

Original Article

Effective and promising feasibility study: Determination of marker ions for the separation of water samples according to regions by chemometric method combined with ion chromatography

Batuhan Yardımcı^a, Sinem Çolak^{b,*}, Jülide Yener^c

^aScience and Technology Application and Research Center (ARTMER), Zonguldak Bülent Ecevit University, Zonguldak, Türkiye

^bDepartment of Chemical and Chemical Processing Technologies, Çaycuma Food and Agriculture Vocational School, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye

^cDepartment of Chemistry, Faculty of Science, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye

ARTICLE INFO

Keywords:
Discriminant analysis
Ion chromatography
Principal component analysis
Tap water

ABSTRACT

Anion and cation concentration data analyzed in drinking water can be very useful in source identification with multivariate statistical methods. In this study, anions and cations were analyzed in samples collected from four regions by ion chromatography (IC). The effects of seasons and regions on ion concentrations were analyzed by variance analysis (ANOVA). Multivariate statistical methods such as principal component analysis (PCA), discriminant analysis (DA), and hierarchical cluster analysis (HCA) were used in the study. In the DA, the percentage of classification of the samples to the original regions was found to be 91.7%. HCA results revealed that Na⁺, Mg²⁺, and Ca²⁺ ions exhibited very similar behavior. In the PCA analysis, PC1 and PC2 explained 68.3% of the total variance for anions, 95.3% for cations, and 99.7% for Ca²⁺, Mg²⁺, and Na⁺. It has been observed that different water supply sources can be distinguished using Ca²⁺, Mg²⁺, and Na⁺ cations, and it is also possible to differentiate regions belonging to the same source. Differences in ion concentrations may be caused by the chemicals used during treatment, the geological structure of the region, and seasonal effects. However, this study has demonstrated that chemometric techniques combined with IC can identify the source of drinking water samples when their regional origin is unknown.

1. Introduction

Water is an essential substance that sustains life and the natural environment. It is also a primary component for industry, a consumable item for humans and animals, and a carrier of domestic and industrial pollution. The regulation of aquatic compounds, bathing water quality, surface and drinking water quality, and effluent control are provided by several directives that are closely related to analytical measurements [1]. Maximum contamination levels (MCLs) are specified for several common inorganic anions, such as fluoride, nitrite, and nitrate, by the US National Primary Drinking Water Standards. To reduce any negative health impacts that could result from consuming these anions in drinking water, the MCLs have been specifically set [2]. For instance, methemoglobinemia, which can be fatal to newborns, can be caused by nitrite and nitrate [3], whereas excessive amounts of fluoride can cause skeletal and dental fluorosis [4]. Hard water is very dangerous to human health and can cause many diseases. Excessive Ca²⁺ and Mg²⁺ intake can increase the risk of coronary artery disease, nephrolithiasis, colorectal cancer, obesity, osteoporosis, hypertension, stroke, and insulin resistance [5]. Contaminants include other common anions like sulfate and chloride. The Secondary Drinking Water Standards

are not federally enforced recommendations and cover taste, odor, color, and some esthetic impacts. Nevertheless, many states generate their enforceable legislation governing these contaminants, and all are advised to pursue them [2].

Contaminants in drinking water do not only come from the external environment. In the treatment process to remove contaminants in drinking water, chemicals are added or in contact with various materials. Therefore, various chemicals used for treatment and their by-products can also be mixed into drinking water [6]. The effects of water pollution and the harmful health effects caused by excessive amounts of some ions have led scientists to develop appropriate methods to identify and measure metal contaminants [7-9].

Typically, chemical wet procedures, including gravimetry, titration, photometry, turbidimetry, and colorimetry, were used to determine the common inorganic anions and cations. Numerous techniques have restricted sensitivity and suffered from interferences; they can require a lot of labor and are sometimes challenging to automate. While analytical methods allow for the specification of performance requirements (accuracy, precision, and limit of detection), it remains challenging to achieve comparable outcomes across various laboratories [1]. While the techniques mentioned above improved the technology

*Corresponding author:

E-mail address: sinem.colak@beun.edu.tr (S. Çolak)

Received: 01 July, 2025 Accepted: 12 October, 2025 Epub Ahead of Print: 16 February, 2026 Published: 02 April, 2026

DOI: 10.25259/AJC_755_2025

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, transform, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

for identifying ions of water sources, they did not account for the information superposition between the identification indicators of water chemicals, which led to issues such as low classification precision and slow response times [1].

Ion chromatography (IC) enables the assessment of the concentration of main anions and cations in water samples [10]. The method depends on the electrostatic interactions of oppositely charged functional groups attached to a stationary phase with charged fragments on the surface of molecules. This surface reaction takes place when an ionic solid and a solution come into contact [11]. One of the main benefits of using IC as an analytical method is that it typically requires minimal time or no sample preparation, and only a small portion of the sample is used [12]. Despite analyzing a complicated matrix like spirit drinks, it demonstrates exceptional selectivity, sensitivity, and repeatability [13]. IC can greatly improve the capacity to examine big datasets, recognize patterns, and distinguish between various water sources according to their ionic content when used in conjunction with chemometric methods [14].

IC data are used for many purposes with multivariate statistical methods. Multivariate statistical analyses, which are frequently used today, are examined in two groups as dependence and independence analyses. In dependence analysis (like discriminant analysis (DA)), one or more dependent variables can be explained by other variables, while in independence analysis (like PCA), there is no dependent variable, and the relationship between variables is examined. With DA, the variables that most cause the differentiation of two or more groups can be determined. In addition, one or more functions can be produced [15,16]. Principal component analysis (PCA) is a statistical technique used to identify patterns and relationships in large data clusters. It is a widely used method to reduce data and dimensions while keeping most of the original information [17,18]. Some scientists have adopted the PCA method to separate water sources and obtained better analysis results [19]. PCA-based data processing helps to efficiently remove high-dimensional data correlation and makes data structures simpler [20].

This study aims to assess the feasibility of using chemometric methods in combination with IC to identify and differentiate water samples from different regions based on their ion profiles. It also demonstrates the usability of the outputs obtained when routine anion and cation concentration data in drinking water are combined with multivariate statistical methods. For this purpose, water samples were collected from 4 different regions, and 13 ions, including Cl^- , Br^- , F^- , SO_4^{2-} , PO_4^{3-} , NO_2^- , NO_3^- , Li^+ , NH_4^+ , Na^+ , K^+ , Mg^{2+} , and Ca^{2+} , were analyzed by IC. By systematically analyzing water samples from

various geographical locations, marker ions for regional differentiation were identified. In addition to descriptive statistical parameters, the relationship between ions was examined using Pearson correlation analysis. The similarity of ion concentrations across regions was determined using hierarchical cluster analysis (HCA). Discriminant and PCA analyses were performed on the data obtained from IC. The results of this research suggest that a better understanding of regional water chemistry can support improved water quality monitoring strategies. It can also contribute to the development of targeted measures for water resources management and pollution control. It is thought that the findings of this study may be useful for source identification of drinking water of uncertain origin with a few types of ion analyses. In the study, marker ions that can provide an idea about the source of drinking water of uncertain origin were also investigated.

2. Materials and Methods

2.1. Chemicals and apparatus

Thermoscientific Dionex Seven anion Standard II (in Deionized water) and Thermoscientific Dionex Six Cation-II standard solutions were used to obtain calibration equations for each ion. Sodium carbonate (0.5 M) standard solution was obtained from Thermoscientific as the mobile phase for anion, and methanesulfonic acid was obtained from Sigma Aldrich as the mobile phase for cation. The IC was equipped with Dionex IonPac CS12A RFIC 4×250 mm column and Dionex IonPac AS9-HC RFIC 4×250 mm column for cation and anion analyses, respectively. DRS 600 Dionex Dynamically regenerated suppressors were used. To protect from high concentrations, Dionex IonPac AG9-HC RFIC 4×50 mm was used as a guard. Ultrapure water was obtained from the Human corporation brand Zeener UP 900 model ultrapure water device.

2.2. Sample collection and preparation

All samples were collected from January 2022 to December 2022. The water samples were collected from treated drinking water from three different municipality stations distributed throughout the Zonguldak city. The regions are labeled A, B, and C. In addition, in municipality C, samples taken at distances from each other were coded as C1 and C2. The map of the regions where the samples were taken has been given in Figure 1 [21]. Three samples were taken from each region in the first, middle, and last week of each month. A total of 144 samples were collected for 12 months. Samples were stored at +4°C until analysis.

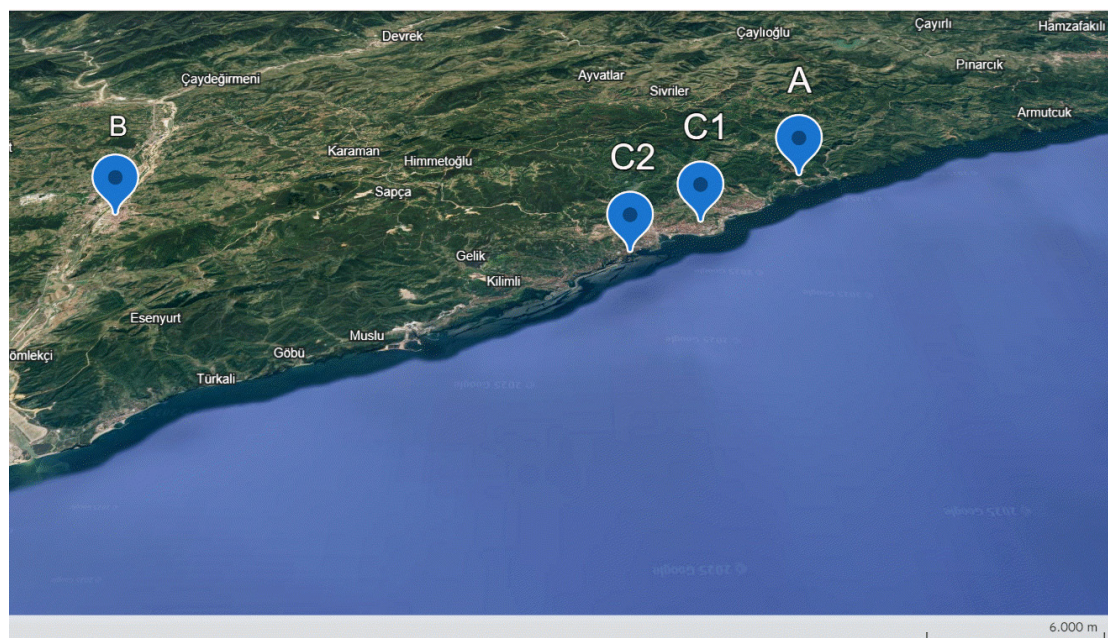


Figure 1. Location map of the sampling area [21].

2.3. Analytical methods

Concentration of anions (Cl^- , Br^- , F^- , SO_4^{2-} , PO_4^{3-} , NO_2^- and NO_3^-) and cations (Li^+ , NH_4^+ , Na^+ , K^+ , Mg^{2+} , and Ca^{2+}) were determined with Thermo Scientific DIONEX ICS-1100. As the mobile phase, 10 mM sodium carbonate was used for anions and 20 mM methanesulfonic acid for cations. The mobile phase was adjusted in isobaric flow with a flow rate of 1 mL min^{-1} . The pressure was 6.87 MPa and 15.88 MPa for cations and anions, respectively. Injection volume was 20 μL , and conductivity detection was performed. The recording time was 18 min for cations and 24 min for anions. The operating temperature was 30°C . Mobile phases and standard solutions were prepared using ultrapure water. Calibration curves were obtained by preparing five different concentrations of standard solutions for each ion. Three replicates were performed for each sample, and the average values were calculated. The instrument's operating range was determined by performing at least three replicate runs on blank and serially diluted standard solutions. The limits of detection (LOD) for anions and cations were determined by analyzing solvent blanks with low analyte concentrations. The analytical parameters (retention time, linear range, equation, regression coefficient, LOD, relative standard deviation (RSD%), and recovery%) for anion and cation analysis have been given in Table 1.

2.4. Statistical analysis

IBM SPSS Statistics 22 and Minitab 19 software packages were used for statistical analysis of the data obtained by IC. The descriptive statistical parameters used in the study were mean (M), standard error of mean (SEM), standard deviation (SD), and data range (minimum-maximum). ANOVA analysis (F test) was used to examine whether months and regions produced significant differences in anion and cation concentrations. Multivariate statistical methods such as PCA and DA were used to evaluate the effect of anion and cation concentrations analyzed in water in distinguishing regions, and HCA was used to determine similarities. DA determined the extent to which anions and cations contributed to the differences between regions, revealing which ions caused these differences. PCA analysis was also used to demonstrate that ion concentrations in water samples could classify regions. Furthermore, the relationships between anion and cation

Table 1. The analytical parameters for anion and cation analysis.

Ions	Retention time (min)	Linear range (mg/L)	Linear equation	Regression coefficient	LOD	RSD%	Recovery %
Anions							
F^-	3.90	0.2-15	$y=0.8943x-0.0815$	99.9262	0.032	0.815	96.72
Cl^-	6.69	1-75	$y=0.6783x-0.4437$	99.9053	0.048	0.472	102.5
NO_2^-	8.75	0.5-75	$y=0.3655x-0.1699$	99.9637	0.024	0.516	97.62
Br^-	11.41	0.5-75	$y=0.2537x-0.0949$	99.9647	0.04	0.322	88.82
NO_3^-	13.58	0.5-75	$y=0.3299x-0.1180$	99.9637	0.08	0.450	96.45
PO_4^{3-}	15.98	1-150	$y=0.1767x-0.2084$	99.9306	0.064	0.375	103.5
SO_4^{2-}	19.75	1-75	$y=0.4596x-0.1860$	99.9702	0.08	0.612	94.63
Cations							
Li^+	3.84	0.5-15	$y=0.7418x$	99.8797	0.06	1.200	86.78
Na^+	4.66	1-75	$y=0.1976x$	99.8762	0.18	0.719	104.3
NH_4^+	5.33	1-75	$y=0.1154x+0.3054$	99.8043	0.16	0.416	99.10
K^+	6.83	1-150	$y=0.1423x$	99.8592	0.04	0.532	92.55
Mg^{2+}	12.35	1-75	$y=0.4015x$	99.7629	0.08	0.645	94.86
Ca^{2+}	15.27	1-150	$y=0.2572x$	99.7924	0.12	0.851	103.2

concentrations were evaluated using two-tailed Pearson correlation coefficients (r).

3. Results and Discussion

3.1. Determination of anion and cation concentrations in water samples

IC was used to analyze water samples taken from the drinking and utility water supply network system of three different regions, and the statistical analysis of the ions obtained from 144 water samples has been given in Table 2. It is seen that the average of the concentrations of all ions except phosphate is in accordance with WHO and EPA standards. Whereas other anions were absent from some samples, Cl^- was present in every sample. Among the cations, NH_4^+ was not detected in any of the water samples, while Li^+ was not detected in some samples. In the study, high values were observed in the standard deviations of some ion concentrations. Ions whose concentrations vary according to regions and months are absent or present in very high amounts in some regions or months. Therefore, there are values far from the average, especially observed in nitrate, sulfate, and phosphate ion concentrations, causing the standard deviation values to be higher than the average.

In the literature, chemical pollutants that generally cause problems in water resources are nitrate, pesticides, industrial solvents, iron, manganese, and hardness-forming ions. Pollutants caused by drinking water treatment plants are fluoride, nitrite, aluminum, chlorine, trihalomethanes, and acrylamide [6]. In addition, some ions in water can also increase to higher values at the outlet of treatment plants. For example, the sulfate concentration in drinking water at the outlet of the treatment can exceed the inlet concentration. Due to the chemicals used in drinking water treatment (aluminum and sulfate), sulfate concentration measured 12.5 mg/L at the source and 22.5 mg/L at the treatment outlet in Ontario, Canada [22]. Cl^- ion is the most common ion found in drinking water and originates from natural sources, sewage, industrial discharges, and urban runoff.

In this study, anion and cation concentrations were analyzed separately according to months and selected regions. Interval plots of average anion and cation concentrations (mg/L) according to months have been given in Figures 2 and 3, and according to regions in Figures 4 and 5.

Temperature is an important factor in anion concentration changes in water. For example, thermal layering (epilimnion and hypolimnion) is caused by the density difference in water. Since water is at its densest at 4°C , the density of water is lower on either side of this temperature. During the summer, the sun heats the surface of the water, and as the density drops, the cooler, denser water remains in the lower layer of the lake or dam. When the water continues to warm, two distinct layers form at the top: the warmer epilimnion and the colder hypolimnion. Since oxygen is completely consumed in the hypolimnion, iron, manganese, ammonia, sulfates, phosphates, and silica pass from the sediment to

Table 2. Summary of the concentrations of anions and cations measured in different municipality water samples collected from Zonguldak city.

Variable	N	Mean	S.E.M.	SD	Minimum	Maximum	WHO	EPA
F^-	144	0.29	0.015	0.18	0	1.09	1.5	0.7-2.4
Cl^-	144	17.41	0.78	9.38	6.42	38.01	250	250
NO_2^-	144	0.287	0.026	0.31	0	1.33	0.5	-
Br^-	144	0.19	0.016	0.20	0	0.52	0.01	-
NO_3^-	144	2.78	0.30	3.67	0	11.95	50	45
SO_4^{2-}	144	15.48	2.29	27.45	0	93.91	250	250
PO_4^{3-}	144	44.31	4.69	56.23	0	195.04	5	-
Li^+	144	0.005	0.0005	0.0064	0	0.025	-	-
Na^+	144	11.07	0.97	11.68	0.32	36.06	-	-
K^+	144	1.93	0.10	1.22	0.08	5.14	-	-
Mg^{2+}	144	8.05	0.70	8.39	0.08	27.10	-	-
Ca^{2+}	144	68.27	2.94	35.25	12.89	153.86	-	-

N: number of samples, SEM: standard error of the mean, SD: Standard deviation.

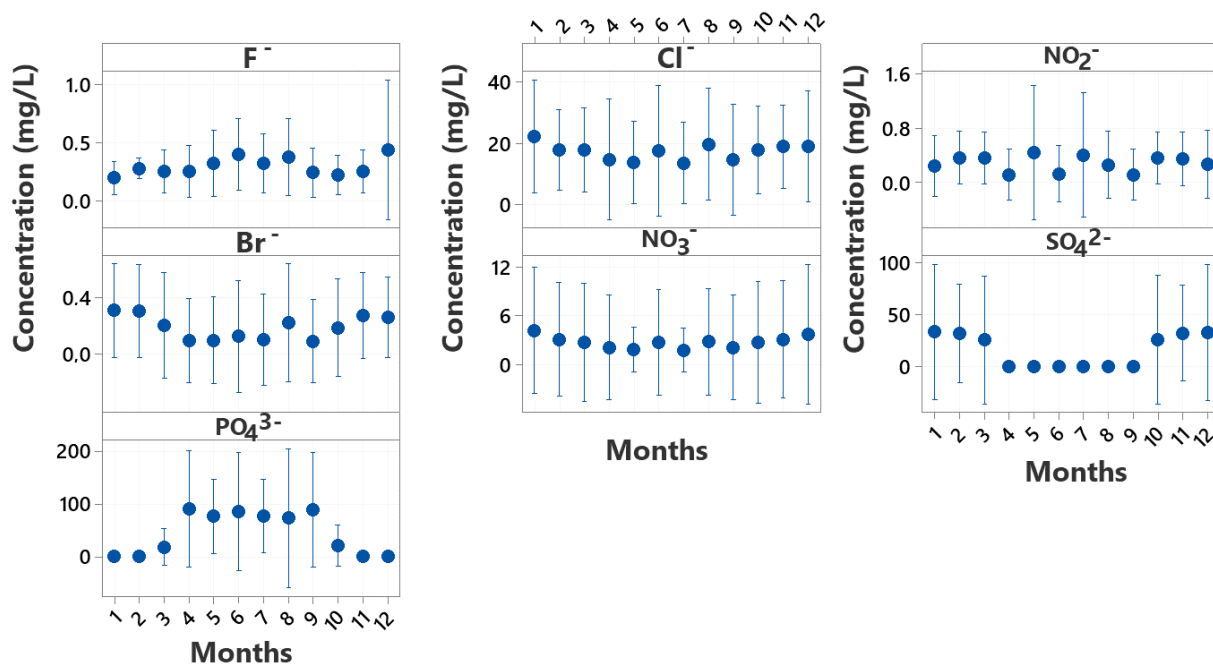


Figure 2. Average anion concentrations (mg/L) vary according to months.

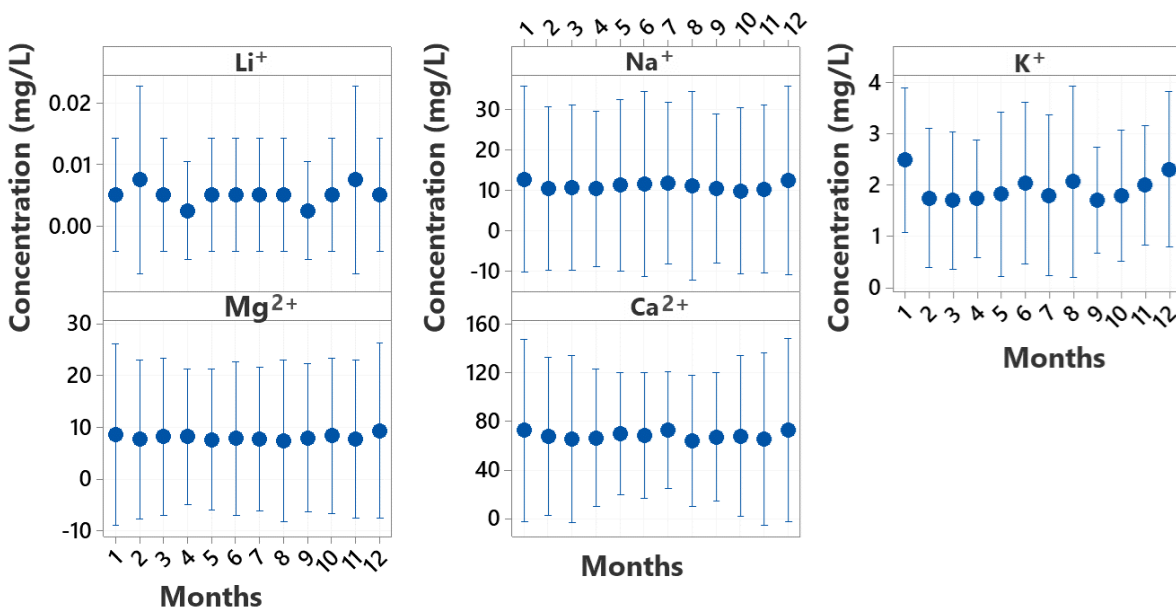


Figure 3. Average cation concentrations (mg/L) vary according to months.

the water under anaerobic conditions. This situation deteriorates the drinking water quality. Ammonia reacts with chlorine and behaves like a nutrient element, causing eutrophication. In addition, sulfates also react with chlorine and cause oxygen depletion [6]. In this study, anion change according to months was found to be quite different, especially for sulfate and phosphate.

When the average anion concentrations according to the regions were analyzed, as seen in Figure 4, it was seen that all ion concentrations belonging to region B were considerably higher than the others. Li⁺ was observed only in regions A and B. Chlorination is a common disinfection method used in drinking water treatment. Chlorine is the most preferred disinfectant in municipal water networks due to its applicability and cheapness [23]. During chlorination, chloroform, trihalomethanes, dibromochloromethane, bromodichloromethane,

bromodichloromethane, and bromoform can be formed due to the reaction of natural organic matter, such as humic and fulvic acid, with chlorine [24]. Although this study focuses on inorganic ions, it should be noted that organic pollutants (e.g., pesticides, humic and fulvic acids, trihalomethanes, and residual treatment byproducts) significantly influence water chemistry and regional variation [25]. According to Figure 4, Cl⁻ concentration was found to be two times higher in region B than in other regions. In 2019, the sulfate value was found to be 223.82 mg/L in a study done on irrigation water samples from region B [9]. In this study, it was observed that the sulfate concentration was 504.10 mg/L in drinking water.

In drinking water treatment systems, the use of chemical coagulation is one of the most important processes for the removal of natural organic matter. Since it is one of the most economical methods for

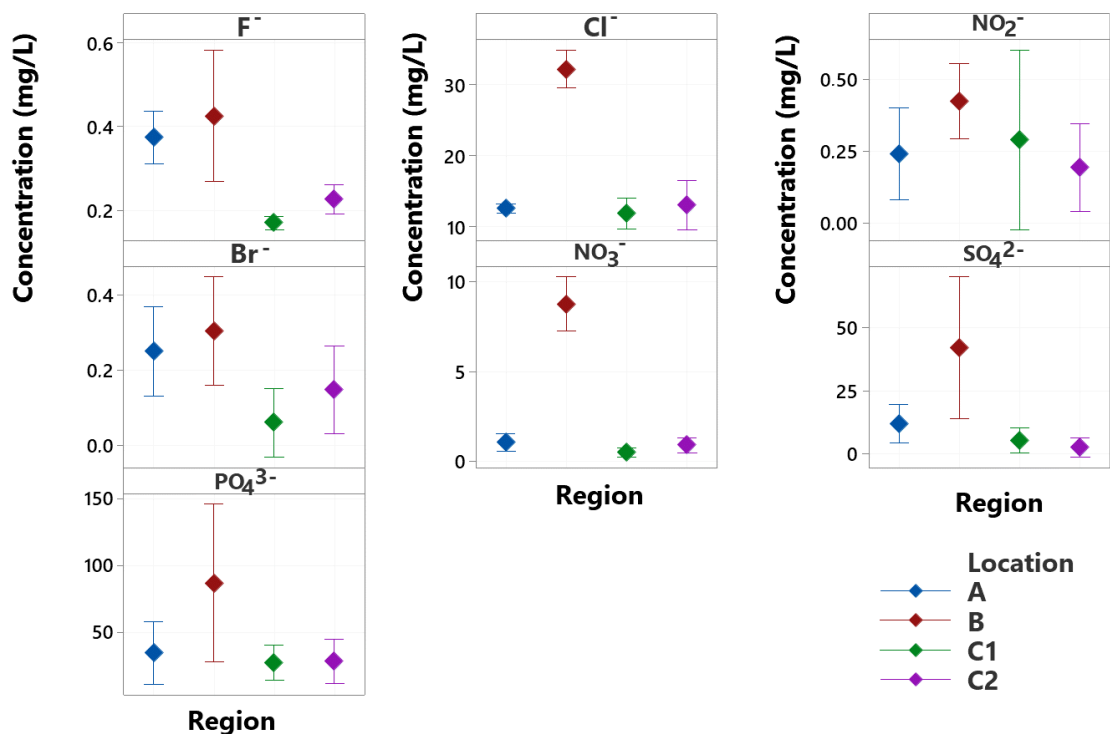


Figure 4. Average anion concentrations (mg/L) vary according to regions.

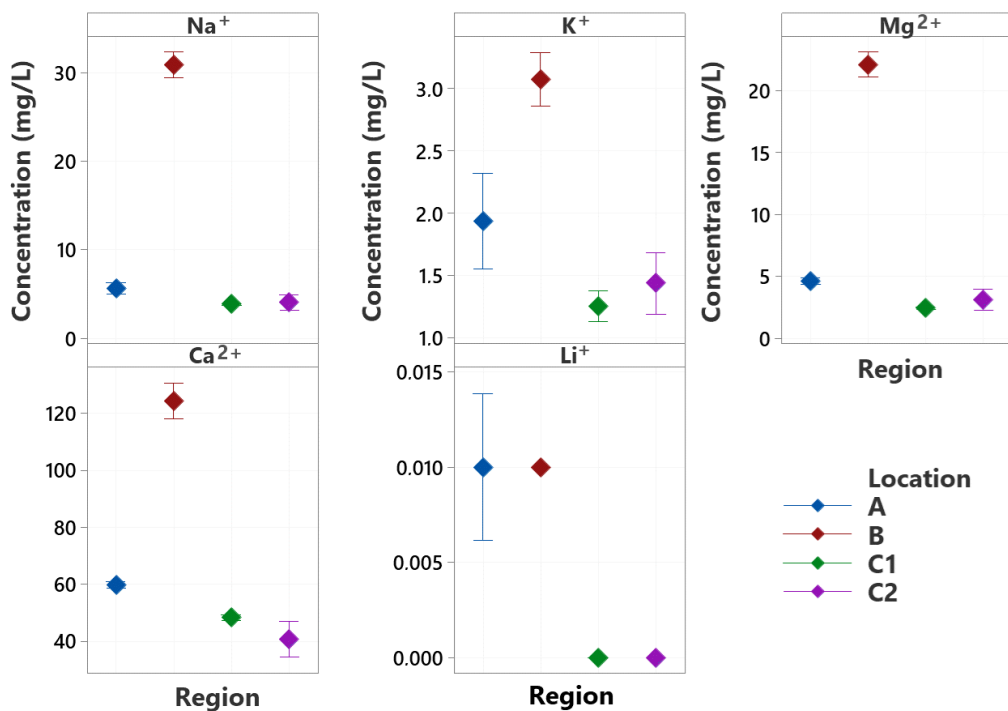


Figure 5. Average cation concentrations (mg/L) vary by region.

large treatment plants, it is still frequently used [26]. In conventional treatment processes, iron chloride or aluminum sulfate is used as an inorganic coagulant in the coagulation/flocculation unit. The high concentrations of sulfate and chlorine ions compared to other anions are thought to be both natural and process-induced in drinking water treatment plants.

The quality and quantity of surface water are also affected by the geology of the basin. In general, for example, hard and clear waters are observed in limestone and limestone basins, while turbid and soft waters are observed in basins where rocks such as granite predominate. A 3D Scatterplot of anions and cations by region has been given in Figure 6. Some ions were found in higher concentrations in the same

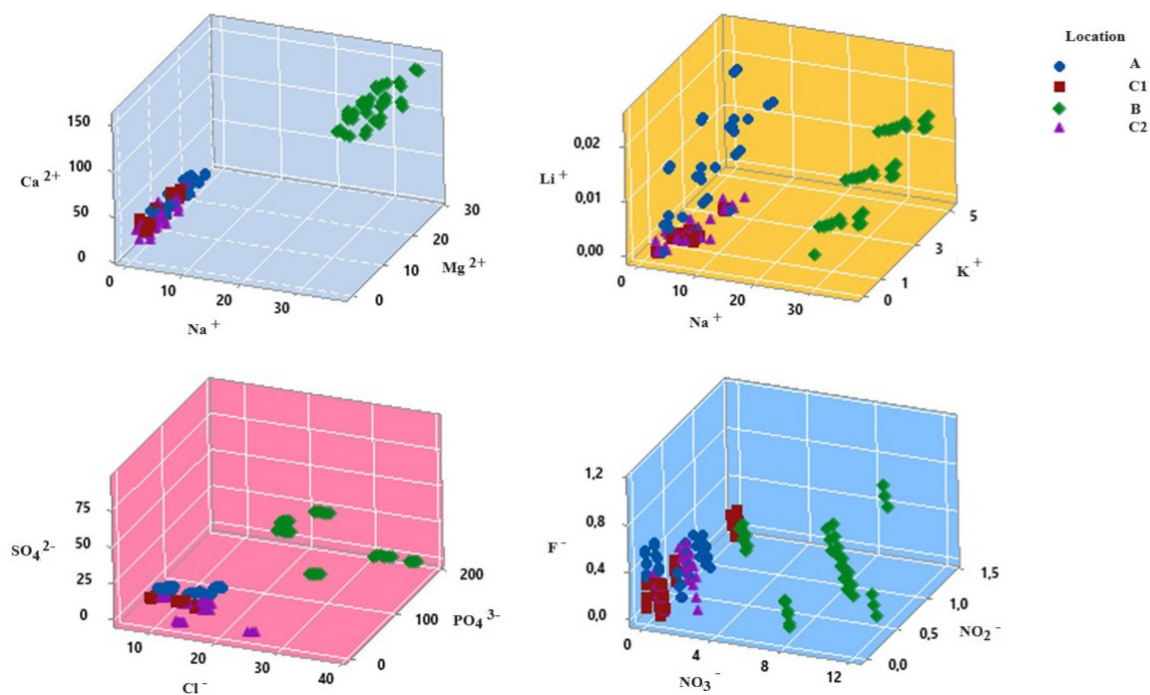


Figure 6. 3D Scatterplot of anions and cations.

regions compared to other regions. This is thought to be caused by geological factors.

3.2. Analysis of variance

The *F*-test was used to analyze whether there were significant differences between months and regions in anion and cation concentrations. The arithmetic means of anion and cation concentrations of the months and regions were compared by analysis of variance (ANOVA), and the results of the analysis have been given in Table 3. Significant differences in anions and cations were determined. The ions with significance values $p < 0.05$ for the results of the *F*-test were found to have statistically significant differences at a 95% confidence interval.

When Table 3 is analyzed, the *p*-value was < 0.05 for all cation concentrations and all anions except phosphate, and it was determined that they did not contribute significantly to the months. Phosphate concentration made a significant difference in the months at the $p < 0.01$ level. In the ANOVA analysis according to the regions, all anions and cations except nitrate were found to be significant at $p < 0.05$ and $p < 0.001$ levels.

3.3. Correlation analysis

The relationships between anion and cation concentrations in water samples from all regions were determined using two-tailed Pearson correlation analysis and have been presented in Table 4. The relationship between the variables is considered weak if the correlation coefficient is < 0.5 , moderate if the correlation coefficient is between 0.5-0.7, and large if the correlation coefficient is > 0.7 [27].

When Table 4 is analyzed, a high positive correlation was observed between NO_3^- and Cl^- ion and Na^+ , Mg^{2+} , and Ca^{2+} cations. A low but negative correlation was found between PO_4^{3-} and SO_4^{2-} . A strong positive correlation was also observed between Na^+ , Ca^{2+} , and Mg^{2+} . The matrix plot for all ions has been given in Figure 7. Similar correlation values were found in a study performed in the irrigation waters of region B. In irrigation water samples, Ca^{2+} and Na^+ ($r = 0.925$), Ca^{2+} and SO_4^{2-} ($r = 0.885$), Ca^{2+} and Mg^{2+} ($r = 0.945$), and Na^+ and Cl^- ($r = 0.912$) [9].

3.4. Discriminant analysis

The main purpose of DA is to predict which group the data from many groups belongs to. The functions created divide the data into groups, and the new data are placed in the most appropriate group by replacing them in the created discriminant function. With DA, it is possible to place the unknown data in the correct group with minimum error [28]. In this study, the level of contribution of anions and cations to the differences between the regions was determined by DA. In addition, it was also found out which ions caused this difference. Wilks' Lambda values and significance of ions have been given in Table 5, which shows the effect of the variables in separating the groups. As the Wilks' Lambda value of the ions is close to zero, the contribution of the variable to the separation of the groups increases, decreases as it approaches 1, and does not contribute when it is equal to 1. On the contrary, as the *F* value of the variable increases, its contribution to the discrimination between groups increases. It was observed that Na^+ (0.014), Mg^{2+} (0.015), and Ca^{2+} (0.040) had the lowest Wilks' Lambda values. In addition, these ions also have the highest *F* values and are the separator ions that make the greatest contribution to the separation between the regions.

Canonical Correlation was found to be 0.996, 0.952, and 0.679 for the 1st, 2nd, and 3rd functions, respectively. In the study, the first function can explain 92.3% of the variation, the second function 7.1%, and the third function 0.6%. When Wilks' Lambda values of the functions are evaluated, 16.5% of the first function and 96.3% of the second function cannot be explained by the differences between groups. For the 1st and 2nd functions, $p < 0.001$, and for the 3rd function, $p < 0.01$ were found to be significant. Table 6 shows the classification of the results obtained by DA according to regions, and also indicates the possible region group for samples where the group is unknown. In Table 6, the percentage of classification of the samples to the original regions was found to be 91.7%.

3.5. PCA

PCA creates combinations of features based on similarity rather than dissimilarity. In PCA, new variables called principal components (PCs) are created. The first PC (PC1) is the linear combination that explains the largest amount of variation in the data. The next PCs are generated sequentially, and each component explains most of the remaining variance. The PCs are uncorrelated, with PC1 explaining most of the total variability and PC2 explaining the remainder [17,29]. PCA analysis was used to show that ion concentrations in water samples can classify regions. The Wilks' Lambda and *F* values of ions were examined by

Table 3. ANOVA results for the effect of months and regions on anions and cations.

Months						Location				
Ions	SS	df	MS	F	S	SS	df	MS	F	S
F ⁻	0.247	11	0.022	0.736	0.697	0.527	3	0.176	9.465	0.000
Cl ⁻	302.637	11	27.512	0.255	0.990	3526.156	3	1175.385	78.081	0.000
NO ₂ ⁻	0.574	11	0.052	0.443	0.925	0.358	3	0.119	1.177	0.329
Br ⁻	0.324	11	0.029	0.650	0.774	0.419	3	0.140	3.998	0.013
NO ₃ ⁻	23.450	11	2.132	0.124	1.000	567.769	3	189.256	111.709	0.000
SO ₄ ²⁻	11500.334	11	1045.485	1.542	0.159	11815.107	3	3938.369	7.192	0.000
PO ₄ ³⁻	71377.367	11	6488.852	2.945	0.007	29365.841	3	9788.614	3.550	0.022
Li ⁺	0.000	11	0.000	0.218	0.995	0.001	3	0.000	44.000	0.000
Na ⁺	36.019	11	3.274	0.019	1.000	6314.432	3	2104.811	1062.632	0.000
K ⁺	2.818	11	0.256	0.322	0.976	24.124	3	8.041	48.382	0.000
Mg ²⁺	12.287	11	1.117	0.012	1.000	3191.994	3	1063.998	981.795	0.000
Ca ²⁺	413.463	11	37.588	0.025	1.000	52421.171	3	17473.724	349.007	0.000

SS: Sum of squares; MS: Mean square, S: Significant

Table 4. Pearson correlation analysis results for anions and cations.

Ions	F ⁻	Cl ⁻	NO ₂ ⁻	Br ⁻	NO ₃ ⁻	SO ₄ ²⁻	PO ₄ ³⁻	Li ⁺	Na ⁺	K ⁺	Mg ²⁺
Cl ⁻	0.426										
NO ₂ ⁻	0.091	0.147									
Br ⁻	0.324	0.489	0.271								
NO ₃ ⁻	0.408	0.938	0.260	0.475							
SO ₄ ²⁻	0.219	0.562	0.324	0.474	0.716						
PO ₄ ³⁻	0.292	0.349	-0.091	-0.128	0.254	-0.444					
Li ⁺	0.487	0.463	0.232	0.366	0.526	0.425	0.172				
Na ⁺	0.495	0.918	0.262	0.366	0.946	0.575	0.432	0.544			
K ⁺	0.546	0.815	0.286	0.467	0.815	0.505	0.342	0.645	0.858		
Mg ²⁺	0.482	0.908	0.257	0.359	0.952	0.615	0.397	0.561	0.992	0.843	
Ca ²⁺	0.496	0.859	0.267	0.367	0.935	0.652	0.353	0.613	0.973	0.824	0.976

DA, and the ions that contributed the most to the separation of regions were found to be Na⁺, Mg²⁺, and Ca²⁺. Therefore, PCA analysis was also performed using only these three cations. PCA score and loading plots generated with cation, anion, and triple cations are given in Figures 8, 9, and 10, respectively.

In Figure 8(a), PC1 (X-axis) and PC2 (Y-axis) explain 83.5% and 11.8% of the total variance, respectively, and 95.3% of the variance in total. When the distribution of the samples was analyzed, it was observed that there was a separation between the three regions, especially on the PC1 axis. However, in two regions where water is supplied from the same source, the desired separation was not realized. The water supply regions are clustered within themselves. It is seen that different water supply sources can be separated by using all the cations, but there will be no separation in the regions belonging to the same source. In the loading graphs of the PCA, there is a strong positive correlation between vectors located close to each other. In the loading graph shown in Figure 8(b), there is a very close relationship between Na⁺, Mg²⁺, and Ca²⁺. Because the angle between the Na⁺ and Mg²⁺ vectors is quite small, the correlation is strong. In contrast, the relationship between Li⁺ and the vectors of other metals is relatively weak. The strong relationship between Na⁺, Mg²⁺, and Ca²⁺ is thought to contribute significantly to the separation.

In Figure 9(a), PC1 (X-axis) and PC2 (Y-axis) explain 45.8% and 22.5% of the total variance, respectively, and 68.3% of the variance in total. When the distribution of the samples was analyzed, the separation between the regions was not sufficiently observed, and no clustering was found. By using all anions, it is seen that there will be no separation between different water supply sources and regions belonging to the same source. In the loading graph shown in Figure 9(b), a very close relationship is observed between the NO₂⁻/SO₄²⁻/Br⁻ anions and the

NO₃⁻/F⁻/Cl⁻ anions. Among the anions, the PO₄³⁻ vector has the largest angle with the other vectors.

In Figure 10(a), PC1 (X-axis) and PC2 (Y-axis) explain 98.7% and 1% of the total variance, respectively, and 99.7% of the variance in total. When the distribution of the samples is analyzed, it can be observed that there is a separation between the three regions, especially in the PC1 axis. In addition, the expected separation was realized in two regions where water was supplied from the same source. The water supply regions are also clustered within themselves. It is seen that different water supply sources can be separated by using Ca²⁺, Mg²⁺, and Na⁺ cations, and it is also seen that there will be separation in the regions belonging to the same source. In the Loading graph shown in Figure 10(b), there is a very close relationship between Na⁺ and Mg²⁺. Furthermore, PC2 values were found to be 0.815 for Ca²⁺, -0.344 for Mg²⁺, and -0.466 for Na⁺.

In a similar study in the literature, water samples were collected from different sources in three different years and seasons to evaluate water quality. The results of parameters such as pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), turbidity, total dissolved solids (TDS), hardness, Ca²⁺, Cl⁻, Fe²⁺, and SO₄²⁻ were determined by PCA analysis. This classification is reported to be useful for planners and field engineers to take precautions in advance to prevent groundwater contamination [30]. Both PCA and PLS-DA methods were used to quickly and accurately identify the sources of mine water flow. The parameters K⁺, Na⁺, Ca²⁺, Mg²⁺, Cl⁻, SO₄²⁻, HCO₃⁻, and TDS were used on 54 original water samples from three main water intake aquifers in the Coal Mine and could distinguish water samples from three different water sources. However, it was reported that the classification effect of PLS-DA was better than PCA because it could strengthen the difference in the chemical composition of water between different water sources [31].

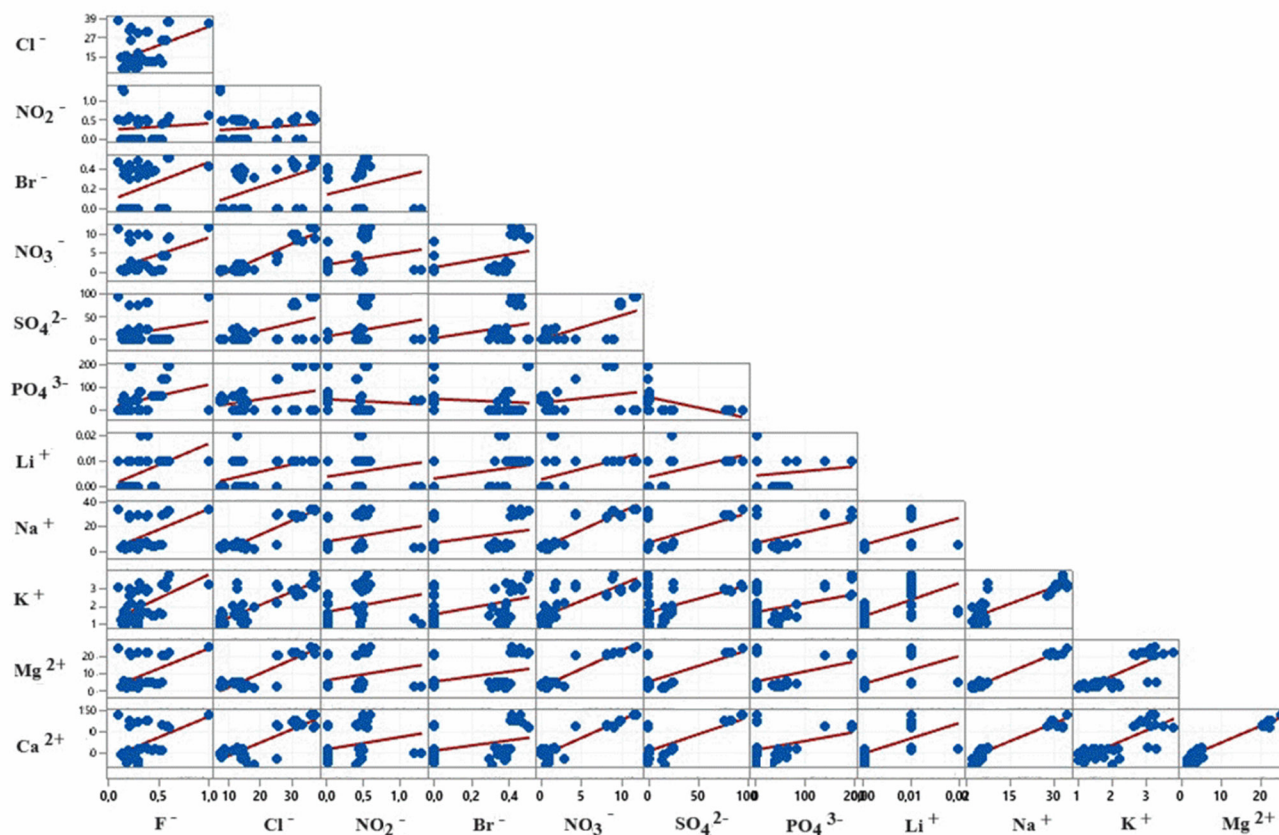


Figure 7. Matrix plot for all ions.

Table 5. Wilks' Lambda values of ions.

Ions	Wilks' Lambda	F	df1	df2	Sig.
F ⁻	0.608	9.465	3	44	0.000
Cl ⁻	0.158	78.081	3	44	0.000
NO ₂ ⁻	0.926	1.177	3	44	0.329
Br ⁻	0.786	3.998	3	44	0.013
NO ₃ ⁻	0.116	111.709	3	44	0.000
SO ₄ ²⁻	0.671	7.192	3	44	0.000
PO ₄ ³⁻	0.805	3.550	3	44	0.022
Li ⁺	0.250	44.000	3	44	0.000
Na ⁺	0.014	1062.632	3	44	0.000
K ⁺	0.233	48.382	3	44	0.000
Mg ²⁺	0.015	981.795	3	44	0.000
Ca ²⁺	0.040	349.007	3	44	0.000

Table 6. Classification results.

Original	Count	Location	Predicted group membership				Total
			A	B	C1	C2	
A	100	A	100	0	0	0	100
B	0	B	0	100	0	0	100
C1	0	C1	0	0	83.3	16.7	100
C2	0	C2	0	0	16.7	83.3	100

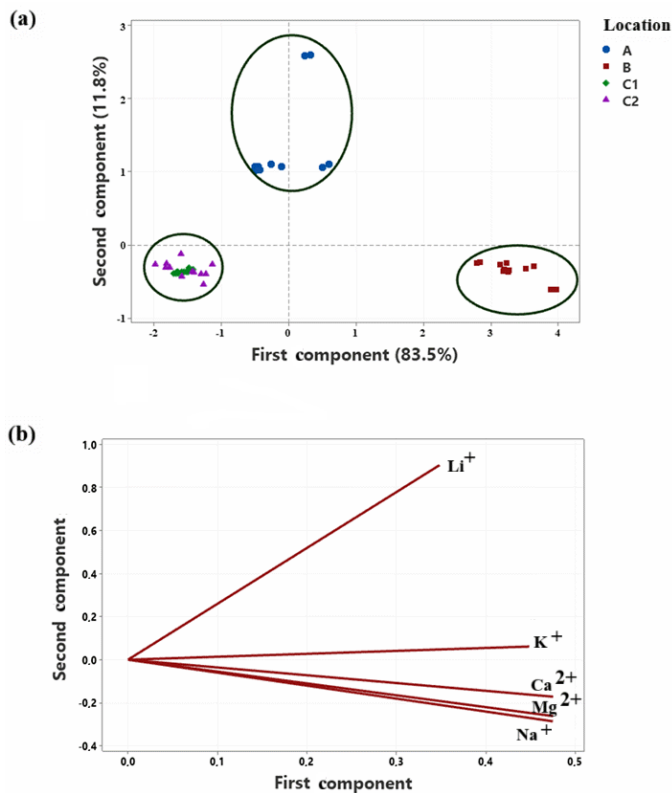


Figure 8. PCA Score plot (a) and loading plot (b) of all cations.

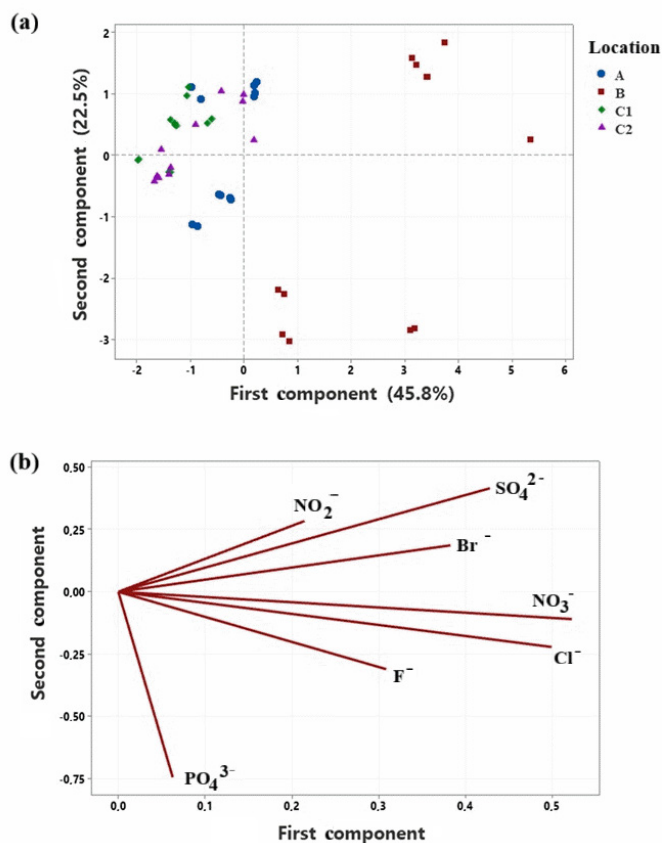


Figure 9. PCA Score plot (a) and loading plot (b) of all anions.

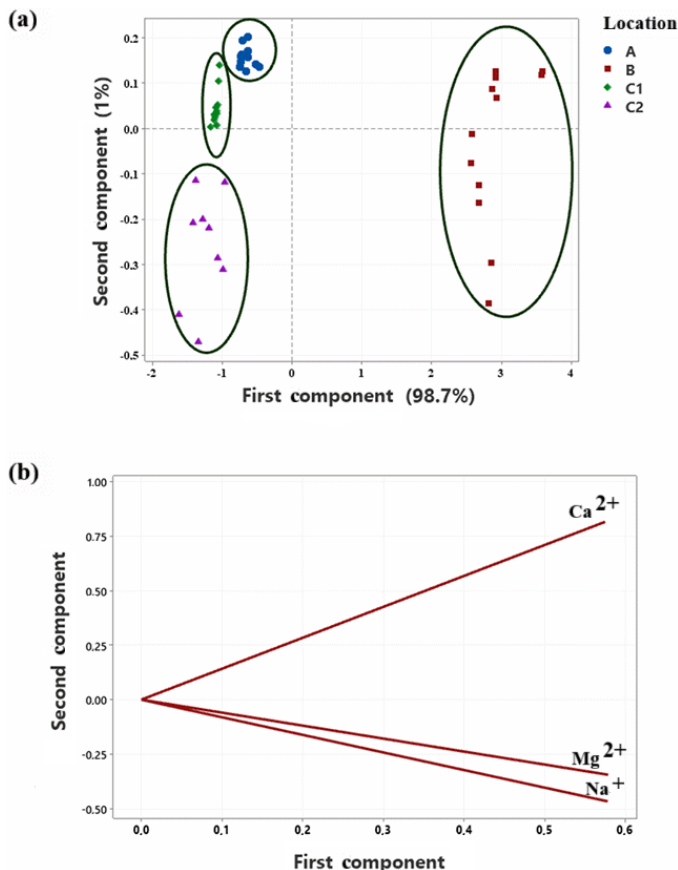
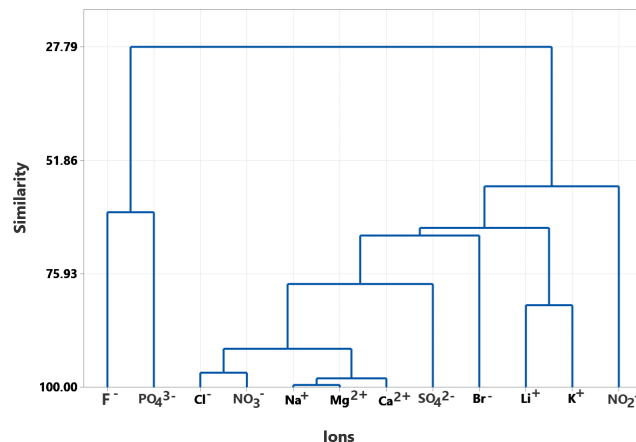
Figure 10. PCA Score plot (a) and loading plot (b) of Ca^{2+} , Mg^{2+} , and Na^{+} cations.

Figure 11. Dendrogram of HCA.

3.6. Hierarchical cluster analysis

HCA was performed to assess the similarity of anions and cations in water samples from all regions included in the study. The dendrogram of the results has been shown in Figure 11. When the HCA results of the anions and cations analyzed in the study were examined, Na^{+} , Mg^{2+} , and Ca^{2+} exhibited similar behavior at the first level. In the second level, these three cations behaved similarly to Cl^{-} and NO_3^{-} anions. F^{-} and PO_4^{3-} ions also exhibited similar behavior above approximately 50%. Similarly, Li^{+} and K^{+} exhibited similar behavior above 75%.

4. Conclusions

For a healthy and safe drinking water supply, it is very important to monitor and control the water from the source until it reaches the end consumer. Many factors affect water quality until it reaches the end consumer. It is very important to evaluate the pollutants in the raw water and the desired water quality together and to select the appropriate treatment method in this way. So, both human health will be protected and costs will be reduced by not using unnecessary chemicals. This study has shown that IC coupled with chemometric methods can be used to identify marker ions to differentiate water samples by their region of origin. By analyzing water samples from 4 regions, this work will add more to the existing knowledge on ionic composition and its variability due to geographical and seasonal influences. It is clear that Ca^{2+} , Mg^{2+} , and Na^{+} are the most important ions in distinguishing the water sources. Indeed, PCA and DA have assigned these parameters as very important for regional separation. This method has great potential in water quality monitoring and management. Identification of marker ions is a powerful tool for tracing the origin of drinking water for water resource management and more effective pollution control. This strategy meets modern environmental monitoring requirements in terms of cost-effective, efficient, and comprehensive water chemistry analysis compared to traditional methods. In short, IC with chemometric analysis is a powerful tool for regional water source discrimination. It will improve understanding of regional water chemistry and justify the use of targeted approaches to protect or improve water quality. More studies should be done to expand the method to other regions and water types and to ensure its applicability and efficiency in different environmental conditions.

CRedit authorship contribution statement

Batuhan Yardımcı: Manuscript preparation, Design, Experimental studies. **Sinem Çolak:** Statistical analysis, Data analysis, Manuscript preparation. **Jülide Yener:** Manuscript editing and review, Concepts, Design.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors confirm that there was no use of artificial intelligence (AI)-assisted technology for assisting in the writing or editing of the manuscript and no images were manipulated using AI.

Acknowledgment

The authors extend their gratitude to the Scientific Research Projects Commission of Zonguldak Bülent Ecevit University for their support (Project Number: 2022-72118496-01). They also thank Zonguldak Bülent Ecevit University, Science and Technology Application and Research Center (ARTMER) for their support.

References

- Michalski, R., 2006. Ion chromatography as a reference method for determination of inorganic ions in water and wastewater. *Critical Reviews in Analytical Chemistry*, **36**, 107-127. <https://doi.org/10.1080/10408340600713678>
- Jackson, P.E., 2001. Determination of inorganic ions in drinking water by ion chromatography. *TrAC Trends in Analytical Chemistry*, **20**, 320-329. [https://doi.org/10.1016/S0165-9936\(01\)00070-X](https://doi.org/10.1016/S0165-9936(01)00070-X)
- Fossen Johnson, S., 2019. Methemoglobinemia: Infants at risk. *Current Problems in Pediatric and Adolescent Health Care*, **49**, 57-67. <https://doi.org/10.1016/j.cppeds.2019.03.002>
- Srivastava, S., Flora, S.J.S., 2020. Fluoride in drinking water and skeletal fluorosis: A review of the global impact. *Current Environmental Health Reports*, **7**, 140-146. <https://doi.org/10.1007/s40572-020-00270-9>
- Alam, M.F., Dafader, N.C., Sultana, S., Rahman, N., Taheri, T., 2017. Physico-chemical analysis of the bottled drinking water available in the Dhaka city of Bangladesh. *Journal of Materials and Environmental Sciences*, **8**, 2076-2083
- Gray, N.F., 2008. Drinking water quality – Problems and solutions, in: 2nd ed. ISBN: 9780521702539. <https://doi.org/10.1017/CBO9780511805387>
- Yardımcı, B., 2024. From biowaste to a green colorimetric agent: Valorization of onion peel for spectrophotometric and smartphone-based environmental monitoring of Al(III) ions. *Sustainable Chemistry and Pharmacy*, **37**, 101391. <https://doi.org/10.1016/j.scp.2023.101391>
- Yardımcı, B., 2023. Spectrophotometric and smartphone-based facile green chemistry approach to determine nitrite ions using green tea extract as a natural source. *Sustainable Chemistry and Pharmacy*, **34**, 101175. <https://doi.org/10.1016/j.scp.2023.101175>
- Çolak, S., Akça, Ş.B., Marım, N., 2019. Evaluation of the irrigation waters of çaycuma district in terms of certain water parameters. *Journal of International Environmental Application and Science*, **14**, 37-45.
- Zeyrek, M., Ertekin, K., Kacmaz, S., Seyis, C., Inan, S., 2010. An ion chromatography method for the determination of major anions in geothermal water samples. *Geostandards and Geoanalytical Research*, **34**, 67-77. <https://doi.org/10.1111/j.1751-908x.2009.00020.x>
- Varão Moura, A., Da Silva, J.É.D.S., Gubert, P., 2022. Ion chromatography: Principles and instrumentation. *Orbital: The Electronic Journal of Chemistry*, **14**, 110-115. <https://doi.org/10.17807/orbital.v14i2.15871>
- Lachenmeier, D.W., Richling, E., López, M.G., Frank, W., Schreier, P., 2005. Multivariate analysis of FTIR and ion chromatographic data for the quality control of tequila. *Journal of Agricultural and Food Chemistry*, **53**, 2151-2157. <https://doi.org/10.1021/jf048637f>
- Lachenmeier, D.W., Attig, R., Frank, W., Athanasakis, C., 2003. The use of ion chromatography to detect adulteration of vodka and rum. *European Food Research and Technology*, **218**, 105-110. <https://doi.org/10.1007/s00217-003-0799-8>
- Biedunkova, O., Kuznetsov, P., 2025. Liquid Ion chromatographic determination of soluble ions in water: Comparison of greenness and comprehensive assessment of irrigation suitability. *Water, Air, & Soil Pollution*, **236**, 315. <https://doi.org/10.1007/s11270-025-07975-3>
- Tsakovski, S., Simeonov, V. Chemometrics as a tool for treatment processing of multiparametric analytical data sets. In: *Analytical chemistry, Analytical measurements in aquatic environments*: CRC Press), 369-387. <https://dergipark.org.tr/en/download/article-file/780994>
- Özdamar, K., 2015. Statistical data analysis with package programs. Nisan Bookstore
- Jolliffe, I., 2002. Choosing a subset of principal components or variables. In: *Principal component analysis. Springer Series in Statistics*. Springer, New York, NY. https://doi.org/10.1007/0-387-22440-8_6
- Biru, E.I., Necolau, M.I., Zainea, A., Iovu, H., 2022. Graphene oxide–protein-based scaffolds for tissue engineering: Recent advances and applications. *Polymers*, **14**, 1032. <https://doi.org/10.3390/polym14051032>
- Li, B., Wu, Q., Liu, Z., 2020. Identification of mine water inrush source based on PCA-FDA: Xiandewang coal mine case. *Geofluids*, **2020**, 1-8. <https://doi.org/10.1155/2020/2584094>
- Farnham, I.M., Stetzenbach, K.J., Singh, A.K., Johannesson, K.H., 2000. Deciphering groundwater flow systems in oasis valley, Nevada, using trace element chemistry, multivariate statistics, and geographical information system. *Mathematical Geology*, **32**, 943-968. <https://doi.org/10.1023/a:1007522519268>
- Google Earth. Location map of the sampling area. [Last accessed 2025 Sep 21]. <https://earth.google.com>
- WHO, 2004. Sulfate in drinking-water background document for development of WHO guidelines for drinking-water quality. [Last accessed 2025 Sep 27]. https://cdn.who.int/media/docs/default-source/wash-documents/wash-chemicals/sulfate.pdf?sfvrsn=b944d584_4
- Gelete, G., Gokcekus, H., Ozsahin, D. U., Uzun, B., Gichamo, T., 2020. Evaluating disinfection techniques of water treatment. *Desalination and Water Treatment*, **177**, 408-415. <https://doi.org/10.5004/dwt.2020.25070>
- Thokchom, B., Radhapyari, K., Dutta, S. Occurrence of trihalomethanes in drinking water of Indian states: A critical review. In: *Disinfection By-products in Drinking Water*: Elsevier), 83-107. <https://doi.org/10.1016/B978-0-08-102977-0.00004-4>
- Kuznetsov, P., Biedunkova, O., 2025. Application of multivariate statistical techniques for assessing spatiotemporal variations of heavy metal pollution in freshwater ecosystems. *Water Conservation Science and Engineering*, **10**, 13. <https://doi.org/10.1007/s41101-025-00341-8>
- Ratnaweera, H., Gjessing, E., Oug, E., 1999. Influence of physical-chemical characteristics of natural organic matter (NOM) on coagulation properties: An analysis of eight norwegian water sources. *Water Science and Technology*, **40**, 89-95. <https://doi.org/10.2166/wst.1999.0450>
- Zou, K.H., Tuncali, K., Silverman, S.G., 2003. Correlation and simple linear regression. *Radiology*, **227**, 617-622. <https://doi.org/10.1148/radiol.2273011499>
- Tabachnick, B.G., Fidell, L.S., Ullman, J.B., 2019. Using multivariate statistics. ISBN-10: 0-13-479054-5, Boston: Pearson Chapter 14.
- Jolliffe, I., 2005. Principal component analysis, in: encyclopedia of statistics in behavioral science, John Wiley Sons, p. 2208. <https://doi.org/10.1017/CBO9780511805387>
- Mahapatra, S.S., Sahu, M., Patel, R.K., Panda, B.N., 2012. Prediction of water quality using principal component analysis. *Water Quality, Exposure and Health*, **4**, 93-104. <https://doi.org/10.1007/s12403-012-0068-9>
- Bi, Y., Wu, J., Zhai, X., Shen, S., Tang, L., Huang, K., Zhang, D., 2021. Application of partial least squares-discriminate analysis model based on water chemical compositions in identifying water inrush sources from multiple Aquifers in Mines. *Geofluids*, **2021**, 1-17. <https://doi.org/10.1155/2021/6663827>