

Physico-chemical parameters and PCA approach to assess river water quality – a case study (Albania)

Belinda Hoxha^{1}, Bederiana Shyti², Blerina Papajani³, Armela Mazrreku¹,
Marilda Osmani¹, Lule Basha⁴*

¹Department of Chemistry, Faculty of Natural Sciences, University of Elbasan, ALBANIA

²Department of Mathematics, Faculty of Natural Sciences, University of Elbasan, ALBANIA

³Department of Physics, Faculty of Natural Sciences, University of Elbasan, ALBANIA

⁴Department of Applied Mathematics, Faculty of Natural Sciences, University of Tirana, ALBANIA

*Corresponding author: belinda.hoxha@uniel.edu.al

Abstract. Water quality refers to the physical, chemical and biological characteristics of water that determine its suitability for a specific use, such as drinking, industry, irrigation or supporting aquatic life. It is assessed by comparing water's properties against established standards for various uses. Factors influencing water quality include natural elements like weather and geology, and human activities like pollution and land use. Water Quality Index (WQI) is a widely used tool for summarizing and communicating water quality information. It is a valuable and unique statistical approach that consolidates the experimental results of various physico-chemical parameters into a single comprehensive and practical term in order to represent the overall quality status of water. The aim of the study is to demonstrate the ability of statistical methods in water quality prediction, specifically by identifying the most important parameters that influence the Water Quality Index (WQI). This involves the use of statistical techniques to analyze water quality data and identify the main factors that influence the overall water quality, potentially leading to a more efficient and accurate water quality management. The study takes in consideration physico-chemical parameters analyzed in Shkumbini River in Albania. The laboratory data from six sampling points during four years are gathered and analyzed based on water quality standards and statistically with PCA (Principal Component Analysis). The chosen parameters to evaluate water quality are TDS, GH, BOD, pH, DO, Cl, HCO₃⁻ and thermotolerant coliforms, which are also used to monitor the suitability of the PCA method in the determination of WQI. The physico-chemical parameters were evaluated against international water standards. Additionally, the PCA method showed that the order of indicators determining the WQI depends on the distance between the variables and the origin. The study reveals that using the PCA method, the recommended nine parameters are sufficient to determine the WQI value, and the cumulative proportion of Component 1, Component 2 and Component 3 explains nearly 63% of the total variance.

Key words: water quality, Shkumbini River, Water Quality Index model, Principal Component Analysis.

Introduction

Water is an essential resource necessary for the survival of all living organisms on Earth. Consequently, the water quality is crucial for sustaining life and for various human activities such as agriculture, industry and recreation (Basha et al., 2024). The assessment of water quality, there-

fore, becomes an essential task for evaluating both the health of the environment and the well-being of the population that depends on it. However, due to increasing industrialisation, urbanisation and pollution from a wide range of sources, both surface and groundwater resources are increasingly at risk of contamination. Natural and anthro-

pogenic activities contribute significantly to water pollution, which poses a threat to ecosystems, human health and economic productivity (Lin et al., 2022; Kucaj et al., 2022). The control and management of water resources requires a multifaceted approach, considering institutional frameworks, regulatory measures, financial capacity and technical expertise (Basha et al., 2024). Agriculture, for instance, depends heavily on water quality to ensure crop safety and productivity, while polluted water sources can degrade aquatic ecosystems and harm biodiversity (Syafudin et al., 2021). Additionally, the treatment of polluted water is often expensive and demands sophisticated technologies (Palansooriya et al., 2020; Osmani et al., 2023).

Water pollution is commonly characterised by alterations in physical, chemical, and biological properties, which can adversely affect both human populations and aquatic environments (Picinini-Zambelli et al., 2025). As such, comprehensive water quality assessments, utilising a variety of methods to measure specific pollutants and contaminants, are crucial for understanding the extent of pollution and guiding effective remediation strategies (Altenburger et al., 2019).

The Water Quality Index (WQI), developed in the early 1970s, serves as a tool to assess the health of water resources. It provides a numerical representation of water quality, allowing for the tracking and analysis of changes in water quality over time (Brown et al., 1972). The WQI is widely regarded as one of the most practical methods for simplifying complex water quality data into an easily understandable score, which can be used for various applications (Uddin et al., 2022). Typically, the WQI ranges from 0 to 100, with higher values signifying better water quality (Uddin et al., 2021).

The WQI is a composite measure that integrates biological, chemical, and physical characteristics of water, taking into account its intended uses and relevant water quality standards (Khalil et al., 2011). Water quality can be assessed by focusing on a specific parameter for a particular purpose, or by selecting critical parameters that represent the overall pollution level of the water body (Gazzaz et al., 2012). The latter approach reflects the broader status of the water quality, making it more applicable for diverse assessments.

Over the years, researchers and environmental agencies across the globe have developed, refined, and validated various WQI models (Uddin

et al., 2021). The calculation of the WQI index simplifies complex data, making it accessible for the general public and decision-makers alike. It facilitates comparisons of water quality over time and across different locations, communicates potential risks to human health and the environment, and aids policymakers in prioritising actions to improve water quality (Guenouche et al., 2024).

Albania is blessed with abundant water resources, featuring a vast network of rivers, lakes, and underground water reserves. The country is home to over 150 rivers and streams, flowing predominantly from east to west, which support urban areas, agriculture, aquaculture, recreation, energy production, and industry (Basha et al., 2024). Despite its abundance of water resources, Albania faces increasing water demand due to economic development, industrialisation, and population growth. The Shkumbini River, located in central Albania, is a significant water body that has become a focal point for water quality monitoring, particularly due to its location in an industrial zone where pollution is a concern. Its quality has become an ongoing environmental issue, with attention given to pollution sources, including industrial and agricultural discharges, as well as untreated sewage (Shyti et al., 2024). Efforts have been made to address the pollution of the Shkumbini River, with various initiatives implemented to mitigate the impact of contaminants (Basha et al., 2024).

The Water Quality Index (WQI), considered in many research studies, has been applied to monitor and assess the water quality in different regions of Albania. Some studies have calculated the WQI using data collected from multiple river locations, while others have focused on specific pollution parameters (Shyti et al., 2024). For instance, Damo & Icka (2013) studied the water quality index for drinking water in the city of Pogradec, finding that the CCME WQI value of 87.81 indicated that the water quality was “good,” with turbidity being the primary issue affecting the water. Zela et al. (2020) conducted a five-year monitoring program of the Seman River in southern Albania, assessing 14 water quality parameters and applying the CCME WQI through a multivariable approach. Similarly, Gega et al. (2022) emphasised the pollution of the Shkumbini River from untreated sewage discharges and the deposition of industrial waste and agricultural

runoff, which contributed to the river's water quality deterioration.

Recent studies, such as that by Gjeci et al. (2024), found that the water quality in the Shkumbini River was generally good, with WQI values ranging from 71 to 90. This was attributed to low water temperatures and high precipitation during the monitoring period. Despite these positive findings, the complexity of water quality data poses challenges for effective interpretation, as the vast number of parameters involved complicates the WQI calculation. To address this, various statistical techniques, such as Principal Component Analysis (PCA) and Factor Analysis, have been applied to simplify the data and reduce subjectivity in WQI modelling (Roy & Roy, 2024). PCA is particularly useful for reducing the dimensionality of complex datasets while retaining essential information, making it an efficient and cost-effective method for selecting critical water quality parameters (Kopeck et al., 2018; Shuquan et al., 2015; Benkov et al., 2023; Aminu et al., 2023).

Multiple Linear Regression (MLR) was previously applied to assess the WQI of the Shkumbini River. Using MLR, we defined an equation for the WQI based on three key variables: Biological Oxygen Demand (BOD), Hydrogen Carbonate (HCO_3), and Total Phosphorus (TP), with an R^2 value of 99.6%, demonstrating a high level of explanatory power (Shyti et al., 2024). Additionally, machine-learning models, including XGBoost, Random Forest, K-Nearest Neighbour, and Naive Bayes, were used to predict water quality, achieving high predictive accuracies (98.61%, 94.44%, 91.22%, and 94.45%, respectively). The models highlighted BOD, HCO_3 , and Total Phosphorus as the most significant factors influencing the Shkumbini River's water quality, with a strong positive correlation ($r = 0.85$) identified between BOD and WQI (Basha et al., 2024).

This study aims to evaluate water quality data from the Shkumbini River using physico-chemical parameters and PCA, a method known for its ability to identify the most influential water quality parameters and reduce the complexity of data interpretation. The findings of this study are expected to support previous conclusions regarding the critical factors influencing water quality in the Shkumbini River, while also providing a more efficient and cost-effective approach to WQI calculation.

Materials and methods

Shkumbini River originates from the Valamar mountain located in the southwest of Albania and flows for approximately 181 km before reaching its delta in the Adriatic Sea. It is one of the most important rivers in the central region of Albania, passing through diverse landscapes such as forests, valleys, and plains (Pano, 1985; Bekteshi et al., 2023). The river supports local agriculture and its water is often used for irrigation in surrounding areas (Miho et al., 2005), thus influencing the environment and economy of the region (Bani et al., 2020). It drains an area of 2,444 km² with a flow direction primarily from east to west and discharges an average of 61.5 m³/second into the Adriatic Sea (Miho et al., 2005). Along its course, Shkumbini River passes through areas rich in minerals, including chrome, nickel and iron, particularly in the industrial area of Elbasan, which has significant industrial activity related to metallurgy of these minerals, as well as cement production (Gjeci et al., 2024). The urban and industrial areas in Elbasan, as well as agricultural practices, significantly affect the river water quality, introducing various manmade materials and thus compromising the ecological health of the river (Bani et al., 2020; Hoxha et al., 2023).

Six monitoring stations along the Shkumbini River, starting from its originating area to its delta in the Adriatic Sea - Qukës, Librazhd, Xibrakë, Papër, Bishqem and Rrogozhinë - were selected for a comprehensive study of water quality over four years, spanning from 2020 to 2023. The selection of these stations was based on various factors, including geographical location, the potential introduction of pollutants from agricultural, urban and industrial activities, as well as considerations of erosion conditions and other environmental variables.

In order to ensure the precision and reliability of water quality assessment, water samples were collected using 2 L, 1.5 L and 0.5 L containers. The sampling protocols adhere to the methodologies outlined by USDA (Musselman, 2012), as well as the standard in ISO 5667-6 (ISO 2014) and ISO 5667-5 (ISO 2006), which precise the procedures for the collection and handling of water samples. Afterwards, the samples were analysed at the laboratory of the Regional Directorate of Public Health in Elbasan, Albania, according to APHA (1998).

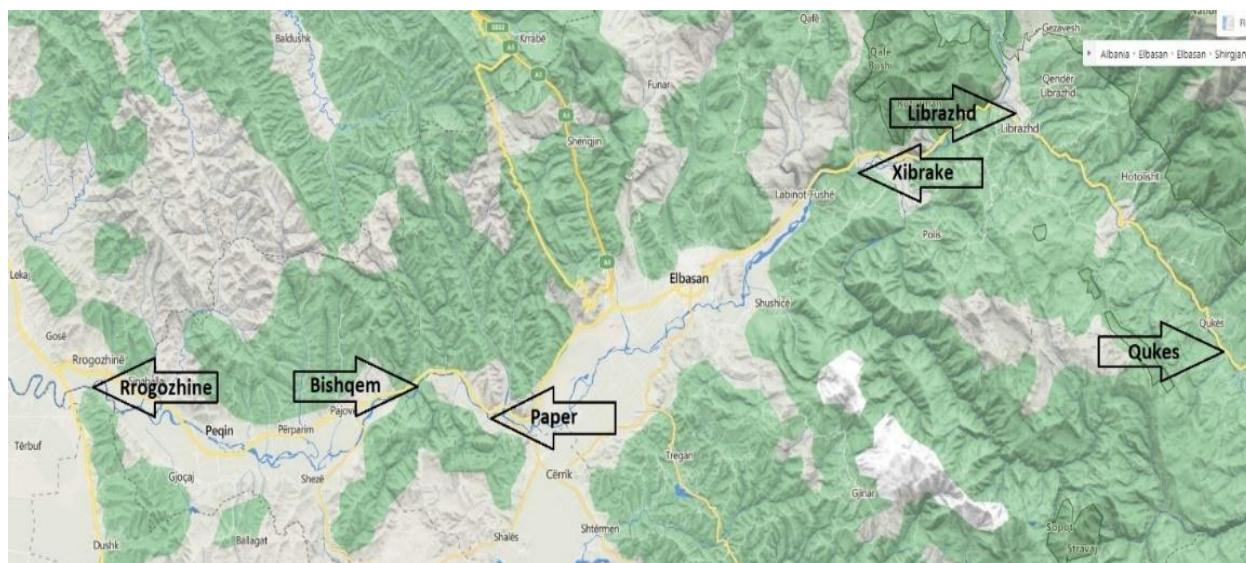


Fig. 1. Map of Sampling Stations Along the Shkumbini River (Shyti et al., 2024).

Principal Component Analysis (PCA) was employed to interpret and group the parameters in a dataset that describe key physicochemical water quality indicators, such as: pH, General Hardness (GH), Total Dissolved Solids (TDS), Biochemical Oxygen Demand (BOD), Dissolved Oxygen (DO), Chloride (Cl), Total Phosphorus (TP), Thermotolerant Coliforms, and Bicarbonate (HCO_3). Each of these parameters plays a significant role in assessing water quality:

- pH serves as an index to indicate the degree of pollution, particularly in cases of contamination by acidic or alkaline wastes (Saalidong et al., 2022; EPA, 2001);

- Water hardness is attributable to the presence of carbonates, bicarbonates, sulfates, calcium, magnesium, and silicates (EPA, 2001);

- High TDS levels are often associated with pollution from industrial discharges and agricultural runoff (WHO, 2003);

- Biochemical Oxygen Demand (BOD) represents the biological oxygen demand due to the decomposition of organic material, which is indicative of organic pollution (Ward et al., 2002; Tchobanoglous et al., 2003; WHO, 2004);

- Low DO levels are expected in polluted waters due to organic matter decomposition, which consumes oxygen (Rajwa et al., 2014);

- Chloride concentrations in freshwater are primarily influenced by evaporation and precipitation, with studies showing that high chloride

levels can be an indicator of pollution in aquatic environments (WHO, 2011; Granato et al., 2015; Guedens et al., 2018);

- Elevated phosphorus levels, often due to agricultural runoff and wastewater discharge, can lead to excessive algae growth and depletion of oxygen in aquatic ecosystems (Vighi & Chiaudani, 1985; Bennett et al., 2001; Johnes, 2007);

- Thermotolerant Coliforms are bacteria commonly used as indicators of faecal contamination in water. Their presence suggests the potential for pathogenic microorganisms, posing a health risk to humans and aquatic life (Edberg et al., 2000; Stevens et al., 2003; Odonkor & Ampofo, 2013);

- High concentrations of bicarbonates typically indicate harder waters and may influence the solubility of other minerals (Grochowska, 2020).

According to the relative significance that each parameter has in determining the water quality, the nine parameters were each assigned a specific weight values (mg/L), as follows: Total Dissolved Solids (TDS) = 6.5 mg/L; General Hardness (GH) = 20 mg/L; Biological Oxygen Demand (BOD) = 1.5 mg/L; pH = 7.5 mg/L; Dissolved Oxygen (DO) = 5.8 mg/L; Chloride (Cl) = 20 mg/L; Bicarbonate (HCO_3) = 200 mg/L; Total Phosphorous (TP) = 0.09 mg/L; and Thermotolerant Coliforms = 600 mg/L. The dataset includes these nine physicochemical indicators, along with their corresponding water quality index values and classification (Basha et al., 2024).

Results and Discussion

WQI calculation

The use of the Water Quality Index (WQI) dates back to the 1960s. The first application of WQI in water quality categorisation was made by Horton in 1965, and followed by Brown in 1970, who developed a general index for evaluating water quality. In 1982, Steinhart et al. introduced a novel environmental quality index to summarise technical information on the status and trends in the Great Lakes ecosystem, and the Water Quality Guidelines Task Group of the Canadian Council of Ministers of the Environment introduced WQI in Canada in the mid-1990s (Uddin et al., 2017).

Over time, the use of WQI has expanded, particularly in developed countries, with many researchers applying the index to represent the

quality status of water following monitoring and analysis (Uddin et al., 2021).

The water quality index (WQI) used in this study is based on nine parameters: pH, TDS, GH, BOD, DO, Cl, HCO₃, TP and Thermotolerant Coliforms. A higher WQI indicates better quality of water. Typically, a WQI model is composed of four components:

- (i) parameter selection
- (ii) sub-indexes process
- (iii) weighting technique
- (iv) aggregation method (Uddin et al., 2021)

In 2022 Uddin et al. proposed a new classification scheme for assessing coastal water quality using the WQI model (Uddin et al., 2022), and in 2023 they proposed a model for water quality rating (Uddin et al., 2023).

Table 1. Water Quality Index (WQI) Classification Scheme for Coastal and River Waters (Uddin et al., 2022).

Classifications scheme	Range of score	Descriptions
(i) Good	80 - 100	Good waterbodies are those that meet the guidelines' values. Water quality is maintained and is suitable for all uses.
(ii) Fair	50 - 79	Waterbodies that have a few indicators meet the guidelines' values; water quality is usually protected with a minor degree of impairment.
(iii) Marginal	30 - 49	The majority of water quality indicators failed to meet the criteria; water quality is unprotected, which may pose a risk for aquatic life.
(iv) Poor	0 - 29	Poor waterbodies are those that fail to meet all of the criteria. Water quality is completely unprotected and unsuitable for many specific uses

As shown in the validation Table 1, the final output of a WQI is a numerical value ranging from 0 to 100. This method is accompanied by a classification scheme that categorises water quality into distinct classes, ranging from "excellent" to "poor". By utilising this scheme, a single numerical value is provided as a representative indicator of the overall water quality status, effectively summarising the data set into a comprehensive measure.

Brown et al. (1972) proposed a method of calculation to increase sensitivity, using the following formulas:

$$WQI = \sum W_i q_i \quad (1)$$

$$W_i = \frac{w_i}{\sum w_i} \text{ and } Q_i = 100 \times \frac{V_i}{S_i} \quad (2)$$

where: W_i is the unit weight;

S_i - recommended standard value of the i^{th} parameter;

Q_i - sub-quality index of the i^{th} parameter or the quality rating scale of each parameter, and for the calculation of its numerical value, we are based on Brown et al. (1972);

V_i - monitored value of the i^{th} parameter.

The WQI, calculated according to the collected data, is then compared with the standards, provided in Table 2.

Table 2. Calculated Water Quality Index (WQI) Values at Six Sampling Stations Along the Shkumbini River (2020–2023).

N.	Period	The sampling points					
		Qukës	Librazhd	Xibrakë	Papër	Bishqem	Rrogozhinë
1	March 2020	66.91	60.09	81.77	75.44	73.82	120.89
2	July 2020	58.93	56.50	74.81	73.72	78.77	109.24
3	October 2020	68.24	88.57	74.81	87.48	75.31	123.73
4	March 2021	67.94	60.97	60.83	92.77	71.82	125.38
5	July 2021	90.02	92.84	87.66	79.83	75.60	114.10
6	October 2021	61.52i	70.74	75.24	81.52	88.49	137.78
7	March 2022	87.48	83.58	76.41	56.58	60.60	126.74
8	July 2022	81.08	96.99	77.53	78.99	71.67	103.40
9	October 2022	68.70	77.51	80.36	76.11	82.07	112.59
10	March 2023	88.26	80.83	78.29	74.07	77.54	112.69
11	July 2023	71.25	86.55	88.03	88.14	82.84	122.72
12	October 2023	67.05	85.48	79.85	91.24	88.33	114.50

Principal Component Analysis (PCA)

PCA is a widely used statistical technique for reducing the dimensionality of data, while retaining as much of the variability as possible (Jolliffe, 2002). By reducing the number of variables, PCA simplifies complex datasets, which facilitates data visualisation and exploration, and enhances the efficiency of subsequent analysis. The primary advantage of PCA lies in its ability to remove redundant or extraneous variables, thus streamlining the dataset and making it more manageable (Wold et al., 1987). In PCA, the principal components are new variables formed as linear combinations of the original variables, designed to capture the maximum variance within the data (Jolliffe, 2002). In this study, nine variables are considered, and PCA analysis is implemented to allocate the majority of the information into the first component, with progressively smaller amounts of variance being captured in the subsequent components, thereby creating a reduced set of components that best represent the original dataset (Zitko, 1994).

The statistical analysis process for the data using Principal Component Analysis (PCA) begins with the construction of the initial data matrix, which consists of n rows and p columns, where n represents the number of data points and p denotes the number of water quality parameters (Jolliffe, 2002). Following this, the collected data

are standardised through z-score standardisation, which is z calculated using the formula:

$$z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s_j}$$

where z_{ij} is the standardised variable, x_j is the mean value of the j parameter, and s_j is the standard deviation for the j^{th} parameter.

This is a crucial step for ensuring that correlations between variables are identified during the PCA. To assess the suitability of the dataset for PCA, the Kaiser–Meyer–Olkin (KMO) test and Bartlett’s test of sphericity are performed. The KMO test evaluates the sampling adequacy, while the Bartlett’s test assesses the homogeneity of the correlation matrix (Bartlett, 1954; Kaiser, 1974). Afterwards, the covariance matrices computed to identify correlations between the variables, and subsequently followed by the calculation of eigenvalues and eigenvectors, to identify the principal components (Jolliffe, 2002).

The next step in PCA involves the creation of a feature vector to determine which principal component should be retained. The final step reorients the data from the original axes to the new axes defined by the principal components. Each principal component explains a percentage of the total variance within the data set, with the first component accounting for the highest variance (Jolliffe, 2002).

To select the appropriate parameters for PCA analysis, the following criteria were employed:

1. The factors that carry the largest amount of the variance of the data are selected.
2. Only those principal components with absolute loadings greater than 30 are selected for further analyses, while components with lower loadings are discarded, as they contribute less to the overall variance (Fraino, 2023).

The plot in Fig. 2 shows the eigenvalues in a downward curve, from highest to lowest. In the Cumulative Proportion (Fig. 2), the first principal component explains almost 26% of the total

variance. The second one explains 21% of the total variance. The cumulative proportion of Comp.1, Comp.2 and Comp3 explains nearly 63% of the total variance.

From the PCA analysis (Fig. 3), it is observed that the nine principal components (PC1 to PC9) have been generated, corresponding to the number of variables in the dataset. For example, PC9 shows a positive correlation with both BOD and HCO_3 , indicating that as the values of BOD and HCO_3 increase, PC9 also increases. This positive correlation suggests that these two parameters contribute significantly to the variance captured by PC9.

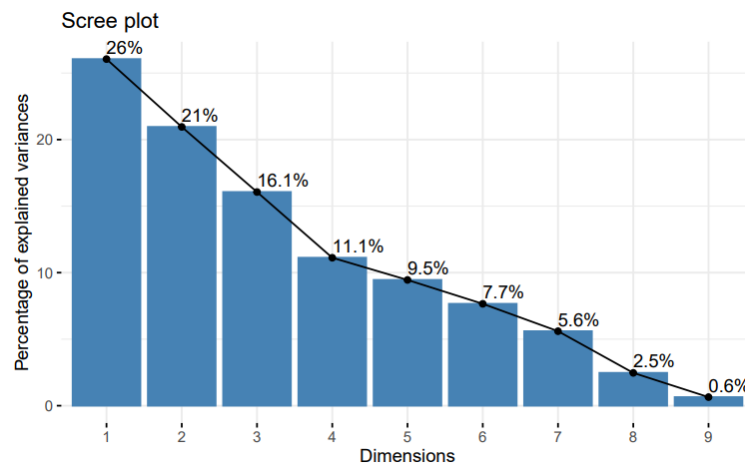


Fig. 2. Scree Plot Showing Eigenvalues and Cumulative Variance Explained by Principal Components.

Variables	Components									Total
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	
BOD	0.31	0.08	0.01	0.00	0.05	0.00	0.08	0.01	0.47	1.00
Cl	0.05	0.19	0.21	0.00	0.04	0.12	0.05	0.33	0.00	1.00
Coliforms	0.06	0.08	0.34	0.00	0.01	0.00	0.32	0.18	0.00	1.00
DO	0.07	0.06	0.03	0.09	0.55	0.12	0.08	0.00	0.00	1.00
GH	0.00	0.11	0.23	0.10	0.15	0.10	0.30	0.00	0.00	1.00
HCO ₃	0.31	0.11	0.00	0.00	0.03	0.00	0.03	0.00	0.52	1.00
p-Total	0.02	0.00	0.11	0.68	0.00	0.07	0.11	0.00	0.00	1.00
pH	0.01	0.15	0.08	0.12	0.16	0.48	0.01	0.00	0.00	1.00
TDS	0.17	0.21	0.00	0.01	0.01	0.10	0.02	0.47	0.00	1.00

Fig. 3. Principal Components Derived from Nine Physico-Chemical Water Quality Parameters.

From the loading matrix, it is evident that the first principal component (PC1) explains 26% of the total variance in the dataset. The parameters contributing to PC1 include BOD (0.31), HCO_3 (0.31) and TDS (0.17), all of which exhibit moderate factor

loadings. In contrast, parameters such as Cl, Thermotolerant Coliforms, DO, GH, TP, and pH show weak factor loadings.

For the second principal component (PC2), which accounts for 21% of the total variance, the

parameters Cl (0.19) and TDS (0.21) exhibit moderate contributions. Similarly, the third principal component (PC3) explains 16.1% of the variance, with Thermotolerant Coliforms (0.34) and GH (0.23) contributing significantly. This pattern of parameter contributions continues for the remaining components.

As can be observed from the biplot in Fig. 4:

1. Variables that are grouped exhibit positive correlation with each other. For example, HCO_3 and BOD are positively correlated. This finding is particularly noteworthy, as these parameters have the highest values in the loading matrix to the ninth principal component (PC9).

2. The greater the distance between a variable and the origin in the biplot, the more effectively that variable is represented by the principal components.

3. Variables that are negatively correlated are positioned on opposite sides of the biplot's origin.

In summary, PCA has extracted information into three factors that are principal components PC1, PC2 and PC3, which together account for approximately 63% of the total variance in the dataset.

This conclusion aligns with the findings of previous studies (Basha et al., 2024; Shyti et al., 2024), where it was demonstrated through various methods that only three of the variables - BOD, TP and HCO_3 - played a significant role in the calculation of the WQI for the water quality assessment of the Shkumbini River. Based on these results, it can be confirmed that PCA is a reliable technique for analysing the water quality of Shkumbini River.

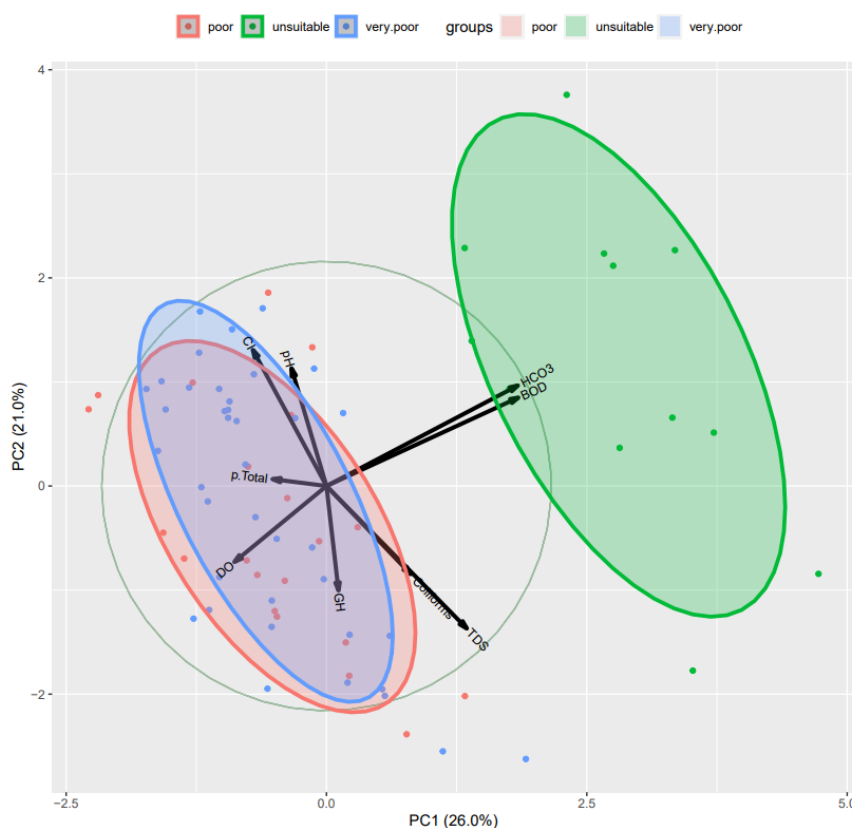


Fig. 4. Biplot of Principal Components 1 and 2 Showing Correlation Among Water Quality Parameters.

Conclusions

As one of the primary rivers in Albania, the Shkumbini River holds significant ecological and economic importance, making the study of its water quality a critical task. This study contributes to the ongoing research focused on the river,

aiming to assess and utilise PCA as an effective tool for identifying the key parameters that influence the WQI model. PCA is a powerful technique that generates new, uncorrelated variables, effectively eliminating components with low variance.

This technique identifies new directions, known as principal components, along which the data exhibit the most significant changes. By projecting the data onto these components, PCA reduces the dimensionality of the dataset, making it easier to process and analyse.

This study demonstrates that the first three principal components together explain approximately 63% of the total variance, and precisely PC1 accounts for 26%, PC2 accounts for 21% and PC3 contributes 16.1%.

The results of this research, alongside other studies on the water quality of the Shkumbini River, provide valuable insights for decision-making bodies. These findings serve as a supportive platform for policymakers, aiding in the development of strategies to improve the water quality of the Shkumbini River.

References

- Altenburger, R., Brack, W., Burgess, R.M., Busch, W., Escher, B.I., Focks, A., & Krauss, M. (2019). Future water quality monitoring: Improving the balance between exposure and toxicity assessments of real-world pollutant mixtures. *Environmental Sciences Europe*, 31(1), 1–17. doi: [10.1186/s12302-019-0193-1](https://doi.org/10.1186/s12302-019-0193-1)
- Aminu, I., Azimah, I., Hafizan, J., Aisha, B.I., Balarabe, T.W., Mustapha, M., & Hassan, A. (2023). Water quality modelling using principal component analysis and artificial neural network. *Marine Pollution Bulletin*, 187, 114493. doi: [10.1016/j.marpolbul.2022.114493](https://doi.org/10.1016/j.marpolbul.2022.114493)
- APHA (1998). *Standard methods for the examination of water and wastewater*. 20th edition, American public health Association, American water works association, Water Environment Federation, Washington DC.
- Bani, A., Shumka, S., Dervishi, O., Duka, I., Kristo, I., & Malollari, I. (2020). Water quality and biodiversity of Shkumbini River. *Journal of Environmental Protection and Ecology*, 21, 2045–2053.
- Bartlett, M.S. (1954). A note on the multiplying factors for various chi-squared approximations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 16(2), 296–298.
- Basha, L., Shyti, B., & Bekteshi, L. (2024). Evaluating the performance of machine learning approaches in predicting Albanian Shkumbini River's waters using water quality index model. *Journal of Environmental Engineering and Landscape Management*, 32(2), 117–127. doi: [10.3846/jeelm.2024.20979](https://doi.org/10.3846/jeelm.2024.20979)
- Bekteshi, L., Hoxha, B., Gega, N., Karamelo, P., & Dauti, A. (2023). *Quality assesment of Shkumbini River water based on physico-chemical parameters*. Proceedings from the 6th International Eurasian Conference on Biological and Chemical Science, October 11-13, 2023, Ankara, Turkey, 1561-1568.
- Benkov, I., Varbanov, M., Venelinov, T., & Tsakovski, S.L. (2023). Principal component analysis and the water quality index – a powerful tool for surface water quality assessment: A case study on Struma River catchment, Bulgaria. *Water*, 15(10), 1961. doi: [10.3390/w15101961](https://doi.org/10.3390/w15101961)
- Bennet, E.M., Carpenter, S.R., & Caraco, N.F. (2001). Human impact on Erodeable Phosphorous Eutrophication: A global perspective: Increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication. *Bioscience*, 51, 227–234. doi: [10.1641/0006-3568\(2001\)051\[0227:HIOEPA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0227:HIOEPA]2.0.CO;2)
- Brown, R.M., McClelland, N.J., Deiniger, R.A., & O'Connor, M.F. (1972). Water quality index-crossing the physical barrier. *Research, Jerusalem*, 6, 787–797.
- Damo, R., & Icka, P. (2013) Evaluation of Water Quality Index for Drinking Water. *Polish Journal of Environmental Studies*, 22, 1045–1051.
- Edberg, S.C.L., Rice, E.W., Karlin, R.J., & Allen, M.J. (2000). *Escherichia coli*: the best biological drinking water indicator for public health protection. *Journal of Applied Microbiology*, 88(S1), 106–116. doi: [10.1111/j.1365-2672.2000.tb05338.x](https://doi.org/10.1111/j.1365-2672.2000.tb05338.x)
- EPA. (2001). *Parameters of water quality, Interpretation and Standards*. EPA, ISBN 1-84096-015-3
- Fraino, P.E. (2023). Using principal component analysis to explore multi-variable relationships. *Nat Rev Earth Environ*, 4, 294. doi: [10.1038/s43017-023-00414-z](https://doi.org/10.1038/s43017-023-00414-z)
- Gazzaz, N.M., Yusoff, M.K., Aris, A.Z., Juahir, H., & Ramli, M.F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, 64(11), 2409–2420. doi: [10.1016/j.marpolbul.2012.08.005](https://doi.org/10.1016/j.marpolbul.2012.08.005)
- Gega, N., Bekteshi, L., Hoxha, B., & Karamelo, P. (2022). *Evaluation of the water quality of the*

- Shkumbini river in the Elbasan area*. In Conference ALBLAKES-4, October 21-22, 2022, Elbasan, Albania.
- Gjeci, V., Hamiti, X., Qarri, F., & Lazo, P. (2024). Water quality of the Shkumbini River, Albania: Evaluation by physicochemical parameters and nutrient content. *Zastita Materijala*, 65, 1270. doi: [10.62638/ZasMat1270](https://doi.org/10.62638/ZasMat1270)
- Granato, G.E., DeSimone, L.A., Barbaro, J.R., & Jeznach, L.C. (2015) *Methods for evaluating potential sources of chloride in surface waters and groundwaters of the conterminous United States*. U.S. Geological Survey Open-File Report 2015-1080, 89 p. doi: [10.3133/ofr20151080](https://doi.org/10.3133/ofr20151080)
- Grochowska, J. (2020). Assessment of Water Buffer Capacity of Two Morphometrically Different, Degraded, Urban Lakes. *Water*, 12(5), 1512. doi: [10.3390/w12051512](https://doi.org/10.3390/w12051512)
- Guedens, W.J., Reynders, M., Van Vinckenroye, K., Yperman, J., & Carleer, R. (2018). Monitoring the Chloride Concentration in International Scheldt River Basin District Water Using a Low-Cost Multifunction Data Acquisition Board. *Water*, 10(8), 1025. doi: [10.3390/w10081025](https://doi.org/10.3390/w10081025)
- Guenouche, F.Z., Mesbahi-Salhi, A., Zegait, R., Chouia, S., Kimourd, M.T., & Bouslamad, Z. (2024). Assessing water quality in North-East Algeria: A comprehensive study using water quality index (WQI) and PCA. *Water Practice & Technology*, 19(4), 1232. doi: [10.2166/wpt.2024.073](https://doi.org/10.2166/wpt.2024.073)
- Horton, R.K. (1965). An index-number system for rating water quality. *Journal of the Water Pollution Control Federation*, 37(3), 300-306.
- Hoxha, B., Mazrreku, A., Osmani, M., & Hajdini, G. (2023). River water quality assessment throughout 10 years – a case study. *Proceedings of the 12th EURASIA Congress on Scientific Researches and Recent Trends*, November 29-30, 2023, Ankara, Turkey, 175-186.
- International Organization for Standardization. (2006). ISO 5667-5:2006 Water quality – Sampling – Part 6: Guidance on sampling of drinking water from treatment works and piped distribution systems.
- International Organization for Standardization. (2014). ISO 5667-6:2014, Water quality – Sampling – Part 12: Guidance on sampling of rivers and streams.
- Johnes, P.J. (2007). Uncertainties in annual riverine phosphorus load estimation: impact of load estimation methodology, sampling frequency, base flow index and catchment population density. *Journal of Hydrology*, 332, 241-258. doi: [10.1016/j.jhydrol.2006.07.006](https://doi.org/10.1016/j.jhydrol.2006.07.006)
- Jolliffe, I.T. (2002). *Principal Component Analysis* (2nd ed.). Springer.
- Khalil, B., Ouarda, T.B.M.J., & St-Hilarie, A. (2011). Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *Journal of Hydrology*, 405(3), 277-287. doi: [10.1016/j.jhydrol.2011.05.024](https://doi.org/10.1016/j.jhydrol.2011.05.024)
- Kopeck, B.G., Feng, X., Posmentier, E.S., Chipman, J.W., & Virginia, R.A. (2018). Use of principal component analysis to extract environmental information from lake water isotopic compositions. *Limnology and Oceanography*, 63, 1340-1354. doi: [10.1002/lno.10776](https://doi.org/10.1002/lno.10776)
- Kucaj, E., Osmani, M., Gjoni, A., Bardhi, A., Kucaj, B., & Bujku, D. (2022). Assessment of Physico Chemical Characteristics of Lana, Tirana and Ishmi Rivers Using IDW Interpolation. *International Journal of Environmental Science and Development*, 13(6), 223-230.
- Lin, L., Yang, H., & Xu, X. (2022). Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. *Front. Environ. Sci.*, 10, 880246. doi: [10.3389/fenvs.2022.8880246](https://doi.org/10.3389/fenvs.2022.8880246)
- Miho, A., Cullaj, A., Hasko, A., Lazo, P., Kupe, L., Bachofen, R., Brandl, H., Schanz, F., & Baraj, B. (2005). The quality of Albanian natural waters and the human impact. *Environment International*, 31(1), 133-146. doi: [10.1016/j.envint.2004.06.008](https://doi.org/10.1016/j.envint.2004.06.008)
- Musselman, R. (2012). *Sampling procedure for lake or stream surface water chemistry*. Research Note RMRS-RN-49. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. Retrieved from: <https://www.fs.usda.gov/>
- Odonkor, S.T., & Ampofo, J.K. (2013). *Escherichia coli* as An Indicator of Bacteriological Quality of Water: An Overview. *Microbiology Research*, 4(1), e2. doi: [10.4081/mr.2013.e2](https://doi.org/10.4081/mr.2013.e2)
- Osmani, M., Hoxha, B., Kucaj, E., Mazrreku, A., & Cinari, R. (2023). Wastewater treatment impact on water quality - a case study. *Periodico di Mineralogia*, 92(1), 33-43.

- Palansooriya, K.N., Yang, Y., Tsang, Y.F., Sarkar, B., Hou, D., Cao, X., & Ok, Y.S. (2020). Occurrence of contaminants in drinking water sources and the potential of biochar for water quality improvement: A review. *Critical Reviews in Environmental Science and Technology*, 50(6), 549–611. doi: [10.1080/10643389.2019.1629803](https://doi.org/10.1080/10643389.2019.1629803)
- Pano, N. (1985). *Hidrologjia e Shqiperise*. Akademia e Shkencave, Instituti Hidrometeorologjik, Tirane, 441 p.
- Picinini-Zambelli, J., Garcia, A.L.H., & Da Silva, J. (2025). Emerging pollutants in the aquatic environments: a review of genotoxic impacts. *Mutation Research - Reviews in Mutation Research*, 795, 108519. doi: [10.1016/j.mrrev.2024.108519](https://doi.org/10.1016/j.mrrev.2024.108519)
- Rajwa, A., Bialik, R.J., Karpiński, M., & Luks, B. (2014). Dissolved Oxygen in Rivers: Concepts and Measuring Techniques. In: Bialik, R., Majdański, M., & Moskalik, M. (Eds.), *Achievements, History and Challenges in Geophysics*. GeoPlanet: Earth and Planetary Sciences. Springer, Cham. doi: [10.1007/978-3-319-07599-0_19](https://doi.org/10.1007/978-3-319-07599-0_19)
- Roy, B.N., & Roy, H. (2024). Principal component analysis incorporated water quality index modeling for Dhaka-based rivers. *City and Environment Interactions*, 23(8), 100150. doi: [10.1016/j.cacint.2024.100150](https://doi.org/10.1016/j.cacint.2024.100150)
- Saalidong, B.M., Aram, S.A., Out, S., & Lartey, P.O. (2022). Examining the dynamics of the relationship between water pH and other water quality parameters in ground and surface water systems. *PLoS One*, 17(1), e0262117. doi: [10.1371/journal.pone.0262117](https://doi.org/10.1371/journal.pone.0262117)
- Shuquan, A., Xiufan, X., & Ying, M. (2015). Evaluation of water quality using Principal Component Analysis. *Nature Environment and Pollution Technology*, 14(4), 855–858.
- Shyti, B., Bekteshi, L., Paraloj, S., & Hila, E. (2024). Remodeling of the WQI Index for the evaluation of the Shkumbini River's water quality in Albania using the statistical method. *Ecologia Balkanica*, 16(1), 58–67.
- Steinhart, C., Schierow, L., & Sonzogno, W. (1982). An environmental quality index for the Great Lakes. *Water Resources Bulletin*, 18(6), 1025–1031. doi: [10.1111/j.1752-1688.1982.tb00110.x](https://doi.org/10.1111/j.1752-1688.1982.tb00110.x)
- Stevens, M., Ashbolt, N., & Cunliffe, D. (2003). Review of coliforms: as microbial indicators of drinking water quality. Australia: Australian government national health and research council, Biotext Pty Ltd., Canberra, 3–16.
- Syafudin, M., Kristanti, R.A., Yuniarto, A., Hadi-barata, T., Rhee, J., Al-Onazi, W.A., & Al-Mohaimed, A.M. (2021). Pesticides in drinking water – a review. *International Journal of Environmental Research and Public Health*, 18(2), 468. doi: [10.3390/ijerph18020468](https://doi.org/10.3390/ijerph18020468)
- Tchobanoglous, G., Burton, F.L., & Stensel, H.D. (2003). *Wastewater engineering: Treatment and reuse* (4th ed.). McGraw-Hill.
- Uddin, M.G., Moniruzzaman, M., & Khan, M. (2017). Evaluation of groundwater quality using CCME water quality index in the Rooppur Nuclear Power Plant area, Ishwardi, Pabna, Bangladesh. *American Journal of Environmental Protection*, 5(2), 33–43. doi: [10.12691/env-5-2-2](https://doi.org/10.12691/env-5-2-2)
- Uddin, M.G., Nash, S., & Olbert, A.I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218. doi: [10.1016/j.ecolind.2020.107218](https://doi.org/10.1016/j.ecolind.2020.107218)
- Uddin, M.G., Nash, S., & Olbert, A.I. (2022). A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment. *Water Research*, 210, 118532. doi: [10.1016/j.watres.2022.118532](https://doi.org/10.1016/j.watres.2022.118532)
- Uddin, M.G., Nash, S., Rahman, A., & Olbert, A.I. (2023). A sophisticated model for rating water quality. *Science of the Total Environment*, 868, 161614. doi: [10.1016/j.scitotenv.2023.161614](https://doi.org/10.1016/j.scitotenv.2023.161614)
- Vighi, M., & Chiaudani, G. (1985). A simple method to estimate lake phosphorus concentrations resulting from natural background loading. *Water Research*, 19(8), 987–991. doi: [10.1016/0043-1354\(85\)90367-7](https://doi.org/10.1016/0043-1354(85)90367-7)
- Ward, J.V., Tockner, K., Arscott, D.B., & Claret, C. (2002). Riverine landscape diversity. *Freshwater Biology*, 47, 517–539. doi: [10.1046/j.1365-2427.2002.00893.x](https://doi.org/10.1046/j.1365-2427.2002.00893.x)
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52.
- World Health Organization (WHO). (2003). Total dissolved solids in drinking water. Background document for development of WHO Guidelines for Drinking-water Quality. Retrieved from <https://cdn.who.int/>
- World Health Organization (WHO). (2004). *Guidelines for Water Quality* (3rd ed.), World Health Organization (WHO), Geneva

- World Health Organization (WHO). (2011). Guidelines for drinking water quality. Retrieved from https://iris.who.int/bitstream/handle/10665/44584/9789241548151_eng.pdf
- Zela, G., Demiraj, E., Marko, O., Gjipalaj, J., Erebara, A., Malltezi, J., Zela, E., & Bani, A. (2020). Assessment of the Water Quality Index in the Semani River in Albania. *Journal of Environmental Protection*, 11(11), 25-39. doi: [10.4236/jep.2020.1111063](https://doi.org/10.4236/jep.2020.1111063)
- Zitko, V. (1994). Principal component analysis in the evaluation of environmental data. *Marine Pollution Bulletin*, 28(12), 718-722. doi: [10.1016/0025-326X\(94\)90329-8](https://doi.org/10.1016/0025-326X(94)90329-8)

Received: 17.05.2025
Accepted: 30.06.2025