Multivariate Statistical Application for the Assessment of Surface Water Quality in Lower Zab River, Kirkuk

ALYA A. MOHAMMED[1], DOAA F. ALMALIKI[1], RAFI M. QASIM[1], MOHAMED A. NAJEMALDEN[2]
[1] Basra Engineering Technical College
Southern Technical University
[2] Environment director in north region
Ministry of health & Environment
IRAQ

Abstract: - Spatiotemporal variation analysis of water quality in the river is very vital for water resources protection and sustainable consumption. In this study, Multivariate statistical methods, i.e., Principal Component Analysis (PCA), Canonical Correlation Analysis (CCA) and Multi Recreation Analysis (MRA) were used to interpret of data matrix in water quality over a period of five years (2015–2019) in the Lower Zab River. The results from PCA matrix showed high positive correlation between total hardness and sulphate (SO_4) at 95% confidence level (0.934) indicating that non-carbonate hardness is a measure in the river. By using Varimax rotation and by extracting the eigenvalues greater than 1 from the correlation matrix, number of significant principal components (PCs) were extracted. Five and four latent factors respectively in Lower Zab station 3 (LZ3) and Lower Zab station 2 (LZ2) were identified as responsible for the data structure, explaining 86.8% of total variance in winter season in LZ3 station and it is strong positive related to the SO_4, nitrate (NO_3), chloride (CL) and pH. All these variables are related to the weathering of minerals component of the river. 87.5% of total variance for winter season in LZ2 station, which is strongly positive, related to the NO_3, pH, sodium (NA) and magnesium (MG) related to the weathering of minerals component. The strong correlations between PC3 and five-day biochemical oxygen demand (BOD_5) as a result of CCA in LZ3 station during summer season, indicates that the high concentration of calcium (CA) and dissolved oxygen (DO) in water cause the low concentration of BOD_5. Among 72 multiple regression model run, only eight dependent variables had statistically significant relationships with independent variables. These results provide may useful information for water quality in the Lower Zab River, which can mainly affected by weathering of minerals component of the river, soil structure and run-off.

Key-Words: - Canonical correlation analysis, multi recreation analysis, principal component analysis, water quality.

1 Introduction

Water is the most nature precious gift for sustaining life and continuing all human activities such as drinking, municipal uses, power generation, and irrigation [1]. In the last decades, there was an increase request of fresh water due to the growth of population and industrialization [2]. Rivers as a fresh surface water play a key role in carrying off industrial and municipal wastewater as well as run-off from farmland, which cause high pollution to the water bodies [3]. This was confirm by many studies which show that rivers are highly susceptible to contamination, therefore it is very important to monitor and control the quality of

water in river basins [4]. In addition, rivers could degrade and damage because of natural processes as well as anthropogenic effects. This indicates that seasonal changes both in natural process like precipitation, hydrological condition, and temperature, and anthropogenic effects lead to several changes in the water quality present in several characters through different seasons [5].

The adverse impacts on water resources could negatively affect the environment and the economy of the country, specially the agricultural sector. It also causes several diseases to the human and animals. Thus, in order to determine highly effective pollution

control and management, it is important to study large number of water quality data. Moreover, it is important to have strong demand to predictions the potential impacts of climate change, which has ramifications on sustaining and managing water resources appropriately to meet water scarcity that has become pronounced [6].

In Iraq, people are mostly depends on surface water represented by the Tigris and Euphrates rivers. Moreover, about half of the available surface water emerges from the upstream rivers from Iran and Turkey through the norther part of Iraq specially Kurdistan region [7]. Lower-Zab River is one of the five major catchments in northern part of Iraq. The other four rivers are Greater-Zab, Ozem, Khabour, and Sirwan. Lower-Zab starts flowing from Iran and enters the Iraqi cities to meet the Tigris river in Iraq. This river covers 20030 Km^2 from its perimeter, which is approximately 1537 Km. The length of the river from its origin (Iran), extends 400 Km to cover an area of 22,250 Km^2. On the Lower-Zab river, the Dukan dam was constructed to control the flow of the river, irrigation, electrical generation, attract tourists and support fish production [8].

The Lower-Zab river water which refilled normally by snowmelt and rainfall, is generally used for irrigation, drinking and fish purposes.

The pollution of the Lower-Zab river is expected from the Alton Kopri wastewater treatment plant, which discharge wastewater directly to the river without any treatment process. Other source of water pollutant comes from the generated slaughter house and domestic wastewater which also discharge to the Lower-Zab river [9].

For effective water resources management and effective pollution control of the river, it is necessary to interpret a huge number of water quality data. The usual technique to evaluate the quality of water is to measure various parameters of contaminants in multiple monitoring positions at different seasonal times in a watershed with different topographical situations. Various statistical methods could be used for analysing complex data sets such as the statistical analysis methods that generated by long term monitoring programs of water quality. This provides good understanding for better conception of variation in water quality and optimizing monitoring network [10].

In this research, multivariate techniques were used to analyse a set of data obtained from the Lower-Zab River. The objective of this study is to apply the Principal Component Analysis (PCA), Canonical Correlation Analysis (CCA) and Multi Regression Analysis (MRA) techniques in the evaluation of correlations of water quality parameters, to identify water quality variables for spatial differences, and further to make sure the impact of the pollution sources on the water quality parameters.

2 Materials and Methods

2.1 Study Area

Lower Zab River (400 Km length) is one of Tigress river estuaries, beginning from the Iraqi-Iranian border at Qandeel mountain crossing three cities (Sulemanyia, Arbel, Kirkuk) meeting Tigress river at the Shag area. Lower ZAB with a total length of about 220 Km represents the main source of water usage in Kirkuk province. This river within Kirkuk city has a multifaceted system that supports a variety of uses, including irrigation systems in agricultural lands, drinking water and several different industries. The return flow from all these users is directly discharged into the river, thus it is greatly essential to assess the temporal and spatial variations of water quality in this region. Kirkuk Environment directorate has collected water quality data from three monitoring sites, Lower Zab station 2 (LZ2), Lower Zab station 3 (LZ3) and Lower Zab station 4 (LZ4) as shown in figure 1. However, there is no much data provide from LZ4 station because of bad security situation in this area.

2.2 Measurement Stations and Parameters

River water data were obtained at summer and winter periods from three different station on river LZ2 located at Altun Cobbry district, LZ3 at Al-Dubass district and LZ4 at Al- Hawija district in
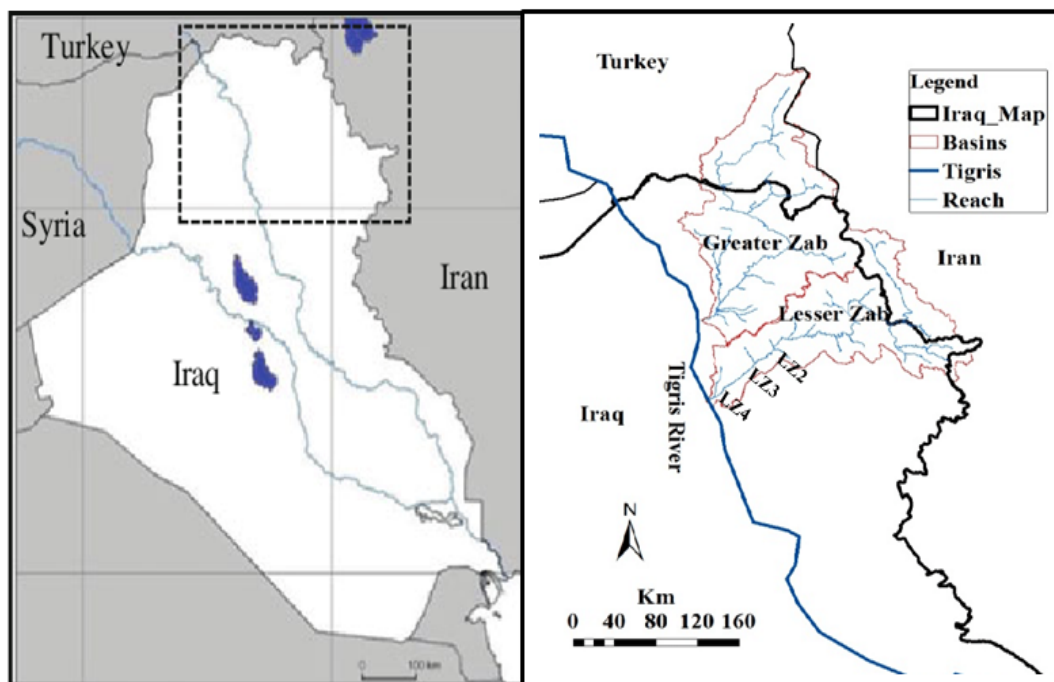
Figure 1 Studying area and its water quality monitoring sites.

Kirkuk city. The data for three water quality monitoring sites, consisting of 23 water quality parameters monitored monthly over five years (2015−2019), were obtained from Kirkuk Environment directorate. Only 17 of the 23 parameters, selected based on their sampling continuity at all the selected monitoring sites, were used in the present analysis. The selected parameters included electrical conductivity (EC, μsim/cm), pH, dissolved oxygen (DO, mg/L), temperature (TEMP, ˚C), total alkalinity (ALK, mg/L), total hardness (TH, mg/L) total dissolved solids (TDS, mg/L), turbidity (TUR, NTU) five-day biochemical oxygen demand (BOD_5, mg/L), nitrate nitrogen (NO_3, mg/L), chloride (CL, mg /L), sulphate (SO_4, mg/L), phosphate (PO_4, mg/L), sodium (NA, mg/L), potassium (K, mg/L), calcium (CA, mg/L), magnesium (MG, mg/L). The sampling, preservation, transportation, and analysis of the water samples were performed according to standard methods [11,12]. DO, pH, EC, TDS and temperature were measured with a multiparameter meter (HANNA HI9828) immediately in situ. BOD_5 was measured with use the WTW Oxitop. NO_3, CL, SO_4, PO_4, NA, K, CA and MG were analyzed using a Hach DR/3900 spectrophotometer according to its standard operating procedures.

The basic statistics of the five-year data set (their mean values and standard deviations of 950 observations) on river water quality are summarized in Table 1.

2.3 Data treatment and statistical methods

Most multivariate statistical methods require variables to conform to the normal distribution.

Thus, before applying of the multivariate statistical techniques, experimental data of each variable were normalized and checked by analysing kurtosis and skewness statistical test [13]. The original data demonstrated values of kurtosis ranging from 0.037 to 4.321 and skewness ranging from 0.068 to 7.594 and, indicating that skewness z-value more than 2.

This means that distributions were far from normal with 95% confidence. Since most of the values of kurtosis or skewness were greater than zero, the original data were transformed in the form $x' = \log10(x)$ [13]. After log-transformation, the kurtosis and skewness values were in the span of -1.96 to +1.96.

2.3.1 Principal Component Analysis (PCA)

PCA is an exploratory data analysis method that uses the correlation structure between multiple constituents to produce a small number of new variables that contain most of the information in the original dataset [14]. PCA is often applied to water sample datasets in order to determine the influencing factors on the water quality. The eigenvalues and eigenvectors were extracted in PCA

technique from the covariance matrix of original variables. The information is transformed into a number of principal components (PCs). These PCs are uncorrelated because of orthogonal transformation of the variables. Moreover, these PCs could achieved by multiplying the original correlated variables with the eigenvector (a list of coefficients (loadings or weightings)). Thus, the PCs are weighted linear combinations of the original variables using the covariance matrix of the data. PC delivers information on the most meaningful parameters. It is describe the total data set and at the same time affording data reduction with a minimum loss of original information [15]. The minimum number of PCs that should be selected to analyse the PCA output is dependent on the amount of explained variance that the researcher wants to account for. For instance, Joliffe [16] recommended that the number of PCs selected should be greater than 70% of the variance in the dataset.

In this study, PCA was used to obtain an overview of the relationships between the of Lower ZAB independent water quality. PCA was performed with Varimax rotation, while Pearson's correlation coefficient, R, measures the strength of a linear relationship between two quantitative variables. In addition, the data were reduced to extract a small number of latent factors (PCs). These factors were used to obtain an overview of the relationships between the physicochemical characteristics of river water. PCA was performed using software package IBM SPSS 24.

2.3.2 Canonical Correlation Analysis (CCA)

Canonical correlation analysis is a multivariate statistical model that facilitates the study of linear interrelationships between two sets of variables. First set of variables represent the independent variable and the second set consider dependent variables [3].

In this study, CCA was applied to quantify correlations between independent variables represented by PCs and the dependent water characteristics (BOD$_5$, ALK, TDS, TUR and EC). CCA was performed using software package IBM SPSS 24.

2.3.3 Multi Regression Analysis (MRA)

Regression modelling use one or more "predictor" variables to predict the value of a "response" variable. Thus, the value of one parameter can predicted if another value was known. For instance, the relation between PCs as predictor variable and BOD$_5$, TDS, EC and ALK as response variable could plot through regression model and then used

this plot to predict BOD$_5$ TDS, EC and ALK value when PCs were measured.

Both simple and multiple regression analysis were conducted on the data using of Microsoft Excel and correlation analysis was carried out on both PCs data and dependent variables (BOD$_5$, ALK, TDS, TUR and EC). Variables were plotted against time on a scatter graph and the coefficient of determination (R^2) was determined from the MRA. The R^2 value is a measure of the strength and direction of the relationship between the two variables, so it is an indicator as to whether a significant relationship exists.

If two or more dependant variables were found to have significant R^2 values in correlation with the same independent variables, they were incorporated into the MRA. A 95% confidence interval was applied in the analysis. Thus, the significance of the F value for each model was checked to guarantee that it was below 0.05 (for 95% confidence) to ensure that the model represented a statistically significant relationship. The F value as well as the R^2 (which was taken as an indication of the fit of the model) to the data is given as part of the output of the regression function in Microsoft Excel.

3 Result and Discussion

3.1 PCA result

PCA was actually achieved on the correlation matrix between thirteen parameters; Pearson's correlation coefficients are shown in tables 2 and 3 for both summer (March to September) and winter (October to February) seasons respectively. According to the results of correlation matrices for LZ2 station during winter season, the highest positive correlation pairs of TH and SO$_4$ at 95% confidence level was 0.934. This result indicates that non-carbonate hardness is a measure in the river and TH is mainly by anions of sulphate (such as calcium sulphate) which is often dissolved from clay [17]. The other significant positive correlation existed between pH and NO$_3$ for LZ2 station during winter season (0.918). However, there is no direct relation between pH and NO$_3$; pH variation may change the degree of solubility and kinetics of other chemical reactions. For instance, it can release or reduce oxygen when convert NO$_3$ to NO$_2$ or vice versa. There was also significant positive correlation appeared between NO$_3$ and NA located in LZ2

Table 1 Statistical descriptive of water quality parameters (as mean and stander deviation) of the Lower Zab River in LZ2 and LZ3 stations

| Parameters | LZ2 | | | | LZ3 | | | |
| | Winter | | Summer | | Winter | | Summer | |
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
PH	7.97650	0.141009	7.99467	0.104616	7.99467	0.104616	7.99467	0.104616
TEMP	18.51500	3.252695	20.86667	3.008717	20.86667	3.008717	20.86667	3.008717
DO	9.86800	0.595890	9.63333	0.554072	9.63333	0.554071	9.63333	0.554071
BOD_5	1.48850	0.438064	1.29600	0.318036	1.29600	0.318036	1.29600	0.318036
PO_4	0.06450	0.032167	0.09063	0.037937	0.09063	0.037937	0.09063	0.037937
NO_3	7.47200	4.181959	5.89343	2.640199	5.89343	2.640198	5.89343	2.640198
Ca	49.80000	5.245104	44.98000	4.187430	44.98000	4.187430	44.98000	4.187430
Mg	14.65100	3.656876	15.25167	3.206276	15.25167	3.206276	15.25167	3.206276
TH	177.95000	21.271329	175.43333	13.107886	175.43333	13.107885	175.43333	13.107885
K	5.83150	12.296043	2.80067	2.782333	2.80067	2.782333	2.80067	2.782333
Na	7.39050	1.518812	6.96767	1.261392	6.96767	1.261392	6.96767	1.261392
SO_4	29.66500	10.042135	25.50367	6.059955	25.50367	6.059955	25.50367	6.059955
CL	15.53500	12.283052	11.28333	1.718769	11.28333	1.718769	11.28333	1.718769
TDS	221.13000	32.977132	220.13333	26.038067	220.13333	26.038067	220.13333	26.038067
EC	326.53000	48.905329	325.50000	38.529210	325.50000	38.529209	325.50000	38.529209
ALK	151.40000	22.395188	147.86667	10.679531	147.86667	10.679531	147.86667	10.679530
TUR	105.12550	148.357526	99.6903	124.20378	99.69033	124.203781	99.69033	124.203781

station during summer season (0.878). Similar to the above, there were significant positive correlation existed between SO_4 and CL (0.812), NO_3 and CL (0.788) and SO_4 and NO_3 (0.814) for LZ3 station during winter indicating the combined effects of geological characteristics and human activities on the water quality [18].

Based on the complexity of relationships between parameters of water quality, it is difficult to draw a clear conclusion directly. Such difficult could be because of interaction among many factors including weather, climate, rainfall, industrial and sewage discharges. For instance there is no clear relation between TEMP and CL concentration in LZ2 during winter (0.891). So that these result followed by Varimax rotation and by extracting the

eigenvalues and eigenvectors from the correlation matrix, number of significant principal components (PCs) and the percentage of variance explained by each of them were calculated.

The results are presented in tables 4 and 5 respectively for both summer and winter season [15]. Eigenvalues greater than 1 were taken as criterion for extraction of the principal components. PCA of the 13 parameters constituted four PCs for LZ2 station explaining about 80.517% and 87.568% during summer and winter season respectively. These components classified the factor loadings as "strong" and "moderate" corresponding to absolute loading values of >0.75 and 0.75–0.60 respectively [15].

Table 2 Correlation matrix of the water quality parameters of the Lower Zab River during summer season.

	CL	K	TUR	DO	TEMP	pH	PO$_4$	NO$_3$	CA	MG	TH	NA	SO$_4$
LZ2- Summer													
CL	1.000												
K	-0.232	1.000											
TUR	-0.252	0.397	1.000										
DO	0.230	0.079	-0.210	1.000									
TEMP	-0.072	-0.414	-0.532	-0.339	1.000								
pH	-0.279	0.540	0.085	-0.275	-0.044	1.000							
PO$_4$	-0.016	-0.021	-0.377	0.303	0.077	0.432	1.000						
NO$_3$	-0.175	0.440	0.445	0.000	-0.562	0.582	0.351	1.000					
CA	0.377	-0.314	-0.234	-0.520	0.294	0.151	-0.029	-0.146	1.000				
MG	0.048	0.396	-0.011	0.494	-0.367	0.066	0.479	0.509	-0.396	1.000			
TH	0.351	0.116	-0.208	0.088	-0.079	0.154	0.467	0.369	0.387	0.689	1.000		
NA	-0.213	0.589	0.420	-0.015	-0.545	0.718	0.310	0.878	-0.059	0.298	0.203	1.000	
SO$_4$	0.191	0.116	0.400	0.235	-0.510	-0.054	0.084	0.634	-0.146	0.460	0.334	0.547	1.000
LZ3- Summer													
CL	1.000												
K	0.264	1.000											
TUR	-0.134	0.672	1.000										
DO	0.325	-0.065	0.093	1.000									
TEMP	-0.202	-0.614	-0.569	-0.251	1.000								
pH	-0.092	0.317	0.015	-0.118	-0.102	1.000							
PO$_4$	-0.001	-0.337	-0.452	0.325	0.174	0.335	1.000						
NO$_3$	-0.145	0.303	0.130	-0.194	-0.435	-0.101	0.042	1.000					
CA	0.229	0.394	0.078	-0.282	0.187	-0.080	-0.334	0.097	1.000				
MG	0.034	0.338	0.659	0.255	-0.316	-0.385	-0.332	0.185	-0.062	1.000			
TH	-0.155	-0.280	-0.388	-0.089	-0.083	-0.276	0.358	0.672	0.008	-0.303	1.000		
NA	-0.002	0.665	0.314	-0.156	-0.536	0.224	0.046	0.869	0.273	0.136	0.402	1.000	
SO$_4$	0.267	0.593	0.650	0.133	-0.501	-0.038	-0.163	0.250	0.089	0.469	-0.112	0.428	1.000

LZ2 in summer accounting for 33.38% of total variance had strong positive loadings on NO_3 and

NA and had moderate positive loadings on MG, K and SO_4. These results could interpreted as a weathering of minerals component of the river. About 18.61% of total variance has moderate positive and negative loadings respectively on TH and TUR. This component represents soil structure

and subsequent run-off. The third component accounting about 16.26% of total variance, which moderate positive loadings on pH and CA and moderate negative loading on DO which describe the natural of river flow. There are high concentrations of soluble oxygen in the water due to speedy flow of water and steepness, which could increase dissolve oxygen in the river water.

Table 3 Correlation matrix of the water quality parameters of the Lower Zab River during winter season.

	CL	K	TUR	DO	TEMP	pH	PO$_4$	NO$_3$	CA	MG	TH	NA	SO$_4$
LZ2- Winter													
CL	1.000												
K	0.174	1.000											
TUR	-0.159	0.281	1.000										
DO	0.026	-0.582	-0.147	1.000									
TEMP	0.891	-0.167	-0.106	0.163	1.000								
pH	0.373	0.378	0.362	0.268	0.192	1.000							
PO$_4$	0.453	-0.352	-0.432	0.286	0.476	0.207	1.000						
NO$_3$	0.332	0.484	0.318	-0.037	0.085	0.918	0.162	1.000					
CA	0.306	0.216	-0.005	0.028	0.045	0.743	0.373	0.881	1.000				
MG	0.564	0.311	-0.274	0.000	0.313	0.663	0.377	0.780	0.829	1.000			
TH	0.481	0.439	-0.025	-0.010	0.360	0.133	0.220	-0.045	-0.239	-0.068	1.000		
NA	0.346	0.705	0.427	-0.212	0.069	0.759	0.085	0.800	0.583	0.445	0.433	1.000	
SO$_4$	0.625	0.468	-0.024	0.046	0.448	0.397	0.398	0.260	0.082	0.231	0.934	0.630	1.000
LZ3- Winter													
CL	1.000												
K	-0.364	1.000											
TUR	0.126	0.293	1.000										
DO	-0.164	-0.066	-0.102	1.000									
TEMP	0.167	-0.205	0.086	0.264	1.000								
pH	0.723	-0.161	0.019	-0.032	0.157	1.000							
PO$_4$	-0.287	-0.414	-0.105	-0.060	0.229	-0.473	1.000						
NO$_3$	0.788	-0.361	0.174	-0.022	0.195	0.700	-0.317	1.000					
CA	0.220	0.208	-0.078	0.102	0.239	0.051	-0.491	0.374	1.000				
MG	0.467	-0.499	0.208	-0.011	0.218	0.283	0.177	0.444	-0.259	1.000			
TH	0.574	-0.391	-0.203	0.034	0.387	0.427	-0.064	0.820	0.591	0.305	1.000		
NA	0.496	0.386	0.141	-0.606	-0.150	0.589	-0.468	0.386	0.265	-0.018	0.244	1.000	
SO$_4$	0.812	-0.136	0.312	-0.239	0.066	0.726	-0.381	0.814	0.304	0.489	0.591	0.672	1.000

In LZ2 station in winter, the first component explains 38.96% of total variance and is strong positive related to the variables NO3, pH, NA and MG and moderate positive to the CA, CL and SO4.

All these variables were related to the weathering of minerals component of the river. Second component account about 21.16% of total variance and it was moderate positive loading on TEMP, PO4, which can interpreted as agricultural activities,

and this resulted in pollution with nutrients coming from fertilizers. The third one account about 17.96% and had strong positive loadings on TH. The last component in this group represent about 9.49% and had moderate positive loading on DO.

While five PCs of the total variance at the LZ3 station were extracted to explain about 85.62% and 86.81% respectively in summer and winter season.

Similar to the results of LZ2, PC1 in this region was also related to common influences of weather on minerals component of the river. Other components were related to the weathering of minerals component, soil structure and subsequent run-off as shown in tables 4 and 5.

Table 4 Principal component analysis of the water quality parameters of the Lower Zab River during summer season.

	PC1	PC2	PC3	PC4	PC5
LZ2- Summer					
NO_3	0.915				
NA	0.878				
TEMP	-0.682		0.425		
MG	0.674	0.546			
K	0.658				
SO_4	0.655			0.470	
TH	0.401	0.707			
TUR	0.459	-0.654			
PO_4		0.580		-0.447	
CL		0.580		0.575	
pH	0.523		0.716		
CA			0.692	0.566	
DO		0.507	-0.635		
Eigenvalue	4.340	2.419	2.114	1.594	
Variability (%)	33.384	18.611	16.260	12.263	
Cumulative (%)	33.384	51.994	68.254	80.517	
LZ3- Summer					
TUR	0.921				
K	0.786				
SO_4	0.753				
TEMP	-0.737				
MG	0.716			-0.502	
NO_3		0.922			
TH		0.876			
NA	0.465	0.780			
CA			0.843		
DO			-0.680		
PO_4	-0.402		-0.607		
pH				0.971	
CL					0.932
Eigenvalue	3.969	2.451	1.768	1.597	1.346
Variability (%)	30.527	18.853	13.603	12.281	10.353
Cumulative (%)	30.527	49.380	62.983	75.264	85.616

3.2 CCA result

When performing the CCA, PCs each from the predictor variables (i.e the independent water quality parameters represented by PCs) and the predicted variable (i.e. dependent water quality parameters represented by TDS, EC, ALK and

BOD_5) were used. The results showed strong summer and PC4 and BOD_5 (0.893) in LZ2

summer, BOD_5 recognized to be an indicator of correlation between PC3, and BOD_5 (0.899) in LZ3 pollution level of water [16]. These results indicate that the higher concentration of CA and CL in water the lower concentration of BOD_5 [10, 19]. In addition, BOD is a measure of the oxygen used by microorganisms to decompose the organic matter in

Table 5 Principal Component Analysis of the water quality parameters of the Lower Zab River during summer season.

	PC1	PC2	PC3	PC4	PC5
LZ2- Winter					
NO_3	0.862	-0.403			
pH	0.856				
NA	0.849				
MG	0.775		-0.460		
CA	0.733		-0.577		
CL	0.689	0.561			
SO_4	0.674		0.566		
TEMP	0.421	0.709			
PO_4		0.662			
TUR		-0.593		0.565	
TH	0.405	0.408	0.771		
K	0.539	-0.503	0.554		
DO		0.419	-0.411	0.698	
Eigenvalue	5.065	2.751	2.334	1.234	
Variability (%)	38.962	21.159	17.957	9.490	
Cumulative (%)	38.962	60.121	78.079	87.568	
LZ3- Winter					
SO_4	0.933				
NO_3	0.923				
CL	0.891				
pH	0.807				
TH	0.749		0.406		
K		-0.833			
PO_4	-0.418	0.652			0.433
NA	0.622	-0.647			
MG	0.459	0.550	-0.443		
TEMP		0.476			0.409
CA	0.401		0.766		
DO			0.573	0.423	-0.518
TUR				0.808	
Eigenvalue	4.820	2.477	1.760	1.227	1.003
Variability (%)	37.073	19.053	13.536	9.435	7.714
Cumulative (%)	37.073	56.126	69.662	79.096	86.810

water. When BOD levels are high, DO levels decrease because the bacteria are consuming the oxygen that is available in the water [20]. other strong correlations were found in LZ3 in summer between PC1 and ALK (0.88), PC4 and ALK (0.858). While, moderate correlation was found between PC5 and ALK (0.611) in LZ2 in winter.

Tseng et al. [21] showed that increase alkalinity can reduce turbidity. It seems that as ALK of water increase, MG, K and SO_4 start to precipitate [22].

Moreover, ALK, TEMP and DO have strongly effect on bacterial activity [23].
EC has a moderate loading on PC1 in LZ2 in summer (0.729), PC2 in LZ2 in winter (0.707) and LZ3 in summer (0.746) and finally on PC5 in LZ3 in winter (0.680). These results indicated that the resource of these loading is probably by anthropogenic sources [24].
The strong loading between TDS and PC3 in LZ3 winter (0.859) confirm that TDS depend mainly on CA [25]. While, there was moderate correlations between TDS and PC2 in LZ2 in summer and winter (0.653 and 0.665 respectively) and PC5 in LZ3 in summer (0.680). These results could be produced by anthropogenic factor such as industrial wastewater discharge.
All CCA results were represented in table 6.

3.3 MRA result

72 multiple regression models were developed from the MRA, producing the relation between independent variables (TDS, EC, ALK, BOD_5) and dependent variables (represented by PC1, PC2, PC3,

PC4 and PC5). There were multiple cases for various parameters where a multiple regression model was created that produced acceptable R^2 value, but F values of one or more of the variables showed them to be statistically insignificant (exceeded 0.05) to the model. Therefore, they were not deemed statistically significant for 95% model confidence. The result from MRA shown only eight predictor variables were had statistically significant relationships with response variables.

TDS and EC were found to be predictor variables for statistically significant relationships with TH and TUR in LZ2 in summer (R^2=0.62, F=0.0325), (R^2=0.63, F=0.0283) respectively. That is mainly because higher concentration of TDS in water, higher TUR and TH.

Table 6 Correlations between independent variables and dependent variables/

|	TDS	EC	ALK	BOD_5
LZ2- Summer				
PC1	-0.502	-0.729	0.334	0.322
PC2	0.653	-0.667	-0.315	-0.164
PC3	0.548	0.100	0.785	0.266
PC4	-0.137	-0.110	0.413	-0.893
LZ2- Winter				
PC1	0.504	-0.480	-0.324	-0.639
PC2	0.665	0.707	0.208	-0.112
PC3	-0.520	0.492	-0.338	-0.609
PC4	-0.176	-0.159	0.858	-0.454
LZ3- Summer				
PC1	-0.315	0.047	0.880	0.326
PC2	-0.498	0.746	-0.333	0.278
PC3	-0.275	0.253	0.224	-0.899
PC4	0.335	0.161	0.004	-0.080
PC5	0.680	0.591	0.252	0.030
LZ3- Winter				
PC1	0.349	-0.079	0.220	0.531
PC2	-0.181	0.610	-0.309	0.665
PC3	0.859	0.001	-0.451	-0.058
PC4	-0.149	0.396	-0.526	-0.441
PC5	0.288	0.680	0.611	-0.277
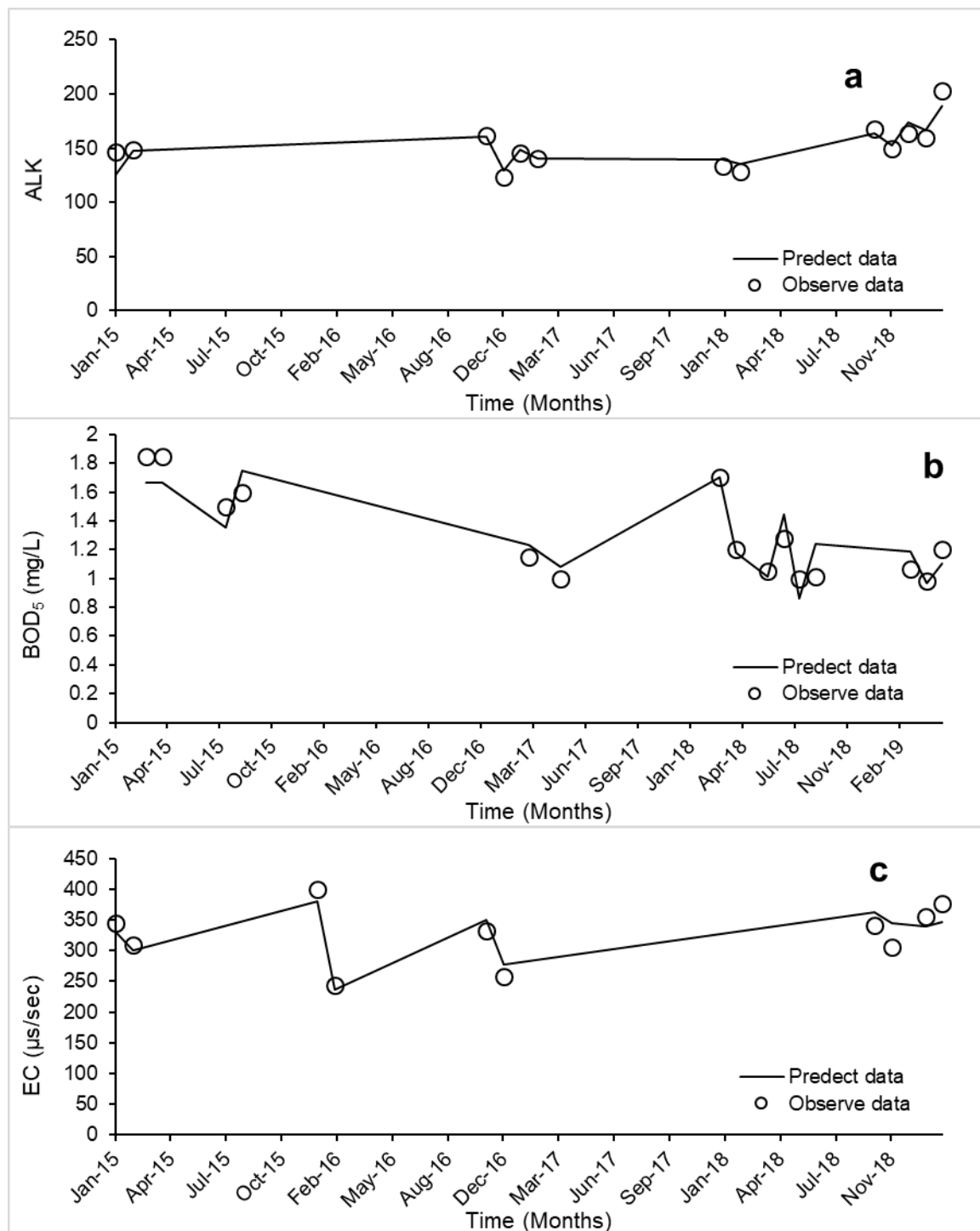
Figure 2 Selected examples of model-fit diagrams: (a) LZ3 winter, (b) LZ2 summer (c) LZ2 winter.

While ALK and BOD_5 were found to be predictor variables with NO_3, NA, MG, K and SO_4 for the same station (R^2=0.8, F=0.0158), (R^2=0.82, F=0.0110) respectively. This result confirm the result from CCA.

In winter, LZ2 with (R^2=0.79, F=0.0164), (R^2=0.77, F=0.0216) and (R^2=0.69, F=0.0501) respectively for TDS, EC and BOD_5, that found to be predictor variables for statistically significant relationships with TEMP and PO_4. This result may be because as

BOD_5 increases, PO_4 could be removed easily by biological nature of phosphate removal [26].

Finally, LZ3 in winter, the model highlighted important relationships between ALK and SO_4,

NO_3, CL, pH and TH (R^2=0.81, F=0. 0.0187). A study done by Maneepong and Wada [27] showed a strong correlation between ALK and SO_4, NO_3, CL, pH.

4 Conclusion

This study presents different multivariate statistical methods of large and complex databases to obtain better information concerning water quality of watercourses in Kirkuk city in Iraq.

PCA is widely used to reduce data and to extract a small number of latent factors. Therefore, PCA allowed a reduction in the dimensionality of the large data set and indicated a few significant parameters responsible for large variations in water quality that could reduce the number of sampling parameters.

Results analysis revealed that Lower Zab River carried high loads of SO_4, NO_3, MG, CL, CA and NA and medium load to PO_4. All contributing to the weathering of minerals component of the river, soil structure, run-off and agricultural activities.

Based on the information obtained from CCA, it is possible to reduce concentration of BOD5 when increase concentration of CA, CL and DO. In addition, increase alkalinity can reduce turbidity. On the other hand, TDS depend mainly on CA, CL, PO_4

and TH, which could cause by anthropogenic factors such as industrial wastewater discharge. The result from MRA confirm the result obtain from CCA. Although seventy-two multiple regression models were run, only eight predictor variables had statistically significant relationships with response variables. Hence, this study demonstrates that multivariate statistical methods are an exceptional exploratory tool to discover the structure of relationships between variables. Therefore, this tool is powerful for water quality management. This water quality management should be supported to ensure water security and sustainable development, including pollution emission reduction, sewage treatment facility, industrial restructuring, scientific sluice regulation and riparian management and so on.

References:

[1] Al-Murib M., "Application of CE-QUAL-W2 on Tigris River in Iraq." *Civil and Environmental Engineering Master's Project Reports.* pdxscholar.library.pdx.edu/cengin_gradprojects/9, vol. 9, 2014.

[2] Yisa J., & Oladejo J. T., "Analytical Studies on Water Quality Index of River Landzu." *American Journal of Applied Sciences*, vol. 7, no. 4, 2010 ISSN 1546-9239, pp. 453-458, doi:10.3844/ajassp.2010.453.458.

[3] Fan X., Cui B., Zhao H., Zhang Z., Zhang H., "Assessment of River Water Quality in Pearl River Delta Using Multivariate Statistical Techniques.". *Procedia Environmental Sciences*, vol. 2, no. 5, 2010, pp. 1220-1234, doi:10.1016/j.proenv.2010.10.133.

[4] Zhao Z., Song X., Wang W., Xiao Y., Gong Z., Wang Y., Zhao Y., Chen Y., Mei M., "Influences of Iron and Calcium Carbonate on Wastewater Treatment Performances of Algae Based Reactors." *Bioresource Technology. Elsevier*, vol. 216, 2016, pp. 1-11, doi:10.1016/j.biortech.2016.05.043.

[5] Yuan Z., Fen G., Wei M., Xi-Qin W., "Water Quality Assessment and Source Identification of Daliao River Basin Using Multivariate Statistical Methods." *Environmental Monitoring and Assessment*, vol. 152, no. 1-4, 2009, pp. 105-121, doi:10.1007/s10661-008-0300-z.

[6] Abbasa N., Wasimia S. A., Al-ansari N., "Assessment of Climate Change Impact on Water Resources of Lesser Zab, Kurdistan, Iraq Using SWAT Model.". *Engineering*, vol. 08, no. 10, 2016, pp. 697-715, doi:10.4236/eng.2016.810064.

[7] Yousuf M. A., Rapantova N., Younis J. H., "Sustainable Water Management in Iraq (Kurdistan) as a Challenge for Governmental Responsibility.". *Water*, vol. 10, no. 11, 2018, pp. 1-19, doi:10.3390/w10111651.

[8] Aziz S. Q., Slewa E. O., Abdullah W. A., "Evaluation of Water Quality for Lesser — Zab River for Various Applications." *Kirkuk University Journal Scientific Studies (KUJSS)*, vol. 12, no. 4, 2017.

[9] Slewa E. O., Abdullah W. A., Aziz S. Q., "Effect of Alton Kopri Wastewater on Lesser-Zab Raw Water." Tikrit Journal of Engineering Sciences, vol. 25, no. 1, 2018, pp. 17-25.

[10] Zhou F., Liu Y., Guo H., "Application of Multivariate Statistical Methods to Water Quality Assessment of the Watercourses in Northwestern New Territories, Hong Kong." *Environmental*

Alya A. Mohammed, Doaa F. Almaliki,
Rafi M. Qasim, Mohamed A. Najemalden

Monitoring and Assessment (May 2014), vol. 132, no. 1-3, 2007, 1-13, doi:10.1007/s10661-006-9497-x.

[11] APHS. "Standard methods for the examination of water and wastewater: Washington." *DC, American Public Health Association, American Water Works Association, and Water Environment Federation*, 1998, PP. 3—37.

[12] ASTM. "Annual Book of ASTM Standards." *ASTM West Conshohocken, PA, USA*, 2001.

[13] Field A. "Discover Statistics Using SPSS.". *Karak, Jordan: Mu'tah Medical E-Library*. 3rd ed. Mu'tah Univ, 2011, pp. 1-854. (2011).

[14] Olsen R. L., Chappell R. W., Loftis, J. C., "Water Quality Sample Collection, Data Treatment and Results Presentation for Principal Components Analysis — Literature Review and Illinois River Watershed Case Study." *Water Research. Elsevier,* vol. 46, no. 9, 2012, pp. 3110-3122, doi:10.1016/j.watres.2012.03.028.

[15] Iscen C. F., Emiroglu Ö., Ilhan S., Arslan N., Yilmaz V., Ahiska S., "Application of Multivariate Statistical Techniques in the Assessment of Surface Water Quality in Uluabat Lake, Turky." *Environmental Monitoring and Assessment*, 2014, doi:10.1007/s10661-007-9989-3.

[16] Jolliffe I., 2002. "'Principal component analysis: Wiley Online Library', Kaiser, HF (1960)." *The Application of Electronic Computers to Factor Analysis*.

[17] Smith P. G., "Dictionary of Water and Waste Management"., *Elsevier*, 2005.

[18] Xu H.S., Xu Z.X., Wu W., Tanga F.F., "Assessment and Spatiotemporal Variation Analysis of Water Quality in the Zhangweinan River Basin, China." *Procedia Environmental Sciences*, vol. 13, 2012, pp. 1641-1652, doi:10.1016/j.proenv.2012.01.157.

[19] Mulyani H., & Pamungkas G. S., "The Influence of Calcium Hypochlorite Dosage Adjustment on Tapioca Wastewater Pre-Chlorination Toward Efficiency of Activated Sludge Treatment.". *Molekul,* vol. 11, no. 2, 2016, pp. 239-247, doi:10.20884/1.jm.2016.11.2.213.

[20] Frick W. E., Ge Z., Zepp R. G., "Nowcasting and Forecasting Concentrations of Biological Contaminants at Beaches: A Feasibility and Case Study.". *Environmental Science and Technology*, vol. 42, no. 13, 2008, pp. 4818-4824, doi:10.1021/es703185p.

[21] Tseng T., Segal B. D., Edwards M., "Increasing Alkalinity to Reduce Turbidity." *Journal –*

American Water Works Association. Wiley Online Library, vol. 92, no. 6, 2000, pp. 44-54, doi:10.1002/j.1551-8833.2000.tb08958.x.

[22] Pujiastuti C., Sumada K., Ngatilah Y., Hadi P., "Removal of Mg^{2+}, K^+, SO_4^{-2} Ions From Seawater by Precipitation Method." *MATEC Web of Conferences conference*/20,165,801,022, 2016, vol. 58, no. 01022, pp. 10-13, doi:10.1051/matec.

[23] Lijklema L. "Factors Affecting pH Change in Alkaline Waste Water Treatment — I." *Water Research. Pergamon*, vol. 3, no. 12, 1969, pp. 913-930, doi:10.1016/0043-1354(69)90075-X.

[24] Ebrahimzadeh S., Bagherabadi A. S., Boustani F., "Pollution Assessment of Groundwater by Nitrate & Phosphate in Corbal Plain." *Inst for Advance Studies in Basic Sciences (IASBS)*, 2018.

[25] Rapant S., Cveˇ V., "Impact of Calcium and Magnesium in Groundwater and Drinking Water on the Health of Inhabitants of the Slovak Republic." *International Journal of Environmental Research and Public Health*, vol. 14, no. 278, 2017, doi:10.3390/ijerph14030278.

[26] Ruzhitskaya O., Gogina E., Shmalko V., "Effective Removal of Nutrients From Wastewater." *MATEC Web of Conferences*, vol. 178, 2018, p. 09015, doi:10.1051/matecconf/201817809015.

[27] Maneepong S., & Wada S., "Soil Science and Plant Nutrition Complexes at Synthetic Goethite/Aqueous Electrolyte Interfaces Complexes at Synthetic Goethite/Aqueous Electrolyte Interfaces." *Soil Scientific and Plant Nutrition*, vol. 37, no. 1, 2012, pp. 141-150, doi:10.1080/00380768.1991.10415019.