

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



## Chapter

# Interpretation of Water Quality Data in uMngeni Basin (South Africa) Using Multivariate Techniques

*Innocent Rangeti and Bloodless Dzwaairo*

## Abstract

The major challenge with regular water quality monitoring programmes is making sense of the large and complex physico-chemical data-sets that are generated in a comparatively short period of time. Consequentially, this presents difficulties for water management practitioners who are expected to make informed decisions based on information extracted from the large data-sets. In addition, the nonlinear nature of water quality data-sets often makes it difficult to interpret the spatio-temporal variations. These reasons necessitated the need for effective methods of interpreting water quality results and drawing meaningful conclusions. Hence, this study applied multivariate techniques, namely Cluster Analysis and Principal Component Analysis, to interpret eight-year (2005–2012) water quality data that was generated from a monitoring exercise at six stations in uMngeni Basin, South Africa. The principal components extracted with eigenvalues of greater than 1 were interpreted while considering the pollution issues in the basin. These extracted components explain 67–76% of the water quality variation among the stations. The derived significant parameters suggest that uMngeni Basin was mainly affected by the catchment's geological processes, surface runoff, domestic sewage effluent, seasonal variation and agricultural waste. Cluster Analysis grouped the sampling six stations into two clusters namely heavy (B) or low (A), based on the degree of pollution. Cluster A mainly consists of water sampling stations that were located in the outflow of the dam (NDO, IDO, MDO and NDI) and its water can be described as of fairly good quality due to dam retention and attenuation effects. Cluster B mainly consist of dam inflow water sampling stations (MDI and IDI), which can be described as polluted if compared to cluster A. The poor quality water observed at Cluster B sampling stations could be attributed to natural and anthropogenic activities through point source and runoff. The findings could assist in determining an appropriate set of water quality parameters that would indicate variation of water quality in the basin, with minimum loss of information. It is, therefore, recommended that this approach be used to assist decision-makers regarding strategies for minimising catchment pollution.

**Keywords:** cluster analysis, multivariate technique, principal component analysis, uMngeni basin, water quality

## 1. Introduction

Water pollution is a global challenge undermining economic growth, health of millions of people as well as the physical status of the environment in both developed and developing countries. The current global water scarcity challenge is not only related to inadequacy in terms of quantity but also related to the progressive deterioration of quality making water unfit for some given uses such as potability. The deterioration of water quality is attributed to both natural (precipitation rate, weathering processes, soil erosion, etc) and anthropogenic (urban, industrial, agricultural activities, etc) factors. Seasonal variations in precipitation, surface run-off, ground water flow, interception and abstraction strongly affect the river discharge and the concentrations of water pollutants in a basin [1]. The effect of contaminant on water depends upon the characteristics of the water itself as well as quantity and characteristics of the contaminant.

Water pollutants, which are usually introduced through surface runoff or direct discharge, may in higher concentration result in rivers failing to provide adequate attenuation of pollutants, resulting in catchments failing to meet minimum compliance of quality for various uses such as potable water production. Furthermore, water quality deterioration is often a slow process not readily noticeable due attenuation effects until an apparent change occurs. Such situations are being exacerbated by the rapid increase in the demand for freshwater in many countries including South Africa. In view of the limited quantity of freshwater resources worldwide and the effect of anthropogenic activities, protection of these resources has become a priority [2–4]. It has, therefore, become imperative to monitor the quality of water in freshwater systems in order to prevent its further deterioration and thus ultimately ensure its continuous availability in a quality that meets various uses including potable water production. Pollutants in water can cause acute or chronic illness in humans especially when polluted water is consumed or when sewage is used to irrigate vegetables meant for human consumption. In specific cases this has resulted in loss of lives. For example, as at 2015, the bacterium *Vibrio cholerae* caused between 1.3 to 4.0 million infections and 21,000 to 143,000 deaths worldwide [5].

With concern of the detrimental effects of pollution, various agencies have been monitoring the quality of raw water within the uMngeni, a 232 km river that is then treated to serve almost 3.8 million people within and around Durban and Pietermaritzburg (South Africa) with potable water [6–10]. The primary objectives of such monitoring exercise have been to identify water quality problems, describe the spatio-temporal water quality trends, determine fitness compliance for specific uses and develop monitoring tools such as water quality indices for enhancing information dissemination. Although such monitoring programs are crucial to a better knowledge of hydrology and pollution problems in catchments such as uMngeni Basin, they tend to produce large amounts of complicated data-sets of various water parameters. The data-sets are often difficult to analyse and extract meaningful information and this makes it difficult to keep the public informed, who are the custodian of the resource. By keeping the public updated, it makes them more participatory in policy formulation and decision making regarding protection of the water resource [11, 12].

The classification and interpretation of monitoring stations are the most important steps in the assessment of water quality. Numerous studies have confirmed multivariate statistical techniques (cluster analysis, principal component analysis, factor analysis and discriminant analysis) as excellent tools for exploring and presenting the bulk and complex water quality data-sets [13–15]. These techniques allow for the determination of spatio-temporal water quality variability,

classification of sampling stations and the identification of pollution sources [15–18]. Furthermore, by eliminating subjective assumptions, multivariate techniques tend to reduce biasness when selecting parameters for developing tools such as water quality index. This ultimately assists in improving the accuracy of such monitoring tool. Nevertheless, the selection of a multivariate technique to apply depends on the nature of data-set and research objectives. While there are a number of multivariate techniques, studies have extensively applied the Principal Component Analysis (PCA) and Cluster Analysis (CA) due to their suitability in extracting information on various situations [15, 19–22].

The application of principal component analysis (PCA) for the interpretation of a large and complex volume of data offers a better understanding of water quality, the ecological status of the basin being studied, while also allowing for the identification of possible factors/sources that influence the surface water systems [16]. Principal Component Analysis (PCA) aims to find combinations for certain variables to determine indices which describe the variation in the data while retaining as much information as possible. This reduction is achieved by transforming original variables into a new set of variables, known as principal components (PCs). These PCs, which are uncorrelated with the first few, retain most of the variation present in the original variables. The PCA technique transforms original variables into new uncorrelated variable known as principal components (PCs) [23–25]. The derived few variables can be used to provide a meaningful description of the entire data-set with a minimal loss of original information. The eigenvalues indicate the significance of each PC and a greater value, indicating the parameter's importance [26]. Correlation of PCs and original variables are given by the loadings [27]. While loadings reflect the relative importance of a variable within the component, it should be highlighted that these values does not show the importance of the component itself [28].

The PCA has been successfully applied on hydrogeological and hydrogeochemical studies. The application of PCA by Razmkhah, Abrishamchi [29] distinguished the anthropogenic and natural polluting activities along Jairood River in Iran. The results identified 5 factors which explained 85% the variation in water quality. Mazlum, Ozer [30] applied the PCA to determine factors causing water quality variability along a tributary, Porsuk, in Turkey. The study identified four PCs which explained 70% of the total water quality variance. The factors were related to the discharge of domestic wastewater, nitrification, industrial wastewater and the seasonal effect. Haag and Westrich [31] applied PCA to analyse the water quality along Neckar River in Germany based on ten parameters monitored from 1993 to 1998. The four principal components extracted accounting for 72% of total variance were interpreted as; (i) dilution by high discharge (ii) biological activity, (iii) seasonal effects and (iv) wastewater impact [31].

The limitations with the PCA technique include ignoring the degree of data dispersion as well as a weakness in processing nonlinear data. The result of PCA can also be influenced by uneven sampling interval, missing values or observations below detection limits of analytical methods, which can be changed during the data collection period. It is thus important to treat water quality data before modelling in order to improve the accuracy [32].

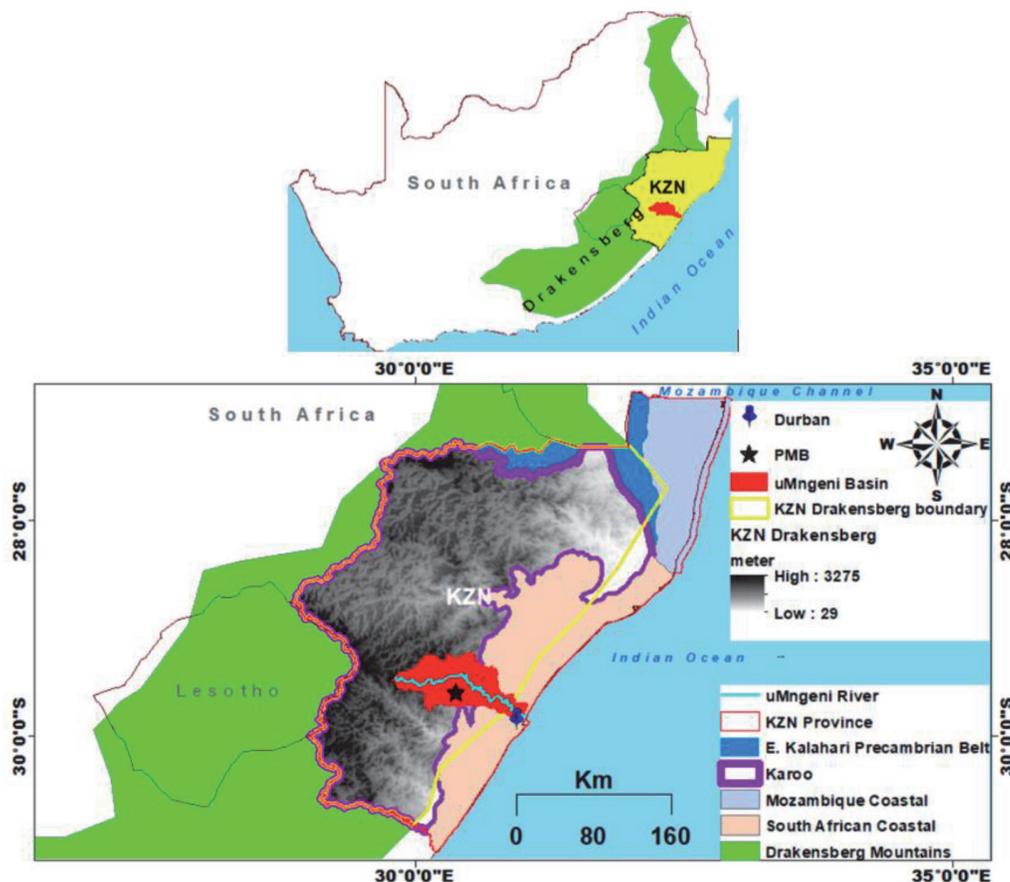
On the other hand, cluster analysis (CA) is an unsupervised pattern recognition multivariate technique which group objects (e.g., water quality variables) based on either their similarities or dissimilarities [33, 34]. Its objective is to sort cases into groups or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. Most studies have applied the hierarchical clustering (HC) technique to sequentially category objects [13, 15, 34, 35]. Based on the hierarchical CA, water quality characteristics of

each sampling location can be classified depending on pollution level. The results of a HC analysis are displayed graphically using a tree diagram commonly known as a dendrogram [13, 14, 36, 37]. The technique firstly groups the objects according to similarity. These groups are further merged according to their similarities or dissimilarities and eventually merge into a single cluster as the similarity among the subgroups decreases. The cluster analysis approach offers a reliable classification of surface water making it possible to design a future spatial sampling strategy that is cost-effective, with reduced number of sampling sites without losing any significant information [38].

## 2. Study area and water monitoring stations

uMngeni Basin, the study area, is situated in KwaZulu-Natal (KZN) Province, which lies along the eastern seaboard of the Republic of South Africa (**Figure 1**). uMngeni River (the main river in the basin), at 232 km long, is the primary source of raw water, which is then treated to serve a population of almost 3.8 million (as at 2013), in and around Durban metro as well as the city of Pietermaritzburg (PMB).

Key activities that generate point and non-point pollution within the catchment include agriculture and animal farming while concentrated urban settlements provide a variety of supportive economic activities that generate solid and liquid waste. A consequence of the concentrated development in the catchment area has been the high levels of pollutants entering the water system, which are eventually flushed out to sea. The basin receives much of its rain in summer, with occasional snow falls in some of its high lying areas such as the Drakensberg Mountain [39]. The geology of uMngeni Basin varies from basalts, granites, sandstones, shale and tillites [40].



**Figure 1.**  
uMngeni Basin, KZN Province, Drakensberg Mountains and South Africa.

About half of uMngeni Basin sits on top of the Karoo in the KZN part of the Drakensberg Mountains. The other portion extends east on top of the South African Coastal Plate (**Figure 2**).

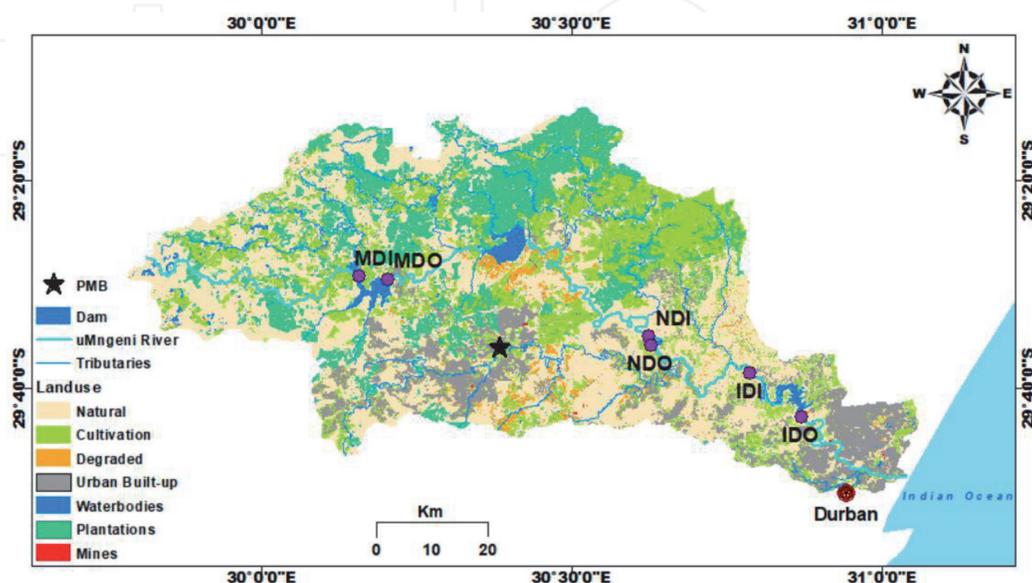
The 2009 Landuse map (**Figure 2**) indicates that there are mixed activities, where cultivation and plantations are located predominantly from central to the north-west of the basin.

## 2.1 Water quality monitoring points considered

Six water quality monitoring points shown in **Figure 2**, namely Midmar Dam Inflow (MDI) (*Upstream point*), Midmar Dam Outflow (MDO), Nagle Dam Inflow (NDI), Nagle Dam Outflow (NDO), Inanda Dam Inflow (IDI) and Inanda Dam Outflow (IDO) (*Downstream point*), were considered in this study. The dam inflow stations were assumed to give a reflection of the pollution activities along the river course while the dam outflow stations were expected to depict the dilution and retention effects.

### 2.1.1 Methods and materials

Multivariate statistical methods have been widely applied in environmental data reduction and interpretation of multi-constituent chemical and physical biological measurements. These techniques have been applied to identify factors that influence water systems, to assist in reliable water resource management as well as determine rapid solutions for pollution problems [16, 41]. This study applied PCA and CA techniques to extract information from the raw data regarding the significant parameters influencing the variation of water quality at each of the six stations studied. The Kaiser-Meyer-Olkin (KMO), which test the sampling adequacy, was used to determine the suitability of water quality data for PCA analysis [42–44]. Kaiser [43] recommended 0.5 as a minimum (barely accepted), values between 0.7–0.8 acceptable, and values above 0.9 as depicting excellence. The current study employed the PCA technique to determine the most significant parameters that would explain the variation in water quality.



**Figure 2.**  
*Human settlements dominate from the central parts of the basin (PMB and its periphery) up to the Indian Ocean.*

PCA is a very powerful multivariate statistical analysis technique used to reduce the dimensionality of a data set consisting multiple inter-related variables, while retaining data variability [45]. The technique extracts primary information representative of the typical characteristics of the water environment from a large amount of data and then represents it as a new set of independent variables of the principal component. PCA reduces the dimensionality of a multivariate data-set to a small number of independent principal components. Each principal component contains all the variable information, thus reducing the omission of information.

The PCA method is composed of five main operational steps, as follows:

1. The original data matrix is shown in Eq. 1:

$$X = (x_{ij})_{n \times p} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

where  $x_{ij}$  is the originally measured data,  $n$  represents the monitoring station, and  $p$  represents each water quality parameter.

2. Standardising the original data with Z-score standardisation formula to eliminate the impact of dimension (Eq. 2).

$$x_{ij}^* = (x_{ij} - \bar{x}_j) / s_j, \quad (2)$$

where  $x_{ij}$  is the standard variable,  $\bar{x}_j$  is the average value for  $j$ th indicator, and  $s_j$  is the standard deviation for the  $j$ th indicator.

3. Calculating the correlation coefficient matrix,  $R$ , with standardised data and determining the correlation between indicators (Eq. 3).

$$R = (r_{ij})_{p \times p} = \frac{1}{n-1} \sum_{t=1}^n x_{ti}^* * x_{tj}^* \quad (i, j = 1, 2, \dots, p) \quad (3)$$

4. Calculating the eigenvalues and eigenvectors of the correlation coefficient matrix,  $R$ , to determine the number of principal components. The eigenvalues of the correlation coefficient matrix,  $R$ , are represented by  $\lambda_i$  ( $i = 1, 2, \dots, n$ ) and their eigenvectors are  $u_i$  ( $U_i = U_{i1}, U_{i2}, \dots, U_{in}$ ) ( $i = 1, 2, \dots, n$ ). The value corresponds to the variance of the principal component, and the value of variance is positively correlated with the contribution rate of the principal components. Further, the cumulated contribution rate of the first  $m$  principal components should be more than 80%, which means that as explained in Eq. 4:

$$\sum_{i=1}^m \lambda_j / \sum_{i=1}^n \lambda_j \geq 0.80. \quad (4)$$

The principal component is represented by Eq. 5.

$$F_i = u_{i1}x_1^* + u_{i2}x_2^* + \cdots + u_{in}x_n^* \quad (i = 1, 2, \dots, n), \quad (5)$$

where  $x_i$  is the standardised indicator variable as shown in Eq. 6:

$$x_i^* = (x_i - \bar{x}_i) / s_i. \quad (6)$$

5. The obtained principal components are weighted and summed to obtain a comprehensive evaluation function, as shown in Eq. 7:

$$F = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \dots + \lambda_n} F_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2 + \dots + \lambda_n} F_2 + \dots \frac{\lambda_n}{\lambda_1 + \lambda_2 + \dots + \lambda_n} F_n \quad (7)$$

Principal components with an eigenvalue greater than 1 were related to the major pollution sources in uMngeni Basin. Water quality parameters with loadings of greater than 0.5 (highlighted in bold in the results tables) were regarded as significantly influencing water quality variation in uMngeni Basin.

Thereafter, cluster analysis (CA) was applied to determine the spatial similarity of the six water sampling stations studied. The hierarchy cluster analysis was employed using the Ward's method with Euclidean distances as a measure of dissimilarity [37, 46]. The number of subgroups for analysis were determined by drawing a line across the dendrogram and examining the main clusters branching out beneath that line [47]. Determination of the subgroups for analysis was subjective based on available information regarding pollution activities the along uMngeni River.

Eight-year (2005–2012) water quality data-sets obtained from then Umgeni Water was used in this study. Since monitoring generally depends on the pollution problem at any given time and space, the number and type of parameters monitored at each of the stations varied. As the study was data-driven, a monthly median was used for in-depth analysis. The adoption of median instead of the mean was in consideration that the latter is normally influenced by the outliers which are common in water quality data-sets while the former is resistant. The period studied was determined in consideration of a criteria explained by Schertz, Alexander [48] and Lettenmaier, Conquest [49]. These studies reported that at least a five-year monthly data and two-year monthly data should be sufficient for a defensible monotonic and step-trend (abrupt shift) study, respectively.

### 3. Results

#### 3.1 PCA analysis results

The Kaiser-Meyer-Olkin (KMO) results for the six stations ranged from 0.610 to 0.786 showing the fitness of the data-sets for PCA analysis. The component matrix tables shown in the different sections of the results only depict PCs with eigenvalue of greater than one (1). Only parameters with a correlation coefficient of great than 0.5 (highlighted in bold black) with its respective principal component were considered as significantly influencing water quality variability at any given station.

##### 3.1.1 Midmar dam inflow (MDI)

**Table 1** show the extracted seven PCs with eigenvalues of greater than 1 that explain 75% of the water quality variation at the Midmar Dam Inflow sampling station. While considering the high positive correlations of nutrient, metal ion and organic related parameters with component 1 (**Table 2**), it can be hypothesised that 20.6% of the water quality variation at this station is a result of both anthropogenic and natural processes. The nutrient and organic related parameters can be explained

Midmar dam inflow: total variance explained									
Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.740	24.955	24.955	5.740	24.955	24.955	4.740	20.607	20.607
2	5.002	21.750	46.705	5.002	21.750	46.705	3.621	15.742	36.349
3	1.903	8.272	54.977	1.903	8.272	54.977	3.011	13.090	49.440
4	1.325	5.760	60.737	1.325	5.760	60.737	1.787	7.771	57.211
5	1.165	5.065	65.802	1.165	5.065	65.802	1.412	6.139	63.350
6	1.103	4.795	70.598	1.103	4.795	70.598	1.390	6.043	69.393
7	1.011	4.394	74.992	1.011	4.394	74.992	1.288	5.598	74.992

*Extraction Method: Principal Component Analysis.*

**Table 1.**  
 Extracted values of the significant components at Midmar dam inflow (MDI).

	Component Matrix <sup>a</sup>						
	Component						
	1	2	3	4	5	6	7
Potassium (K)	<b>0.890</b>	-0.094	-0.027	0.111	0.090	-0.054	-0.074
Sulphate (SO <sub>4</sub> )	<b>0.852</b>	-0.027	-0.213	-0.078	-0.128	-0.018	0.072
Chloride (Cl)	<b>0.824</b>	-0.322	-0.169	-0.099	-0.129	-0.008	-0.071
Total Dissolved Solid	<b>0.743</b>	-0.199	-0.161	-0.197	-0.161	-0.001	-0.106
Nitrate (NO <sub>3</sub> )	<b>0.666</b>	-0.243	-0.231	-0.097	-0.295	0.049	-0.014
Total Organic Carbon(TOC)	<b>0.662</b>	0.464	0.012	-0.053	0.074	0.003	-0.156
<i>Escherichia coli</i> ( <i>E. coli</i> )	0.440	0.356	0.035	0.431	-0.153	0.135	-0.085
Suspended Solid	0.416	<b>0.741</b>	0.005	0.172	0.181	0.113	0.319
Iron (Fe)	0.401	<b>0.713</b>	-0.084	0.037	0.133	-0.101	0.152
Turbidity	0.400	<b>0.692</b>	0.064	0.209	0.190	0.141	0.404
Calcium (Ca)	<b>0.503</b>	<b>-0.687</b>	0.304	0.091	0.266	0.098	0.045
Magnesium (Mg)	<b>0.602</b>	<b>-0.681</b>	0.168	0.025	0.089	0.073	0.009
Sodium (Na)	<b>0.581</b>	<b>-0.666</b>	0.156	0.010	0.207	0.085	0.059
Alkalinity (Alk)	-0.036	<b>-0.632</b>	0.451	0.090	0.338	0.157	0.149
Total Phosphate	0.192	0.556	0.415	-0.092	-0.089	0.260	-0.284
Colour	0.246	0.545	0.225	-0.183	0.355	-0.198	0.026
Dissolved Oxygen (DO)	-0.048	-0.075	<b>-0.591</b>	0.410	0.017	0.262	0.093
Temperature (Temp)	0.230	0.453	0.460	-0.383	0.017	-0.222	-0.137
Silicon (Si)	0.347	0.321	-0.379	-0.460	-0.026	0.059	0.121
pH	-0.059	-0.147	-0.433	-0.345	<b>0.550</b>	0.222	-0.199
Ammonia (NH <sub>3</sub> )	0.085	0.160	0.470	-0.089	-0.376	0.619	0.009
Conductivity	0.286	-0.324	0.225	0.099	-0.252	-0.528	0.339
Soluble Reactive Phosphate (SRP)	0.313	0.250	0.019	<b>0.513</b>	0.131	-0.223	<b>-0.599</b>

Extraction Method: Principal Component Analysis.

<sup>a</sup>7 Components extracted.

Bold: Significant contributors to the respective principal component in their respective there column.

**Table 2.**

The correlation among the parameters measured and the extracted significant components at MDI.

by piggery, dairy and maize farming activities surrounding Midmar Dam [6]. Animal manure enters surface water, both accidentally and deliberately, from households, villages, communal farms and feedlots. Without treatment, manure runoff tends to result in algae blooms which can lead to human health problems if consumed. The second component explained 15.7% (**Table 1**) of the total variance at MDI and showed a positive correlation to Suspended Solids (SS), iron (Fe) and turbidity (**Table 2**). Turbidity and suspended solids can be related to surface runoff from agricultural activities along uMngeni River whilst Iron (Fe) can be attributed to weathering processes.

Agriculture, a sector responsible for the usage of 70% of water being abstracted globally, plays a major role in water pollution [50, 51]. Runoff from agricultural activities such as agrochemicals, organic matter, drug residues, sediments and saline drainage into water bodies can lead to nutrient enrichment and eutrophication. The resultant water pollution poses a risk to aquatic ecosystems, human health and

productive activities. Poor land management practises and deforestation can also explain the water quality variation at MDI. It is important for communities to practise improved land management through planting vegetation such as trees and plants to cover the ground. The negative relationship of metal ions with component 2 can be explained by the seasonality effect. Increased flow in the wet season turns to reduce the concentration of mineral salt content in a river system as a result of the dilution effect.

### 3.1.2 Midmar dam outflow (MDO)

The results in **Table 3** indicate that the first seven principal components with eigenvalues of greater than one (1) account for 73.7% of the total variance in the water-quality data set at Midmar Dam Outflow station. Component 1 which explains 27.1% of the total variance at MDO (**Table 3**) is mainly influenced by parameters related to human activities (turbidity and ammonia) as well as natural geological processes (silicon and calcium) (**Table 4**). Silicon is part of various essential plant minerals and it is released during weathering processes. Sodium and potassium which showed a moderate positive correlation to component 5 reflects the natural processes such as weathering.

### 3.1.3 Nagle dam inflow

The PCA technique identified five components, which cumulatively explained 67.4% of the total variance at NDI (**Table 5**). Component 1 which explained 24.7% of the total variance is significantly affected by parameters (highlighted in bold black) which normally originate from surface runoff of agriculture areas and effluent from wastewater treatment plants (**Table 6**). The pollution of water bodies when practicing agriculture is mainly due to fertiliser runoff after rainfall, nutrients (such as nitrogen) that percolate through the soil and contaminated groundwater, as well as sediment that is eroded from fields and washed into watercourses during and after rainfall. While studying the limnology of South Africa's major impoundments, Walmsely and Butty [52] described Nagle Dam as a phosphate limited oligotrophic system.

### 3.1.4 Nagle dam outflow

Seven significant components which explained 75% of the total variance were extracted at Nagle Dam Outflow station (**Table 7**). The first component which contributed 23% (**Table 8**) of the total variance is dominated with metal ions which can be related on natural geological processes such as weathering. The second component which explains 15.5% of the total variation is dominated by *E. coli*, suspended solids and turbidity (highlighted in bold black) (**Table 8**). These pollutants can be related to the discharge of sewage effluent and runoff from a community practicing open defecation.

### 3.1.5 Inanda dam inflow

At Inanda Dam Inflow station, seven components which explained 76% of the total variance were extracted (**Table 9**). Component 1 (**Table 10**) which is mainly metal ions and explains 11% of the total variance was comprised of metal ions (highlighted in bold black) which suggest that geological processes in the area could be significantly attributing to the water quality variation. The high positive correlation of chloride and component 1 also reflects the effect of anthropogenic pollutants

Midmar dam outflow: total variance explained									
Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.729	32.044	32.044	6.729	32.044	32.044	5.690	27.096	27.096
2	2.042	9.726	41.770	2.042	9.726	41.770	1.871	8.910	36.006
3	1.759	8.378	50.148	1.759	8.378	50.148	1.794	8.542	44.548
4	1.399	6.660	56.808	1.399	6.660	56.808	1.769	8.424	52.971
5	1.271	6.053	62.861	1.271	6.053	62.861	1.659	7.902	60.873
6	1.196	5.694	68.555	1.196	5.694	68.555	1.508	7.181	68.054
7	1.078	5.133	73.687	1.078	5.133	73.687	1.183	5.634	73.687

*Extraction Method: Principal Component Analysis.*

**Table 3.**  
*Extracted values of the significant components at Midmar dam outflow.*

Component Matrix <sup>a</sup>							
	Component						
	1	2	3	4	5	6	7
Calcium (Ca)	<b>0.906</b>	-0.100	0.117	-0.076	0.134	-0.065	0.054
Conductivity	<b>0.869</b>	0.028	0.083	0.114	0.030	0.060	-0.050
Magnesium (Mg)	<b>0.856</b>	-0.131	0.109	-0.236	0.171	-0.045	0.039
Alkalinity	<b>0.821</b>	-0.208	0.103	-0.103	-0.092	0.080	0.075
Ammonia (NH <sub>3</sub> )	<b>0.783</b>	-0.136	0.201	-0.051	-0.181	0.072	-0.036
Turbidity	<b>0.736</b>	0.442	-0.045	-0.171	-0.183	-0.146	0.062
Suspended Solids (SS)	<b>0.679</b>	<b>0.533</b>	-0.039	-0.187	-0.118	-0.163	0.079
Sulphates (SO <sub>4</sub> )	<b>-0.663</b>	0.052	-0.267	0.088	0.314	0.107	0.315
Silicon (Si)	<b>0.613</b>	0.253	0.099	0.390	0.123	-0.164	0.086
Dissolved Oxygen (DO)	<b>-0.544</b>	0.260	0.062	-0.368	0.195	-0.143	-0.019
Nitrate (NO <sub>3</sub> )	<b>0.519</b>	-0.334	0.068	0.130	-0.303	0.288	-0.061
Total Phosphate (TP)	0.205	<b>0.532</b>	-0.285	-0.332	0.016	0.291	-0.048
% NH <sub>3</sub> %	-0.247	0.292	<b>0.791</b>	0.065	0.097	0.314	0.154
pH	-0.336	0.463	<b>0.732</b>	0.005	0.061	0.103	0.130
Total Organic Carbon (TOC)	0.212	0.412	-0.169	0.459	-0.116	-0.355	0.306
Potassium (K)	0.399	0.046	-0.1608	0.308	<b>0.605</b>	-0.105	-0.160
Sodium (Na)	0.278	-0.360	0.245	-0.421	<b>0.533</b>	-0.148	-0.084
Chloride (Cl)	0.384	0.277	-0.303	-0.104	0.390	0.211	0.193
Temperature °C	0.320	-0.159	-0.057	0.437	0.209	0.558	0.160
SRP	0.063	0.434	-0.279	-0.095	-0.048	0.488	-0.443
<i>Escherichia coli</i> ( <i>E. coli</i> )	-0.031	0.230	0.225	0.355	0.127	-0.195	<b>-0.718</b>

Extraction Method: Principal Component Analysis.

<sup>a</sup>7 Components extracted.

Bold: Significant contributors to the respective principal component in their respective there column.

**Table 4.**

The correlation among the parameters measured and the extracted significant components at MDO.

on this station. Both legal and illegal effluent discharges from industrial areas such as Willowton, are the predominant pollution sources that affect Inanda Dam [53]. In developing countries, 70 percent of industrial wastes are dumped untreated into waters, impacting on the usability of the water resource [54].

Given the high positive correlation coefficient of Soluble Reactive Phosphate (SRP), Suspended Solids (SS), Total Organic Carbon (TOC), Total Phosphate (TP) and turbidity on the component 2, it can be claimed that pollutants in this group which explained 9% of the variation in water quality are emanating from agricultural activities (Table 10). The negative correlation noted between dissolved oxygen and component 5 indicates deterioration in the water quality. Since component 5 also exhibited a positive correlation to temperature, it can be deduced that climatic conditions could explain the 7.7% variation noted at this station. Component 6 which is mainly influenced by *E. coli* suggest the effects of sewage effluent into the water system. This could be attributed to the effluent from DV Wastewater Works which treat domestic effluent from Pietermaritzburg.

Nagle dam inflow: total variance explained									
Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.960	30.998	30.998	4.960	30.998	30.998	3.962	24.760	24.760
2	2.164	13.523	44.521	2.164	13.523	44.521	2.073	12.957	37.717
3	1.429	8.934	53.455	1.429	8.934	53.455	1.973	12.332	50.049
4	1.161	7.257	60.713	1.161	7.257	60.713	1.558	9.736	59.785
5	1.074	6.710	67.422	1.074	6.710	67.422	1.222	7.638	67.422

*Extraction Method: Principal Component Analysis.*

**Table 5.**  
 Extracted values of the significant components at Nagle dam inflow (NDI).

	Component Matrix <sup>a</sup>				
	Component				
	1	2	3	4	5
Turbidity	<b>0.822</b>	-0.253	-0.036	-0.214	0.052
Total Phosphate	<b>0.776</b>	-0.084	-0.350	0.307	0.068
Soluble Reactive Phosphate (SRP)	<b>0.757</b>	-0.207	-0.323	0.299	0.080
Suspended Solids (SS)	<b>0.741</b>	0.001	-0.213	-0.333	0.093
Total Organic Carbon (TOC)	<b>0.721</b>	0.099	-0.132	-0.410	0.152
<i>Escherichia coli</i> ( <i>E. coli</i> )	<b>0.718</b>	-0.404	0.047	0.273	0.234
Nitrate (NO <sub>3</sub> )	<b>0.677</b>	-0.278	0.337	-0.076	-0.251
Conductivity	<b>0.662</b>	0.020	0.383	-0.180	-0.336
Temperature	<b>0.566</b>	0.395	0.205	0.131	-0.325
% NH <sub>3</sub> %	0.363	<b>0.844</b>	-0.137	-0.088	0.152
pH	0.145	<b>0.819</b>	-0.094	0.010	0.228
Scenedesmus	0.160	0.336	-0.119	0.246	-0.283
Nitzschia	0.206	0.153	<b>0.711</b>	-0.130	0.087
Algal count cell	0.253	0.301	0.093	0.486	-0.334
Navicula	0.143	0.016	0.517	0.409	<b>0.612</b>
Ammonia- N (NH <sub>3</sub> )	-0.051	-0.185	-0.106	0.183	-0.206

Extraction Method: Principal Component Analysis.

<sup>a</sup>5 Components extracted.

Bold: Significant contributors to the respective principal component in their respective there column.

**Table 6.**

The correlation among the parameters and the significant components at NDI.

Nagle dam outflow: total variance explained						
Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.832	23.011	23.011	4.832	23.011	23.011
2	3.262	15.532	38.542	3.262	15.532	38.542
3	2.186	10.409	48.951	2.186	10.409	48.951
4	1.879	8.946	57.897	1.879	8.946	57.897
5	1.390	6.618	64.515	1.390	6.618	64.515
6	1.185	5.643	70.158	1.185	5.643	70.158
7	1.072	5.103	75.261	1.072	5.103	75.261

Extraction Method: Principal Component Analysis.

**Table 7.**

Extracted values of the significant components at Nagle dam outflow.

### 3.1.6 Inanda dam outflow

At Inanda Dam Outflow station, eight components explaining 75% of the total variance were extracted as depicted in **Table 11**. We hypothesised that pollutants in component 1 depicted in **Table 12** (11.8% and highlighted in bold black) were

	Component Matrix <sup>a</sup>						
	Component						
	1	2	3	4	5	6	7
% NH <sub>3</sub> %	-.158	.051	.364	<b>.690</b>	.483	-.092	.164
Alkalinity	.342	-.099	.339	-.377	.102	.149	.346
Calcium (Ca)	<b>.650</b>	-.537	-.454	.154	.103	-.007	.030
Chloride (Cl)	<b>.673</b>	-.154	.499	-.032	.055	.037	-.281
Conductivity	<b>.782</b>	-.017	.430	-.134	.114	.024	-.039
Dissolved Oxygen (DO)	-.496	-.280	.224	-.228	.016	.255	-.213
<i>Escherichia coli</i> ( <i>E. coli</i> )	.330	<b>.630</b>	-.300	-.071	.287	.227	-.175
Potassium (K)	<b>.707</b>	-.235	-.219	.259	-.269	-.213	.031
Magnesium (Mg)	<b>.774</b>	-.557	-.120	.036	.126	.087	.050
Sodium (Na)	.549	-.591	-.482	.257	.017	-.040	-.050
Ammonia (NH <sub>3</sub> )	-.027	.091	-.381	.117	-.047	.453	.426
Nitrate (NO <sub>3</sub> )	.498	.053	<b>.511</b>	.232	-.337	.048	-.408
pH	-.350	-.039	.181	<b>.671</b>	<b>.524</b>	-.032	-.052
Silicon (Si)	.442	.404	.255	-.162	.043	.011	.364
Sulphate (SO <sub>4</sub> )	.326	-.363	.139	-.445	<b>.583</b>	.266	.000
Soluble Reactive Phosphate (SRP)	.056	.148	-.166	-.385	.302	<b>-.597</b>	-.028
Suspended Solids (SS)	.312	<b>.618</b>	-.238	.193	.046	.404	-.197
Temperature	<b>.506</b>	.429	.292	.184	-.183	-.177	.441
Total Organic Carbon	.489	.497	.031	.116	-.047	-.133	-.031
Total Phosphate (TP)	.127	.354	-.303	-.227	.283	-.347	-.130
Turbidity	.402	<b>.692</b>	-.282	.006	.084	.109	-.165

Extraction Method: Principal Component Analysis.

<sup>a</sup>7 Components extracted.

**Table 8.**

The correlation among the parameters and the significant components at NDO.

mainly contributed by metal ions which reflects the geology of a catchment area. The positive correlation of sulphate and component 1 reflects the effect of anthropogenic polluting activities. Component 3 (**Table 12**) is mainly attributable to agricultural pollutant sources due to moderate positive high correlations with turbidity, nitrate and suspended solids. It is most plausible to suggest that turbidity and suspended solids is a result of surface runoff due to rainfall. *Escherichia coli* (*E. coli*) which dominates in component 6 normally can be attributed to wastewater treatment plants in Pietermaritzburg.

### 3.2 Cluster analysis

Cluster analysis was used to detect similarities among the sampling stations in the study area. The dendrogram shows that the six sampling stations in the area studied could be grouped into two significant clusters (A and B) as illustrated by **Figure 3**. Such is the case of the relatively large linkage distance at which the two groups combine, which indicates the Euclidean distances [47]. Cluster A mainly

Inanda dam inflow: total variance explained									
Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.005	31.840	31.840	7.005	31.840	31.840	5.946	27.025	27.025
2	2.888	13.127	44.967	2.888	13.127	44.967	2.443	11.104	38.129
3	1.865	8.478	53.445	1.865	8.478	53.445	2.019	9.179	47.309
4	1.427	6.485	59.930	1.427	6.485	59.930	1.922	8.737	56.045
5	1.363	6.196	66.127	1.363	6.196	66.127	1.695	7.704	63.749
6	1.171	5.325	71.451	1.171	5.325	71.451	1.385	6.297	70.046
7	1.012	4.599	76.050	1.012	4.599	76.050	1.321	6.004	76.050

*Extraction Method: Principal Component Analysis.*

**Table 9.**  
*Extracted values of the significant components at Inanda dam inflow.*

	Component Matrix <sup>a</sup>						
	Component						
	1	2	3	4	5	6	7
% NH <sub>3</sub> %	-0.155	0.269	<b>0.718</b>	0.193	0.209	-0.185	-0.170
Alkalinity	<b>0.746</b>	-0.186	0.182	-0.241	0.254	0.045	-0.065
Calcium (Ca)	<b>0.859</b>	0.215	-0.059	-0.247	0.030	0.093	0.063
Chloride (Cl)	<b>0.911</b>	0.140	0.089	-0.054	0.122	0.024	-0.090
Conductivity	<b>0.889</b>	0.169	0.009	-0.048	0.108	0.027	-0.096
Dissolved Oxygen (DO)	0.313	-0.033	0.128	0.464	<b>-0.640</b>	0.168	0.017
<i>Escherichia coli</i>	-0.297	0.400	-0.055	0.080	-0.122	<b>0.538</b>	-0.355
Flourine (F)	0.204	0.307	0.318	0.329	-0.067	0.236	-0.319
Potassium (K)	<b>.514</b>	0.496	-0.065	-0.122	-0.012	-0.240	0.116
Magnesium (Mg)	<b>0.734</b>	0.145	0.075	-0.161	0.074	0.411	0.176
Sodium (Na)	<b>0.882</b>	0.174	0.100	-0.017	0.045	0.009	-0.006
Ammonia (NH <sub>3</sub> )	0.163	0.330	0.194	0.239	-0.195	0.012	<b>0.681</b>
Nitrate (NO <sub>3</sub> )	0.493	0.194	-0.375	0.336	-0.191	-0.131	0.144
pH	-0.075	0.227	<b>0.717</b>	0.019	-0.260	-0.408	-0.102
Silicon (Si)	-0.581	0.216	0.178	0.174	0.322	0.192	0.306
Sulphate (SO <sub>4</sub> <sup>2+</sup> )	<b>0.799</b>	.339	0.088	0.076	0.071	0.061	0.005
Soluble Reactive Phosphate (SRP)	0.086	<b>0.553</b>	-0.461	0.396	0.271	-0.278	-0.102
Suspended Solids	-.522	<b>0.569</b>	-0.048	-0.382	-0.219	0.006	0.108
Temperature	-0.506	0.187	0.303	0.088	<b>0.544</b>	0.184	0.215
Total Organic Carbon (TOC)	-0.190	<b>0.557</b>	0.036	<b>-0.516</b>	-0.192	-0.249	-0.049
Total Phosphate (TP)	-0.303	<b>0.645</b>	-0.351	0.229	0.227	-0.142	-0.158
Turbidity	-0.585	<b>0.620</b>	-0.028	-0.241	-0.184	0.328	-0.010

Extraction Method: Principal Component Analysis.  
<sup>a</sup>7 Components extracted.  
 Bold: Significant contributors to the respective principal component in their respective there column.

**Table 10.**  
 The correlation among the parameters and the significant components at IDI.

consists of four sampling stations that were located mostly in the outflow of the river (NDO, IDO, MDO and NDI) while Cluster B mainly consist of two stations mainly dam inflow stations (MDI and IDI). Except for Nagle Dam Inflow, Cluster A basically comprised of dam outflow stations. These stations can be described as less polluted due to the dilution and retention effect. On the other hand, Cluster B composed of dam inflow stations. These stations can be described as more polluted as a result of activities along the river course. The PC results explained Section 4.1 of this chapter showed that poor agriculture practises resulting in runoff of agro-chemicals, organic matter, drug residues, sediments and saline drainage as well as sewage and industrial effluent discharges are key factors being reflected by the poor water quality results of the dam's inflow stations (Cluster B). These practices pose a risk to aquatic ecosystems, human health and productive activities. The significant presence of *E. coli*, suspended solids and turbidity in both Cluster A and B sampling stations indicates that raw water along uMngeni Basin is not fit for potable use before treatment.

Inanda dam outflow: total variance explained						
Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.575	25.342	25.342	5.575	25.342	25.342
2	2.597	11.804	37.146	2.597	11.804	37.146
3	2.066	9.390	46.536	2.066	9.390	46.536
4	1.582	7.192	53.728	1.582	7.192	53.728
5	1.360	6.183	59.911	1.360	6.183	59.911
6	1.217	5.534	65.445	1.217	5.534	65.445
7	1.135	5.161	70.606	1.135	5.161	70.606
8	1.009	4.589	75.194	1.009	4.589	75.194

Extraction Method: Principal Component Analysis.

**Table 11.**  
Extracted values of the significant components at Inanda dam outflow.

Component Matrix <sup>a</sup>								
	Component							
	1	2	3	4	5	6	7	8
Magnesium (Mg)	<b>0.903</b>	-0.053	-0.074	-0.034	-0.141	0.023	0.013	-0.014
Sodium (Na)	<b>0.881</b>	0.007	-0.096	-0.078	-0.102	0.030	-0.107	-0.029
Calcium (Ca)	<b>0.877</b>	-0.033	0.056	0.108	-0.146	0.174	-0.189	0.000
Chloride (Cl)	<b>0.841</b>	-0.050	-0.031	0.029	0.071	0.170	0.257	0.001
Potassium (K)	<b>0.800</b>	-0.076	-0.093	-0.100	-0.138	-0.411	-0.045	-0.011
Conductivity	<b>0.671</b>	-0.096	-0.186	0.042	0.241	0.182	0.046	-0.131
Alkalinity	<b>0.664</b>	-0.334	-0.367	0.268	0.138	0.059	-0.009	0.158
Sulphate (SO <sub>4</sub> )	<b>0.612</b>	0.374	0.080	-0.262	-0.118	0.055	0.095	-0.324
% NH <sub>3</sub> %	0.165	<b>0.816</b>	-0.177	-0.033	-0.212	-0.037	-0.011	0.376
pH	0.212	<b>0.781</b>	-0.046	-0.084	-0.295	0.039	-0.090	0.310
Temperature	0.027	<b>0.596</b>	-0.134	-0.120	0.347	-0.151	0.470	0.041
Turbidity	0.327	0.099	<b>0.785</b>	0.153	0.143	-0.057	0.028	0.046
Nitrate (NO <sub>3</sub> )	-0.009	0.192	<b>0.721</b>	0.039	-0.279	0.294	-0.290	-0.197
Suspended Solids	0.325	0.347	<b>0.531</b>	0.121	0.317	-0.194	0.170	-0.178
Ammonia (NH <sub>3</sub> )	0.271	-0.384	0.430	-0.071	-0.094	0.011	0.004	0.290
Total Phosphate (TP)	-0.008	-0.226	0.106	<b>-0.636</b>	-0.032	0.172	-0.051	0.217
Soluble Reactive Phosphate (SRP)	0.017	-0.366	0.232	<b>-0.605</b>	-0.017	-0.043	0.160	0.371
Flourine (F)	0.469	-0.136	-0.003	-.103	0.538	-0.089	-0.300	0.130
Dissolved Oxygen (DO)	0.041	-0.213	-0.176	0.398	-0.508	0.042	0.083	0.045
<i>Escherichia coli</i> ( <i>E coli</i> )	-0.094	0.149	-0.026	0.171	0.282	<b>0.832</b>	0.139	0.199
Silicon (Si)	0.095	-0.283	0.327	0.458	-0.161	-0.146	0.524	0.300

Component Matrix <sup>a</sup>								
	Component							
	1	2	3	4	5	6	7	8
TOC	-0.047	0.169	0.064	0.407	0.283	-0.203	-0.496	0.363

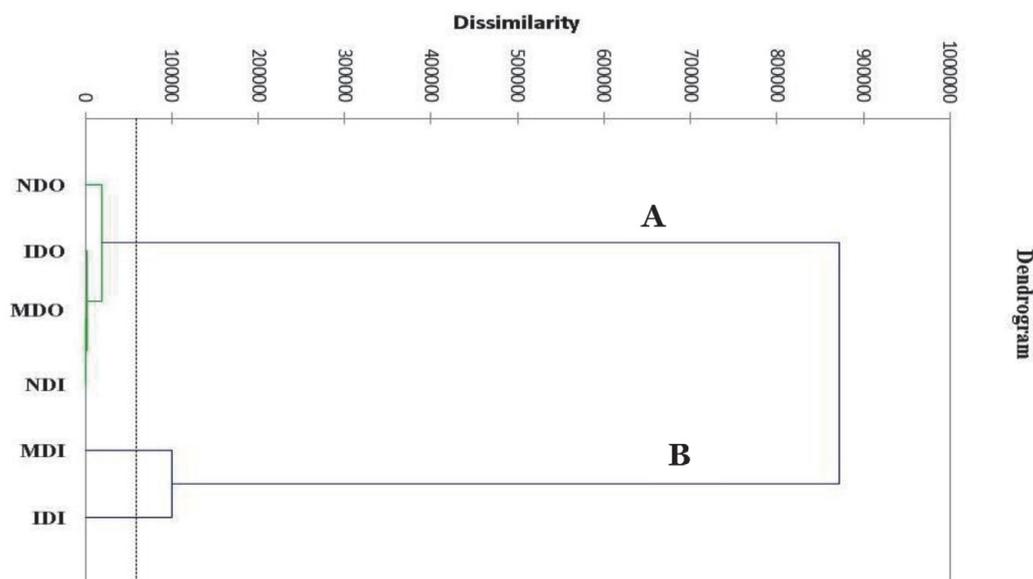
Extraction Method: Principal Component Analysis.

<sup>a</sup>8 Components extracted.

**Bold:** Significant contributors to the respective principal component in their respective there column.

**Table 12.**

The correlation among the parameters and the significant components extracted at IDO.



**Figure 3.**

Dendrogram of the stations along uMngeni basin.

## 4. Conclusions

Understanding the primary effects of anthropogenic activities and natural factors on river water quality is important in the study and efficient management of water resources. Hence, the PCA method assisted in the identification of significant parameters influencing water quality variations at the six stations studied in uMngeni Basin. The PCs extracted suggest that pollution sources along uMngeni Basin can be attributed to geological processes, sewage effluent, agricultural runoff and surface runoff pollutants. The findings could assist in reducing the number of parameters being monitored at any station and thus ultimately reducing the associated cost monitoring cost. It is recommended that, effluents be treated before discharge into the river. Additionally, it is recommended that buffer zone policies be enforced.

The result of the cluster analysis should also assist in categorising sampling sites according to pollution levels. Classification of sampling stations based on pollution level can assist in the designing of an optimal sampling strategy, which could reduce the number of sampling stations and associated costs. This study highlights the usefulness of multivariate statistical assessment such as PCA and CA in analysing complex databases, especially in the identification of pollution sources and to better

comprehend the spatial and temporal variations for effective river water-quality management. It is worthwhile to conclude that PCA and CA are better tools for better understanding concealed information about parameter variance and datasets. The study recommends the application of PCA and CA for interpreting bulk surface water quality data-sets.

## **Acknowledgements**

The authors gratefully acknowledge Durban University of Technology for hosting and funding the main author during his Master's Degree study.

## **Author details**

Innocent Rangeti<sup>1\*</sup> and Bloodless Dzwairo<sup>2</sup>

1 Department of Community Health Studies, Faculty of Health Sciences, Durban University of Technology, Durban, South Africa

2 Civil Engineering Midlands, Durban University of Technology, Imbali, South Africa

\*Address all correspondence to: [innoranger@gmail.com](mailto:innoranger@gmail.com)

## **IntechOpen**

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Vega M, Pardo R, Barrado E, Debán L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*. 1998;**32**(12):3581-3592
- [2] Dinar S. Water, security, conflict, and cooperation. *SAIS Review*. 2002;**22**(2):229-253
- [3] Chinhanga JR. Impact of industrial effluent from an iron and steel company on the physico-chemical quality of Kwekwe River water in Zimbabwe. *International Journal of Engineering, Science and Technology*. 2010;**2**(7)
- [4] Wei S, Wang Y, Lam JC, Zheng GJ, So M, Yueng LW, et al. Historical trends of organic pollutants in sediment cores from Hong Kong. *Marine pollution bulletin*. 2008;**57**(6):758-766
- [5] Ali M, Nelson AR, Lopez AL, Sack DA. Updated global burden of cholera in endemic countries. *PLoS neglected tropical diseases*. 2015;**9**(6): e0003832
- [6] Department of Water Affairs and Forestry. uMngeni River and Neighbouring Rivers and Streams. Forestry DoWAA: State of the Rivers Report Pretoria; 2003
- [7] Umgeni Water. Umgeni Water Infrastructure Master Plan. (2012/2013–2042-2043). Pietermaritzburg, South Africa: Umgeni Water. Planning Services EaSSD: UW; 2012. p. 2012
- [8] VanDerZel DW. Umgeni River Catchment Analysis. *Water SA*. 1975;**1**(2).
- [9] Hemens J, Simpson DE, Warwick RJ. Nitrogen and phosphorus input to the Midmar dam. Natal. *Water SA*. 1977;**3**(4):193-201
- [10] Statement UWUWP, Volume 2 F-YBP. Business Plan. 2012/13 to 2016/17. Pietermaritzburg. Water U: Umgeni Water; 2012
- [11] Couillard D, Lefebvre Y. Analysis of water indices. *Journal of Environmental Management*. 1985;**21**(2):161-179
- [12] Wepener V, Cyrus DP, Vermeulen LA, O'Brien GC, Wade P. Development of a Water Quality Index for Estuarine Water Quality Management in South Africa. Commission WR, editor. Pretoria; 2006
- [13] Yerel S, Ankara H. Application of multivariate statistical techniques in the assessment of water quality in Sakarya River, Turkey. *Journal of the Geological Society of India*. 2011;**78**(6):1-5
- [14] Alkarkhi AM, Ahmad A, Easa A. Assessment of surface water quality of selected estuaries of Malaysia: Multivariate statistical techniques. *Environmentalist*. 2009;**29**(3):255-262
- [15] Singh KP, Malik A, Mohan D, Sinha S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research*. 2004;**38**(18): 3980-3992
- [16] Simeonov V, Stratis J, Samara C, Zachariadis G, Voutsas D, Anthemidis A, et al. Assessment of the surface water quality in northern Greece. *Water Research*. 2003;**37**(17):4119-4124
- [17] Boyacioglu H. Spatial differentiation of water quality between reservoirs under anthropogenic and natural factors based on statistical approach. *Archives of Environmental Protection*. 2014;**40**(1):41-50
- [18] Varol M, Gökot B, Bekleyen A, Şen B. Water quality assessment and apportionment of pollution sources of Tigris River (Turkey) using multivariate

statistical techniques—A case study. *River research and applications*. 2012; **28**(9):1428-1438

[19] Ouyang Y. Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*. 2005;**39**(12):2621-2635

[20] Bengraïne K, Marhaba TF. Using principal component analysis to monitor spatial and temporal changes in water quality. *Journal of Hazardous Materials*. 2003;**100**(1–3):179-195

[21] Kazi TG, Arain MB, Jamali MK, Jalbani N, Afridi HI, Sarfraz RA, et al. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicology and Environmental Safety*. 2009;**72**(2): 301-309

[22] Juntunen P, Liukkonen M, Lehtola M, Hiltunen Y. Cluster analysis by self-organizing maps: An application to the modelling of water quality in a treatment process. *Applied Soft Computing*. 2013;**13**(7):3191-3196

[23] Jolliffe I. *Principal component analysis*: Wiley online. Library. 2005

[24] Jackson JE. *A user's Guide to Principal Components*: John Wiley & Sons; 2005.

[25] Lawless HT, Data Relationships HH. *Multivariate Applications*. Springer; 2010

[26] Kim J-O, Mueller CW. *Introduction to Factor Analysis: What it Is and How to do it*. SAGE Publications; 1978

[27] Cadima J, Jolliffe IT. Loading and correlations in the interpretation of principle compenents. *Journal of Applied Statistics*. 1995;**22**(2):203-214

[28] Statistics DJC. *Data Analysis in Geology*. John Wiley & Sons; 1986

[29] Razmkhah H, Abrishamchi A, Torkian A. Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: A case study on Jajrood River (Tehran, Iran). *Journal of Environmental Management*. 2010;**91**(4):852-860

[30] Mazlum N, Ozer A, Mazlum S. Interpretation of water quality data by principal components analysis. *Turkish Journal of Engineering & Environmental Sciences/Turk Muhendislik ve Cevre Bilimleri Dergisi*. 1999;**23**(1):19-26

[31] Haag I, Westrich B. Processes governing river water quality identified by principal component analysis. *Hydrological Processes*. 2002;**16**(16): 3113-3130

[32] Rangeti I. *Determinants of key drivers for potable water treatment cost in uMngeni basin*. Durban: Durban Universty of. Technology. 2015

[33] Panda UC, Sundaray SK, Rath P, Nayak BB, Bhatta D. Application of factor and cluster analysis for characterization of river and estuarine water systems – A case study: Mahanadi River (India). *Journal of Hydrology*. 2006;**331**(3–4):434-445

[34] Ge J, Ran G, Miao W, Cao H, Wu S, cheng L. Water quality assessment of Gufu River in three gorges reservoir (China) using multivariable statistical methods. *Advance Journal of Food Science and Technology*. 2013;**5**(7): 908-920

[35] Ragno G, Luca MD, Ioele G. An application of cluster analysis and multivariate classification methods to spring water monitoring Data. *Microchemical Journal*. 2007;**87**:119-127

[36] Johnson RA, Wichern DW. *Education P*. Prentice hall Englewood Cliffs, NJ: Applied multivariate statistical analysis; 1992

- [37] Otto M. *Multivariate Methods. Analytical chemistry* Weinheim: Wiley-VCH; 1998
- [38] Shrestha S, Kazama F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*. 2007;**22**(4):464-475
- [39] Basson MS, Rossow JD. Mvoti to Umzimkulu Water Management Area Overview of Water Resource Availability and Utilisation. Department of Water Affairs (DWA); 2003. Report No.: PWMA 11/000/00/0203 Contract No.: PWMA 11/000/00/0203.
- [40] Wilson AJ. Mvoti to Mzimkulu water management area situational assessment South Africa: A.J. Wilson & Associates International cc. In: cc AJWAI. 2001
- [41] Reghunath R, Murthy TS, Raghavan B. The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka. India. *Water research*. 2002;**36**(10):2437-2442
- [42] Williams B, Brown T, Onsman A. Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*. 2012;**8**(3):1
- [43] Kaiser HF. An index of factorial simplicity. *Psychometrika*. 1974;**39**(1): 31-36
- [44] Parinet B, Lhote A, Legube B. Principal component analysis: An appropriate tool for water quality evaluation and management— Application to a tropical lake system. *Ecological Modelling*. 2004;**178**(3-4): 295-311
- [45] Jianqin M, Jingjing G, Xiaojie L. Water quality evaluation model based on principal component analysis and information entropy: Application in Jinshui River. *Journal of resources and ecology*. 2010;**1**(3):249-252
- [46] Ryberg KR. Continuous Water-Quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in the Red River of the North, Fargo, North Dakota, 2003–05. 2006.
- [47] Ryberg KR. Cluster Analysis of Water-Quality Data for Lake Sakakawea, Audubon Lake, and McClusky Canal, Central North Dakota, 1990–2003. 2007.
- [48] Schertz TL, Alexander RB, Ohe DJ. The Computer Program Estimate Trend (Estrend), a System for the Detection of Trends in Water-Quality Data. U.S. Geological Survey, Water-Resources Investigations: Reston, Virginia; 1991
- [49] Lettenmaier DP, Conquest LL, Hughes JP. Routine streams and rivers water quality trend monitoring review. Washington: University of Washington; 1982. Contract No.: Technical Report 75.
- [50] Mateo-Sagasta J, Zadeh SM, Turrall H, Burke J. Water Pollution from Agriculture: A Global Review. Executive Summary: Rome, Italy: FAO Colombo. Sri Lanka: International Water Management ...; 2017
- [51] Dzwaairo B. Modelling raw water quality variability in order to predict cost of water treatment [DTech thesis]. Pretoria: Tshwane university of Technology. 2011
- [52] Walmsely RD, Butty M. Limnology of some selected south African impoundments. Pretoria: Water Research Commission/National Institute for water research Council for Scientific and Industrial Research. 1980
- [53] Neysmith J. Investigating Non-regulatory Barriers and Incentives to Stakeholder Participation in Reducing

Water Pollution in Pietermaritzburg's  
Baynespruit [Masters Degree].  
Pietermaritzburg: University of  
KwaZulu-Natal; 2008

[54] Geographic N. National Geographic.  
línea] Available: [http://www.nationalgeographic.com/es/historia/grandes-reportajes/los-barcos-de-losfaraones\\_8270/3](http://www.nationalgeographic.com/es/historia/grandes-reportajes/los-barcos-de-losfaraones_8270/3)[Último acceso: 20 Enero 2017]. 2017.

IntechOpen