Indonesian Journal of Environmental Management and Sustainability
e-ISSN:2598-6279 p-ISSN:2598-6260

Research Paper

Evaluation of Surface Water Quality Indices in Mthatha River Using Multivariate Statistical Techniques

Makupula Nombuyiselo[1], Amoo Oseni Taiwo[1]*, Abayomi Abdultaofeek[2], Akinola Ikudayisi[3]

[1] *Risk and Vulnerable Science Centre, Walter Sisulu University, Mthatha, Eastern Cape, South Africa*
[2] *Department of Information and Communication Technology, Mangosuthu University of Technology, Durban, South Africa*
[3] *Civil Engineering Department, Walter Sisulu University, Buffalo City Campus, East London, Eastern Cape, South Africa*
***Corresponding author e-mail**: oamoo@wsu.ac.za

Abstract

Mthatha town of Eastern Cape Province, South Africa has been challenged to address the pollutant issues that are coming from rampant densification and effluent concentration discharge from the Mthatha Correctional Services Centre and the Efata School for the Blind and Deaf which have caused ineffable impaired damage to the Mthatha River Catchment (MRC). This paper is aimed at identifying drivers of poor water quality in the catchment and classified the River's water quality into different cluster groups for proper pollutant source control measures. Water quality parameters data comprising of pH; conductivity; Phosphorus; Ammonia (NH_4-N); Feacals; and *E-coli* covering 95 percent and 105 percent of the upstream and downstream sections of the River were available at ten monitored sites of the river catchment. These datasets covering eight years 2012-2020 were analysed in this study. Factor analysis as a choice of principal component analysis (PCA) and Agglomerative Hierarchical Clustering (AHC) was used to deduce inferences for the pollutants' subsequent classification. The results classified the catchment into three different clusters of lower pollutant (LP), medium pollutant (MP), and high pollutant (HP) areas, with PC1 accounting for 84.54% of the total variance from the three components classification. Adaptive catchment managers would find usefulness in the employed statistical tools in ensuring real-time measures for river non-point pollutants sources control that could offer additional benefits in maintaining a safe life above and below water in the preservation of their public values benefit. The study recommends the issuance of compliance notices and non-point pollutant source control measures to improve the water quality (WQ) parameters.

Keywords

Agglomerative Hierarchical Cluster, Principal Component Analysis, Water Quality Index, Public Value Benefit, Pollutants

Received: 7 September 2022, Accepted: 5 January 2023
https://doi.org/10.26554/ijems.2023.7.1.1-12

1. INTRODUCTION

A deeper understanding of river flow management and its interactions with water quality parameters could permeate the effective adoption of the most cost-effective strategy for controlling pollutants and maintaining the river's health (Diamantini et al., 2018; Sagan et al., 2020). The Mthatha River Catchment (MRC) authority has been challenged to address the pollutant issues that are coming from the rampant effluent discharge from the Mthatha Prison and the Efata School for the Blind and Deaf which have caused ineffable damage to the MRC water quality. This has triggered different studies on both preventive and curative ways of regulating and maintaining the river's health for the general populace's uses (Achieng et al., 2017; Ajayan et al., 2018).

Hence, among the different multidisciplinary factors influencing flow regime and water quality are human activities such as water abstraction, excess water disposal, irrigation, and vegetation clearing. These activities have the potential to change both the quantity and quality of flow (Rostami et al., 2020). Climate change has also caused significant extreme flow events which had produced changes in aquatic and riparian vegetation, aquatic connectivity, water quality, erosion, and sedimentation processes (Diamantini et al., 2018). This is not limited to the timing, length, and seasonal pattern of flow and varying water quality parameters, but all of which have an impact on the river ecosystem (O'Briain, 2019). In streams and rivers, water quality may vary significantly, depending on the water magnitude flow, a high flow serves to dilute the waste concentration while when there is a low flow, concentration may become undesirably high. It is therefore common practice to pick a flow condition for judging whether ambient water quality standards are being met. Also, contaminants may even be induced through the various hydrological variations in flow regimes, and their biological or chemical interactions with underlying geological rock and soil processes may also impact the waste concentration.

The quantity and qualities of the inflowing and receiving waters may amplitude the flow's influence on the altered water quality (Hallouin et al., 2018). Furthermore, other underlying features that could influence streamflow as drivers of water quality parameters include geology, vegetation cover, and rainfall characteristics such as magnitude, intensity, and frequency. The interaction between these attributes and their response varies spatially and temporally (Snelder et al., 2009). Moreso, the vitality and health of aquatic species depend on the natural processes in the drainage areas. Likewise, the natural process of precipitation, runoff, and percolation can cause significant changes in ecosystem structure and function, as well as the number and types of organisms that can live in the new environment (Hallouin et al., 2018).

Water quality classification methods are preventive and curative ways of maintaining required effluent standards in degraded river courses due to wastewater treatment plants discharged into the river. Most of the water quality classification procedures and approaches consist of three main methods which include: water quality index (Abbasnia et al., 2019; Bora and Goswami, 2017; Misaghi et al., 2017; Wu et al., 2018); the use of trophic status index (Kulshreshtha and Shanmugam, 2017 ; Robert et al., 2016; Thakur and Jindal, 2017); and the statistical analysis approaches (Rakotondrabe et al., 2018; Ustaoğlu et al., 2020).

Many of the water quality index methods have been addressed through the weighted arithmetic index approach. This method consists of average computation of the water quality index based on the physical-chemical and biological quality parameters that have been measured over an interval in a region while the trophic state index is a classification system designed to rate water bodies based on the amount of biological productivity they sustain. The existing entropic and mesotrophic environment for the aquatic ecosystem in the catchment calls for other biodiversity status classifications, particularly for South African rivers and lakes which had not been researched. Thus, depending on the available or measured water quality parameters, the expected benefits and aim of the research, the various existing methods for water quality assessment can be used to aggregate the diverse parameters, and harness their varied benefits into a single score in representing the historical water quality status of a river or catchment.

However, most of these approaches had been inexplicably complex due to the dynamic nature of the river/streamflow and the non-linearity in the water quality parameters. Only a few studies have paid attention to rainfall characteristics such as magnitude, intensity, and frequency to alter streamflow impacts on water quality (Richter et al., 1998; Poff and Zimmerman, 2010; Likens, 2013; Allen et al., 2020). Majority of these address water quality classification on the basis of precautionary average weighted arithmetic index and segmented rated water bodies based on the amount of biological productivity life it can sustain. Furthermore, depending on other biodiversity status classifications, integrated statistical analysis could serve as a standard for evaluating the water quality status and ensuring its public benefit for ecosystem survival.

Among the techniques aimed at reducing the drivers of poor water quality in any catchment are the application of statistical causal inference methods, the use of graphical models, neural networks, extreme event algorithms, time series modeling, non-linear dynamics-inspired methods, and surrogate modelling (Rakotondrabe et al., 2018; Ustaoğlu et al., 2020). Hence, the classification of the river water quality into different clusters for proper pollutant control, suggests new ways analyse and possibly model how the altered water quality (WQ) parameters and flow regime intersect towards maintaining a healthy state of the river for prompt mitigation measures in controlling pollutants from the effluence discharge. Thus, statistical analysis approaches offer a universal basis for proactive preventive methods depending on the available or measured water quality parameters and the main benefit the research aimed at achieving. The main contribution of this study lies in the understanding of the effects of altered streamflow patterns on some selected water quality parameters from the effluent discharge in the MRC. The novel feature of this study lies in its ability to mimic and reveal the interrelationships that exist among streamflow and pollutants as a common component of a water system.

Furthermore, because of unit disparity, the normal and standardised representation of water quality sample had proved to be effective for statistical analyses. Among the non-parametric methods which are not limited to cluster analysis (CA), factor analysis (FA), and principal component analysis (PCA) had been used to investigate the dimensionality of a measurement instrument by identifying the fewest number of interpretable factors (Jolliffe and Cadima, 2016). The PCA had enjoyed wide usage for reducing and interpreting large multivariate data sets with underlying linear structures, and for discovering previously unsuspected relationships (Taherdoost et al., 2022; Panaretos et al., 2017). PCA contributions are scale-independent and less sensitive to extreme values since it places a unit on the diagonal of all the variances signifying common and error variance among variables in the matrix (Brown et al., 2019). In addition, it may be used to identify the main pollutants and to aid in the interpretation of complicated data matrices to gain a better understanding of the water quality and ecological status of the examined systems (Tripathi and Singal, 2019; Wang et al., 2017). Thus, the most essential components that describe natural and anthropogenic impacts can be found using PCA.

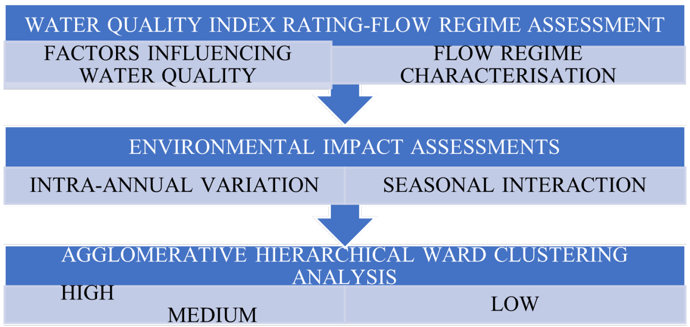

Figure 1. Research Methodology Block Diagram

Cluster analysis (CA) is a statistical technique for classifying datasets based on similarities without the use of prior assumptions or supervision. Among the various clustering algorithms, the hierarchical clustering has been mostly employed to analyze water quality data (Banda and Kumarasamy, 2020; Varol, 2020). Thus, the need to minimize the hazardous health impacts and aesthetically offensive odour which could pose danger to the lives and health of the community where river water quality is not properly managed cannot be over-emphasised. The South African Water Quality Guidelines (Volume 8) have been the main source of information for determining the water quality status of any river in the country depending on its water uses. This document is a compilation of all the different Targeted Acceptable Water Quality Ranges (TWQR), which have been previously discussed in Volumes one to seven of the gazettes for the protection and maintenance of healthy aquatic ecosystems.

Hence, the current study examined the interaction of streamflow and measured water quality parameters including pH, conductivity, phosphorus, Ammonia (NH_4-N), feacals, and *E-coli* in the MRC. It further highlights the inter and intra-annual seasonal trends in the monitored water quality parameters while also correlating the relationships between the altered streamflow and the measured water quality parameters. The categorised flow regime from the upper and lower sections of the River was used for the pollutants' strength classification into good, poor, and very poor rating index. This clustered classification provides possible control measures for managing the River catchment.

2. EXPERIMENTAL SECTION

2.1 Materials and Methods

Figure 1 depicts the employed research methodology block diagram employed in this study.

The study employed a mixed design methodology which entails a desktop review of documents and fieldwork for data collection. Thereafter, different statistical analysis was performed on the collected dataset. Correlation was used to establish the relationship between the water quality parameters and streamflow while the t-test was used to compare the statistical difference between the upper and lower reach of the River water quality and streamflow. The Kruskal–Wallis (KW) statistical tool test was used to determine if there is a significant difference between the medians of each dataset between the upper and lower grouped flow regime and water quality. Finally, the Agglomerative Hierarchical Ward Clustering Analysis was used to classify the pollutants into high, medium, and low pollutants strength.

2.2 Water Quality Index Rating Procedural Method

The water quality index was assessed using established procedures (DWAF, 1996; American Public Health Association, 2012), and the results were compared to the South African Water Quality Guidelines Volume 8 and WHO (2018). For the calculation of the water quality index, the weighted arithmetic index approach was utilized (WQI). Ewaid et al. (2018) detailed the steps for calculating the WQI as follows:
i) Determine a weighting rate based on expert advice
ii) Using Equation 1, calculate relative weight (RW)

$$RW = \frac{AW_i}{\sum_{i=1}^{n} AW_i} \tag{1}$$

where AW is each parameter's assigned weight, and n denotes the number of parameters.
iii) All of the parameters were then given a quality rating, except for pH. As shown in Equations 2 and 3, this was accomplished by increasing the result of each sample water quality parameter acquired from the laboratory analysis by a standard suggested by the WHO (2018) or the South African Water Quality Guidelines Volume 1 (DWAF, 1996), then multi1plying by 100.

$$Q_i = \left[\frac{C_i}{S_i}\right] \times 100 \tag{2}$$

$$Q_{pH} = \left[\frac{C_i - V_i}{S_i - V_i}\right] \times 100 \tag{3}$$

where Q_i signifies the quality rating, C_i denotes the value of the water quality parameter obtained from the suggested WHO, S_i is the ideal value of 7.0 for pH, and V_i means the measured value above or below 7.0 for pH. Equations 2 and 3 ensure that $Q_i = 0$ when there is no pollution in the water sample and $Q_i = 100$ when the parameter's value is barely above the permitted range. As a result, the water becomes more contaminated as the Q_i value rises (Şener et al., 2017). Finally, using Equations 4 and 1, the computed WQI (SI_i) was determined for each parameter.

$$SI_i = RW \times Q_i \tag{4}$$

$$WQI = \sum_{i=1}^{n} SI_i \tag{5}$$

The WQI scores were assigned to the following categories: <50 = Excellent, ≥50-≤100 = Good, >100-≤200 = Poor, >200-≤300 = Very Poor, and >300 = Unsuitable (Şener et al., 2017).

2.3 Data Sampling Design and Study Area

A preliminary survey was carried out to identify the general condition of the environment as well as the possible sources of pollutants that could impact water resources negatively. This helps in sampling location, from which water quality samples were collected twice before being transported to the laboratory to measure physio-chemical and bacteriological parameters. Water quality parameters comprising the pH, electrical conductivity-EC (mS/m) of the wastewater, Ammonia (mg/L), Phosphate (mg/L), Faecal coliforms (per 100 mL) and *E-coli* (count/mL) were used for the study. An average of the measured samples ensures an adequate design sample was used while the collected samples were transported to the Talbot and Talbot laboratory for analysis using the standard procedure as outlined in American Public Health Association (2012). A total of 1680 water samples were collected for the catchment between the period 2012 to 2020, with the year 2018 dataset missing due to data storage corruption. Thereafter, the geographic information system (GIS) was used to map the ten-identified monitoring points (coordinate) along the Mthatha River as shown in Figure 2.

2.4 Dataset–Sample Collection and Analyses

The sampled dataset comprises six water quality parameters which include the pH, conductivity (mS/m), PO_4-P (mg/L), NH_4-N (mg/L), Faecals (per 100 mL) and *E-coli* (count/mL) for both the upper and lower reach along the river catchment were used. Organic substances that have dissolved in water are measured as conductivity. Table 1 depicts the collected measured water quality parameter at Efata School (Upper River reach) and Mthatha Prison (Lower River reach) for the final effluent discharge in comparison to the required standard effluent compliance requirement.

From Table 1, the various water quality effluence parameters results showed that the Efata School effluent had not complied with the generally accepted standards as stated in Section 24 of the South Africa Constitution, and Section 19 of the National Water Act, therefore, effort is required for the effluent treatment before been discharged into rivers water resource.

2.5 Data Analyses

Table 2 summarizes the basic statistics for the studied water quality parameters at upper and lower monitoring sites for the period 2012-2020 while Figure 3 depicts the streamflow hydrographs at the upper and lower reaches of the Mthatha River with mean values of 57 ± 30 and 51 ± 8 respectively.

The mean streamflow as illustrated in Table 2 are 51.48 and 57.99 (m^3/s) respectively at the upper reach and lower reach of the Mthatha River. These values indicate that the catchment characteristics are closely related to the upstream and downstream flow regimes while the mean pH was 7.43 at upper reaches and 7.46 at lower reaches. With the mean values of 11.90 mS/m, 0.035 mg/L, 0.24 mg/L and 462.53 (100/mL) respectively for the conductivity, phosphate, ammonia and faecal coliforms parameters in the upper reaches of the MRC as against the lower reaches parameters' mean values of 14.92 mS/m, 0.28 mg/L, 0.41 mg/L and 468.06 (100/mL) respectively, the result reveals that most of the lower reaches water quality parameters are higher than those of the upper reaches. This implies a higher water quality matrix with streamflow's hydrological indices.

3. RESULTS AND DISCUSSION

The various results obtained in this study are hereby presented in this section.

3.1 Preliminary Data Analysis Results and Discussion

The current study was made with an assumption that the River maintains a steady state, indicating no change over time in the water quality concentrations with time. Thus, a trend analysis of the seasonal water quality (WQ) discharge of the catchment allows us to better understand the prevailing underlying physio-chemical mechanism occurring in the catchment. This can be used as a starting point for planning, and in controlling the pollutants trend. Thus, using the Standard Normal Homogeneity Test (SNHT) as depicted in Figure 4, shows that most of the water quality parameters dataset were not from the same source (homogeneity) and that the water quality parameters are not consistent in their trend, because their p-observed values was less than the significance level of 0.05. Thus, there was a significant difference between the upper reaches and the lower reaches.

The pre-processing phase is very important in data analysis to improve data quality. Since the various water quality parameters are of different units, there is a need to normalise and standardise the data to ensure superior accuracy and minimum bias. The z-score method has been used for the dataset normalisation. This method transformed the dataset into a non-dimensional scalar value. Figure 5 depicts the normalised and standardised versions of the water quality parameters.

In rating the surface water quality index into pollutant strength for their environmental and public values benefit preservation, the pre-data process helped to minimise accumulated errors in the different water quality parameters. Thus, contributing to higher accuracy and aggregating the effects of the altered streamflow pattern on the selected water quality parameters.

3.2 Classification into Pollutant Strength

Table 3 illustrates the statistical summary of the water quality parameters correlation with the entire catchment

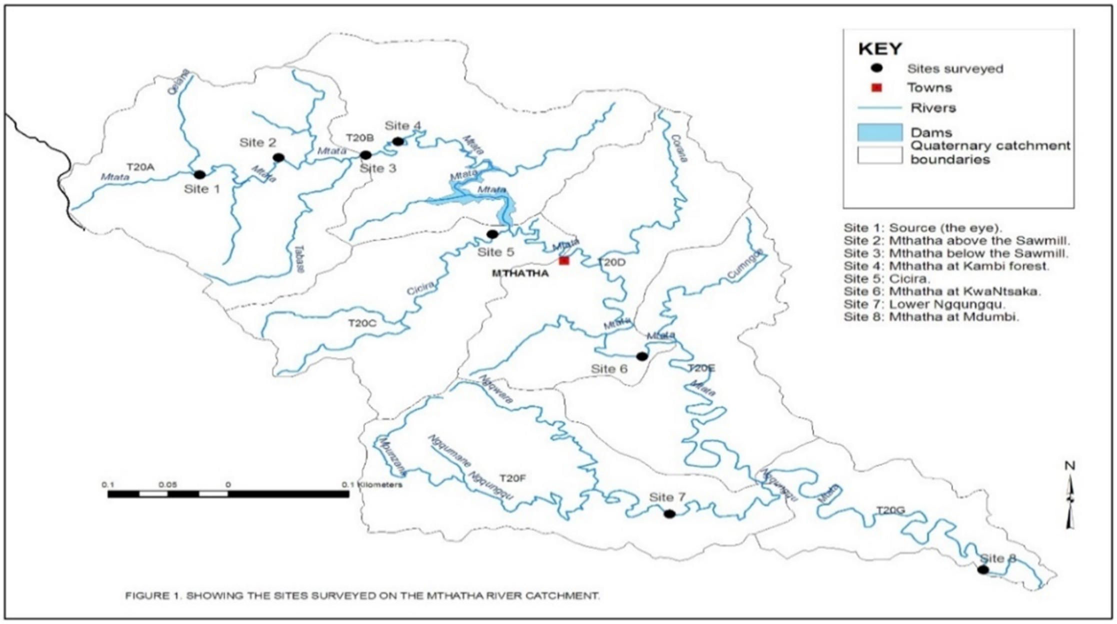

Figure 2. Map of the Mthatha River Catchment Study Area Sampling Sites

Table 1. Efata School and Mthatha Prison Measured Statistical Analysis (2012-2020)

Parameter	Efafa School	Mthatha Prison	Standard Discharge Limit
pH	6.6-8.1	5.47-7.20	5.5-9.5
Conductivity (mS/m)	25.9-62.0	25-84.6	70-100
Phosphate(mg/L)	21-293	20-317	75
Ammonia (mg/L)	1.04-6.6	0.240-39	10
Faecals (per 100 mL)	1.38-35.3	4.29-55	6
E-coli (Count/mL)	26 - 61300	2400-77000	1000

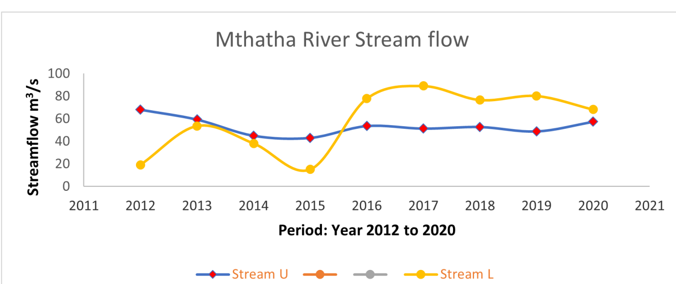

Figure 3. Mean Annual Streamflow Hydrograph for Mthatha River Upper and Lower Reaches

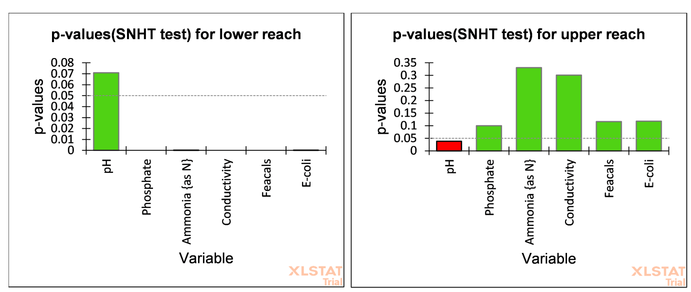

Figure 4. Homogenous and Abrupt Change Test on MRC Water Quality Parameters at the Lower and Upper Reaches

streamflow. This analysis helps to determine the dependence of water quality on flow for the catchment.

A negative relationship with the water quality parameters for the MRC implies that the likelihood of the streamflow yield decreases. The correlation coefficient of variation in the streamflow magnitude reveals that increasing streamflow from storms may dilute water quality and reduce pollution rate time, as well as change water quality parameters. This may also impact the aquatic ecosystem by lowering the water temperature through reducing the water quality disturbance effects. In all, the water quality parameters correlation with streamflow suggests that the river's water quality status has rapidly deteriorated going by the negative correlation values

Table 2. Statistics of Water Quality Parameters at MRC Upper and Lower Reach

	Upper Reach				Lower Reach			
	Min	Max	Mean	Std. Deviation	Min	Max	Mean	Std. Deviation
Stream Flow	43.18	59.37	51.48	5.62	15.09	89.08	57.99	27.55
pH	6.50	8.10	7.43	0.33	6.70	8.50	7.46	0.34
Conductivity	4.00	82.00	11.90	11.71	5.00	60.00	14.92	11.002
Phosphate	0.002	0.250	0.04	0.05	0.001	10.00	0.28	1.37
Ammonia	0.07	1.37	0.24	0.26	0.08	3.08	0.41	0.63
Faecals	0.00	3600.00	462.53	772.72	0.00	4100.00	468.04	728.04
E-coli	0.00	242.40	323.20	635.24	0.00	1986.00	268.67	385.26

Table 3. MRC Water Quality Variables Correlation with Streamflow

	Stream Flow	pH	Conductivity	Phosphate	Ammonia	Faecals	*E-coli*
Stream Flow	1						
pH	-0.0178114	1					
Conductivity	-0.026936	0.001029	1				
Phosphate	-0.0017228	0.000917	0.018476401	1			
Ammonia	-0.0314364	-0.02566	0.223397513	0.0424139	1		
Faecals	-0.0569992	-0.00314	0.008206518	0.010088	0.000623	1	
E-coli	-0.0541755	-0.0047	0.006950648	-0.0084705	0.000461	0.969333	1

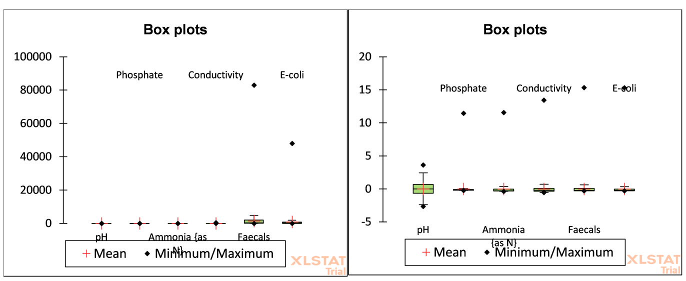

Figure 5. Normal and Standardised Representation of Water Quality Parameters Data

observed. This could be attributed to low rainfall, rising population, and human activities' impacts on the catchment.

3.3 Seasonal Water Quality Parameters Trends with Streamflow

The seasonal water quality parameters trend could assist in identifying period drivers of the poor water quality parameters witnessed in the catchment. This could also be useful in other seasonal periods to proffer mitigation strategies against emerging pollutants and classification ratings when optimal measures are needed for control. Using the intra-annual season's dependence of water quality parameters on the altered seasonal streamflow, Figure 6 shows the annual seasonal pattern for the observed water quality parameters. The indicated bracket water quality parameters account for the remaining 2%, 3%, 3%, and 2% respectively for the different seasonal variations.

Figure 6 depicts that, the pH was observed as being stable throughout the seasons with a ranged value between 6.6-8.1, this may be due to none effect of relative rainfall distribution to cause greater dilution in the acidity or alkalinity of a water resource. As the pH is a log scale, a one-unit change would mean a ten-fold change in the hydrogen ion concentration (Wanda et al., 2016). In terms of the water quality guidelines, the pH should not change by more than 0.5 or less than 0.5 unit change to the background pH.

Conductivity is a measurement of organic substances that have dissolved in water, high conductivity peaks were observed during the dry seasons, which indicated that there was no dilution happening instead the high evaporation increased the organic substances/solutes in water. Naturally, the River has its assimilative capacity indicating the ability to clean itself, which becomes much more possible during high rainfall while also noting that high solutes in water prohibit plants' growth.

Ammonia showed some high peaks (1.38-55 mg/L) in all the seasons (Table 1) equally as shown in Figure 6. Likewise, Phosphate pollutants may be a result of the presence of high algae growth and eutrophication. The Mthatha River downstream was noted to have thick layers of water hyacinth and algae growth. Faecal coliforms and *E-coli* were equally very high, and these are an indicator of the presence of organisms and that the water is Faecal contaminated with organic materials from both humans and animals. The poorly treated sewer discharged by the waste stabilization ponds was also responsible for the high Faecal matter.

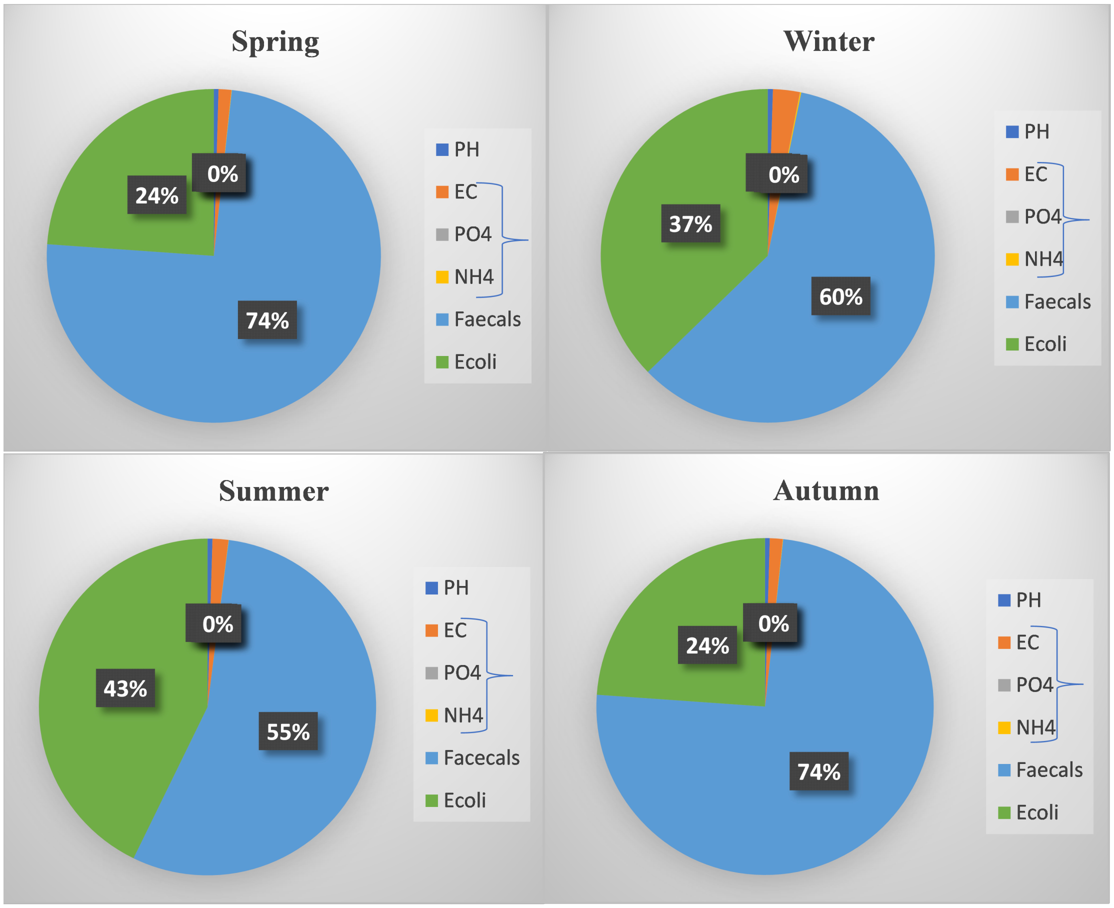

Figure 6. Mthatha River Seasonal Variation Measured Water Quality (2012-2020)

3.4 Surface Water Quality Seasonal Index Rating

Tables 4 to 7 illustrate the statistical t-test comparison at the upper reach of the River to the lower reach based on the seasonal water quality parameter and streamflow magnitude. Table 4 depicts the t-tests used to determine whether the group means are statistically heterogeneous or whether they differ by group variability. The paired-sample t-test between water quality parameters at the MRC and the River's reaches in Summer is shown in Table 4.

From Table 4, the observed p-value range of 0.084-0.957 was greater than the significance value of 0.05 for streamflow, pH, EC, ammonia, phosphates, faecal coliforms, and *E-coli*. As a result, the null hypothesis that there is meaningful difference among the pollutants can thus be drawn that there was no meaningful difference in streamflow, pH, EC, ammonia, phosphates, faecal coliforms, and *E-coli* in summer.

Tables 5 to 7 illustrate the statistical summary for the seasonal water quality parameter and streamflow correlation. The t-test hypothesis tests were used to determine whether the group means are statistically heterogeneous or whether they differ by group variability. The paired-sample t-test between water quality parameters at the Mthatha River's upper reaches and lower reaches in Autumn was shown in Table 5.

The result from Table 5 depicts that, there was no noticeable difference in the parameters in Autumn. The null hypothesis that the grouped means water quality parameters and streamflow differ by group variability was rejected because the streamflow p-value was less than the significance level of 0.05. Thus, there was a significant difference in streamflow between the upper reaches and the lower reaches. Table 6 shows the t-test paired-sample comparison between water quality parameters at the Mthatha River's upper reaches and lower reaches in Winter.

As depicted in Table 6, the result shows that the drivers of poor water quality in the catchment, in Winter, show there was no meaningful difference in the parameters. Thus, the null hypothesis was rejected since the p-value for the streamflow-water quality was less than the 0.05 significance level. Thus, there was a significant difference in streamflow-water quality parameters between the upper reaches and the lower reaches. Table 7 shows the paired-sample t-test between water quality parameters in Spring.

As depicted in Table 7, the result of the streamflow-water quality drivers of poor water quality in the catchment

Table 4. Paired Sample t-test Results of Mthatha River Water Quality Parameters in Summer

	t	df	P-sig(2 tailed)
Mthatha River stream flow upper_Mthatha River flow lower	-1.483	40	0.146
pH upper_pH lower	-1.543	46	0.130
EC upper_EC lower	-0.175	38	0.862
Ammonia upper_Ammonia lower	-1.769	46	0.084
Phosphates upper_Phosphates lower	-1.543	46	0.130
Feacals upper_Faecals lower	-0.806	46	0.424
E-coli upper_*E-coli* lower	-0.054	46	0.957

Table 5. Paired Sample t-test of Mthatha River Water Quality Parameters in Autumn

	t	df	P-sig(2 tailed)
Mthatha River stream flow upper_Mthatha River flow lower	-3.252	40	0.002
pH upper_pH lower	-0.274	50	0.785
EC upper_EC lower	0.733	50	0.467
Ammonia upper_Ammonia lower	1.286	50	0.204
Phosphates upper_Phosphates lower	-1.362	50	0.179
Feacals upper_Faecals lower	-0.695	50	0.490
E-coli upper_*E-coli* lower	0.055	50	0.956

Table 6. Paired Sample t-test of Mthatha River Water Quality Parameters in Winter

	t	df	P-sig(2 tailed)
Mthatha River stream flow upper_Mthatha River flow lower	-5.059	40	< 0.0001
pH upper_pH lower	-1.581	70	0.118
EC upper_EC lower	0.155	70	0.877
Ammonia upper_Ammonia lower	-0.313	70	0.755
Phosphates upper_Phosphates lower	-0.006	70	0.995
Feacals upper_Faecals lower	-0.246	62	0.807
E-coli upper_*E-coli* lower	-0.385	66	0.701

Table 7. Paired Sample t-test of Mthatha River Water Quality Parameters in Spring

	t	df	P-sig(2 tailed)
Mthatha River stream flow upper_Mthatha River flow lower	-1.2963	40	0.050
pH upper_pH lower	-0.568	48	0.573
EC upper_EC lower	0.635	50	0.528
Ammonia upper_Ammonia lower	-1.260	50	0.213
Phosphates upper_Phosphates lower	-1.821	50	0.075
Feacals upper_Faecals lower	0.141	42	0.888
E-coli upper_*E-coli* lower	-1.275	50	0.208

shows there was no noticeable difference in the parameters in Spring. The null hypothesis was rejected because the p-value was less than the significance level of 0.05. Thus, there was a significant difference in water quality between the upper reaches and the lower reaches in Spring. It may then be inferred that streamflow, pH, electrical conductivity, ammonia, phosphates, faecal coliforms, and *E-coli* do change much. Thus, the intra-annual dependence of water quality on the seasonal flow regime, shows the witnessed annual temporal variability of streamflow is due to the complex interaction of catchment water quality concentration and rainfall interaction which may induce variation witnessed in the study area.

3.5 Water Quality and Streamflow Classification Result

Using the Euclidean distance in CA, the catchment driver and the water quality parameters had been broadly grouped into three classes namely high, medium, and low pollutants as illustrated in Figure 7. The Agglomerative hierarchical cluster (AHC) in Figure 7 illustrates the dendrogram classification for the water quality sample parameters and their driver (flow). The results by class indicate that cluster C1 represents high pollutant (HP); cluster C2 represents medium pollutant (MP) while cluster C3 represents low pollutant (LP) classification.

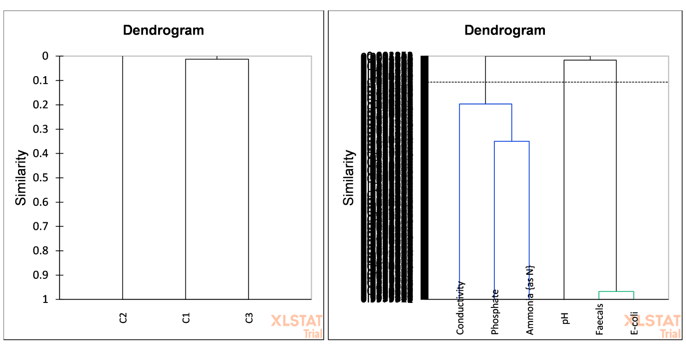

Figure 7. The Agglomerative Hierarchical Clusters - HP, MP, and LP Classification

Table 8 depicts the object dendrogram classification by distance. Going by the correlations with the centroids factor degree, the factor pattern squared cosine's contribution classified the water quality observation into class 1, 2 and 3.

The factor analysis effect was quite evident from the grouped Agglomerative hierarchical cluster classification result of the catchment with a correlation of 1.00 for class 1, 0.163 -0.995 for class 2, and 0.987-0.996 for class 3 respectively. This grouped the flow of the water quality parameters into three major clusters. Table 9 shows the grouped Eigen factor classification that depicts the degree of the factor pattern squared cosine's contribution. The cluster factors F1, F2, and F3 represent the most important contribution to the water quality index rating. Factor F1 shows the degree

Table 8. Classification by Distance Results

Observation	Class	Distance to Centroid	Correlations with Centroids
pH	1	0.000	1.000
Phosphate	2	247.007	0.163
Ammonia {as N}	2	231.216	0.384
Conductivity	2	475.829	0.995
Faecals	3	27752.723	0.996
E-coli	3	27752.723	0.987

of reliance on the water quality variance of 0.984 (faecal) and 0.982 (*E-coli*) as the most contributing variables in the group while 0.339 (Phosphate) and 0.999 (Ammonia) shows significant contribution in factor (F2). The negative 0.51 and 0.31 in grouped factor (F3) resulted in three cluster groups: highly polluted (HP), moderately polluted (MP), and less polluted (LP).

Table 9. The Factor Pattern Squared Cosine's Contribution

	F1	F2	F3
pH	0.018	-0.178	**-0.510**
Phosphate	-0.003	**0.339**	-0.039
Ammonia {as N}	-0.035	**0.999**	0.012
Conductivity	-0.014	0.292	**-0.307**
Faecals	**0.984**	0.025	0.009
E-coli	**0.982**	0.019	-0.004

The bolded values correspond to the factor for which the squared cosine is the largest for each variable

Table 9 shows the relative catchment cluster summary based on a supervised factor analysis classification. The results suggest increased streamflow causes lower water temperature and the subsequent increase in pH, Conductivity, Phosphate {as P}, Ammonia {N}, Faecals, and *E-coli*). This is useful in order to proffer mitigation strategies against the emerging pollutants rating between the water quality parameters and the streamflow.

3.6 Water Quality Index (WQI) Results and Discussion

The use of AHC to group the water quality drivers into pollutant strength provides adequate intervention measures in rating the water quality index (WQI) for different usage. According to Ustaoğlu et al. (2020), the general status of the environment's water quality accounts for upstream tributary inputs that could have a considerable impact on downstream water quality. Thus, the analysis of the water quality index in the Mthatha River catchment was subsequently carried out. The WQI technique has been adequately described in Section 3.1 using Equations 1 and 5. Table 10 shows

Table 10. Analysis of the Water Quality Index in the Mthatha River Catchment

Parameters	Water Quality Standard	Assigned Weight	Relative Weight	Quality Rating	Sub-Indices
pH (pH Unit)	6.50	2.50	0.19	15.03	2.89
NO_3+NO_2-N (mg/L)	1.50	2.40	0.18	123.83	22.86
NH_4-N (mg/L)	2.00	2.00	0.15	52.33	8.05
F (mg/L)	1.00	1.60	0.12	156.36	19.24
PO_4-P (mg/L)	1.00	1.00	0.08	135.97	10.46
SO_4 (mg/L)	1.50	1.50	0.12	123.94	14.30
TDS (mg/L)	3.00	2.00	0.15	112.61	17.32
Total		13.00		WQI=Sum(SI)	95.13

Table 11. Mthatha River Monitoring Segment Water Quality Index Rating

Sub Basin	Drinking Water		Irrigation		Aquatic Life	
A	88	Good	67	Good	56	Good
B	167	Poor	83	Good	78	Good
C	78	Good	126	Poor	98	Good
D	54	Good	63	Good	79	Good
E	168	Poor	67	Good	56	Good
F	164	Poor	63	Good	108	Poor
G	176	Poor	63	Good	78	Good
H	57	Good	76	Good	79	Good
I	157	Poor	74	Good	98	Good
J	178	Poor	65	Good	90	Good

the results of the Mthatha River Catchment water quality index analysis while Table 11 depicts the water quality index rating for aquatic life, drinking, and irrigation use. The water quality index rating figures were computed using Equations 1–5.

Where the sub-catchments were labeled alphabetically as follows: (A) upstream of Langeni forest; (B) downstream Langeni WWTW; (C) below Langeni Sawmill (D) upstream of Cicira; (E) downstream of Cicira; (F) above Mthatha prison (G) below Mthatha prison (H) below Mthatha wastewater treatment plant (wwtp); (I) below First Falls and (J) below Ngangelizwe township. These intervention measures portrayed the water quality index when compared to the South African Water Quality Guidelines and the World Health Organization (WHO) guideline for aquatic and biodiversity survival restrictions.

The results of the water quality index rating at river sections D to G imply that irrigation and aquatic usage have only a little impact as the values of these parameters suggest that they are not harmful to human health. These values have not exceeded the threshold permissible value stipulated in (American Public Health Association, 2012; DWAF, 1996; WHO, 2018) standard values. The high value of the Water Quality Index rating in sub-basin (J), on the other hand, shows that the ecosystem is entropic. The greater the entropy, the greater the losses, wastes, and environmental impacts- including heated waterways and degraded air quality to land contamination.

3.7 Limitations of the Study

This study was based on modeling the different sections of the Mthatha River into upper and lower reach, which may differ slightly from the simulation of the whole catchment. Also, the assigned magnitude index from an expert in calculating the WQI may be of varying degrees to the water body's pollutant rate classification. Furthermore, although one could argue that all portions upstream of a river's section have a significant impact on water pollutants downstream, this was not factored into the WQI calculation and subsequent classification for different usage. Similarly, the analysis ignores mechanisms like self-purification, pollution retention for instance through sedimentation, and/or dilution due to the inflow of cleaner tributaries. As a result, the model's applicability is limited. Although an entropic and mesotrophic environment exists for the aquatic ecosystem in the catchment, other biodiversity status classifications for South African rivers and lakes should be further researched.

4. CONCLUSIONS

Sustainable maintenance of rivers' water quality drivers provide a standard for healthy, economic development, and poverty alleviation in a catchment. This study examined seasonal trends and drivers of altered water quality parameters on a seasonal basis. The application of the statistical Agglomerative Hierarchical Clustering factor in classifying the river pollutants' strength into low pollutant (LP), medium pollutant (MP), and high pollutant (HP) clustered groups could help provides adequate intervention measures in rating the water quality index (WQI) for different usage. The correlation of the water quality parameters with the upper and lower sections of the Mthatha River depicts the effect of the drivers on the water quality status in the catchment. Furthermore, the results revealed that most of the failed sections of the River contain moderate pollutants, which implies the presence of high nuisance algae/aquatic plants which should be treated with caution. A similar report by DWAF (2004), corroborates with part of this current study outcome, which suggests the need for water resource managers to recognize the complexities of water quality pollutants sources and designate priority areas classification that needs immediate attention and intervention. Also, the exceedance of effluent limits standard for the area called for the issuance of non-compliance notices, enforcement of the polluter pays principle, and query of the government officers in charge of the waste stabilization pond for the treatment of the effluent in the catchment before being discharged into the Mthatha River.

5. ACKNOWLEDGEMENT

The authors would like to express their gratitude to the Department of Water and Sanitation for providing financial support as well as the data needed for this study. The Walter Sisulu University through the Risk and Vulnerable Science Centre is appreciated for the necessary infrastructural support.

REFERENCES

Abbasnia, A., N. Yousefi, A. H. Mahvi, R. Nabizadeh, M. Radfard, M. Yousefi, and M. Alimohammadi (2019). Evaluation of Groundwater Quality Using Water Quality Index and its Suitability for Assessing Water for Drinking and Irrigation Purposes: Case Study of Sistan and Baluchistan Province (Iran). *Human and Ecological Risk Assessment: An International Journal*, **25**(4); 988–1005

Achieng, A., P. Raburu, E. Kipkorir, S. Ngodhe, K. Obiero, and J. Ani-Sabwa (2017). Assessment of Water Quality Using Multivariate Techniques in River Sosiani, Kenya. *Environmental Monitoring and Assessment*, **189**(6); 1–13

Ajayan, A. P., J. W. Rijstenbil, and K. A. Kumar (2018). *Water Quality Assessment of an Unexplored Tropical Freshwater System in Thiruvananthapuram, India: A Multivariate Statistical Approach*. Springer

Allen, D. C., T. Datry, K. S. Boersma, M. T. Bogan, A. J. Boulton, D. Bruno, M. H. Busch, K. H. Costigan, W. K. Dodds, and K. M. Fritz (2020). River Ecosystem Conceptual Models and Non-perennial Rivers: A Critical Review. *Wiley Interdisciplinary Reviews: Water*, **7**(5); 1473

American Public Health Association (2012). *Standard Methods for Examination of Water and Wastewater. 23rd Edition*. Washington D.C.

Banda, T. D. and M. Kumarasamy (2020). Application of Multivariate Statistical Analysis in the Development of a Surrogate Water Quality Index (WQI) for South African Watersheds. *Water*, **12**(6); 1584

Bora, M. and D. C. Goswami (2017). Water Quality Assessment in Terms of Water Quality Index (WQI): Case Study of the Kolong River, Assam, India. *Applied Water Science*, **7**(6); 3125–3135

Brown, V., D. I. Smith, R. Wiseman, and J. Handmer (2019). *Risks and Opportunities: Managing Environmental Conflict and Change*. Routledge

Diamantini, E., S. R. Lutz, S. Mallucci, B. Majone, R. Merz, and A. Bellin (2018). Driver Detection of Water Quality Trends in Three Large European River Basins. *Science of the Total Environment*, **612**; 49–62

DWAF (1996). *South African Water Quality Guidelines (First Edition)*. Field Guide

DWAF (2004). *Department of Water Affairs and Forestry*. Mthatha Technical Report

Ewaid, S. H., S. A. Abed, and S. A. Kadhum (2018). Predicting the Tigris River Water Quality within Baghdad, Iraq by Using Water Quality Index and Regression Analysis. *Environmental Technology and Innovation*, **11**; 390–398

Hallouin, T., M. Bruen, M. Christie, C. Bullock, and M. Kelly-Quinn (2018). Challenges in Using Hydrology and Water Quality Models for Assessing Freshwater Ecosystem Services: A Review. *Geosciences*, **8**(2); 1–19

Jolliffe, I. T. and J. Cadima (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374**(2065); 20150202

Kulshreshtha, A. and P. Shanmugam (2017). Development of Optical Models for Assessing the Trophic Status of Coastal Waters. *OCEANS 2017-Aberdeen*; 1–7

Likens, G. E. (2013). *Biogeochemistry of a Forested Ecosystem*. Springer Science and Business Media

Misaghi, F., F. Delgosha, M. Razzaghmanesh, and B. Myers (2017). Introducing a Water Quality Index for Assessing Water for Irrigation Purposes: A Case Study of the Ghezel Ozan River. *Science of the Total Environment*, **589**; 107–116

O'Briain, R. (2019). Climate Change and European Rivers: An Eco-hydromorphological Perspective. *Ecohydrology*, **12**(5); 2099

Panaretos, D., G. Tzavelas, M. Vamvakari, and D. Panagiotakos (2017). Factor Analysis as a Tool for Pattern Recognition in Biomedical Research; A Review with Application in R Software. *Journal of Data Science*, **16**; 615–630

Poff, N. L. and J. K. Zimmerman (2010). Ecological Responses to Altered Flow Regimes: A Literature Review to Inform the Science and Management of Environmental Flows. *Freshwater Biology*, **55**(1); 194–205

Rakotondrabe, F., J. R. N. Ngoupayou, Z. Mfonka, E. H. Rasolomanana, A. J. N. Abolo, and A. A. Ako (2018). Water Quality Assessment in the Bétaré-Oya Gold Mining Area (East-Cameroon): Multivariate Statistical Analysis Approach. *Science of the Total Environment*, **610**; 831–844

Richter, B. D., J. V. Baumgartner, D. P. Braun, and J. Powell (1998). A Spatial Assessment of Hydrologic Alteration within A River Network. *Regulated Rivers: Research and Management: An International Journal Devoted to River Research and Management*, **14**(4); 329–340

Robert, O., S. Ngodhe, and O. Omondi (2016). *Assessment of Primary Productivity, Nutrient Levels and Trophic Status of Kuinet Dam, Kenya*. IJFAS Publishers

Rostami, S., J. He, and Q. K. Hassan (2020). Water Quality Response to River Flow Regime at Three Major Rivers in Alberta. *Water Quality Research Journal*, **55**(1); 79–92

Sagan, V., K. T. Peterson, M. Maimaitijiang, P. Sidike, J. Sloan, B. A. Greeling, S. Maalouf, and C. Adams (2020). Monitoring Inland Water Quality Using Remote Sensing: Potential and Limitations of Spectral Indices, Bio-optical Simulations, Machine Learning, and Cloud Computing. *Earth-Science Reviews*, **205**; 103187

Şener, Ş., E. Şener, and A. Davraz (2017). Evaluation of Water Quality Using Water Quality Index (WQI) Method and GIS in Aksu River (SW-Turkey). *Science of the Total Environment*, **584**; 131–144

Snelder, T. H., N. Lamouroux, J. R. Leathwick, H. Pella, E. Sauquet, and U. Shankar (2009). Predictive Mapping of the Natural Flow Regimes of France. *Journal of Hydrology*, **373**(1-2); 57–67

Taherdoost, H., S. Sahibuddin, and N. Jalaliyoon (2022). Exploratory Factor Analysis; Concepts and Theory. *Advances in Applied and Pure Mathematics*, **27**; 375–382

Thakur, R. and R. Jindal (2017). Principal Component Analysis: Productivity and Trophic Status of Some Selected Lakes of Western Himalaya. *SSRN*; 1–18

Tripathi, M. and S. K. Singal (2019). Use of Principal Component Analysis for Parameter Selection for Development of a Novel Water Quality Index: A Case Study of River Ganga India. *Ecological Indicators*, **96**; 430–436

Ustaoğlu, F., Y. Tepe, and B. Taş (2020). Assessment of Stream Quality and Health Risk in A Subtropical Turkey River System: A Combined Approach using Statistical Analysis and Water Quality Index. *Ecological indicators*, **113**; 105815

Varol, M. (2020). Use of Water Quality Index and Multivariate Statistical Methods for the Evaluation of Water Quality of A Stream Affected by Multiple Stressors: A Case Study. *Environmental Pollution*, **266**; 115417

Wanda, E. M., B. B. Mamba, and T. A. Msagati (2016). Determination of the Water Quality Index Ratings of Water in the Mpumalanga and North West Provinces, South Africa. *Physics and Chemistry of the Earth, Parts a/b/c*, **92**; 70–78

Wang, J., G. Liu, H. Liu, and P. K. Lam (2017). Multivariate Statistical Evaluation of Dissolved Trace Elements and A Water Quality Assessment in the Middle Reaches of Huaihe River, Anhui, China. *Science of the Total Environment*, **583**; 421–431

WHO (2018). *International Standards for Drinking Water*. World Health Organisation, Geneva

Wu, Z., X. Wang, Y. Chen, Y. Cai, and J. Deng (2018). Assessing River Water Quality Using Water Quality Index in Lake Taihu Basin, China. *Science of the Total Environment*, **612**; 914–922

