Original Scientific paper
10.7251/AGREN2503101S
UDC 556.535:543.3

AN INTEGRATED ASSESSMENT OF WATER QUALITY IN THE RIVA (ÇAYAĞZI) STREAM USING MULTIVARIATE STATISTICAL AND SPATIAL APPROACHES

Reyhan SAĞLAM*, Ferhat GÖKBULAK

Istanbul University-Cerrahpaşa, Faculty of Forestry, Department of Watershed Management, Bahçeköy, 34473, Istanbul, Turkey
*Corresponding author: reyhan.demir@iuc.edu.tr

ABSTRACT

This study evaluates the spatial variability of water quality and identifies pollutant sources in the Riva (Çayağzı) Stream watershed, located in Istanbul, Türkiye. Monthly, water samples were collected from five strategically selected locations throughout 2024. In situ measurements included physicochemical parameters (pH, electrical conductivity, dissolved oxygen, water temperature, and stream flow velocity). At the same time, laboratory analyses included major cations, nutrients, and heavy metals (Pb, Cd, Cr, Cu, Ni, Zn). A multivariate statistical approach—comprising Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and Pearson correlation—was employed to uncover pollutant patterns and site similarities. These analyses were further integrated with land use classification derived from Landsat-8 imagery, using the Random Forest algorithm. PCA results revealed distinct pollution signatures: Öyümce was characterized by pronounced heavy metal contamination (Pb, Cd, Cr, Cu, Ni, Zn); Ömerli exhibited moderate metal levels; Riva was dominated by high ionic content (EC, Na, Mg) and suspended solids; Pasamandıra displayed intermediate pollution; and Ömerli Dam Outlet represented the least-impacted, reference site with high dissolved oxygen and minimal contaminants. HCA supported these spatial groupings, clustering sites by their dominant pollution profile. Land use assessments indicated significant urban and industrial pressure in the downstream Riva area, agricultural influence in Öyümce and Ömerli, and forest dominance in the dam upstream watershed. Strong correlations among pollutants suggested that urban runoff and agricultural activities were the primary drivers of contamination. These findings highlight the value of integrating statistical and remote sensing-based spatial analyses for source apportionment.

Keywords: *Water quality, Pollution source, Multivariate statistical analysis, Remote sensing.*

INTRODUCTION

Freshwater ecosystems are essential for sustaining life on Earth, supporting economic growth, energy production, and enabling climate adaptation. Yet, these systems are increasingly at risk due to anthropogenic pressures and climate-induced changes (Kumari and Tripathi, 2015; Bao et al., 2024). Globally, freshwater security is undermined by rapid urbanization, the intensification of agriculture, aging infrastructure, and shifting hydrological regimes (Sharma, 2018). More than 80% of global wastewater is discharged into rivers and streams without adequate treatment, rendering them unsuitable for human use, aquatic life, and recreational activities (UN Water, 2023). This widespread pollution has triggered a global water quality crisis, with severe socio-ecological implications. Istanbul, a rapidly expanding megacity, exemplifies these challenges. The city is particularly vulnerable to climate-related risks such as irregular precipitation, flash floods, prolonged droughts, and increasing urban heat island effects (Yilmaz et al., 2015). These factors exacerbate pressures on existing water resources and further complicate water quality management. Within this context, green infrastructure approaches have gained prominence in urban planning. Stream corridors, including the Riva Stream, are critical components of this infrastructure due to their role in maintaining ecological connectivity, regulating water quality, and providing diverse ecosystem services across urban–rural gradients. However, urban sprawl, industrial discharge, agricultural runoff, and tourism-related activities have increasingly degraded the water quality of the Riva stream.

Several previous investigations have documented the presence of key pollution indicators in the Riva Stream, including elevated nutrient concentrations, heavy metal accumulation, and microbial contamination (Yilmaz et al., 2015; Çiftçi-Türetken et al., 2024). Although these studies have contributed valuable baseline data regarding the stream's environmental status, they have predominantly employed conventional monitoring techniques and descriptive statistical evaluations, which limit the ability to trace pollution origins or evaluate complex source–parameter relationships. Traditional monitoring and descriptive methods have provided valuable insights into surface water quality; however, their capacity to differentiate among diverse pollution sources can be limited. Particularly in heterogeneous watersheds where both point and non-point sources coexist, accurately identifying and quantifying pollutant contributions poses a significant analytical challenge (Hou et al., 2022). While the importance of source apportionment is increasingly acknowledged, the adoption of integrated, multivariate, and data-driven approaches remains relatively limited in many riverine systems. Expanding the application of such methodologies can enhance the understanding of spatial and temporal pollution patterns and support the development of more effective, site-specific watershed management strategies.

This study aims to address this gap by conducting a year-long assessment of water quality in the Riva Stream using a combination of field-based measurements and multivariate statistical techniques. By providing a robust interpretation of pollution dynamics and spatial differentiation within the basin, this study offers actionable insights for targeted interventions. The integration of statistical tools enables the identification of pollution hotspots and their possible links to anthropogenic or natural drivers. Ultimately, the findings contribute to a scientific foundation for water-sensitive urban planning and sustainable watershed management in Istanbul, where developmental needs and ecological protection must be balanced.

MATERIALS AND METHODS

Study Area and Sampling Design

The Çayağzı Stream, which constitutes the focal point of our study, drains a watershed area of approximately 859 km² (Tarkan, 2007). Originating from the Ömerli Dam, this significant watercourse generally flows in a northwesterly direction and ultimately discharges into the Black Sea near the village of Riva. The Riva watershed exhibits heterogeneous land-use patterns, encompassing urbanized areas, agricultural land, forests, and recreation zones. Five sampling stations were strategically determined along the stream to represent the spatial variability of potential pollution pressures: Ömerli Dam Outlet, Ömerli Site, Öyümce Site, Paşamandıra Site, and Riva Site (Fig. 1). Streamwater sampling was conducted monthly throughout 2024, yielding a total of 60 water samples (5 stations × 12 months). All field sampling followed standard protocols and quality assurance procedures recommended by APHA (2017).

At each sampling point, key physicochemical parameters including pH, electrical conductivity (EC), dissolved oxygen (DO), temperature, and flow velocity were measured in situ using portable field equipments. Following field measurements, water samples were collected in clean, sterilized bottles, preserved under appropriate conditions (cooling at 4°C), and promptly transported to the laboratory for comprehensive analysis. In the laboratory, a range of optical, physical, and chemical parameters was assessed to characterize water quality. These included color, turbidity, and total suspended solids (TSS) as indicators of physical and visual clarity. Water samples were analyzed for major cations such as calcium (Ca), magnesium (Mg), sodium (Na), and potassium (K) to evaluate mineral content and for heavy metals including iron (Fe), lead (Pb), cadmium (Cd), chromium (Cr), copper (Cu), nickel (Ni), and zinc (Zn) to assess potential toxicological risks. Additionally, key nutrients and anions such as sulfate (SO_4^{2-}), phosphorus (P), and boron (B) were analyzed to understand eutrophication potential and geochemical characteristics. Elemental concentrations were determined using Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES), providing high-precision multi-elemental analysis. Optical properties and colorimetric determinations were performed using standard photometric methods following APHA protocols, ensuring methodological consistency and comparability.

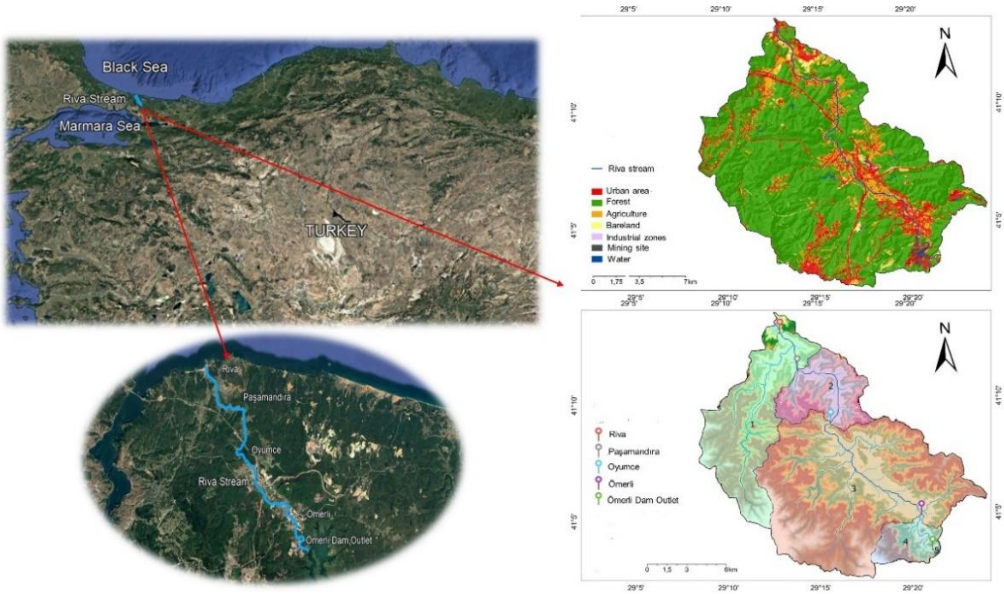

Fig 1. Study site and sampling locations

To analyze spatial pollution patterns, land use within the Riva watershed was classified using remote sensing techniques. A cloud-free Landsat-8 OLI/TIRS image dated August 10, 2024, was obtained from the USGS Earth Explorer platform (www.earthexplorer.usgs.gov). Image processing and classification were performed using ArcGIS Pro and Google Earth Pro software. A supervised classification was implemented with the 5-4-3 band combination (near-infrared, red, green), and the Random Forest machine learning algorithm was chosen for its high accuracy and robustness against overfitting. The model was trained with ground truth data and validated through cross-validation. According to the CORINE Land Cover system, seven land use classes were identified: urban areas, forest, agriculture, bareland, industrial areas, mining sites, and water bodies (Fig. 1, Table 1). These spatial maps were then used to identify pollution-prone zones and evaluate the proximity and influence of point-source pollution (e.g., industrial facilities, settlements) on stream water quality.

Table 1. Spatial distribution (ha) of upstream watershed land use types for each sampling site

Land use	**Riva**		**Paşamandıra**		**Öyümce**		**Ömerli**		**Ömerli Dam Outlet**	
	ha	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**
Forest	12932.37	10.65	3805.11	3.13	44599.59	36.71	29327.49	24.14	2561.13	2.11
Urban	700.20	0.58	345.51	0.28	10580.58	8.71	9925.65	8.17	7.92	0.01
Agriculture	675.27	0.56	513.09	0.42	1431.63	1.18	102.87	0.08	0.09	0.00

Water	9.45	0.01	7.11	0.01	27.54	0.02	1526.04	1.26	1512.81	1.25
Bareland	193.14	0.16	122.94	0.10	356.67	0.29	98.01	0.08	2.25	0.00
Mining sites	1.53	0.00	0.45	0.00	23.94	0.02	56.52	0.05	0	0.00
Industrial	0.27	0.00	0.36	0.00	24.75	0.02	4.59	0.00	0	0.00

Data Preparation and Statistical Analysis

Data were first screened for missing values and outliers. Non-normally distributed variables were transformed (log or square root), and z-score normalization was applied for multivariate analysis. An Exploratory Data Analysis was performed to identify spatial patterns and variable distributions. Kolmogorov–Smirnov tests were used to evaluate normality. Pearson correlation coefficients were calculated to explore relationships among parameters. To detect underlying pollution gradients, PCA and HCA were used. One-way ANOVA followed by Tukey's HSD test was applied to identify significant differences in parameter levels across stations. All statistical operations were carried out using Python (v3.11) and its scientific computing libraries, including pandas, scipy, statsmodels, matplotlib, and scikit-learn.

RESULTS AND DISCUSSION

Descriptive statistical analysis was performed to evaluate the overall distribution and variability of water quality parameters measured across all sampling stations throughout the study period. The observed ranges reveal considerable spatial heterogeneity, reflecting the influence of multiple land-use types and pollution sources within the watershed. Notably, pH and water temperature levels did not show statistically significant differences across the sampling stations ($p>0.05$), indicating relatively stable acid–base conditions and seasonal thermal patterns throughout the stream (Table 2). Riva showed distinctly elevated electrical conductivity (4769.25 µS/cm) and total suspended solids (2183.75 mg/L), suggesting high ionic and particulate loads likely due to upstream domestic and industrial discharges. This station also recorded the highest levels of magnesium, sodium, and sulfate. Ömerli Dam Outlet and Ömerli generally showed lower concentrations of heavy metals and nutrients, reflecting less impacted upstream conditions. In contrast, Öyümce and Paşamandıra showed increased levels of turbidity, color, and Fe, indicating moderate to high sediment and pollutant input, possibly from surrounding agricultural or settlement activities. Heavy metals such as Cr, Ni, Zn, and Pb were predominantly detected at Ömerli and Öyümce, albeit at low concentrations. Nutrient enrichment was evident across all stations but was more pronounced at Paşamandıra and Riva, with higher phosphorus and boron levels (Table 2).

Table 2.Descriptive statistics (mean ± standard deviation) for physicochemical water quality parameters at each sampling site along the Riva Stream

Parameter (mg/lt)	**Sampling Site**				
	Ömerli Dam Outlet	**Ömerli**	**Öyümce**	**Pasamandıra**	**Riva**
pH	8.01^{a}±0.66	7.97^{a}±0.71	7.9^{a}±0.72	7.99^{a}±0.53	7.65^{a}±0.61
EC	579.92^{a}±87.54	959.17^{a}±43.63	918.25^{a}±119.43	1316.25^{a}±971.03	4769.25^{b}±2722.71
DO	5.96^{a} ± 1.48	4.16^{ac} ± 1.42	2.71^{bc} ± 1.65	5.15^{a} ± 2.81	3.24^{bc} ± 2.22
Temperature	17.28^{a} ± 5.54	21.02^{a} ± 4.73	19.22^{a} ± 5.57	18.32^{a} ± 6.36	18.53^{a} ± 5.93
Ca	85.73^{a} ± 14.18	66.89^{b} ± 10.91	64.7^{b} ± 6.55	62.28^{b} ± 12.44	78.44^{ab} ± 18.07
Mg	15.34^{a} ± 3.53	11.82^{a} ± 1.8	12.39^{a} ± 0.9	24.78^{a} ± 14.22	74.96^{b} ± 34.6
Na	22.72^{a}±5.7	102.18^{b}±11.15	84.67^{b}±27.72	136.41^{b}±103.62	386.2^{c}±161.91
K	2.35^{a} ± 1.78	19.16^{b} ± 2.43	15.36^{b} ± 6.07	14.43^{b} ± 12.79	44.05^{c} ± 17.08
Cd	0.00^{a} ± 0.0	0.00^{b} ± 0.0	0.02^{b} ± 0.0	0.00^{a} ± 0.0	0.00^{a} ± 0.0
Cr	0.0^{a} ± 0.0	0.05^{b} ± 0.03	0.05^{b} ± 0.03	0.01^{c} ± 0.01	0.01^{c} ± 0.01
Cu	0.0^{a} ± 0.0	0.01^{b} ± 0.01	0.01^{b} ± 0.01	0.0^{a} ± 0.0	0.0^{a} ± 0.0
Fe	0.07^{a} ± 0.02	0.59^{b} ± 0.14	0.83^{c} ± 0.11	0.34^{d} ± 0.12	0.30^{d} ± 0.06
Ni	0.0^{a} ± 0.0	0.02^{b} ± 0.01	0.02^{b} ± 0.01	0.0^{c} ± 0.0	0.0^{c} ± 0.0
Pb	0.0^{a} ± 0.0	0.04^{b} ± 0.02	0.05^{b} ± 0.03	0.01^{c} ± 0.01	0.02^{c} ± 0.01
Zn	0.0^{a} ± 0.0	0.22^{b} ± 0.11	0.28^{b} ± 0.14	0.02^{ac} ± 0.01	0.04^{c} ± 0.02
P	0.05^{a} ± 0.04	1.57^{b} ± 0.72	1.18^{bc} ± 0.85	0.61^{ac} ± 0.54	0.59^{ac} ± 0.38
B	0.55^{a} ± 0.16	2.72^{b} ± 0.58	2.51^{b} ± 0.62	1.49^{c} ± 0.6	4.36^{d} ± 1.35
SO_4^{2-}	49.42^{a}±16.19	169.42^{b}±47.61	143.77^{bc}±40.72	121.25^{c}±37.35	232.67^{d}±55.01
Flow velocity	0.0^{a} ± 0.0	0.33^{b} ± 0.07	0.19^{c} ± 0.02	0.0^{a} ± 0.0	0.0^{a} ± 0.0
Color	14.75^{a} ± 11.3	48.33^{b} ± 21.45	58.0^{b} ± 93.73	21.08^{a} ± 15.41	20.67^{a} ± 11.19
Turbidity	5.61^{a} ± 5.38	20.78^{bc}±19.17	25.0^{b} ± 14.05	15.94^{acb}±9.02	9.14^{ac} ± 12.79
TSS	281.83^{a} ± 83.42	554.5^{b}±147.36	527.58^{b}±90.03	731.58^{b}±562.83	2183.75^{c}±1219.9

Pollution Source Apportionment

The PCA biplot indicated that the first two principal components (PC1: 36.4%, PC2: 28.2%) explained a total of 64.6% of the variance in the dataset. The Öyümce exhibited the highest PC1 score (+6) and slightly negative PC2 values, indicating the strongest association with heavy metal pollution (Pb, Cd, Cr, Cu, Ni, Zn), while ionic loadings and TSS remained relatively moderate. Ömerli Site, with moderately high PC1 (+2 to +4), also showed a heavy metal influence but to a lesser extent than Öyümce and a similar PC2 profile indicating intermediate levels of ionic and suspended solids. In contrast, the Riva Site had lower PC1 values (~0 to +2) but strongly positive PC2 scores (~+4 to +9), reflecting high EC, Na, Mg, and TSS, with moderate metal contamination (Fig. 2a). This suggests a predominant influence of urban runoff, salinization, and sediment input, rather than industrial or agricultural point sources. Pasamandıra Site clustered near the origin in PCA space, indicating moderate pollution levels, with no single dominant pollutant class. Ömerli Dam Outlet, located in the lowest PC1 and PC2 quadrant (–4 to –1), showed the cleanest water quality profile, characterized by high DO and the lowest concentrations of metals, ions, and solids, rendering it suitable as a reference point for "clean water" conditions.

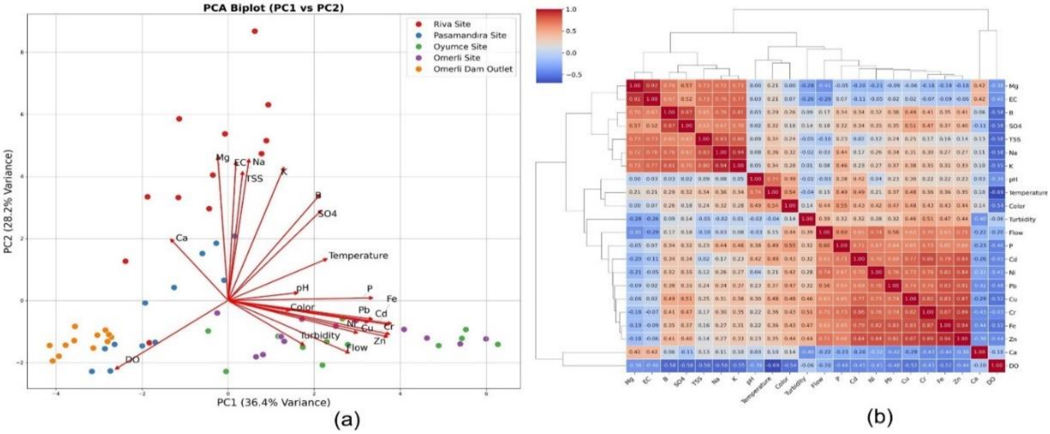

Fig 2. (**a**) PCA biplot of sampling sites and water quality parameters. (**b**) Clustered Pearson correlation heatmap of physicochemical parameters, highlighting co-occurrence trends and pollutant groupings along the Riva Stream

In addition to the PCA biplot, a hierarchical clustered correlation heatmap (Fig. 2b) was generated to explore the underlying relationships between variables. Strong positive correlations were observed among EC, Na, Mg, SO_4^{2-}, and TSS, reinforcing the ionic and conductivity-related influence, particularly evident at the Riva site. Likewise, heavy metals such as Cd, Pb, Ni, Zn, and Cu clustered tightly with strong intercorrelations, which corresponds to their high loading on PC1, particularly for samples from the Öyümce and Ömerli sites. Conversely, DO and Ca exhibited strong negative associations with most pollutants, further supporting their role as indicators of cleaner water conditions, as observed at the Ömerli Dam Outlet. The strong correlation cluster consisting of EC, Na, Mg, SO_4^{2-}, B, and TSS suggests dominant ionic and sediment-related pollution, commonly associated with runoff originating from urban sites and wastewater discharges (Varol et al., 2012). Similar associations between high ionic concentrations and urban land uses have been documented by previous studies conducted in rapidly urbanizing watersheds (Shrestha and Kazama, 2007). Moreover, heavy metals including Cd, Pb, Ni, Cu, Cr, Fe, and nutrient P formed a distinct correlation group, consistent with typical agricultural or industrial sources (Wang et al., 2017). These contaminants often enter water bodies through diffuse agricultural runoff, industrial effluents, or resuspended sediments. The negative correlations between DO and both ionic loads and heavy metals confirm DO as an effective indicator of water quality degradation, aligning well with the findings reported by Li et al. (2019) and Varol and Şen (2012). Intermediate parameters such as turbidity, color, and flow displayed moderate positive correlations with multiple pollutant groups, likely reflecting their general sensitivity to watershed-wide pollution intensity rather than specific sources.

Clustering patterns indicated that Ömerli and Öyümce formed a distinct group characterized by elevated concentrations of heavy metals, suggesting localized contamination likely linked to agricultural activities. Riva exhibited a separate pollution profile dominated by high EC and TSS, aligning with the influence of urban and industrial land uses in its watershed. In contrast, Pasamandıra and the Ömerli Dam Outlet sites were grouped as low- to moderate-impact areas, reflecting relatively lower pollution loads. These groupings were further corroborated by land use analysis: while Öyümce and Ömerli watersheds included extensive forested areas (44599 ha and 29327 ha, respectively) (Table 1), the presence of significant agricultural land likely contributed to nutrient and metal runoff. Riva's watershed featured notable urbanization and industrial activity, consistent with its ionic and particulate pollution indicators. Meanwhile, the Dam Outlet watershed was predominantly forested (>60%), with minimal anthropogenic disturbance, corresponding to its cleanest water quality profile among all sites.

CONCLUSIONS

The Riva Stream shows a complex pollution landscape, with spatial differentiation driven by land use intensity and anthropogenic activities. Upstream sites—particularly Öyümce and Ömerli—are burdened with heavy metal contamination linked to agricultural inputs. The downstream Riva station is influenced predominantly by urban runoff, exhibiting elevated EC and ionic loads. In contrast, the Ömerli Dam Outlet represents a minimally impacted site, marked by superior water quality. This study underscores the effectiveness of integrating multivariate statistical analyses and remote sensing-derived land use classifications to reveal pollutant patterns and inform targeted water resource management strategies. By delineating pollution sources with high spatial resolution, the findings contribute to evidence-based decision-making for sustainable watershed management in rapidly urbanizing regions.

ACKNOWLEDGEMENT

This article is based on Reyhan Sağlam's PhD thesis titled "Water quality and herbaceous plant interaction in stream corridors: Example of Riva Stream", conducted under the supervision of Prof. Dr. Ferhat Gökbulak. This study was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under project number 124Y042.

REFERENCES

Bao, V.Q., Van Toan, P., Van Tuyen, N., Hoang, H.M., Du, L. Van, Downes, N.K., Tri, V.P. D. (2024). Understanding watershed sources of pollution in Vinh Long Province, Vietnamese Mekong Delta. Discover Applied Sciences, 6(6).

Çiftçi Türetken, P.S., Altuğ, G., Öztaş, M., Baş, S.D.K., Şahin, S. K. (2024). Bio-indicator Bacteria Levels in Riva Stream, an Important Stream in İstanbul, Türkiye. Aquatic Sciences and Engineering, 39(2), 88–94.

Dökümcü, N., Koşal Şahin, S., Taş Divrik, M., Odabaşı, S. (2024). Investigation of seasonal changes in Annelida fauna and some physicochemical parameters of Riva stream (Istanbul). Aquatic Research, 7(1), 39–50.

Hou, X., Gao, W., Zhang, M., Xia, R., Chen, X., Deng, Y. (2022). Source apportionment of water pollutants in Poyang Lake Basin in China using absolute principal component score – multiple linear regression model combined with land-use parameters. (August), 1–12.

Kumari, M., Tripathi, B.D. (2015). Efficiency of Phragmites australis and Typha latifolia for heavy metal removal from wastewater. Ecotoxicology and Environmental Safety, 112, 80–86.

Li, P., Wu, J., Qian, H., Lyu, X. (2019). Origin and assessment of groundwater pollution and associated health risk: a case study in an industrial park, northwest China. *Environmental Geochemistry and Health*, 41(3), 967-980.

Sharma, H.R. (2018). Water-General Introduction . Module prepared for ePG Pathshala (a MHRD Project) for Environmental Sciences subject , paper Water Resources and. (January).

Shrestha, S., Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River basin, Japan. *Environmental Modelling & Software*, 22(4), 464-475.

Tarkan, A.S. (2010). Effects of streams on drinkable water reservoir: an assessment of water quality, physical habitat and some biological features of the streams. Journal of Fisheries Sciences, May.

United Nations, The United Nations World Water Development Report 2023. Partnerships and Cooperation for Water. UNESCO, Paris. https://unesdoc.unesco.org/ark:/48223/pf0000384655.

Varol, M., Şen, B. (2012). Assessment of nutrient and heavy metal contamination in surface water and sediments of the upper Tigris River, Turkey. *Catena*, *92*, 1-10.

Yilmaz, N., Ozyigit, I.I., Demir, G., Yalcin, I.E. (2015). Determination of phytoplankton density, and study of the variation of nutrients and heavy metals in the surface water of Riva Stream. Desalination and Water Treatment, 55(3), 810–820

Wang, J., Liu, G., Liu, H., Lam, P.K. (2017). Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the Middle Reach of Huaihe River, Anhui, China. *Science of the Total Environment*, 583, 421-431.