

# 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)



# Application of Principal Component Analysis in Surface Water Quality Monitoring

Yared Kassahun Kebede and Tesfu Kebedee  
*Ethiopian Institute of Agricultural Research  
Ethiopia*

## 1. Introduction

Surface water systems such as rivers, lakes and ground water are affected by the natural processes such as erosion of minerals and dissolution of nutrients from the overlying rocks as well as anthropogenic influences from urban, industrial and agricultural activities. These degradation of surface water quality resulted in altered species composition and decreased overall health of aquatic communities (Ouyang et al., 2002). Therefore, in view of the spatial and temporal variations in the physico-chemical, hydrological and biological attributes of surface water systems, regular monitoring programs are required for reliable estimates of the water quality.

Rivers are the only ecosystem characterized by strong and predominantly unidirectional flows of materials that intimately connect the upstream and downstream reaches (Thompson & Lake, 2010) and thus, rivers play a major role in assimilation or carrying off the municipal and industrial wastewater and nutrient removal from agricultural fields and mineral rocks by surface runoff are responsible for river pollution. The municipal and industrial wastewater discharge constitutes the constant polluting source, whereas, the surface run-off is a seasonal phenomenon, largely affected by climate in the basin (Sing et al. 2004; Vega et al., 1998). Therefore, since rivers are the most important inland water resources for human consumption, it is imperative to have reliable information on characteristics and trends of water quality for effective water management.

The usual program of water quality assessment is the periodic measurement of multiple parameters in different monitoring stations which resulted in a complex data matrix of a large number of physico-chemical parameters that should be assessed to evaluate water quality (Chapman, 1992; Dixon & Chiswell, 1996). To simplify the problem of data reduction and to draw meaningful conclusion, several researchers used water quality indices (WQI) to verify the influence of waste discharges on water quality of streams and rivers (e.g. Cao et al., 1996; Pesce & Wundelin, 2000). However, WQI are often specific to the type of pollution or the geographical area involved (Rosenbeg & Resh, 1993) and has difficulty in universal applications. In addition, they do not provide evidences on the pollution sources (Pesce & Wundelin, 2000). Similarly, univariate procedure is a common technique applied in river water quality monitoring which does not adequately characterize simultaneous similarities and differences between samples or variables (Dixon & Chiswell, 1996). In addition, the

intrinsic values of analytical data are inadequate for the investigation of multivariate data table as the variables are correlated (Vega et al. 1998). Therefore, the indirect relationship between analytical parameters should be taken in to account for complete understanding of surface water quality.

Multivariate statistical methods have been widely applied in environmental data reduction and interpretation of multiconstituent chemical, physical biological measurements (Ramzakh et al., 2010; ter Braak & Verdonschot, 1995; Wenning and Erickson, 1994). These techniques have been applied for identification factors that influence water systems for reliable management of water resources as well as for rapid solution for pollution problems (Reghunath et al., 2002; Simeonov et al., 2004). In this regard, PCA is a very powerful multivariate statistical analysis method technique which is applied to reduce the dimensionality of a data set consisting of a large number of inter-related variables, while retaining as much as possible the variability present in data set (Jianqin et al. 2010; Sing et al. 2004). In addition, it allows to assess the association between variables, since they indicate participation of individual chemicals in several influence factors (Vega et al., 1998).

Since certain correlations exist among multi-indicators, PCA attempts to transform a large set of inter-correlated indicators into a smaller set of composite indicators, uncorrelated (orthogonal) variables called principal components (PCs), and simplifies the structure of the statistical analysis system (Jianqin et al. 2010). In this way, the correlation coefficient matrix measures how well the variance of each constituent can be explained by relationship with each of the others (Liu et al., 2003) and PC provides information on the most meaningful parameters, which describe the whole data set affording data reduction with minimum loss of original information (Helena et al., 2000; Vega et al., 1998). The characteristic root (eigenvalues) of the PCs is a measure of associated variances and the sum of the eigenvalues coincides with the total number of variables (Razmkhah et al. 2010). Correlation of PCs and original variables is given by loadings, and individual transformed observations are called scores (Wunderlin et al., 2001). Liu et al. (2003) classified the factor loadings as 'strong', 'moderate' and 'weak', corresponding to absolute loading values of >0.75, 0.75-0.50 and 0.50-0.30, respectively. However, loading reflects the relative importance of a variable within the component and does not reflect the importance of the component itself (Davis, 1986 cited in Ouyang, 2005).

A rotation of principal components can achieve a simpler and more meaningful representation of the underlying factors by decreasing the contribution to PCs of variables with minor significance and increasing the more significant ones (Vega et al. 1998). However, rotation might have resulted in an increase of the number of factors necessary to explain the same amount of variance of the original data set. However, it allows the association of small groups of variables and individual rotated factors with a clearer hydrochemical meaning (Vega et al., 1998) which greatly helps in data interpretation (Helena et al., 2000; Morales et al., 1999; Simeonov et al., 2003; Vega et al., 1998).

## 2. Application of PCA in physico-chemical water quality monitoring

The main problem in regular water quality monitoring programmes is the generation of large physico-chemical data matrix in a relatively short period of time which necessitate effective data handling mechanism for interpretation of results, association of variables and meaningful conclusion.

## 2.1 Anthropogenic and seasonal effects on water quality

Vega et al. (1998) applied exploratory data analysis for the assessment of the seasonal and polluting effects on water quality of Pisuerga river (Duero basin, Spain). These authors reported that the overall component loadings (i.e., no seasonal loading provided) for 22 experimental variables. PC1 explained 46.1% of the variance and PC2 explains 19.0% of the variance and it is highly participated by the variables related to anthropogenic pollution like BOD, COD, phosphorous or nitrogen. However, VF1 explained 37.2% of the total variance and is highly participated by mineral component of the river water (calcium, chloride, conductivity, dissolved solids, hardness, bicarbonate, magnesium, sodium and sulphate). VF2 contained 16.7% of the variance and include BOD, COD and ammonia. Therefore, these two verifactors identified the natural (mineral) component of pollution from anthropogenic organic pollution.

However, the study conducted by Vega et al. (1998) has not address seasonal effects which are known to have significant effect on water quality and might have affected the results of the study when interpreted for the different seasons. Ouyang et al. (2006) addressed the seasonal changes in surface water quality of the Lower Saint John River (LSJR) by the application of PCA. The authors found that PC1 explained 56.8% of the total variance measure the preponderance of physical (i.e., color, DO, and BOD) and organic-related (i.e., TKN, TOC, and DOC) water quality parameters over the mineral (i.e., alkalinity, salinity, and EC) and inorganic nutrient (i.e.,  $\text{TNH}_3$ ,  $\text{DNO}_x$ , TP, and  $\text{PO}_4^{3-}$ ) related water quality parameters while component 2 explained 26.8% of the total variance distinguished the importance of anthropogenic inputs and physical parameters (e.g., temperature and turbidity) over the natural inputs (e.g., pH, alkalinity, and salinity) during the spring season. However, unlike the cases for PC1 in spring and summer, the PC1 in fall which explained 54.2% of the total variance was positively contributed by mineral inorganic nutrient-related parameters and was negatively participated by the physical and organic-related parameters. Overall, their results revealed a high seasonal variation of water quality parameters in the dynamic river system and showed their significance (seasonal variation) when establishing the pollutant load reduction goals (PLRGs) and developing the total maximum daily loads (TMDLs). In a separate study, Ouyang (2005) applied PCA for evaluation of river water quality monitoring stations (22) in the same stream. Results showed that the first component accounted for about 94.6% and the second component accounted for about 4.5% of the total variance in the data set. PC1 is, therefore, the only one major source of data variation and 3 monitoring stations were identified as less important (non-principal) in explaining the annual variance of the data set. The authors attributed the very high variation explained by PC1 to the use of monitoring stations rather than water quality parameters in. In fact, it is expected that the water quality parameters which are controlled by hydrological, chemical and biological conditions to have higher correlations than water quality parameters which are solely controlled by hydrological conditions.

Numerous studies also confirmed that multivariate statistical techniques served as an excellent exploratory tool in understanding their temporal and spatial variations on water quality (Sing et al, 2004). The application of PCA by Razmkhah et al. (2010) discriminate the anthropogenic and "natural" influences on Jajrood river in Iran. PCA has allowed identification of a reduced number of mean 5 varifactors, pointing out 85% of both temporal and spatial changes. Rotation of the selected factors explained that VF1 (mineral contents)

and VF2 (mineral and anthropogenic contamination) in spring time identified sites with worst water quality (high mineral content and higher organic pollution) and 43% of the existing variance was briefly contributed by minerals (temporal variations) whereas 26% by anthropogenic factors (spatial variations). However, PCA extracted 3 VFs for the summer season and 15.5% of the variance is contributed to by organic factors, 15.5% by minerals and 55% by both and identified the most polluted rivers. Performing PCA in the autumn season found that 39% of the variation was due to temporal factors, 10% due to organic factors and 36% was the result of both sources. PCA extracted 4 VFs for winter season and 38% was due to minerals, 12% due to organics and 31% developed by both sources. Overall, every season revealed the presence of either mineral or anthropogenic or both sources of pollution.

Recently, Fan et al. (2010) applied PCA for spatial water quality assessment and pollution sources identification in Northern, Western and Eastern part of Pearl River delta (China). The results of the PCA suggested the parameters responsible for water quality variations in North River region was mainly related to organic related parameters (DO and  $\text{COD}_{\text{Mn}}$ ), inorganic nutrients ( $\text{NH}_3\text{-N}$  and TP) and metal Hg; but in East River region, it was mainly related to organic related parameters ( $\text{BOD}_5$ ) and inorganic nutrients ( $\text{NH}_3\text{-N}$  and TP), and in West River Region, mainly related to organic related parameters ( $\text{COD}_{\text{Mn}}$ ) and inorganic nutrients ( $\text{NH}_3\text{-N}$  and TP). Therefore, PCA offer a useful tool for assessment of water quality and management of water resources in some regions with a large number complex water quality datasets involved. Similarly, Sing et al. (2004) applied multivariate statistical techniques for evaluation of temporal and spatial variations in water quality in Gomti river (India). A varimax rotation (raw) of the PCs to six different VFs of eigenvalue  $> 1$  which are considered significant (Kim & Mueller, 1987; Liu et al., 2003) explained about 71% of the total variance. VF1 explained 17.6% of total variance and has strong positive loadings on EC, chloride, potassium and sodium and this VF represents a mineral component of the river water. VF2, explained 16.2% of total variance and has strong positive loadings on BOD and COD which represent anthropogenic pollution sources. Overall, these results from temporal PCA suggested that most of the variations is explained by the set of soluble salts (natural) and organic pollutants (anthropogenic). However, this finding is in contrast to other studies (e.g. Fan et al., 2010; Vega et al., 1998) as PCA does not result in much data reduction, as it still need 14 parameters (about 60% of the 24 parameters) to explain 71% of the data variance. Parinet et al. (2004) successfully reduced the number of analytical parameters from 18 to 4 (pH, conductivity, UV absorbance at 254 nm and permanganate index for raw water) in a study of 10 tropical lakes in Ivory Coast without notably impairing the quality of the PCA representation. However, this difference might be related to the difference in the water system (river vs. lake) and geographical factors. Overall, simplification of water quality parameters to easily quantifiable ones eases water quality monitoring programmes.

### 3. Changes in biological community structure

Kebede et al (2010) applied PCA to detect changes in community composition of macroinvertebrates arising from wet coffee processing effluents in major coffee producing region of Ethiopia by comparing upstream sites (control sites without any impact from the effluent and other possible pollutants because of their location above processing stations) with downstream locations (locations below coffee processing stations which are effluent

receivers). The relationships between the environmental and biological data were assessed using canonical multivariate analysis with the software program CANOCO 4.5 (ter Braak, Smilauer 2002). First, detrended correspondence analysis (DCA) of square-root transformed taxa abundance, with down weighting of rare taxa, detrending by segments and non-linear rescaling was used to determine the biological turnover, or gradient length, of the species data set. DCA for taxa (species) abundance of the first axis was less than 3, implying that taxa abundance exhibit linear response to environmental gradients (Leps & Smilauer, 2003). Therefore, principal component analysis (PCA) was used in the ordination of taxa abundance, and sampling sites by focusing scaling on inter-sample distances, standardization the species score (species score divided by the standard deviation), log transformation and centring by the species. In addition, to check the influence of environmental variable in explaining the variation among response variables (species), a PCA analysis of the response variable was run by using the physico-chemical data as supplementary environmental variable.

### 3.1 Physico-chemical parameters

There was a highly significant variation between BOD values of the study sites ( $p < 0.01$ ). BOD levels extend from 0.8 mg/l at upstream site of Urgessa river to 1900 mg/l and 1700 mg/l at downstream sites of Bore and Fite rivers, respectively. Similarly, there was a significant variation between DO values of the study sites as expected ( $p < 0.05$ ). The upstream sites showed good oxygen content as the DO values were above 5mg/l. However, DO value is totally depleted at the downstream site of Bore river which is in agreement with the high BOD value recorded for the site. Conversely, a slight reduction in pH values on average (7.03 to 6.74) might be attributed to the high assimilation capacity of water. The relatively higher amount of TDS at the upper site of Chiseche river might be attributed by the high mucilage coming out from coffee processing stations which are located around Chiseche river while the variation in temperature could be related to daily temperature variation during sampling period. A general pattern of  $\text{NO}_3$  and  $\text{NH}_4^+$  increment at downstream sites compared the upstream sites was observed during the sampling periods. Overall, the result of the physico-chemical analysis supports similar findings such that the main ecological effect of organic pollution in a watercourse is the decrease in oxygen content (Murthy et al., 2004; von Enden & Calbert, 2002).

### 3.2 Descriptive analysis of macroinvertebrates

To assess the downstream water quality (river water that receive discharge from wet coffee processing stations), 6047 macroinvertebrate individuals representing 27 different taxa were collected and identified from riffle sampling sites at the upstream and downstream locations (Table 2).

Species abundance is generally believed to be a useful measure of the severity of pollution (Sheehan, 1984). The total number of individuals found at the downstream sites was 5459 which was compared to 588 individuals collected from their respective upstream sites. The highest number of individuals was found at the lower course of Guracho river (2166 individuals) while the lowest number was recorded at the upper course of Chore river with a total of 3 individuals. Almost all the sampling sites displayed higher number of individuals at the downstream sites than the upper ones (Table 3). The highest number of

Rivers	No.		BOD		DO		TDS		pH		Temp	
	U	L	U*	L*	U	L	U	L	U	L	U	L
Sugo	1	2	1.5	600	5.98	5.98	120	50	7.06	7.06	18.3	18.3
Sunde	3	4	4	330	6.82	6.54	50	245	7.17	6.56	15.4	15.3
Chiseche	5	6	1	2.3	6.6	6	295	210	6.8	7.35	17.5	18.3
Songa	7	8	270	150	5.4	5.4	195	220	6.73	6.63	17.4	17.2
Urgessa	9	10	0.8	200	7.64	7.13	145	275	7.8	6.99	16.1	16.1
Janje	11	12	1	600	5.15	6.87	95	130	6.73	7.39	19	17.4
Fite	13	14	1.3	1700	7.02	6.35	95	140	7.2	6.99	17.8	18.5
Guracho	15	16	6	5.3	6.15	5.71	60	90	6.83	6.75	18.3	18.5
Funtule	17	18	110	2.5	5.15	3.93	120	135	6.73	5.91	19	19.5
Bore	19	20	2.3	1900	6.49	0	160	255	6.96	5.87	17.2	18.1
Chore	21	22	1.9	3.2	6.04	3	120	125	7.27	6.59	18.5	19.6
Average			36.3	499.4	6.22	5.17	132.3	170.5	7.03	6.74	17.7	17.9

Table 1. Mean physico-chemical characteristics of water samples of rivers of Jimma zone at locations of effluent discharge points. All units except pH and temperature (°C) are in mg/l. (Data from Kebede)

IntechOpen

Family name	Code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Atericidae	ATR	-	-	-	-	-	3	1	1	2	-	-	-	-	-	1	-
Baetidae	BAE	2	-	-	-	4	2	5	6	-	1	40	15	-	1	22	6
Ceratopogenidae	CER	25	70	-	10	-	-	-	-	13	45	-	21	-	58	-	355
Chaoboridae	CHA	19	46	-	-	-	-	-	-	12	28	-	16	-	29	-	200
Chironomidae	CHR	107	871	1	42	10	3	5	11	73	157	2	224	55	506	2	1602
Corydalidae	COR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Elmide	ELM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
Ephemeridae	EMP	2	-	-	-	-	-	-	-	11	-	-	-	-	-	-	-
Ephemellidae	EPH	-	-	-	-	-	-	-	-	1	-	-	1	-	-	3	-
Fiddler Crab	FDC	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-
Gyrinidae	GYR	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
Heptageniidae	HEP	-	-	-	-	-	-	1	-	1	-	1	-	-	-	-	-
Hydrophilidae	HYR	1	-	-	-	-	-	-	-	-	-	-	-	9	-	-	-
Hydropsychidae	HYD	-	-	1	-	-	13	7	7	19	23	6	3	-	-	9	-
Hydroptilidae	HYP	-	-	-	-	-	-	-	-	8	-	-	-	-	-	-	-
Libellulidae	LIB	-	-	-	-	1	1	2	-	-	-	-	-	-	-	-	-
Macromiidae	MAC	1	-	-	-	-	2	2	-	-	-	2	-	-	-	6	-
Muscidae	MUS	-	-	-	-	-	-	2	-	-	-	4	9	-	-	5	-
Perlidae	PEL	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1	-
Perlodidae	PER	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-
Planarium	PLA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
Polycentropidae	POL	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-
Psychomidae	PSY	-	-	-	-	-	-	-	2	5	-	-	-	-	-	-	-
Round worms	RW	-	-	-	-	-	8	1	-	2	-	-	2	-	-	-	-
Oligochaeta	OLI	19	131	-	5	-	-	-	5	-	-	-	1	-	-	6	-
Syrphidae	SYR	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-
Tipulidae	TIP	-	-	2	-	-	2	1	-	-	-	-	-	-	-	3	-

Table 2. Macroinvertebrate community identified from each sampling sites with their given codes. M given in Table 1. -indicates absence of the species from the site during the sampling period.

IntechOpen

Names	Species Richness		Total Species		Maximum		Shannon Index		Equitability Index		EPT	
	U	L	U	L	U	L	0.6688	0.4056	0.4069	0.2467	5	0
Sugo	177	1118	9	4	107	871	0.4515	0.323	0.2747	0.1965	1	0
Sunde	4	57	3	3	2	42	0.3488	0.7905	0.2122	0.4809	4	15
Chiseche	15	35	3	9	10	13	1.000	0.6883	0.6085	0.4188	17	13
Songa	32	32	13	6	7	11	0.8225	0.5866	0.5005	0.3569	40	24
Urgessa	147	254	11	5	73	157	0.4757	0.5116	0.2894	0.3113	49	19
Janje	57	292	7	9	40	224	0.2775	0.3139	0.1688	0.1910	9	1
Fite	68	594	4	4	55	506	0.8485	0.4356	0.5163	0.2651	35	6
Gurracho	59	2166	11	5	22	1602	0.2606	0.3132	0.1586	0.1905	0	1
Funtule	11	37	3	4	9	28	0.4429	0.0489	0.2695	0.0297	10	2
Bore	15	476	4	5	9	467	0.4771	0.4439	0.2903	0.2701	1	2
Chore	3	398	3	7	1	329	0.6688	0.4056	0.4069	0.2467	5	0
Total	588	5459									171	8
Average							0.5725	0.44192	0.3359	0.2688		

Table 3. Descriptive analysis of the macroinvertebrate community, summary of diversity, equitability study sites

IntechOpen

species found was 13 in Songa lower with 32 individuals. The upper and lower course of Sunde river and the upper course of Chiseche and Chore rivers had only 3 species, respectively. The maximum number of individuals which belong to the same species was found at the downstream site of Gurracho river (1602 individuals of the Chironomidae family). The general increment in total number of individuals at the downstream sites might be attributed to the enrichment of downstream sites from coffee processing wastewater, which provides physiologically adapted organisms to exploit the excess nutrients available despite low oxygen tension and this response of macroinvertebrates to organic pollution has been well documented (Cao et al., 1996; Hilsenhoff, 1988; Johnson et al., 1993). However, the general increment in species richness at the downstream sites was not supported with species diversity. Hence, the downstream site of Gurracho river which had 2166 individuals were found to contain only 5 species while the upstream sites of Chore river which had 3 individuals were belong to different species. This finding is consistent with Johnson et al. (1993) in that pollution tolerant species like Chironomidae species dominate the impacted sites.

### 3.3 Biotic and diversity indices

In order to understand the effect of wet coffee processing discharge on the biotic environment of the rivers, different diversity indices were tested (Table 3). These indices would indicate the environmental impact of coffee processing activities on the surrounding environment. The highest Shannon index was found at the upstream site of Songa river while the lowest Shannon index was found at the downstream sites of Bore and Urgessa river (Table 3). Similarly, Shannon index decreased from 0.57 at the upstream sites to 0.44 at the downstream sites. But the Equitability index which describes the evenness of species distribution within the site showed a different pattern. The highest and lowest Equitability indices were found at the upstream and downstream sites of Songa and Bore rivers, respectively. High values of diversity indices and equitability at the downstream sites of Songa river despite its presence below coffee discharge points, were unimpacted by coffee processing effluent. This finding of biological data is also supported by physicochemical data. Its good river water quality might be related to efficient coffee effluent control mechanism (lagoon) and the relative placement of the stations at a higher distance from the rivers i.e. large area possessed by the station enabled the construction of efficient pits to the containment of both coffee wastewater and pulp. In general, Equitability index decreased from 0.336 at the upstream sites to 0.268 at the downstream sites and more than 50% of the sites had an Equitability index of less than 0.3 and, thus, few species had dominated these sites.

The highest number of Ephemeroptera, Plecoptera and Trichoptera (EPT) individuals was found at the upper course of Janje river followed by Urgessa river, while the lowest EPT taxa were found in almost all of the downstream sites and upstream site of Funtule rivers. The total number of EPT taxa at the upstream sites was 171 and declined to 83 at the downstream sites. However, the number of Chironomidae (CHR) sharply increased from 264 at the upstream sites to 4240 at the downstream sites (Table 3), which is about 16 times higher than the upstream sites. The low diversity at the upstream sites of Fite and Funtule rivers was mainly related to the dominance of these sites by Chironomidae family. It comprises more than 76% of all the individuals found in at these sites. The situation is also the same at the upper course of Funtule river. It had three species which was highly dominated by Chironomidae taxa. Many reports have shown a significant decrease in species diversity indices associated with pollution

(Norris & Georges, 1993; Sheehan, 1984). Similarly the high EPT taxa at the upstream sites of Urgessa and Janje is related to their low BOD values.

With regard to the Hiselhof Family Biotic Index (HFBI) evaluation, the upper course of Bore (1.933) river had the lowest HFBI score; whereas Funtule river had the highest HFBI value followed by the lower course of Sugo rivers. Similarly, the family biotic index increased from 5.585 at the upstream sites to 7.306. In fact 14 of the 22 downstream sites had HFBI values above 6.51. . The abundance of pollution tolerant Oligochaeta taxa and Syrphidae species at the downstream sites of Sugo and Funtule rivers resulted in high FBI score. Blood-red chironomids and other dipterans (e.g., Ceratopogenidae and Chaoboridae) that are considered indicators of severely polluted sites were found at the downstream sites in high abundance than upstream sites. In fact, a benthic community dominated by one or few taxa is often indicative of environmental stress.

### 3.4 Principal component analysis

The separate analysis of environmental variables and biological indicators resulted in a slight difference results in depicting the pollution gradient along the study rivers. In fact this is expected since the abundance of species at a given site is more a reflection of past environmental conditions (Richard et al., 1997), while measurements of physical and chemical factors may be more of an indication of present conditions. However, the combination of both environmental variables (Table 1) and biological indicators (Table 2) by PCA was able to clearly define and explain the pollution gradients.

#### 3.4.1 Community composition of macroinvertebrates

Based on the species data of the PCA analysis (Fig.1), the first axis described 56.7% of the variation in the species composition while the second axis described 11.7% of the variation. Therefore, the first two axes were responsible for 68.4% of the variation in the dataset.

The dissimilarity between sites ( 12-2) dominated by Chaoboridae (CHA), Ceratopogenidae (CER), Chironomidae (CHR), Oligochaeta (OLI) on the right side of the first axis with those sites (15-8) at the top left of the diagram was high as shown by the distance in the PCA analysis.

The distribution of Chaoboridae, Ceratopogenidae Chironomidae were centered at sites 10 (Urgessa lower), 22 (Chore lower), 14 (Fite lower), 2 (Sugo lower) and 16 (Gurracho lower). The distribution of Chironomidae, the most abundant taxa, at the downstream sites indicated the impact of coffee processing effluent on the community composition of these particular sites. Similarly, many reports have documented a significant increase by Chironomidae taxa following toxic exposure (eg. Clements, 1994). On the contrary, Hydropsychidae (HYD), Baetidae (BAE), Heptagenidae (HEP), Perlidae (PER), and Psychodidae (PSY) was found at the negative of the first axis. These species have been used as indicators of good water quality (Cao et al., 1996; Legesse, 2001). In general, the PCA rendered three classifications: the extremely impacted downstream sites at the right side of the first axis and the two classes located at the right and negative side of the second axis which mostly constitute the upstream sites and their separation is attributed to their difference in species composition.

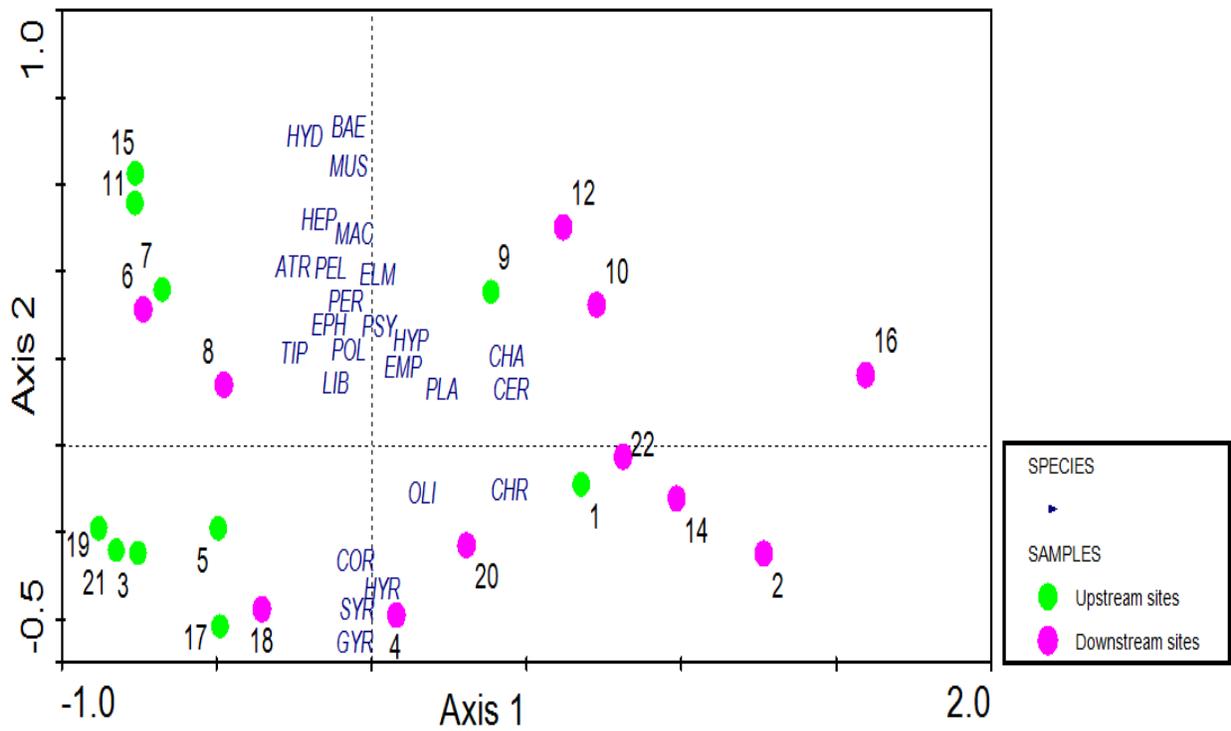


Fig. 1. PCA biplot of samples and macroinvertebrates based on the first two axes. Name of sampling points are given in Table 1 and full name of abbreviation codes of macroinvertebrate taxa is given in Table 2.

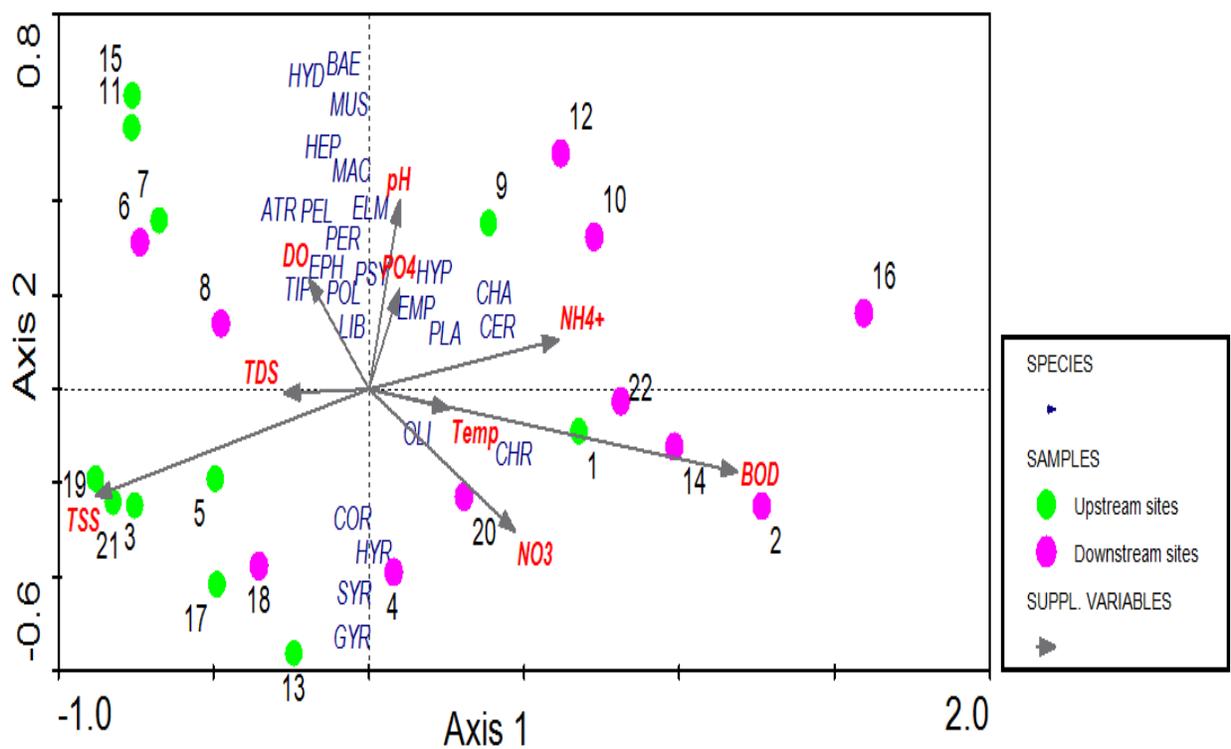


Fig. 2. PCA triplot of samples, macroinvertebrates and environmental variables based on the first two axes.

The species and environmental variables together described about 41% of the variation in the data set and the first two axes are responsible for 70% of these variations or 28.7% of the total variation (70% X 41%) suggesting that the relative importance of the first two axes in describing the variation in taxa separation and environmental variables.

In the ranking of the projection points, the origin (0, 0) indicates the global average of the variable. Therefore, TDS, Temperature and PO<sub>4</sub> did not change much (short arrow length) and hence their influence were minimal as compared to BOD, TSS, NO<sub>3</sub>, pH, and DO environmental lines which displayed the maximum rate of change across the diagram. Therefore, variations in macroinvertebrate composition were strongly correlated with these environmental variables. The high oxygen weighted average for oxygen loving species of Hydropsychidae (HYD), Baetidae (BAE), Heptagenidae (HEP), Perlidae (PER) and Psychodidae (PSY) is as expected since Heptagenidae is in Ephemeroptera family while Hydropsychidae, Perlidae and Psychodidae are in Trichoptera family. Ephemeroptera, Trichoptera and Plecoptera are considered to be abundant in oxygen rich water, very sensitive for pollution and used as a water quality monitoring index in Ethiopia (Kebede et al., 2010; Legesse, 2001). Therefore sites 8 (Sugo lower), 6 (Chiseche lower), 7 (Songa upper), 11 (Janje upper), and 15 (Gurracho upper) had higher DO values than the rest sites.

On the contrary, organically polluted sites were found to be dominated by pollution tolerant species of Chaoboridae (CHA), Oligochaeta (OLI), Ceratopogonida (CER) and Chironomidae (CHR) taxa and except the taxa of Oligochaeta, all the species are in the Diptera family. The special characteristics of the above mentioned taxa are their ability to flourish in very low oxygen (high BOD) level and dominate the entire site with great abundance. Sites 14 (Fite lower) and 2 (Sugo lower) and 20 (Bore lower) had the highest BOD values than the rest of the sites. Therefore the species-site-environment PCA triplot diagram clearly discriminated unimpacted sites from the severely impacted ones.

### 3.4.2 Sporulation pattern of aquatic hyphomycetes

Recently, Kebede applied PCA for discrimination of the sporulation pattern of aquatic hyphomycetes (aquatic fungi) communities on oak leaves immersed in a low order stream for up to 12 weeks in Candal stream (Central Portugal; 40° 4' 44" N and 8° 12' 10" W). In this study, the significance of the sporulation temperature at the peak of sporulation (Table 4), day 28 at 20°C (T4\_20), day 42 at 10°C (T6\_10) & 15°C (T6\_15) and day 56 at 5°C (T8\_5), was evaluated using PCA.

The non metric dimensional scaling (MDS) was applied in Winkyst program by using the Bray-Curtis distance matrix for producing the species axes (coordinates) that were used as a species data in principal component analysis (CANOCO 4.5; ter Braak & Smilauer, 2002).

The response of the dominant species at the four temperature values is shown in Fig. 3. The stress value (fit of regression) of 0.038 depicted the excellent representation of the coordinate axes in discriminating the samples (sampling dates and incubation temperature) in the ordination diagram. The supplementary species data indicate the species of *Clavariopsis aquatica* (AQU) and *Mycocentrospora acerina* (ACE) were associated with the discrimination of the positive side of the axis (right side) by over-dominating the sporulation at 5°C.

Conversely, the species of *Tetrachaetum elegans* (ELE) and *Articulospora tetracladia* (TET) and to a lesser extent *Alatospora acuminata* (ACU) and *Clavatospora longibrachiatum* (LON) were associated with the discrimination of the negative side of the axis (left side). These species highly dominate the sporulation at 10, 15 and 20°C. In addition, the PCA diagram shows a pattern of species dominance (longer arrow length) mainly due to ELE and ACE during the study period. Overall, the PCA analysis showed the species specific response to temperature variation by the dominant species. This is particularly pronounced during the peak period of sporulation with ACE, AQU, ELE and TET preferentially sporulate at 5°C, 10°C, 15°C and 20°C, respectively. This pattern revealed by PCA appears to be due primarily to the differential response by the late colonizer ACE and early colonizer ELE to low and high temperatures, respectively. This finding adds to similar changes in community composition induced by temperature (Bärlocher & Kendrick 1974, Suberkropp, 1984; Gessner et al. 1993; Bärlocher et al. 2008; Dang et al., 2009; Fernandes et al., 2009; Ferreira & Chauvet, 2011).

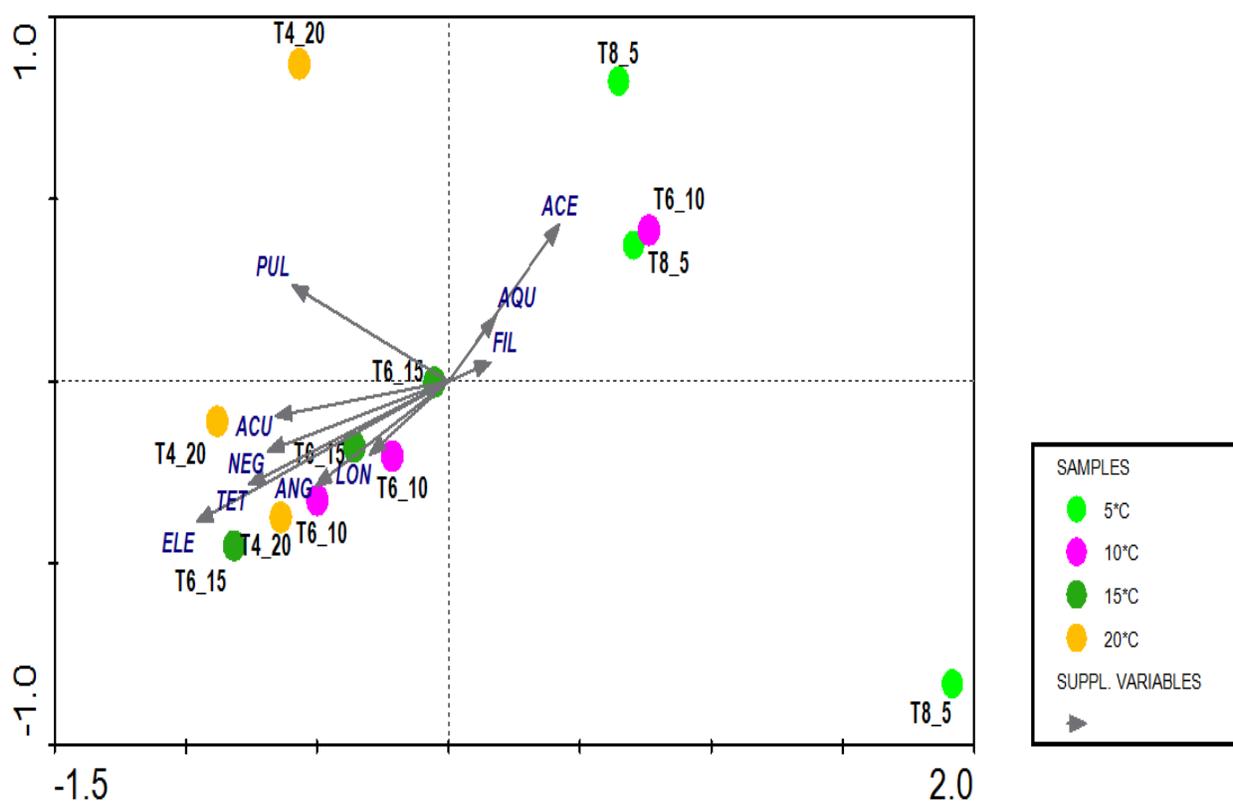


Fig. 3. PCA diagram of the sporulation temperature constructed on the species axes. The original species data was used as a supplementary variable.

Species	T8_5			T6_10			T6_15			T4_20		
	1	2	3	1	2	3	1	2	3	1	2	3
<i>Anguillospora filiformis</i> (FIL)	0	0	3	1	2	0	0	0	0	0	0	0
<i>Alatospora acuminata</i> (ACU)	5	0	1	22	32	1	0	22	7	12	65	10
<i>Alatospora pulchella</i> (PUL)	0	0	0	0	0	0	3	0	2	8	5	0
<i>Articulospora tetracladia</i> (TET)	0	1	1	3	13	0	15	8	8	4	32	40
<i>Clavariopsis aquatica</i> (AQU)	33	3	24	17	17	33	3	15	79	0	9	13
<i>Clavatospora longibrachiatum</i> (LON)	4	0	5	60	48	4	13	7	5	0	0	0
<i>Mycocentrospora acerina</i> (ACE)	70	1	27	0	2	0	0	2	0	0	0	0
<i>Stenoclaadiella neglecta</i> (NEG)	0	0	0	0	0	0	5	1	0	2	1	5
<i>Tetrachaetum elegans</i> (ELE)	10	0	13	89	38	10	157	83	58	17	106	134
<i>Tricladium angulatum</i> (ANG)	0	0	0	0	0	0	1	0	0	0	0	1

Table 4. Abundance of aquatic hyphomycetes conidia associated with oak leaves during the peak period of sporulation at three replications. T4, T6 and T8 represent 4, 6 and 8 weeks of incubation period in weeks while 5, 10, 15 and 20 represent incubation temperatures (°C).

#### 4. Conclusions

The usual program of water quality assessment depends on the periodic measurement of multiple parameters in different monitoring stations which resulted in a complex data matrix of a large number of physico-chemical parameters. Therefore, to simplify the problem of data reduction and draw meaningful conclusion multivariate statistical techniques have been widely used. In this regard, PCA, a powerful multivariate statistical technique, is applied to reduce the dimensionality of a data set while retaining as much as possible the variability present in data set and allows to assess associations between variables. PCA has been applied for the assessment of the seasonal and polluting effects on water quality, evaluation of monitoring stations, temporal and spatial variations on water quality and pollution source identification with in the river basin. Recently, PCA was applied to assess the water quality of rivers impacted by organic pollution. The combination of both environmental variables and biological indicators in PCA was able to clearly define and explain the pollution gradients. Similarly, PCA enabled to identify the distinct sporulation pattern of aquatic hyphomycetes communities incubated at four different temperatures. Therefore, the use of PCA to detect changes in community structure, ecological integrity of surface water systems and environmental impact assessment will enable to understand an integrative health condition of surface water systems due to chemical, physical and biological stressors.

## 5. Acknowledgments

I would like to thank Professor Felix Bärlocher and Professor Cristina Canhoto for their guidance and constructive comments during the sporulation study in Portugal. I am also grateful to Dr. Fassil Assefa for his valuable comments and guidance during the study in Jimma zone, Ethiopia.

## 6. References

- Bärlocher, F. & Kendrick, B. (1974). Dynamics of fungal populations on leaves in a stream. *Journal of Ecology*, Vol.62, No. 3, pp. 761-791.
- Bärlocher, F.; Seena, S.; Wilson, K. & Williams, D. (2008). Raised water temperature lowers diversity of hyporheic aquatic hyphomycetes. *Freshwater Biology*, Vol. 53, No. 2, pp. 368-379.
- Cao, Y.; Bark, A. & Williams, W. (1996). Measuring the responses of macroinvertebrate communities to water pollution: a comparison of multivariate approaches, biotic and diversity indices. *Hydrobiologia*, Vol. 341, No. 1, pp. 1-19.
- Chapman, D. (1992). *Water Quality Assessment; A guide to the use of biota, sediments and water in environmental monitoring*. University Press, Cambridge.
- Clements, W. (1994). Benthic invertebrate community responses to heavy metals in the Upper Arkansas River Basin, Colorado. *Journal of the North American Benthological Society* Vol. 13, No. 1, pp. 30-44.
- Dang, C.; Schindler, M.; Chauvet E. & Gessner, M. (2009). Temperature oscillation coupled with fungal community shifts can modulate warming effects on litter decomposition. *Ecology* Vol 90, No. 1, pp. 122-131.
- Dixon, W. & Chiswell, B. (1996). Review of aquatic monitoring program design. *Water Research*, Vol. 30, No. 9, pp. 1935-1948.
- Fan, x.; Cui, B.; Zhao, H.; Zhang, Z. & Zhang, H. (2010). Assessment of river water quality in Pearl River Delta using multivariate statistical techniques, *International Society for Environmental Information Sciences 2010 Annual Conference (ISEIS), Procedia Environmental Sciences*, Vol. 2, pp. 1220-1234.
- Fernandes, I.; Uzun, B.; Pascoal, C. & Cássio, F. (2009). Responses of aquatic fungal communities on leaf litter to temperature change events. *International Review of Hydrobiology*, Vol. 94, No. 4, pp. 410-418.
- Ferreira, V. & Chauvet, E. (2011). Future increase in temperature more than decrease in litter quality can affect microbial litter decomposition in streams. *Oecologia*, Vol. 64, No. 1, pp. 279-291.
- Gessner, M.; Thomas, M.; Jean-Louis, A. & Chauvet, E. (1993). Stable successional patterns of aquatic hyphomycetes on leaves decaying in a summer cool stream. *Mycological Research*, Vol. 97, No. 2, pp. 163-172.
- Helena, B.; Pardo, R.; Vega, M.; Barrado, E.; Fernandez, J. & Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research*, Vol. 34, No. 3, Pp. 807-816.
- Hilsenhoff, W. (1988). Rapid field assessment of organic pollution with a family-level biotic index. *Journal of the North American Benthological Society*, Vol. 7, No. 1, pp. 65-68.

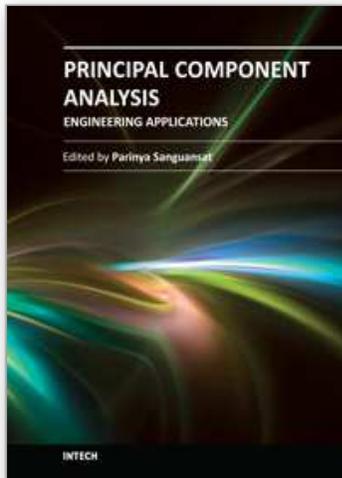
- Jianqin, M.; Jingjing, G. & Xiaojie, L. (2010). Water quality evaluation model based on principal component analysis and information entropy: Application in Jinshui River. *Journal of Resources and Ecology*, Vol. 1, No. 3, pp. 249-252.
- Johnson, R.; Wiederholm, T. & Rosenberg, D. (1993). Freshwater biomonitoring using individual organisms, populations, and species assemblages of benthic macro- - invertebrates, In: *Freshwater biomonitoring and benthic macroinvertebrates*, Rosenberg, D. & Resh, V., pp (40-158), Chapman & Hall, New York.
- Kebede, Y.; Kebede, T.; Assefa, F. & Amsalu, A. (2010). Environmental impact of coffee processing effluent on the ecological integrity of rivers found in Gomma woreda of Jimma zone, Ethiopia. *Ecology and Hydrobiology*, Vol. 10, No. 2, pp. 259-270.
- Kim, J. & Mueller, C. (1987). *Introduction to factor analysis: what it is and how to do it. Quantitative Applications in the Social Sciences Series*. Sage University Press, Newbury Park.
- Legesse, W. (2001). *Sensitivity of biotic indices to natural disturbances*. PhD thesis, Department of Zoology and Animal Ecology, University College Cork, Republic of Ireland.
- Leps, J. & Smilauer, P. (2003). *Multivariate analysis of ecological data using CANOCO*, Cambridge University Press, New York.
- Liu, C.; Lin, K. & Kuo, Y. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment*, Vol. 313, No. 1-3, pp. 77-89.
- Morales, M.; Marti, P.; Llopis, A.; Campos, L. & Sagrado, S. (1999). An environmental study by factor analysis of surface seawaters in the gulf of Valencia (Western Mediterranean). *Analytica Chimica Acta*, Vol. 394, No. 1, pp. 109-117.
- Murthy, K.; D'Sa, A. & Kapur, G. (July 2011). An effluent treatment-cum-electricity generation option at coffee estates: is it financially feasible?, In: *International Energy Initiative, Bangalore*. < <http://iei-asia.org/IEIBLR-Clean-Coffee.pdf> >
- Norris, R. & Georges, A. (1993). Analysis and interpretation of benthic macroinvertebrate survey. In: *Freshwater biomonitoring and benthic macroinvertebrates* Rosenberg, D. & Resh, V., pp. (234-286), Chapman & Hall, New York.
- Ougang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*, Vol. 39, No. 12; pp. 2621-2635.
- Ouyang, T.; Zhu Z. & Kuang, Y. (2006). Assessing impact of urbanization on river water quality in the Pearl River Delta Economic Zone, China. *Environmental Monitoring Assessment*, Vol. 120, No. 1-3, pp. 313-325.
- Ouyang, Y.; Higman, J.; Thompson, J.; O'Toole, T. & Campbell, D. (2002). Characterization and spatial distribution of heavy metals in sediment from cedar and ortega rivers Basin. *Journal of Contaminant Hydrology*, Vol. 54, No. 1-2, pp. 19-35.
- Parinet, B.; Lhote A. & Legube, B. (2004). Principal component analysis: an appropriate tool for water quality evaluation and management—application to a tropical lake system. *Ecological Modelling*, Vol. 178, No. 3-4, pp. 295-311.
- Pesce, S. & Wunderlin, D. (2000). Use of water quality indices to verify the impact of Córdoba City (Argentina) on Suquía River. *Water Research* Vol. 34, No. 11, pp. 2915-2926.

- Razmkhah, H.; Abrishamchi, A. & Torkian, A. (2010). 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*, Vol. 91, No. 4, pp. 852-860.
- Reghunath, R.; Murthy, T. & Raghavan, B. (2002). The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Research*, Vol. 36, No. 10, pp. 2437-2442.
- Richard, S.; Thorne, J. & Williams, W. (1997). The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. *Fresh water biology*, Vol. 37, No. 3, pp. 671-686.
- Rosenberg, D. & Resh, V. (Eds). (1993). *Freshwater Biomonitoring and Benthic Macroinvertebrates* Chapman and Hall, New York.
- Sheehan, P. (1984). Effects on community and ecosystem structure and dynamics, In: *Effects of pollutants at the ecosystem level*, Sheehan, P.; Miller, D.; Butler, G. & Bourdeau, pp. (51-100), John Wiley & Sons, Chichester, New York.
- Simeonov, V.; Stratis, J.; Samara, C.; Zachariadis, G.; Voutsas, D.; Anthemidis, A.; Sofoniou, M. & Kouimtzis, T. (2003). Assessment of the surface water quality in Northern Greece. *Water Research*, Vol. 37, No. 17, pp. 4119-4124.
- Singh, K.; Malik, A.; Mohan, D. & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-a case study. *Water Research*, Vol. 38, No. 18, pp. 3980-3992.
- Suberkropp, K. (1984). Effect of temperature on seasonal occurrence of aquatic hyphomycetes. *Transactions of the British Mycological Society*, Vol. 82, No. 1, pp. 53-62.
- ter Braak, C. & Smilauer, P. (2002). *CANOCO reference manual and CanocoDraw for windows user's guide: Software for canonical community ordination (Version 4.5)*. Microcomputer power Ithaca, New York.
- ter Braak, C. & Verdonschot, P. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences*, Vol. 57, No. 3, pp. 255-289.
- Thompson R. & P. Lake (2010). Reconciling theory and practise: the role of stream ecology. *River Research and Applications*, Vol. 26, No. 1, pp. 5-14.
- Vega M., Pardo R., Barrado E. & L. Deban (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, Vol. 32: No. 12, pp. 3581-3592.
- Von Enden, J. & Calvert, K. (June 2010). Review of coffee waste water characteristics and approaches to treatment. PPP Project, In: *Improvement of Coffee Quality and Sustainability of Coffee Production in Vietnam*, German Technical Cooperation Agency (GTZ), Available from <http://coffee.20m.com/CoffeeProcessing/CoffeeWasteWater.pdf> >
- Wenning, R. & Erickson, G. (1994). Interpretation and analysis of complex environmental data using chemometric methods. *Trends in Analytical Chemistry*, Vol. 13, No. 10, pp. 446-457.

Wunderlin, D.; Diaz, M.; Ame, M.; Pesce, S.; Hued, A. & Bistoni, M. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality, a case study: Suquia river basin (Cordoba Argentina). *Water Research*, Vol. 35, No. 12, pp. 2881-2894.

IntechOpen

IntechOpen



## **Principal Component Analysis - Engineering Applications**

Edited by Dr. Parinya Sanguansat

ISBN 978-953-51-0182-6

Hard cover, 230 pages

**Publisher** InTech

**Published online** 07, March, 2012

**Published in print edition** March, 2012

This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as energy, multi-sensor data fusion, materials science, gas chromatographic analysis, ecology, video and image processing, agriculture, color coating, climate and automatic target recognition.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yared Kassahun Kebede and Tesfu Kebedee (2012). Application of Principal Component Analysis in Surface Water Quality Monitoring, Principal Component Analysis - Engineering Applications, Dr. Parinya Sanguansat (Ed.), ISBN: 978-953-51-0182-6, InTech, Available from: <http://www.intechopen.com/books/principal-component-analysis-engineering-applications/application-of-principal-component-analysis-in-surface-water-quality-monitoring>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen