vertical axis displays scores from the second sparse principal components. Colored dots correspond to scores from each month.

Next, the monthly values of the WQI for each station were plotted for both the untreated and treated water. These results are in Figs. 3 and 4. The graphs do appear to display some variability in water quality throughout the year. However, it is difficult to decipher a clear, meaningful pattern to the temporal variation. The plot for untreated water shows a noticeable difference in water quality between Station 1 and the other stations. For the majority of the year, the value of the WQI for Station 1 is discernably higher than the corresponding values for the eleven other stations. Station 1 also shows unique behavior after treatment with WQI values displaying a much larger variance making large jumps between months.

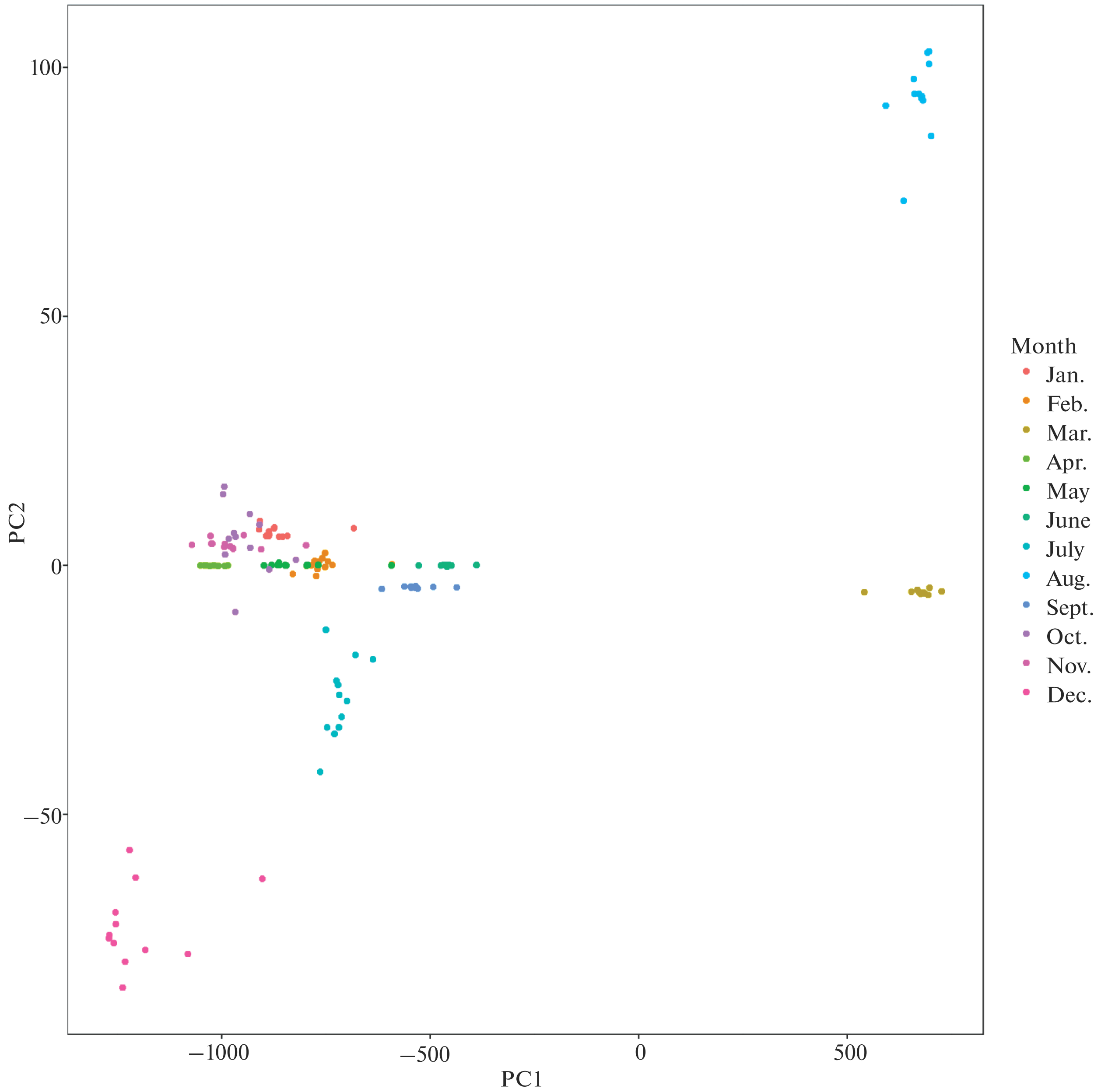


**Fig. 2.** The plot of the first two sparse principal component scores for each station by month for treated water. The horizontal axis contains scores from the first sparse principal components while the vertical axis displays scores from the second sparse principal components. Colored dots correspond to scores from each month.

The WQI measurements for each station were also used to perform hierarchical clustering. This procedure groups stations with similar water qualities. The "hclust" function in R was used to perform the clustering using complete linkage. The resulting dendrograms for untreated and treated water are in Figs. 5 and 6. From the dendrograms, it is clear that Station 1 had unique water quality characteristics as that station appears to be isolated from the others both before and after treatment. The clustering results support the previous conclusion that the water quality at Station 1 appears to be significantly different than the other stations. Station 1 standing out different is not surprising due to higher human activities in that area. Millions of people visit this area every year due to religious importance. Also, this area has various industries such as oil

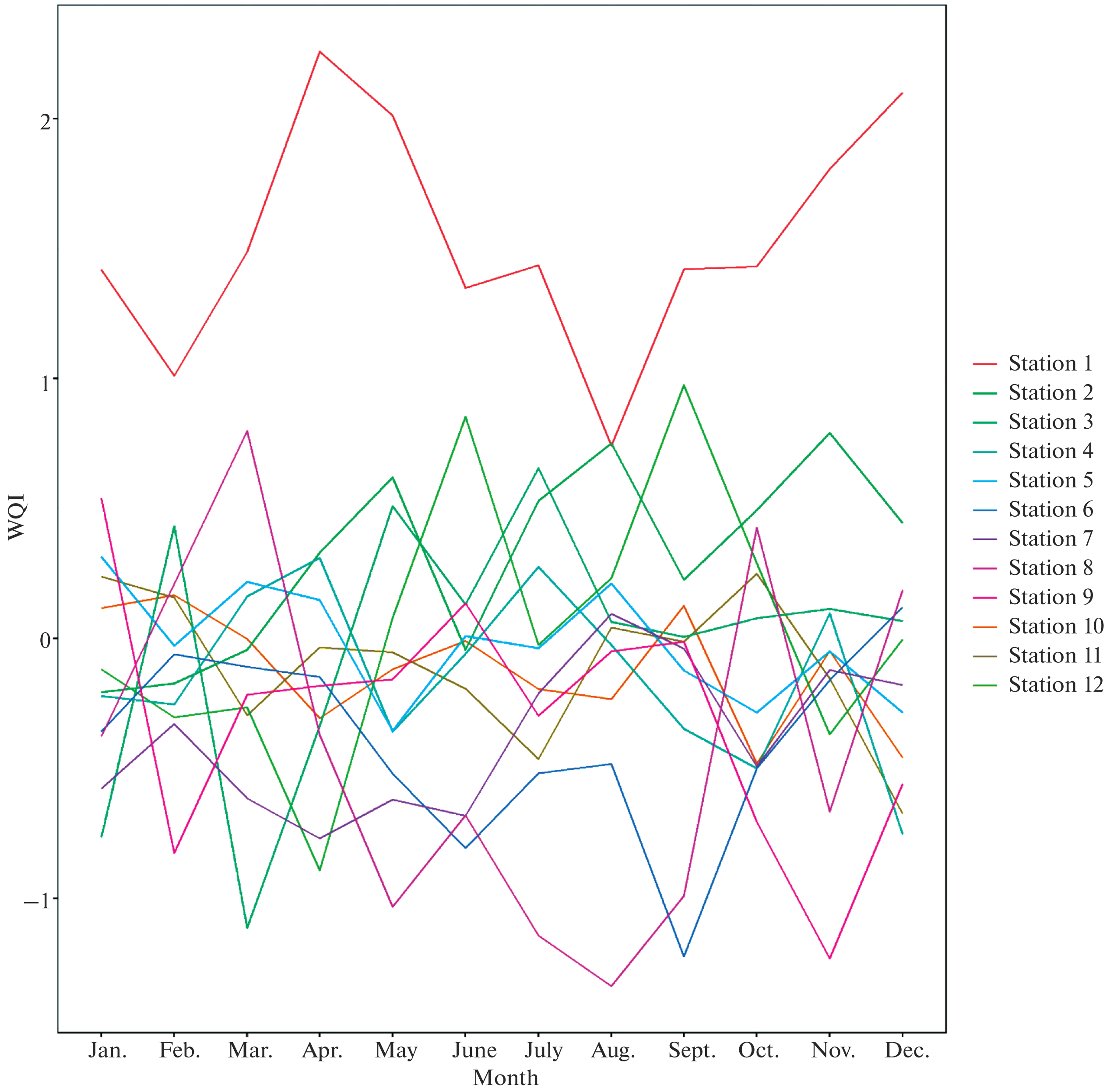


**Fig. 3.** Monthly WQI values for each station for untreated water. Each month is represented on the horizontal axis and the WQI values are shown on the vertical axis. Values for each station are represented by colored lines.

refining and plastic manufacturing. Smaller clusters of similar stations can also be seen for both untreated and treated WQI values, but these groups are more difficult to interpret as a set of clustered stations changes before and after treatment.

Finally, hypothesis tests were performed in order to determine if the means of variables contributing to water quality showed significant changes after treatment. Two sample *t*-tests with a one-sided alternative hypothesis were conducted for all variables except temperature assuming unequal variances. The null hypothesis was the means of values for untreated and treated water was the same, and the alternative hypothesis stated that the means were lower for treated water. All tests were performed using a significance level of 0.05. Turbidity ($p < 0.001$), alkalinity ($p < 0.001$), pH ($p < 0.001$), suspended solids ($p < 0.001$),

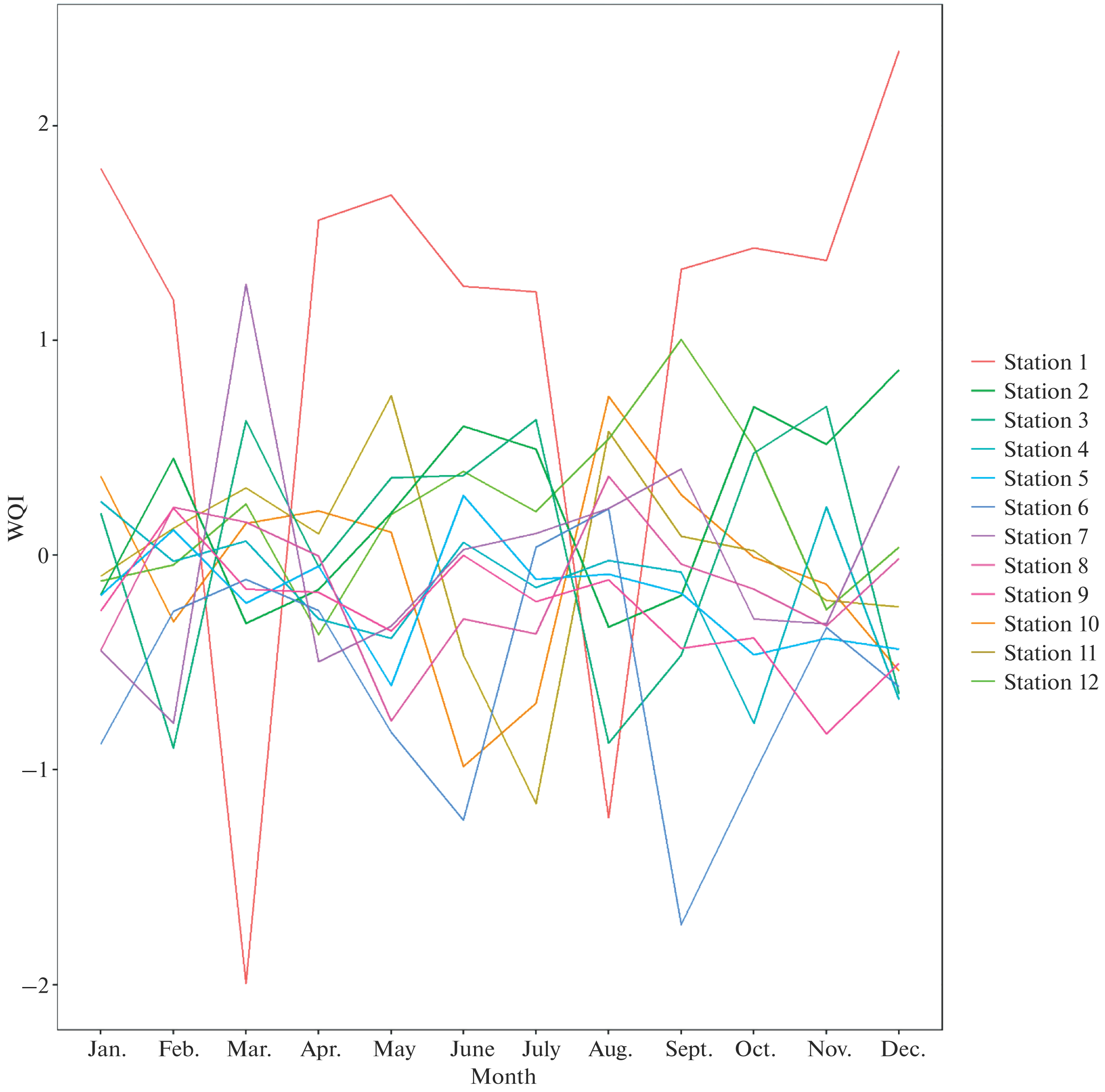


**Fig. 4.** Monthly WQI values for each station for treated water. Each month is represented on the horizontal axis and the WQI values are shown on the vertical axis. Values for each station are represented by colored lines.

$Fe^{+3}$ ($p = 0.002$), $F^-$ ($p < 0.001$) $NO_2^-$ ($p < 0.001$), $PO_4^{-3}$ ($p = 0.012$), and $NH_3$ ($p < 0.001$) all showed statistically significant differences in means between untreated and treated water. *P*-values for all the tests can be seen in Table 5. Typical methods for water treatment in all water treatment stations are suspensions precipitation, sand filtration, and chlorination 0.5 mg/L. The hypothesis testing results show that the above parameters are successfully reduced at the water treatment facilities.

Biological Oxygen Demand (**BOD**) and Dissolved Oxygen (**DO**) are useful parameters in water quality measurements. The amount of oxygen required by bacteria and other microorganisms while decomposing organic matter under aerobic (oxygen present) conditions at a specific temperature is referred to as

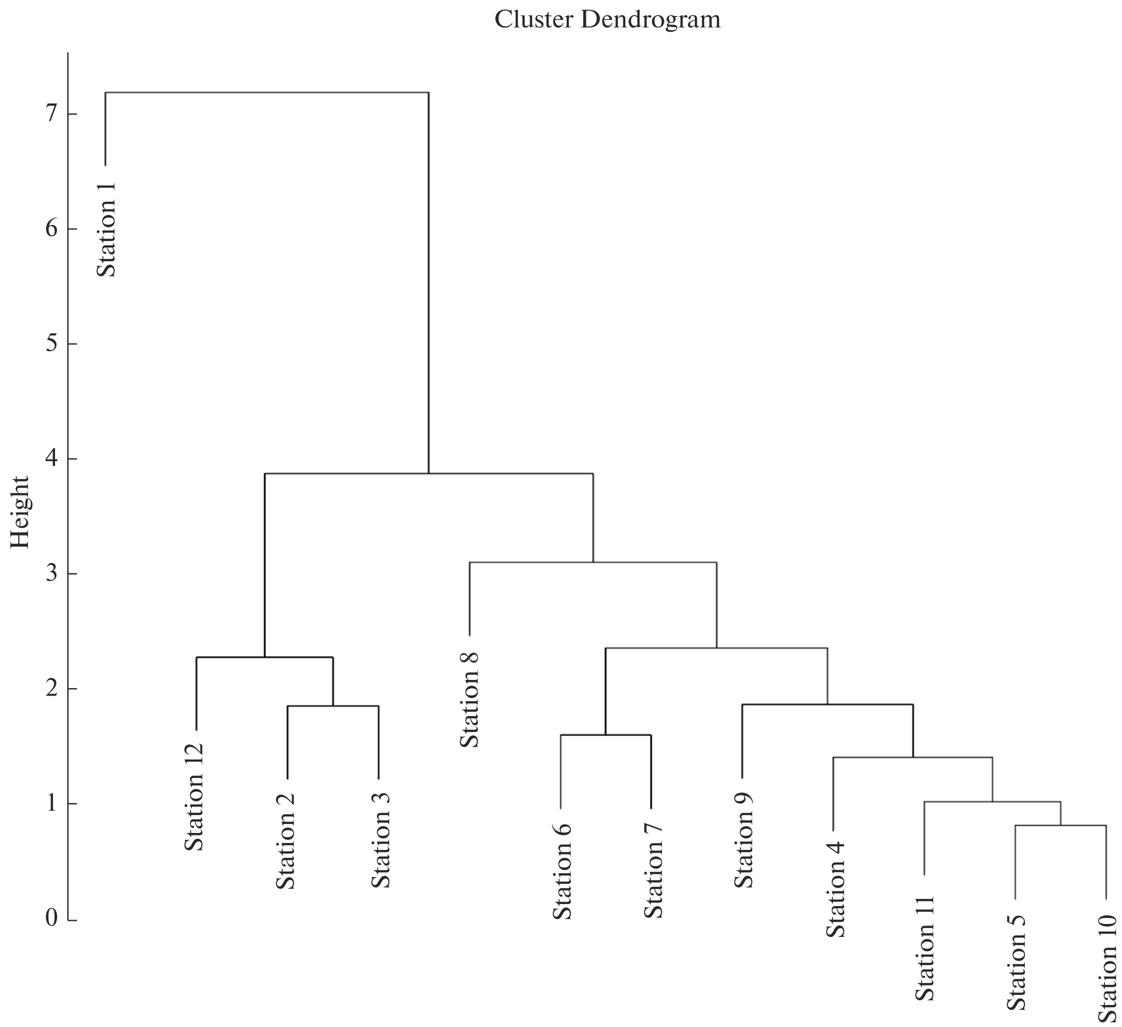


**Fig. 5.** Dendrogram for clustering the stations based on WQI calculated from untreated water.

BOD. The DO is a measurement of the amount of oxygen dissolved in water and available to living aquatic organisms. These parameters were not measured in this study. However, it won't affect the objectives of this project. The main goal of this study is to demonstrate the use of SPCA in water quality measurements. Therefore, in the future if BOD, and DO data are collected with the other parameters, they can be easily integrated into this SPCA–WQI computation. In addition, there are several water quality indexes reported in literature without the BOD and DO measurements such as in WAWQI used by Akoteyon et al. [4] and Singh et.al. [35], Bhargava's WQI used by Al-Musawi et al. [5] and NSFWQI used by Misaghi et al. [25].

Every year water quality monitoring stations measure thousands of chemical, physical, and biological parameters of water. One cannot predict the overall condition of a water stream without considering these data collectively, because these individual parameters do not give any trends in water quality over time and across geographical areas. Water quality indices such as the one proposed in this manuscript provide a strategy to process hundreds of water quality parameter data into meaningful values that indicate the condition of water resources. National Sanitation Foundation Water Quality Index (NSF–WQI) is the well-known and most widely used index in the world. However, indices such as the NSF–WQI rely on subjective metrics derived from expert opinion rather than an unbiased analysis of the data. In addition, the unified

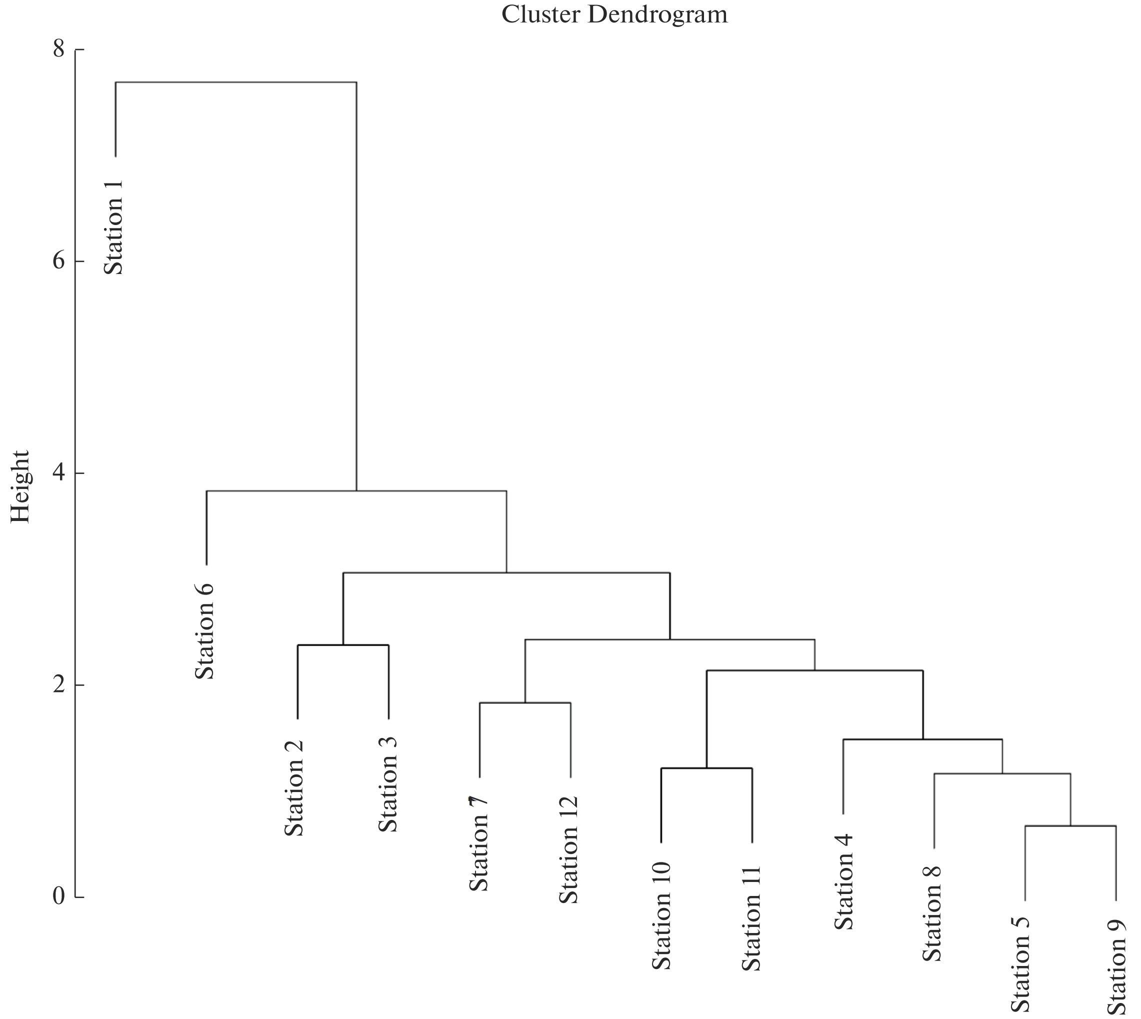


**Fig. 6.** Dendrogram for clustering the stations based on WQI calculated from treated water.

nature of this NSF–WQI can sometimes cause regional water quality concerns to be left unnoticed. The SPCA WQI suggested in this manuscript has several potential benefits over methods such as the NSF–WQI that would make it useful for applications in a variety of other contexts. For example, it can reflect regional water quality variations and concerns more clearly since the variable weights in the principal components are determined from the data. Importantly, SPCA can also identify a small subset of important variables contributing to water quality from a large number of variables, and any analysis focused on identifying a small collection of variables that influence water quality could benefit from this feature.

## CONCLUSIONS

According to the above results, the sparse principal component analysis successfully identified a small subset of important variables that contribute to water quality. This helps improve the interpretability of the resulting WQI. The calculated WQIs showed differences in water quality between stations and through time. The WQI can be effectively used to cluster the stations based on water quality. The above two points could help officials when monitoring the water quality at various sites throughout the year in order to maintain safe, drinkable water standards.

## ACKNOWLEDGMENTS

We gratefully acknowledge support from the Water Directorate of Baghdad, Iraq. This study was supported by

**Table 5.** Results of one-tailed hypothesis tests to determine if mean values are lower for treated water

| Variable | *p*-value |
|---|---|
| Tur | <0.001 |
| TA | <0.001 |
| Hard | 0.568 |
| $Ca^{+2}$ | 0.666 |
| $Cl^-$ | 0.834 |
| $Mg^{+2}$ | 0.526 |
| pH | <0.001 |
| EC | 0.639 |
| $SO_4^{-2}$ | 0.715 |
| TS | 0.560 |
| SS | <0.001 |
| $Fe^{+3}$ | <0.001 |
| $F^-$ | <0.001 |
| $Al^{+3}$ | >0.999 |
| $NO_2^-$ | <0.001 |
| $NO_3^-$ | 0.977 |
| $NH_3$ | <0.001 |
| $SiO_2$ | 0.091 |
| $PO_4^{-3}$ | 0.006 |

the Thi-Qar University, the Environmental Center of Al-Shatrah Technical Institute, the Office of Research and Sponsored Programs at the University of Central Oklahoma, and the Fulbright Visiting Scholar Program for Iraq.

## REFERENCES

1. Awomeso, J.A., Taiwo, A.M., Gbadebo, A.M., and Arimoro, A.O., Waste disposal and pollution management in urban areas: a workable remedy for the environment in developing countries, *Am. J. Environ. Sci.*, 2010, vol. 6, no. 1, pp. 26–32.
2. Abbasi, T. and Abbasi, S.A., *Water Quality Indices*, Elsevier, 2012, 1st Ed.
3. AL-Dulaimi, G.A. and Younes, M.K., Assessment of potable water quality in Baghdad City, Iraq, *Air, Soil Water Res.*, 2017, vol. 10, pp. 1–5.
4. Akoteyon, I., Omotayo, A., Soladoye, O., and Olaoye, H.O., Determination of water quality index and suitability of urban river for municipal water supply in Lagos-Nigeria, *Eur. J. Sci. Res.*, 2011, vol. 54, pp. 263–271.
5. Al-Musawi, N., Evaluation Water Quality of Diyala River in Iraq using Bhargava Method, *MATEC Web Conf.*, 2018, vol. 162.
6. Brown, R.M., McClelland, N.I., Deininger, R.A., and Landwehr, J.M., A water quality index — do we dare?, *Water Sewage Works*, 1970, vol. 117, pp. 339–343.
7. Benidis, K., Sun, Y., Babu, P., and Palomar, D.P., Orthogonal sparse PCA and covariance estimation via procrustes reformulation, *IEEE Trans. Signal Process.*, 2016, vol. 64, no. 23, pp. 6211–6226.
8. Clarke, J.I., Contemporary urban growth in the Middle East, in *Change and Development in the Middle East (Routledge Revivals)*, John, C.I., Howard, B.J., Eds., Essays in Honour of WB Fisher. Taylor & Francis, London, England, 2013, pp. 131–154.
9. Clesceri, L.S., Greenberg, A.E., and Trussell, R.R., *Standard methods for the examination of water and waste water*, 17th Ed., American Public Health Association, Baltimore, Md., USA, 1989.
10. Eltawil, M.A., Zhengming, Z., and Yuan, L., A review of renewable energy technologies integrated with desalination systems, *Renew. Sust. Energ. Rev.*, 2009, vol. 13, no. 9, pp. 2245–2262.
11. Fathy, S.A., Abdel Hamid, F.F., Shreadah, M.A., Mohamed, L.A., and El-Gazar, M.G., Application of principal component analysis for developing water quality index for selected coastal areas of Alexandra, Egypt, *Res. Environ. J.*, 2012, vol. 2, no. 6, pp. 297–305.
12. Glińska-Lewczuk, K., Effect of land use and lake presence on chemical diversity of the Łyna River system, *Pol. J. Environ. Stud.*, 2006, vol. 15, no. 2, pp. 259–269.
13. Gupta, P.K., *Methods in Environmental Analysis Water, Soil and Air*, Agrobios, India, 2009, 1st Ed.
14. Greenberg, A.E. and Clesceri, L.S., *Standard Methods for the Examination of Water and Wastewater*, Washington, DC: American Public Health Association, 1992.
15. Harkins, R.D., An objective water quality index, *J. Water Pollut. Control Fed.*, 1974, vol. 46, no. 3, pp. 588–591.
16. Gupta, S. and Gupta, S.K., A critical review on water quality index tool: Genesis, evolution and future directions, *Ecol. Informatics*, 2021, vol. 63, 101299.
17. Gazzaz, N.M., Yusoff, M.K., Aris, A.Z., Juahir, H., and Ramli, M.F., Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors, *Mar. Pollut. Bull.*, 2012, vol. 64, no. 11, pp. 2409–2420.
18. Horton, R.K., An index number system for rating water quality, *J. Water Pollut. Control Fed.*, 1965, vol. 37, no. 3, pp. 300–306.
19. Kalogirou, S.A., Seawater desalination using renewable energy sources, *Prog. Energ. Combust. Sci.*, 2005, vol. 31, no. 3, pp. 242–281.
20. Kašiarová, S. and Feszterová, M., Changes in stream water contamination in select Slovakian settlements, *Pol. J. Environ. Stud.*, 2010, vol. 19, no. 2, pp. 343–349.
21. Krishnan, R.R., Dharmaraj, K., and Kumari, B.R., A comparative study on the physicochemical and bacterial analysis of drinking, borewell and sewage water in the three different places of Sivakasi, *J. Environ. Bio.*, 2007, vol. 28, no. 1, pp. 105–108.
22. Lermontov, A., Yokoyama, L., Lermontov, M., and Machado, M.A.S., River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil, *Ecol. Indic.*, 2009, vol. 9, no. 6, pp. 1188–1197
23. Lu, R.S., Lo, S.L., and Hu, J.Y., Analysis of reservoir water quality using fuzzy synthetic evaluation, *Stoch.*

*Environ. Res. Risk Assess.*, 1999, vol. 13, no. 5, pp. 327–336.

24. Muniz, D.H.D.F., Moraes, A.S. Freire, I.D.S., Cruz, C.J.D.D., Lima, J.E.F.W., and Oliveira-Filho, E.C., Evaluation of water quality parameters for monitoring natural, urban, and agricultural areas in the Brazilian Cerrado, *Acta Limnol. Bras.*, 2011, vol. 23, no. 3, pp. 307–317.
25. Misaghi, F., Delgosha, F., Razzaghmanesh, M., and Myers, B., Science of the Total environment introducing a water quality index for assessing water for irrigation purposes: a case study of the Ghezel Ozan River, *Sci. Total Environ.*, 2017, vol. 589, pp. 107–116.
26. Nas, S.S., Bayram, A., Nas, E., and Bulut, V.N., Effects of some water quality parameters on the dissolved oxygen balance of streams, *Pol. J. Environ. Stud.*, 2008, vol. 17, no. 4, pp. 531–538.
27. Nbatista, N.J.C., Cavalcante, A.A.D.C.M., de Oliveira, M.G., Medeiros, E.C.N., Machado, J.L., Evangelista, S.R., Dias, J.F., Dos Santos, C.E., Duarte, A., da Silva, F.R., and da Silva, J., Genotoxic and mutagenic evaluation of water samples from a river under the influence of different anthropogenic activities, *Chemosphere*, 2016, vol. 164, pp. 134–141.
28. Nikanorov, A.M. and Yemelyanova, V.P., Comprehensive evaluation of continental surface water quality, *Water Resour.*, 2005, vol. 32, pp. 56–64.
29. Ocampo-Duque, W., Osorio, C., Piamba, C., Schuhmacher, M., and Domingo, J.L., Water quality analysis in rivers with non-parametric probability distributions and fuzzy inference systems: Application to the Cauca River, Colombia, *Environ. Int.,* 2013, vol. 52, pp. 17–28.
30. Ocampo-Duque, W., Ferre-Huguet, N., Domingo, J.L., and Schuhmacher, M., Assessing water quality in rivers with fuzzy inference systems: A case study. *Environ. Int.,* 2006, vol. 32, no. 6, pp. 733–742.
31. Parinet, B., Lhote, A., and Legube, B., Principal component analysis: an appropriate tool for water quality evaluation and management – application to a tropical lake system, *Ecol. Model.*, 2004, vol. 178, nos. 3–4, pp. 295–311.
32. Rahmanian, N., Ali, S.H.B., Homayoonfard, M., Ali, N.J., Rehan, M., Sadef, Y., and Nizami, A.S., Analysis of physiochemical parameters to evaluate the drinking water quality in the State of Perak, Malaysia, *J. Chem.*, 2015, pp. 1–10.
33. Rice, E.W. Baird, R.B. Eaton, A.D., and Clesceri, L.S., Eds., *Standard Methods for the Examination of Water and Wastewater*, 22$^{nd}$ Ed., American Public Health Association, American Water Works, Water Environment Federation, Washington DC, 2012.
34. Shiklomanov, I.A., World freshwater resources, in: *Water in Crisis*: *A Guide to the World's Freshwater Resources*, Gleick, P.H., Ed., New York, Oxford University Press, 1993, pp. 13–24.
35. Singh, R.K., Chaturvedi, A., and Kumari, K., Water-quality assessment of Damodar River and its tributaries and subtributaries in Dhanbad Coal mining areas of India based on WQI, *Sustain. Water Resour. Manag.*, 2019, vol. 5, pp. 381–386.
36. Uddin, M.G., Nash, S., and Olbert, A.I., A review of water quality index models and their use for assessing surface water quality, *Ecol. Indic.*, 2021, vol. 122, 107218.
37. Wang, C., Gamagedara, S., Shi, H., Adams, C.D., Timmons, T., and Ma, Y., Investigation of occurrence and removal of pharmaceuticals and personal care products in natural water by using liquid chromatography tandem mass spectrometry, *Water Res.*, 2010, vol. 45, pp. 1818–1828.
38. World Health Organization-WHO, *New Country-by-Country Data Show in Detail the Impact of Environmental Factors on Health*, 2007.
39. West, D.M., Mu, R., Gamagedara, S., Ma, Y., Adams, C., Eichholz, T., Burken, J.G., and Shi, H., Simultaneous detection of perchlorate and bromate using rapid high-performance ion exchange chromatography–tandem mass spectrometry and perchlorate removal in drinking water, *Environ. Sci. Pollut. Res.,* 2015, vol. 22, pp. 8594–8602.
40. Zou, H., Hasite, T., and Tibshirani, R., Spare principal component analysis, *J. Comput. Graph. Stat.,* 2006, vol. 15, no. 2, pp. 265–286.

SPELL: 1. OK