Pol. J. Environ. Stud. Vol. 21, No. 5 (2012), 1359-1367

Original Research

Multivariate Statistical Analysis and Environmental Modeling of Heavy Metals Pollution by Industries

Adamu Mustapha[1, 2], Ahmad Zaharin Aris[1]*

[1]Environmental Forensics Research Centre, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Department of Geography, Kano University of Science and Technology, Wudil, Nigeria

Received: 7 November 2011
Accepted: 5 April 2012

Abstract

This study presents the application of some selected multivariate statistical techniques, prediction method, and confirmatory analysis to identify spatial variation and pollution sources of the Jakara-Getsi river system in Kano, Nigeria. Two-hundred and forty water samples were collected from eight different sampling sites along the river system. Fifteen physico-chemical parameters were analyzed: pH, electrical conductivity, turbidity, hardness, total dissolved solids, dissolved solids, dissolved oxygen, biochemical oxygen demand, chemical oxygen demand, mercury, lead, chromium, cadmium, iron, and nickel. Correlation analysis showed that the mean concentration of heavy metals in the river water samples were significantly positive correlated values. Principal component analysis and factor analysis (PCA/FA) investigated the origin of the water quality parameters as due to various anthropogenic activities: five principal components were obtained with 81.84% total variance. Standard, forward, and backward stepwise discriminant analysis (DA) effectively discriminate thirteen (92.5%), nine (90.1%), and six (88.5%) parameters, respectively. Multiple linear regression yielded multiple correlation coefficient R value of 0.98 and R-square value of 0.97 with significant value 0.0001 (p <0.05) showing that water qualities in Jakara-Getsi can be predicted due to high concentration of heavy metals. Structural equation modeling (SEM) confirmed the finding of multivariate and multiple linear regression analysis. This study provides a new technique of confirming exploratory data analysis using SEM in water resources management.

Keywords: discriminate analysis, heavy metals, multiple linear regression analysis, river Jakara-Getsi, structural equation modeling

Introduction

Surface water pollution problems are not only confined to industrialized countries alone, although developing countries have a relatively small proportion of world industrial production. There are a number of third-world cities and city regions with high concentrations of industries and significant industrial output [1]. It is desirable for every country to develop industrially. Industries are known through their various processing and manufacturing activities. No doubt industries provide the necessities of life in contemporary communities. However, key productions from industries are accompanied with undesirable toxic effluent that more often than not is discharged into the environment [2]. Urbanization and industrial development in developing countries during the last decade have provoked

*e-mail: zaharin@env.upm.edu.my

some serious concern for the environment [3]. Human activities such as intensive agriculture, urbanization and industrialization contribute to river water deterioration [4]. Rivers, due to their role in carrying off domestic and industrial wastewater and runoff from agricultural land in their vast drainage basin, are among the most vulnerable water bodies to pollution [5]. Surface water quality in a river is a matter of serious concern.

A river system comprising both the main courses and the tributaries carryies the one-way of a significant load of matter from both natural and anthropogenic sources [5, 6]. Rivers are heterogeneous at a different spatial scale, which may be attributed to a number of factors, including anthropogenic input, biomass characteristic, soil erosion, weathering of crustal minerals, local environmental conditions, water discharge, water velocity, and degree of surface water chemistry [7, 8]. River flow is highly variable with respect to the climatic condition and drainage pattern [9]. The assessment of water quality is based upon physical, chemical, and biological properties of water.

Kano metropolitan, located between Latitude 12º30 N and Longitude 8º30 E, is the largest and most populous city in Nigeria. The population of the metropolitan area is estimated at over 5 million during the 2006 national population census [10]. It is the commercial nerve center of northern Nigeria. The climate of the region is tropical dry-and-wet type, classified by Koppen's as Aw climate. The seasonal migration of the inter-tropical discontinuity (ITD), gives rise to two seasons, one dry and the other wet. The wet season lasts from June to September, although May is sometimes humid. The dry season extends properly from mid-October of one calendar year to mid-May of the next. The annual mean rainfall in the region is between 800mm to 900 mm. Variation about the mean value is up to +30 or -30 per cent. More than 300mm of the rainfall is received in August alone, while the truly wet season lasts from June to September. In addition, the mean monthly temperature is 21ºC and 23ºC with diurnal range of 12-14ºC [11, 12].

Kano has grown into an industrial city with over 320 industrial establishments [13]. Urban growth is rapid, as is industrialization, both of which have made water pollution inevitable. The wastes generated are so enormous and the liquid waste is evacuated through a network drainage system both lined and unlined, natural and artificial. There are three earlier industrial areas in Kano metropolitan: the old city industrial area, where several large factories operate together with small firms, the township industrial estates where factories of colonial origin of medium size produce consumer goods, and the Bompai industrial estate where the majority of the largest and most of the advance industrial producers are located. Industrial estates have virtually encircled the city and its water resources. A sizable number of these industries are tanneries, textile, and food processing factories that used large quantities of water and produce large quantities of waste water. Kano metropolitan has more than 37 tanneries, 24 textiles, and over 43 food processing industries [13].

The study area is located adjacent to the industrial area. The types of industry presently in operation are: leather tanning and processing, including footwear, plastic and plastic products, including footwear; food, beverages, vegetable oil, mineral water, packed juices, and spirits; paper, paper products, and stationery; foam, rubber, and rubber products; glass products; electrical and communications industry; cardboard packing materials; metal processing and fabrication; motor vehicle and bicycle assembly; textiles, weaving, knitting, and spinning; steel and steel products; candles; paints; pharmaceuticals; batteries and chemical industries; sweets and confectionaries; burn bricks; floor, wall tiles; and ceramic wares; and cement, asbestos, and concrete products.

The primary aim of the study is to characterize the relationship and spatial variability of physio-chemical parameters in the Getsi stream (a tributary of the Jakara River) using multivariate statistical techniques to identify the natural and anthropogenic factors controlling the distribution of these parameters and to predict the pollution loading the Jakara-Getsi river system.

Materials and Methods

Water Sampling and Analytical Procedures

Thirty water samples were collected from eight randomly selected sites, making the total number of 240 samples along the Getsi at approximately the same sampling locations during the dry season in the year 2011, when there were no natural inputs into the river. Every collected sample was kept in a plastic bottle clearly marked and labeled with references to the sampling points. Similarly, to eliminate differences in the water quality that could arise due to variation in the timing, all the samples were collected during the same period of the day. Fifteen physio-chemical parameters were selected for analysis, which represent the quality of the river water. A detailed reconnaissance of the study area was conducted to ascertain the sampling sites. A survey was made by tracing the Getsi stream up to where it joins the Jakara River system. During the survey, all sites where wastewater is discharged into the river were noted and sampling stations designed as SB 01 to SB 08 were established.

SB 01 is located upstream of Getsi Stream at longitude12º02 N and latitude 8º33 E. This sampling site is located close to foam, rubber, and rubber products industryies; SB 02 is located close to the metal processing and fabricating industries; SB 03 is located around the leather and tanning industry; SB 04 is closely located to food, beverage, and vegetable oil; SB 05 is paper, paper products, and stationery; SB 06 soap, perfume, toiletries, and cosmetics; SB 07 glass and steel industries; and SB 08 is located around sweets and confectioneries and also the textile, weaving, knitting, and spinning industries.

The samples were filtered through Whatman filter paper number 1 and was preserved at about 4ºC. All samples were determined according to American Public Health Association Standard Method for Water and Wastewater Analysis [14]. pH, electrical conductivity, and dissolved oxygen of water samples were measured in the field immediately after collection of the samples using pH, conductivity, and dissolved oxygen meter, respectively. Biochemical oxygen demand (BOD_5) was determined by conventional methods, a sample of the solution was placed into a 500 mL BOD bottle and filled to the mark with previously prepared dilution water. Turbidity was measured using a turbidity meter [15], and total solids (TS) were determined using the evaporation method. The total metal concentration of chromium (Cr), cadmium (Cd), lead (Pb), iron (Fe), Mercury (Hg), and nickel (Ni) in the filtered and digested samples were determined in $mg \cdot L^{-1}$ using flame atomic absorption spectrophotometry (AAS).

Statistical Analysis

Principal Component Analysis/Factor Analysis

Principal component analysis and factor analysis PCA/FA provides information on the most statistically significance parameters and at the same time reduces the data with minimum loss of information. Factor analysis offers a powerful means of identifying the similarities among variables that represent water chemistry [16]. PCA also identified the likely factors that cause variation and also reveal relative significance of the combination of the parameters under study.

In our study, PCA/FA was applied to extract the most significant PCs and to reduce the contribution of variables with minimum significance. The PCs obtained were further subjected to varimax rotation to maximize differences between the variables and facilitate easy interpretation of the data [6].

Discriminant Analysis

Discriminant analysis (DA) is used to provide statistical classification of data. DA helps in grouping the most significance variables that share common properties [5, 7, 17]. DA is used to determine the variable that discriminate between two or more naturally occurring groups, and this technique constructs a discriminant function for each parameter/variable.

In our study DA was employed to bring out the most statistically significant parameters that result in high variation in the water quality in the Jakara-Getsi river system.

Multiple Linear Regression Models

Multiple linear regression analysis (MLR) is a statistical tool for determining the relationship between single depended variable and a set of independent variables to best represent a relationship in a population [18]. Regression analysis is a way of predicting an outcome variable from a predictor variable; the technique is both used as predictive and explanatory purposes within experimental and non-experimental designs [19]. The prediction equation is derived below:

$$Y = \beta o + \beta 1x1 + \beta 2x2 + \beta 3x3 + ... \beta mxm + eij \quad (1)$$

...where Y represents the dependent variable, $x1$ to xm represent the different set of independent variables, βo to βm represents the regression coefficient, and e represents the random error [18]. Multiple regression analysis fit a model to our data and used it to predict a value of the dependent variable from set of independent variables.

In this study the MLR model was used to predict the pollution load in the Jakara-Getsi river system. Water quality index was used as a dependent variable while heavy metals measured was used as an independent variable.

Structural Equation Modeling Using AMOS

Structural equation modeling (SEM) is use as a confirmatory technique rather than exploratory. It is used to confirm a model rather than to discover a new model. SEM is a statistical methodology that takes a confirmatory approach and it uses structural relation and modeled processes pictorially to enable a clearer conceptualization of a theory understudy. SEM uses latent variables, exogenous latent variables (synonymous with independent variables), and endogenous (synonymous with dependent) variables.

Confirmatory factor analysis (CFA) is appropriately used when one has some prior knowledge of the underlying latent variable's structure [20]. Based on the knowledge of the theory of empirical research, one needs to postulates a relationship between the observed measurements and the underlying factors prior and then test this hypothesized structure statistically. SEM provides overall tests of model fit, and individual parameters estimate tests simultaneously, and regression coefficients, means, and variance may be compared simultaneously. It also improves statistical estimation by incorporating measurement errors.

SEM models have two types: measurement models, which assess the measurement instrument which involves confirmatory factor analysis, and structural model, which test the relationship between variables involving regression and path analysis. A SEM specifies the indicators/items for each construct and enables an assessment of construct validity. A number of fit indices can be used to assess the model fit. These fit indices are direct measures of how well the model specified by a researcher reproduces the observed data [21].

In our study, SEM was applied using analysis of moment structure (AMOS) to test a confirmatory model theory on how heavy metals contribute significantly to overall water quality in the Jakara-Getsi river system. The study uses a single construct to test for the fitness of the model.

Result and Discussion

Descriptive Statistic

The statistical summary of the selected metal concentration in the water sample is presented in Table 1.

These data represent a total of 240 samples that correspond to 8 sampling sites. From the descriptive statistics, it is clear that heavy metals dominate the water samples with average concentrations of Ni 46.21 $mg \cdot L^{-1}$, Cr 24.92 $mg \cdot L^{-1}$,

Table 1. Statistical summary of selected metals concentration in mg·L^{-1} (n=240).

Parameters	Minimum	Maximum	Mean	S.E Mean	SD
Hg	0	28.1	12.20	0.547	8.488
Cr	0	48	24.92	1.057	16.381
Cd	0	28.6	15.62	0.614	9.525
Pb	0	46.2	24.21	0.983	15.232
Fe	1.2	36.9	23.38	0.800	12.401
Ni	2	190	46.21	3.417	52.945

SD – Standard deviation
S.E Mean – Standard error of mean

Table 2. Pearson correlation matrix of Heavy metals in mg·L^{-1}.

	Hg	Cr	Cd	Pb	Fe	Ni
Hg	1					
Cr	0.837**	1				
Cd	0.788**	0.951**	1			
Pb	0.810**	0.946**	0.984**	1		
Fe	0.591**	0.649**	0.703**	0.658**	1	
Ni	0.476**	0.522**	0.558**	0.656**	0.124	1

**Correlation is significant at the 0.01 level (2-tailed).

Pb 24.21 mg·L^{-1}, Fe 23.38 mg·L^{-1}, Cd 15.62 mg·L^{-1}, and Hg 12.20 mg·L^{-1}. The order of distribution is Ni > Cr > Pb> Fe > Cd > Hg. The high concentration of metals in the sampled water may be attributed to the release of effluent directly into the river by the industries in the study area, service stations, the natural enrichment process, wood, and low-grade coal combustion in homes [22, 23].

Pearson Product Moment Correlation Coefficient

The Pearson correlation coefficient for heavy metals in the surface of the Jakara-Getsi river system is presented in Table 2. The relationship between the heavy metals studied offer remarkable information on the sources and pathway of the heavy metals. Ni was significantly correlated with Pb (r = 0.656), Cd (r = 0.558), and Cr (r = 0.522). Fe was significantly correlated with Pb (r = 0.658), Cd (r = 0.703), Cr (r = 0.649), and Hg (r = 0.591). Cr in turn was strongly correlated with Ni (r = 0.522), Fe (r = 0.649), Pb (r = 0.946), Cd (r = 0.951), and Hg (r = 0.837).

The highly significant positive correlation between the heavy metals indicates that their compounds are used in various industries for various purposes [24]. This also suggest the possibility of common sources of origins that are anthropogenic [25]. This is obvious considering the large amount of industries located around the study area that release their effluent directly into the stream without any form of treatment and significantly contribute to the pollution of the Jakara-Getsi stream. Similar studies by Bichi and Anyata [26] and Mustapha [27] reveal that the concentration of heavy metals have exceeded the limit in the Jakara basin. The correlation matrix provides a justification for the use of principal component analysis to simplify the data.

Source Identification

Prior to the application of the principal component analysis (PCA), the Kaiser Meyer Olkin test (KMO) of the sampling adequacy and Bartlett's test of sphericity were checked. The KMO test is a helpful measurement of whether is suitable and adequate for factor analysis. As a rule of thumb, if the KMO test comes out at 0.5 or higher for a satisfactory factor analysis to proceed, we can then continue with the factor analysis suitable for our data. The Bartlett test of significance indicates it is worth continuing with the factor analysis as there are relationships to investigate. The KMO result was 0.687, and the Bartlett sphericity test was significant (0.0001, $p<0.05$), showing that PCA could be considered appropriate and useful to provide significant reduction in the data dimensionality [28, 29]. PCA was applied on the data set to identify the spatial sources of pollution in the Jakara-Getsi. According to eigen value criterion, only PCs with eigen value greater than one are considered essential and important. Five PCs were obtained with eigen value greater than one with total variance of 81.8%. These are considered responsible for the variation in the Jakara-Getsi (Fig. 1).

Table 3 summarizes The PCA results, including the loading, eigen value, and variance contribution rate. VF1 explained 36.9% of the total variance, and had strong loading on heavy metals Cd, Cr, Pb, Hg, Ni, and Fe (Table 3).

This result of PCA suggested that most of the variations in Jakara-Getsi River water quality are explained by heavy metals. Similar studies conducted by Hani and Pazira [30] have revealed that heavy metals in agricultural soils are due to anthropogenic sources related to the use of urban and industrial wastewater. Heavy metals such as Cd, Cr, Pb, Hg, Ni, and Fe originate from wet industries and are associated with traffic and other diffuse sources [31, 32]. The high loading of these metals may be related to wastewater and industrial effluent as shown by several previous studies [30-35].

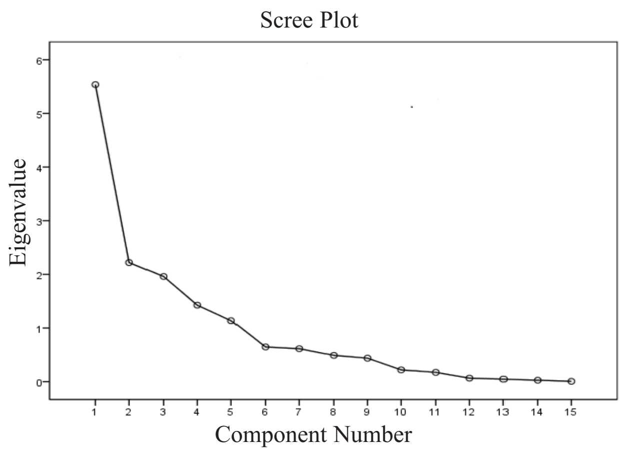


Fig. 1. Scree plot for principal component analysis.

Table 3. Rotated components matrix.

Parameters	Unit	VF1	VF2	VF3	VF4	VF5
Pb	$mg \cdot L^{-1}$	0.960	-0.111	0.105	0.060	-0.098
Cd	$mg \cdot L^{-1}$	0.953	-0.101	0.145	-0.049	-0.040
Cr	$mg \cdot L^{-1}$	0.940	-0.037	0.148	-0.049	-0.060
Hg	$mg \cdot L^{-1}$	0.854	-0.070	0.020	-0.086	-0.299
Fe	$mg \cdot L^{-1}$	0.706	0.172	0.327	-0.353	-0.078
EC	µS/cm	-0.659	-0.139	-0.181	0.094	-0.144
Ni	$mg \cdot L^{-1}$	0.620	-0.014	-0.331	0.602	-0.234
BOD5	$mg \cdot L^{-1}$	0.537	0.594	-0.114	-0.399	0.399
DS	$mg \cdot L^{-1}$	-0.156	0.835	0.557	0.074	-0.167
TS	$mg \cdot L^{-1}$	0.042	0.670	0.514	0.133	-0.036
pH		-0.418	0.260	0.633	0.217	-0.073
DO	$mg \cdot L^{-1}$	0.166	-0.583	-0.617	-0.017	0.240
COD	$mg \cdot L^{-1}$	0.332	0.527	0.565	0.187	0.090
Turbidity	NTU	0.342	-0.009	0.126	0.788	0.191
Hardness	$mg \cdot L^{-1}$	0.178	-0.257	0.367	0.162	0.809
Eigen value		5.53	2.21	1.96	1.42	1.13
% Variance		36.91	14.79	13.06	9.49	7.57
Cumulative variance		36.91	51.70	64.77	74.27	81.84

This cannot be a surprise that the study area is surrounded by various categories of wet industries, including leather and tanning, metal, glass, cosmetics industries, among others, that use these compound in their production.

Moreover, previous studies at this site show that the Jakara-Getsi is under pressure from industrial estate and traffic activities. For example, Dan'azumi and Bichi [10] observed that heavy metals in the Challawa and Jakara Rivers exceeded the maximum permissible limits established by the Federal Environmental Protection Agency of Nigeria (FEPA) and WHO. Lynch et al. [36] observed that pollutants such as Hg, Cr, Pb, and Cd were found to be in excess of the recommended permissible limits.

VF2 has strong positive loading on BOD_5, COD, DS, TS, and DO, which explained about 14.7% of total variance (Table 3). This factor represents domestic or municipal sewage [37]. Furthermore, COD measurements of chemical oxidizable organic matter indicates high organic matter content in the study area, corresponding to what was obtained in BOD_5, which shows that there is trans-boundary introduction of organic matter pollution through anthropogenic activities [38]. Reza and Singh [3] reported that the abnormally high BOD_5 and COD content indicated the presence of pollutants other than usual domestic sewage. This is not unconnected with the fact that effluent from the food processing industry at sampling site SB 04 are organic compound and some lost product as the substances undergo oxidation that they combine with some of the amount of dissolved oxygen in water. The amount of oxygen use is therefore a good indicator of the amount of organic waste present in the Jakara-Getsi.

VF3 has positive loading on DS, TS, pH, and COD and a negative loading on DO (Table 3), and explains 13% of the total variance. This represents the solid group; the high loading on DS and TS point out a common source of river water variation. Natural processes, such as precipitation rate, weathering processes and soil erosion occurring in the vicinity of the study area are responsible for high variation in water quality [7]. The factor represents erosion from upland areas during rainfall [6].

VF4 has positive loading on turbidity and Ni and explains 9.49% of the total variance of Jakara-Getsi water quality. Turbidity is a measure of the degree to which the water loses its transparency due to the presence of suspended particulates in the water. The more total suspended solids in the water, the murkier it seems and the higher the turbidity level. Turbidity is considered a good indicator of water quality [17]. The suspended particles absorb heat from sunlight, making turbid water become warmer and thus reducing the concentration of DO. This might be a reason for the negative high loading of DO in factor 3.

VF5 explains 7.57% of the total variance of water quality and has strong positive loading on hardness. Water hardness is determined by the concentrations of multivalent

Table 4. Model summary.

Model	R	R-Square	Adjusted R-Square	SE of the Estimate	R Square Change	Change Statistics				
						F Change	df1	df2	Sig. F Change	Durbin-Watson
1	0.98	0.97	0.84	2.331	0.97	7.382	23	5	0.018	2.651

df – degree of freedom, F – F statistics, SE of Estimate – Standard error of estimates, R – Correlation coefficient Value

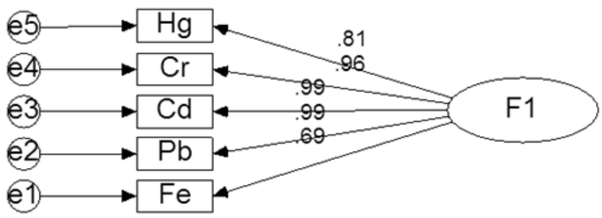
cations in the water. Multivalent cations are positively charged metal complexes with a charge greater than 1^+. Usually, the cations have a charge of 2^+ such as Ca^{2+}, Mg^{2+}, and metals cations. Hardness in water is a measure of the capacity of water to precipitate soap, due to the presence of calcium and magnesium ions in the water [12, 15]. This can be supported because there are a number of soap, perfume, toiletry, and cosmetics industries around sampling site SB 06, which could explain the hardness of the river water.

Spatial Variation in Water Quality

Spatial variation in water quality was investigated through discriminant analysis (DA). DA was applied via standard, forward stepwise, and backward stepwise mode with the aim of finding the most statistically significant parameters that result in variation in the Jakara-Getsi river system. The accuracy of classification matrix using standard, forward stepwise, and backward stepwise were 92.5% (13 parameters), 90.1% (9 parameters), and 88.5% (6 parameters), respectively.

The wilk's lambda value for standard mode (Rao approximation) gives lambda value 0.0001 and $p < 0.0001$. The null hypothesis states that the mean of the parameters under study are equal. The alternative hypotheses state that at least one of the mean of the parameter's understudy is different from another. From the result, since the computed p value is less than alpha ($\alpha = 0.05$), one should reject the null hypothesis; and accept the research or alternative hypothesis, the risk to reject the null hypothesis while it is true is lower than 5%.

Using the stepwise backward mode of discriminant analysis, variables are removed step by step beginning with less significant variables until no significant changes are obtained. Six variables were found to be the most significant, bringing variation in the river water; these variables are Cd, Cr, Pb, Hg, Fe, and Ni. This shows that heavy metals have high variation in terms of their spatial distribution in the Jakara-Getsi. This is supported by PCA/FA analysis, as these parameters were found to have higher factor loading and were also grouped in factor 1 of the rotated components and contribute about 36.9% of the total variance of water quality in the Jakara-Getsi.



Chisquere = 9.408
DF=5
P=0.094
Cfi=0.974
Gfi=/gfi
Agfi=/agfi
Nfi=0.948
Tli=0.949
RMSEA=0.196

Fig. 2. Structural model of the study.

Predicting Water Quality in the Jakara-Getsi River System

Multiple linear regression models were used on the data set to predict the pollution load in the Jakara-Getsi. Before interpreting the model, we checked for the basic assumptions of the regression model. Cross validation was checked using Stein formula [18]. The formula provides a better cross validation and indicates how well the regression model would predict an entire set of data. The formula is derived using the equation below:

$$R_C^2=1-[(n-1)(n-2)(n+1)](1-R) \quad (2)$$

...where R_C^2 is the cross validation, n represents the number of samples, and R is the R-square value.

The value of cross validation is similar to R^2 (0.97), indicating that cross validation using this model is very good and respectable. All other assumptions of the model that may bias or affect the value of the regression, such as outliers, influential cases, linearity, normality and multicollinearity, were checked and found to be conforming to the assumptions.

Water quality Index was used as a dependent variable while heavy metals provided by PCA/FA and discriminate by DA were used as independent variables. Enter method of regression procedure was used. The result of regression analysis (Table 4) shows that R-square value obtained was 0.97 with significant value ($p < 0.0001$), which indicates that 97% of the water quality in the Jakara-Getsi can be predicted and reported is as a result of the concentration of heavy metals.

From the coefficient table, the result shows that each individual heavy metal contributes to the variation in water quality in the Jakara Getsi with significant value 0.0001 ($p < 0.01$). The change statistics provided the value of significant F statistics, degree of freedom (df) and Durbin Watson statistics. These values tell us whether the change in R-square is significant. The significance of R-square can actually be tested using F-ratio from the change statistic values.

The Durbin Watson statistics inform us about whether the assumption of independent errors is tenable. As a conservative rule, a value of Durbin Watson less than 1 or greater than 3 should definitely raise an alarm. The closer to 2 the value is, the better the model [18]. This model got Durbin Watson 2.651, which is closer to the recommended value. This shows that the assumptions have been met.

Confirmation of Impact of Heavy Metals in Water Quality

Structural equation modeling (SEM) using analysis of moment structures (AMOS) Version 16 was used to confirm the grouping of heavy metals revealed by PCA/FA and DA as the most statistically significant variables that result in variations in the Jakara-Getsi. The model is tested using SEM goodness-of-fit, to determine if the pattern of variance and covariance in the data is consistent with the structural path. SEM may test two or more causal models to determine the best fit. There are many goodness of fit indices to test for model fit. Kline [39] reported and recommended at least four goodness of fit to test for the model fit: Chi-square value (Cmin), root means square error of approximation (RMSEA), comparative fit indices (CFI), and Turker Lewis indices (TLI), which are similar to normed fit indices (NFI).

In this study, a research model is based on the heavy metals to explain the water quality variation in the Jakara-Getsi. It consists of water quality perceived attributes: Cr, Cd, Pb, Fe, and Hg. The Null hypothesis is stated: Heavy metals have no effect on water quality variation while the alternative hypothesis is stated as: Heavy metals have a positive effect on water quality variation. The graphical relationship between exogenous and endogenous is presented in Fig. 2.

The regression weight of the exogenous variables shows that they are significant (Table 5), considering the p value. The result shows that RMSEA is 0.196 (Table 6) and based on these results the fit is good. By convention there is a good model fit if RMSEA is less than or equal to 0.05, the model whose RMSEA is 0.10 or more has poor fit [40]. Hu and Bentler [41] suggested RMSEA < 0. 2 as a cut-off for a good model fit. Fan et al. [42] reported that exact values for the cutoff are arbitrary.

The chi-square value is divided by the degree of freedom in an attempt to make it less dependent on sample size [43]. The result of chi-square (Cmin) is 9.408/5 =1.88 and it shows that the model fits (Table 7).

Kline [39] says a value of 3 or less is acceptable, while Ullman and Bentler [43] reported that a value of 2 or less reflect a good fit. Some researchers, for example Carmin and Mclever [44], allow value as high as 5 to consider a model of adequate fit, while others (Shumacker and lomax [40]) insist that relative chi-square value be 2 or less and also consider that less than 1 is poor model fit.

Bentler comparative fit indices (CFI) varies from 0-1, CFI close to 1 indicates a very good fit [42]. By convention, CFI should be equal to or greater than 0.90 to accept the model. The result shows that the value of CFI is 0.974 (Table 8).

Table 5. Regression weights of heavy metals.

		Estimate	S.E.	C.R.	*p*
Fe	F1	1			
Pb	F1	1.76	0.389	4.522	***
Cd	F1	1.105	0.243	4.538	***
Cr	F1	1.832	0.417	4.389	***
Hg	F1	0.801	0.214	3.748	***

***p < 0.01, SE – Standard Error

Table 6. Value of RMSEA.

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	0.196	0	0.387	0.113
Independent model	0.865	0.758	0.978	0

RMSEA – root mean square error of approximation, PCLOSE – P of close fit

Table 7. Value of C_{MIN}.

Model	NPAR	C_{MIN}	df	*p*	CMIN/DF
Default model	15	9.408	5	0.094	1.882
Saturated model	20	0	0		
Independent model	10	182.279	10	0	18.228

NPAR – Number of parameters, C_{min} – minimum discrepancy, df – degree of freedom

Table 8. Value of baseline comparisons.

Model	NFI	GFI	IFI	TLI	CFI
Default model	0.948	0.897	0.975	0.949	0.974

NFI – normed fit indices, GFI – goodness-of-fit indices, IFI – incremental fit indices, TLI – Tucker Lewis indices, CFI – comparative fit indices

CFI value indicates that 97.4% of the covariation in the data can be reproduced by the given model. Raykov [45], and Bollen and Curran [46] have argued that based on noncentrality of CFI, it is biased as a model fit measure.

Turker Lewis indices (TLI) are relatively independent of sample size. TLI ranges from 0-1.TLI indicates that chi-square/df ratio for the null hypothesis is less than the ratio for a given model. TLI close to 1 indicates a good fit, some researchers use a cut-off mark value as low as 0.80. However, Hu and Bentler [41] suggested that TLI > 0.80 as a cut-off for good model fit and this is widely accepted by some researchers such as Schumaker and Lomax [40] as the cut-off value. TLI value of less than 0.80 indicates a need to re-specify the model [47]. The result of TLI of this model is 0.949 (Table 8), which shows a good fit.

Conclusions

Descriptive statistics of all the parameter's understudy revealed that the main water quality pollution in the studied area can be attributed mainly to the anthropogenic activities through effluent discharged by the industries. PCA/FA was proven as a feasible technique in source's apportionment: it is a useful method that could assist decision makers in determining the extent of pollution via practical pollution indicators. PCA/FA generated five significant factors. VF 1 was correlated with heavy metals, explaining 36.9% of the total variance, VF 2 have strong loading of DS, TS, COD and negative loading on DO, explaining 14.79% of the total variance. VF 3 explained 13.06% of the total variance and has loading on TS, pH, DO, and COD. VF 4 have strong loading on turbidity and nickel and explains 9.49% of the total variance. VF 5 explains 7.57% of the total variance and has positive loading on hardness. Land use pattern of the basin, which was dominated by industrial activities, was concluded as the major water threat in the study area. DA gave the best results and supported PCA/FA; it shows that heavy metals parameters are responsible for a large variation in the water quality in the basin. MLR predicted parameters that result in water quality variation in conforming to DA. SEM revealed good fit indices, confirming that the variation in water quality in the Jakara-Getsi is by heavy metals. This study provides the reduction in dimensionality of the large data set and usefulness of multivariate statistical tools in revealing sources of water quality pollutants in the study area.

References

1. BICHI M. H. Environmental Pollution in Kano, Nigeria: The contribution of wastewater discharge from Bompai, Sharada and Challawa industrial estates. Sci. Assoc. Nigeria, **1**, 33, **1993**.
2. MILUKAITE A., SAKALYS J., KVIETKUS K., VOSYLIENE M.Z., KAZLAUSKIENE N., KARLAVICIENE N. Physico-chemical and ecotoxicological characterizations of urban storm water runoff. Pol. J. Environ. Stud. **19**, 1279, **2010**.
3. REZA R., SINGH G. Heavy metal contamination and its indexing approach for river water. Int. J. Environ. Sci. Technol. **7**, 785, **2010**.
4. CACHADA A., PEREIRA M.E., FERREIRA S.E., DUARTE A.C. Sources of potentially toxic elements and organic pollutants in an urban area subjected to an industrial impact. Environ. Monit. Assess. **184**, 15, **2011**.
5. SINGH K.P., MALIK A., MOHAN D., SINHA S., SINGH V.K. Chemometric data analysis of pollutants in wastewater—a case study. Anal. Chimi. Acta. **532**, 15, **2005**.
6. SHRESTHA S., KAZAMA F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environ. Modell. Softw. **22**, 464, **2007**.
7. SINGH K.P. 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. Res. **38**, 3980, **2004**.
8. QADIR A., MALIK R.N., HUSAIN S.Z. Spatio-temporal variations in water quality of Nullah Aik-tributary of the river Chenab, Pakistan. Environ. Monit. Assess. **140**, 43, **2007**.
9. SHARMA M., SINGAL S., PATRA S. Water Quality Profile of Yamuna River, India. J. Water. Energy. Environ. **3**, 19, **2008**.
10. DAN'AZUMI S., BICHI M. Industrial pollution and heavy metals profile of Challawa river in Kano, Nigeria. J. Appl. Sci. Environ. Sanit. **5**, 23, **2010**.
11. MUSTAPHA A., ARIS A.Z. Spatial aspects of surface water quality in the Jakara Basin using chemometic analysis. J. Environ.Sci. Heal. A. **47**, (10), 1455, **2012**.
12. MUSTAPHA A. NABEGU A.B. Surface water pollution source identification using principal component and factor analysis in Getsi River, Kano, Nigeria. Aust. J. Basic Appl. Sci. **5**, 1512, **2011**.
13. OLOFIN E.A, NABEGU A. B. DANBAZAU A. M. (Ed), Wudil within Kano region: A geographical synthesis, Adamu Joji Publishers, Kano, Nigeria, pp. 5, **2008**.
14. APHA. Standard methods for the examination of water and wastewater: Washington DC, pp. 135, **1998**.
15. CLESCERI L S., GREENBERG A. E., EATON A. ANDREW D. Standard methods for the examination of water and wastewater, Washington, DC. pp. 115, **1992**.
16. ARIS, A.Z., PRAVEENA S.M., ABDULLAH M.H., RADOJEVIC M. Statistical approaches and hydrochemical modeling of ground water system in small tropical island. J. Hydroinf. **14**, 206, **2011**.
17. JUAHIR H., ZAIN S.M., ARIS A.Z., YUSOFF M.K., MOKHTAR M. Spatial assessment of Langat River water quality using chemometrics. J. Environ. Monit. **12**, 287, **2009**.
18. FIELD A.P. Discovering statistics using SPSS, 3rd ed.; Sage publications, London. pp. 197-263, **2009**.
19. PALLANT J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15, Open University Press. Berkshire, UK. pp. 165, **2007**.
20. BYRNE B.M. Structural equation modeling with EQS: Basic concepts, applications, and programming, Lawrence Erlbaum. pp. 252, **2006**.
21. KENNY D A., MC-COACH D.B. Effect of the number of variables on measures of fit in structural equation modeling. Struct. Equa. Modeling. **10**, 333, **2003**.
22. PRAVEENA S.M., AHMED A., RADOJEVIC M., ABDULLAH M.H., ARIS A.Z. Heavy metals in mangrove surface sediment of Mengkabong Lagoon, Sabah: multivariate and Geo-Accumulation index approaches. Int. J. Environ. Res. **2**, 139, **2008**.
23. ISHAQ M., JAN F.A., IHSANULLAH, I., ASIM S. Multivariate statistical analysis of heavy metals pollution in industrial area and its comparison with relatively less area: A case study from the city of Peshawar and district Dir Lower. J. Hazard. Mater. **176**, 609, **2010**.
24. SEVCENKOS T. Pollution of Klaipe ports water. Pol. J. Environ. Stud. **20**, 445, **2011**.
25. ARMAH F.A., OBIRI S., YAWSON D.O., ONUMAH E. E., YENGOH G. T., AFRIFA E. K.A. ODOI J O. Anthropogenic sources and environmentally relevant concentrations of heavy metals in surface water of a mining district in Ghana: a multivariate statistical approach. J. Environ. Sci. Heal. A. **45**, 1804, **2010**.
26. BICHI M.H., ANYATA B. Industrial waste pollution in the Kano river Basin. Environ. Manage. Heal. **10**, 112, **1999**.

27. MUSTAPHA A. Identification of anthropogenic influences on water quality of Jakara River, northwestern, Nigeria. J. Appl. Sci. Environ. Sanit. **7**, 11, **2012**.
28. HUTCHESON G.D., SOFRONIOU N. The multivariate social scientist, Sage Publications London. pp. 126, **1999**.
29. PRAVEENA S.M., KWAN O.I., ARIS A.Z. Effects of data pre-treatment procedures on principal component analysis: a case study for mangrove surface sediment datasets. Environ. Monit. Assess., DOI 10.1007/s10661-011-2463-2, **2011**.
30. HANI A., PAZIRA E. Heavy metals assessment and identification of their sources in agricultural soils of Southern Tehran, Iran. Environ. Monit. Assess. **176**, 677, **2011**.
31. PHIRI O., MUMBA P., MOYO B., KADEWA W. Assessment of the impact of industrial effluents on water quality of receiving rivers in urban areas of Malawi. Inter. J. Environ. Sci. Tech. **2**, 237, **2005**.
32. NAUMCZYK J., RUSINIAK M. Physico-chemical and chemical purification of tannery wastewater. Pol. J. Environ. Stud. **14**, 789, **2005**.
33. AKTAR M., PARAMASIVAM M., GANGULY M., PURKAIT S., SENGUPTA D. Assessment and occurrence of various heavy metals in surface water of Ganga River around Kolkata: a study for toxicity and ecological impact. Environ. Monit. Assess. **160**, 207, **2010**.
34. LOHANI M.B., SINGH A., RUPAINWAR D., DHAR D. Seasonal variations of heavy metal contamination in river Gomti of Lucknow city region. Environ .Monit. Assess. **147**, 253, **2008**.
35. SZYNKOWSKA M.J., PAWLACZYK A., LEŠNIEWSKA E., PARYJCZAK T. Toxic metal distribution in rural and urban soil samples affected by industry and traffic. Pol. J. Environ. Stud. **18**, 1141, **2009**.
36. LYNCH K., BINNS T., OLOFIN E.A. Urban agriculture under threat: the land security question in Kano, Nigeria. The Land. **2**, 22, **2000**.
37. ZHOU F., LIU Y., GUO H. Application of Multivariate Statistical Methods to Water Quality Assessment of the Watercourses in Northwestern New Territories, Hong Kong. J. Environ. Monit. Assess. **132**, 1, **2007**.
38. MORENO C., MENDIGUCHIA C., GALINDO-RIANO M., GARCIA-VARGAS M. Using chemometric tools to assess anthropogenic effects in water A case study: Guadalquivir River (Spain). Anal. Chim. Acta. **515**, 143, **2004**.
39. KLINE, R. Principles and practice of structural equation modeling, 3rd ed.; Guilford, New York, pp. 209, **1998**.
40. SCHUMACKER R.E., LOMAX R.G. A beginner's guide to structural equation modeling, 3rd ed.; Lawrence Erlbaum. pp. 79-93, **2004**.
41. HU L., BENTLER P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct. Equ. Modell. **6**, 1, **1999**.
42. FAN X., THOMPSON B., WANG L. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Struct. Equa. Model Multi. J. **6**, 56, **1999**.
43. ULLMAN J.B., BENTLER P. M. Structural equation modeling: Handbook of psychology. John Wiley and sons Inc. pp. 224, **1996**.
44. CARMINES E. G. MCIVER J. Analyzing models with unobserved variables: Analysis of covariance structures in Social measurement: Current issues, Sage publication, London, pp. 65-115, **1981**.
45. RAYKOV T. Analysis of longitudinal studies with missing data using covariance structure modeling with full-information maximum likelihood. Struct. Equa. Modell. **12**, 493, **2005**.
46. BOLLEN K.A., CURRAN P.J. A structural equation approach: Hoboken, New Jersy, Wiley. pp. 185, **2006**.
47. SCHMUKLE S.C., HARDT J. A cautionary note on incremental fit indices reported by LISREL Methodology. Euro. J. Res. Meth Behavi. Soc. Sci. **1**, 81, **2005**.