

Statistical Analysis of Water Quality Monitoring Network Case Study: Gharbia Drainage Catchments Area

H.A.A. El Gammal

National Water Research Center, P. O. Box 74, Cairo, Egypt.

Address For Correspondence:

H.A.A. El Gammal National Water Research Center, P. O. Box 74, Cairo, Egypt.
E-mail: elgammalhussein@gmail.com

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Received 12 February 2016; Accepted 28 April 2016; Available online 15 May 2016

ABSTRACT

Gharbia Drainage catchments area (GDA) is one of the most important and biggest drainage area in the Nile Delta and Valley. The sub-catchment areas of GDA pumps the drainage water to Gharbia main drain. The drain flows from North to South by gravity with an average flow of 1.9 billion m³/year; most of its water is reused for irrigation and fish farms. The agricultural drainage, domestic, and industrial water represents 75, 23, 2 percent of the main drain discharge. The water quality of GDA is deteriorated due to illegal discharges of untreated or partially treated domestic and industrial wastewater; in addition to improper farming practices. Data on water quality variables for GDA have been collected monthly for 15 years through the National Water Quality Monitoring Network supported by the National Water Research Center (NWRC) in Egypt. Water quality monitoring programs need to be properly designed and integrated in decision making. Water quality monitoring is costly; especially in far areas from laboratory facilities like GDA. Statistical methods like Factor Analysis and ANOVA are used to redesign of the water quality monitoring network of GDA in terms of measured water quality variables and sampling frequency. Results showed that five water quality variables; TDS, BOD, TN, TSS, FC can be measured out of the fifteen parameters. These parameters represent the different types of pollution sources; agricultural, domestic, and industrial pollution. The research successfully concluded that the water quality variables can be measured from four to eight times per year instead of 12 times. Reduction in the number of the measured water quality parameters and sampling frequency will secure sufficient fund that can be used to improve the resources in the water quality-monitoring network. Finally, it is recommended to intensify the measurement of the water quality variables in the concluded sampling months. In addition, reassessment of the water quality monitoring network should be carried out and redesign the water quality monitoring network if it is necessary.

KEYWORDS: Gharbia, Water Quality, Monitoring, Network, Drainage, Redesign, Factor Analysis, Nile Delta.

INTRODUCTION

As Egypt lies in arid and semi-arid region with uncountable rainfall across the country, limitation of water resources with increasing rate of population impose major challenges to water resources management in Egypt. Over 97 % of these resources come from upstream of the Nile; the remaining water comes from groundwater and reclaimed wastewater [1].

In addition, human activities impose threats to the water resources in terms of quantity and quality especially in the drainage catchments area. Water quality monitoring is one of the most important components in environmental management of aquatic ecosystems. Monitoring of water quality provides water managers with the necessary information for sustainable water resources management and provides insight into complex dynamic environmental processes. Reliable, consistent and appropriate information is necessary to understand water resources; therefore, water quality monitoring programs need to be properly designed and integrated in decision making.

The rest of drain braches is distributed along the Eastern and Western parts of the main drain namely are Hafir Shehab El-Din, drain No.3, drain No.4, drain No.5, Drain No.6 and Samatay drain (Figure 2).

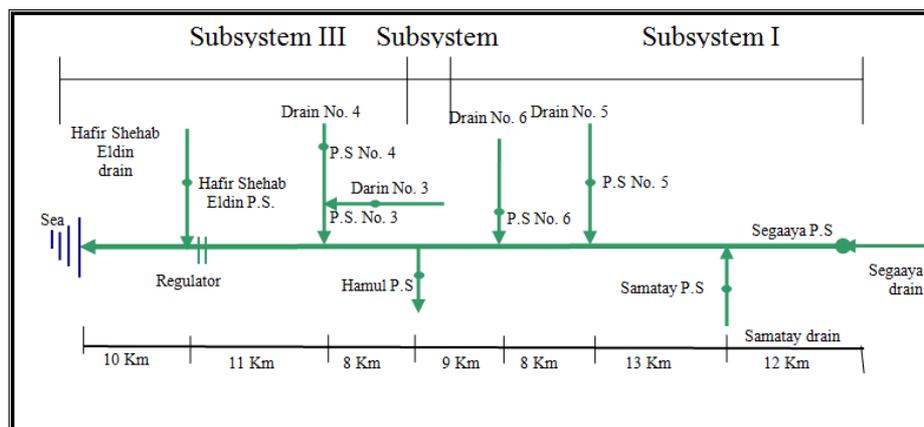


Fig. 2: Layout of Garbia Drainage System

2.2 Water Quality Monitoring Sites:

Ten water samples are collected from GDA; Seven from monitoring locations on the tributary agricultural drains flowing into the main drain and three from the main drain. These locations are chosen according to the noticeable changes expected in water quality in GDA due to the effluent of the drainage pumping stations into the drain and the spatial characteristics of the monitoring locations (Table 1).

Table 1: Water Quality Monitoring sites in Gharbia Drainage Catchments Area

NO	Site	Description	Notes
1	MG02	U.S Segaaya lifting drainage P.S at Km70	Segaaya drain
2	MG04	U.S Samaty lifting drainage P.S at Km 60	Samaty Drain
3	MG28	D.S Segaaya and Samatay drainage P.S's at Km 56	Gharbia main Drain
4	MG05	U.S of P.S No.5 to Gharbia drain at Km 46	Drain No.5
5	MG07	U.S of P.S No. 6 to Gharbia drain at Km 38	Drain No.6
6	MG08	U.S of Hamul P.S to Bahr Tira Canal at Km 29	Gharbia main Drain
7	MG09	U.S of P.S No.4 to Gharbia drain at Km 29	Drain No.4
8	MG10	U.S of P.S No.3 to Gharbia drain at KM 21	Drain No.3
9	MG12	U.S of Hafer P.S to Gharbia drain at KM 10	Hafer Drain
10	MG14	Outfall of Gharbia drain and the delivery of Hafer P.S	Gharbia main drain

U.S : Up Stream P.S Pumping Stations KM : Kilometer distance to drain outfall

The water samples are collected monthly by Drainage Research Institute (DRI) and analyzed by Central Laboratory of Environmental Quality Monitoring (CLEQM) under the umbrella of National Water Research Center (NWRC) in Egypt. The water quality samples are collected and analyzed according to the standard methods for examination of water and wastewater [5].

2.3 Water Quality Monitoring Parameters:

Water quality can be described in terms of physical, chemical and bacteriological variables. Water quality monitored parameters are selected according to the requirements for the reused drainage water in irrigation purposes and the sources of pollution at GDA, taking into consideration the laboratory facilities and the knowledge and experience of the field staff.

Some water quality monitoring parameters are measured in the field using portable meter such as Temperature (T), Hydrogen Ion Concentration (pH), Electric Conductivity (EC), Total Dissolved Salts (TDS), and Dissolved Oxygen (DO). The remaining water quality parameters are measured in the laboratory such as Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Nitrate (NO₃), Ammonium (NH₄), Total Nitrogen (TN), Total Phosphors (TP), Heavy metals, and Fecal Coliform (FC).

2.4 Statistical Analysis:

Normality tests are used to determine whether a random variable is normally distributed or not. Many data analysis methods (e.g. t-test, ANOVA, regression, or multivariate techniques) depend on the assumption that data are normally distributed.

Principal Component Analysis/Factor Analysis (PCA/FA):

Principal Component Analysis/Factor Analysis (PCA/FA) is a very powerful technique which provides information on the meaningful parameters (water quality variables) to describe the whole data set rendering data reduction with minimum loss of information [6]. In addition, the PCA/FA is a quantification of the significance of variables that explain the observed grouping and patterns of the inherent properties of the individual objects and allows an explaining of related parameters by only one factor which responsible for variation in river water quality and eventually leads to sources identification of river water pollution [7], [8].

In this research, PCA/FA was applied to extract the most significance PC's (Water Quality variables) and to reduce the contribution of less significant variables to simplify even more of the data structure coming from PCA/FA. The obtained PC's were further subjected to varimax rotation according to well established rules to maximize differences between variables and facilitate smooth interpretation of the data [9]. The rotating axis is defined by PCA/FA generates Varimax Factor (VF) through factor analysis which can further reduce the contribution of variable with minor significance. One of the most important steps of factor analysis is to determine number of factors that need to be extracted for accurate data analyses. In this regard, the rotation of the factor axis is performed to enhance a simple structure such as factors indicated by high loadings for some variables and low loadings for the others. In order to determine the number of factors to be used, variances and co-variances of the variables are computed. Then, eigenvalues and eigenvectors could be evaluated for the covariance matrix and the data is transformed into factors. All mathematical and statistical calculations were implemented using Minitab and SPSS software packages.

One-Way Analysis of Variance (ANOVA):

One-way ANOVA is used to compare means from at least three groups from one variable. The null hypothesis is that all the population group means are equal versus the alternative that at least one of the population means differs from the others.

Hypotheses Statements and Assumptions for One-Way ANOVA

The hypothesis test for analysis of variance for g populations:

$H_0: \mu_1 = \mu_2 = \dots = \mu_g$ Where μ_g is the mean value of group g

H_a : not all μ_i ($i = 1, \dots, g$) are equal

In this research Analysis of Variance (ANOVA) test was carried out to detect the temporal variation of the water quality parameters at Gharbia Drainage Catchments Area (GDA).

Tukey's test:

Tukey's test, is a single-step multiple comparison procedure and statistical test. It can be used on raw data or in conjunction with an ANOVA (Post-hoc analysis) to find means that are significantly different from each other.

Tukey's test compares the means of every treatment (the means of water quality variables of each month for all years,) to the means of every other treatment; that is, it applies simultaneously to the set of all pairwise comparisons and identifies any difference between two means that is greater than the expected standard error.

In this work Tukey's Test is used to formally test whether the difference between a pair of groups (different months all years) is statistically significant.

RESULTS AND DISCUSSIONS

Preliminary analysis is carried out for field and lab data by preliminary screening to remove the outliers and check the data consistency, i.e. normality and dependency. Raw data is screened to check if they are satisfying the general physics law of surface water quality bodies. Values that do not comply with these requirements are eliminated [10].

Principal Component Analysis / Factor Analysis:

Principal component analysis (PCA) /factor analysis (FA) is done for 15 water quality variables of GDA (pH, EC, TDS, DO, BOD, COD, NO₃, NH₄, TN, TP, TVS, TSS, Boron, FC, Total coliform) through different years to identify the important water quality parameters. These water parameters represent the water quality of the GDA and the pollution sources in the area. The Scree plot; figure 3, is used to identify the number of principal components to be retained to comprehend the underlying data structure [11]. An eigenvalue gives a measure of the significance of the factor; the factors with the highest eigenvalues are the most significant. Eigenvalues of one or greater are considered significant [9]. There are five factors out of fifteen have Eigenvalues greater than one (Figure 3).

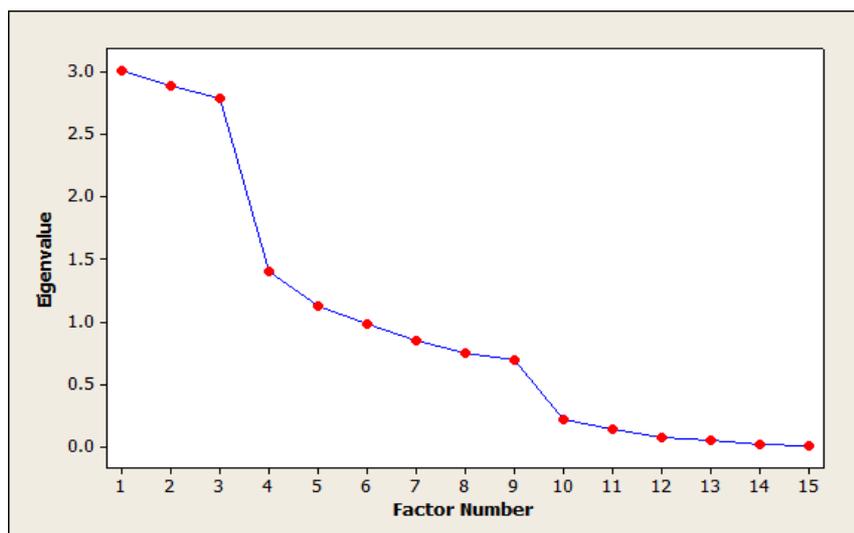


Fig. 3: Scree plot graph for components with its eigenvalues in GDA

The interpretation of un-rotated principal components is relatively more difficult to explain physically. Therefore, the five components were rotated using Varimax normalized rotation in order to make interpretation easier. In this work, PCA by the unrotated component matrix and PCA by Varimax rotation are both performed. Corresponding, variable loadings (before and after Varimax) and explained variance are presented in tables 2 and 3 and strong loading values have been highlighted (greater than 0.8).

Table 2: Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Fecal	-0.160	0.439	-0.639	0.472	-0.201	0.889
BOD	-0.582	0.665	0.029	0.021	0.319	0.884
COD	-0.635	0.623	0.077	-0.004	0.332	0.908
TSS	-0.100	0.592	0.559	-0.226	-0.491	0.965
TVS	0.004	0.555	0.577	-0.157	-0.548	0.966
NO ₃	0.785	0.261	-0.421	-0.092	-0.016	0.871
NH ₄	0.558	0.550	-0.352	-0.339	0.140	0.872
TP	-0.279	0.394	-0.182	0.001	0.132	0.284
TN	0.687	0.498	-0.346	-0.304	0.100	0.941
Boron	0.080	0.311	0.130	-0.209	0.212	0.209
pH	0.462	-0.173	-0.008	0.246	-0.158	0.329
EC	0.472	0.256	0.613	0.493	0.197	0.945
TDS	0.477	0.251	0.595	0.515	0.191	0.946
DO	0.281	0.026	0.410	-0.035	0.268	0.321
Coliform	-0.062	0.417	-0.597	0.519	-0.260	0.870
eigenvalue	3.0071	2.8846	2.7843	1.3979	1.1257	11.1995
% Var	20	19.2	18.6	9.3	7.5	74.7

Table 3: Rotated Factor Loadings and Communalities Varimax Rotation

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Fecal	0.128	0.198	-0.092	0.906	0.065	0.889
BOD	-0.045	0.914	0.045	0.165	-0.128	0.884
COD	-0.107	0.931	0.029	0.107	-0.131	0.908
TSS	0.020	0.210	0.106	-0.068	-0.951	0.965
TVS	0.030	0.089	0.174	-0.035	-0.962	0.966
NO ₃	0.846	-0.318	0.097	0.194	0.082	0.871
NH ₄	0.925	0.102	0.009	0.059	-0.046	0.872
TP	0.089	0.471	-0.080	0.218	-0.018	0.284
TN	0.965	-0.030	0.061	0.063	-0.038	0.941
Boron	0.267	0.275	0.119	-0.189	-0.108	0.209
pH	0.112	-0.490	0.246	0.116	0.052	0.329
EC	0.048	-0.070	0.954	-0.061	-0.153	0.945
TDS	0.046	-0.082	0.957	-0.037	-0.141	0.946
DO	0.102	-0.006	0.413	-0.374	-0.024	0.321
Coliform	0.141	0.092	-0.025	0.917	0.028	0.870
eigenvalue	2.6604	2.4634	2.1454	1.9928	1.9376	11.1995
% Var	17.7	16.4	14.3	13.3	12.9	74.7

From table 3, it is observed that the five varifactors are accounted for 74.7% of the total variance. The first varifactor, with an eigenvalue of 2.66, explained 17.7% of the total variance which is clearly dominated by the nitrogen nutrients; NO_3 , NH_4 and TN (positive loading) as shown in figure 4. This varifactor reflect the nutrients related to agricultural and untreated or partially treated domestic wastewater discharges, which reflect inorganic contamination. The second varifactor; with an eigenvalue of 2.46, explained 16.4% of the total variance and it is clearly dominated by BOD and COD which reflect the organic load effects. The third varifactor explain 14.3% of the total variance and dominated by TDS, which reflect the agricultural drainage water. The fourth varifactor explain 13.3% of the total variance and dominated by Fecal contamination which reflect the significant of domestic wastewater discharges and could be identified as domestic contamination. The last factor explain 12.9% of the total variance and dominated by TSS and TVS, which reflect the domestic wastewater and agricultural runoff.

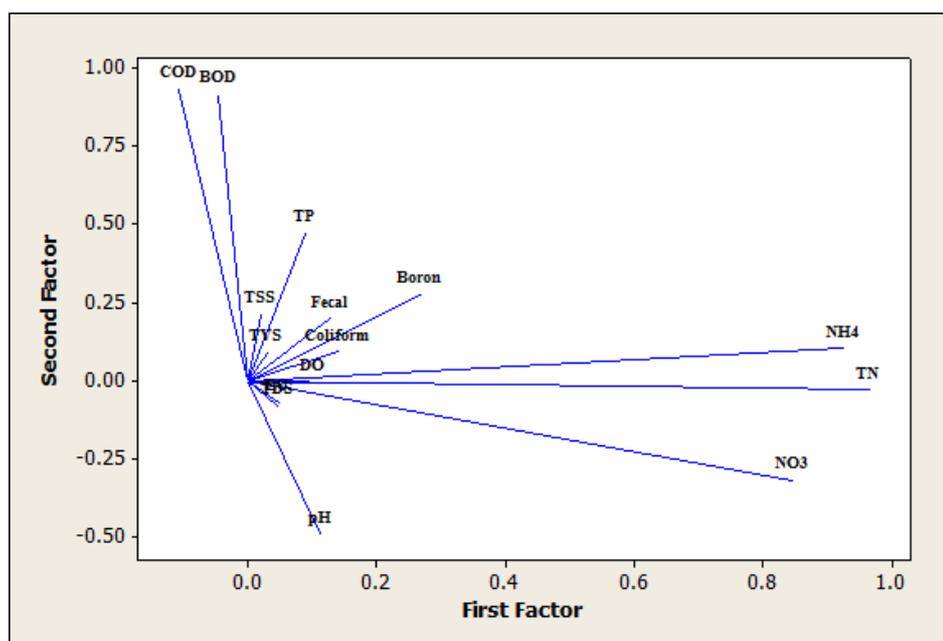


Fig. 4: Varimax normalized rotated factors in Gharbia Drainage Catchment Area

In general, the most important (significant) water quality parameters can be summarized in the following table.

Table 4: Summary of most important Water Quality Parameters

Factors	Parameters with strong positive loading	Parameters with strong negative loading
F1	NO_3 , NH_4 and TN	
F2	BOD and COD	
F3	EC and TDS	
F4	FC and Total Coliform	
F5		TSS and TVS

Table 4 shows that eleven water quality parameters out of fifteen are more important and significant. Also, these parameters in each loading factor are strongly correlated; physically, chemically, and biologically. For example, Nitrate (NO_3) and Ammonium (NH_4) are part of the Total Nitrogen (TN), Biological Oxygen Demand (BOD) is a part of Chemical Oxygen Demand (COD), Electric Conductivity (EC) and Total dissolved Salts (TDS) give the same meaning, Faecal coliform (FC) is a part of total faecal coliform (Total coliform), Total volatile solids (TVS) is a part of Total Suspended Solids (TSS). Therefore, five water quality parameters can be measured out of the 11 parameters. These parameters are Total Nitrogen (TN), Biological Oxygen Demand (BOD), Total Dissolved Salts (TDS), Faecal Coliform (FC), and Total suspended solids (TSS). Finally, five parameters out of 15 water quality parameters can be used to present the water quality of GDA. These parameters represent the different types of pollution sources; agricultural and domestic pollution. Decreasing the number of the measured water quality parameters will ensure the reduction in the cost of the monitoring network that can be used to improve the resources in the water quality-monitoring network, i.e. Staff, equipment, facilities, etc.

One-Way Analysis of Variance (ANOVA):

ANOVA is carried out for the eleven water quality variables resulting from the five loading factors. The results show that the p-value for FC, BOD, Nitrate, Ammonia, TSS, TDS and TN ANOVA is less than 0.05. This result indicates that the seasonal variations for each variable are statistically significant. The following figures (Ln value of water quality variables vs measured months all years) and table show the significant variation of some of water quality variables.

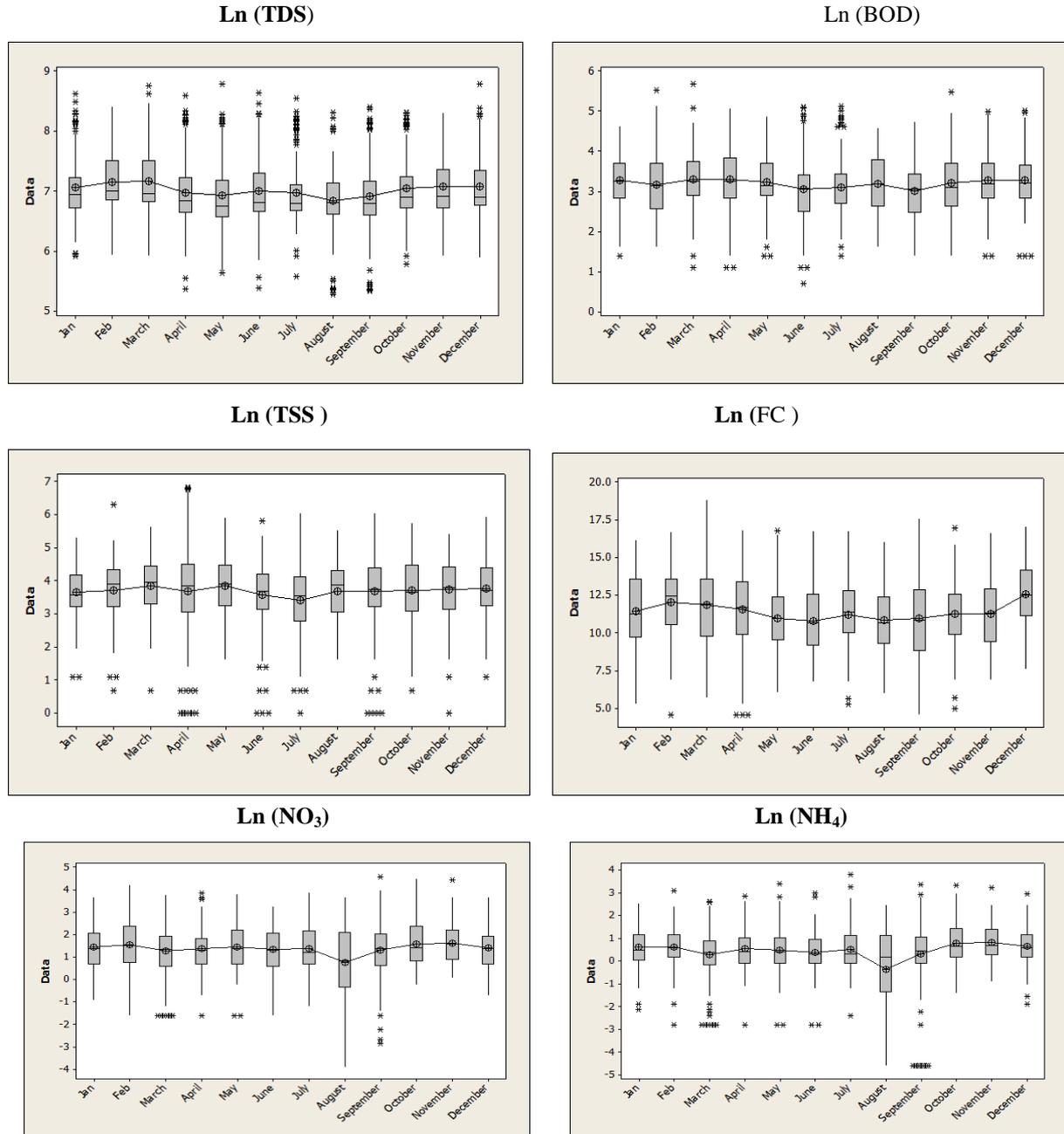


Fig. 5: Monthly Variation of Water Quality Variables all years

Table 5: ANOVA test results for Water Quality Variables

Parameters	P-Value	Seasonal Variation
FC	0.000	Significant difference
BOD	0.000	Significant difference
TSS	0.001	Significant difference
TDS	0.000	Significant difference
Nitrate	0.000	Significant difference
Ammonium	0.000	Significant difference
TN	0.000	Significant difference

Based on Tukey test, table 6 shows the months that have significant values different than other months for water quality variables. Most of water quality variables have different significant values in months January, February, March, May, June, July, August, and September. So; Generally, the measurement of water quality variables can be intensified in the previous eight months and reassessment of the monitoring water quality network is carried out again to ensures the obtained results.

Table 6: Summary of Months that have significant difference values for Water Quality Variables based on Tukey test

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
Jan						=	=	^=	+ =		=	+ #
Feb					*#	# =	=	*#^	*#			
Mar					*#	# =	*#--- =	*#^ =	*+ # =		=	
Apr								^	+			#
May							---	^				#
Jun								^		=		#
Jul								^		=		#---
Aug									^	^	^	#^
Sept												#
Oct												#
Nov												#
Dec												#

* TDS + BOD # FC --- TSS ^ TN = TP

Table 7 shows the months that have different (significant) values for each water quality variables. The results show that the water quality parameters should be measures at least three times per year in months March, July, September, One in winter and twice in summer. It is possible to add another measurement in winter season; either in December or January. So, the water quality variables; TDS, BOD, FC, TSS, TN, TP; should be measured intensively four times per year.

Table 7: Significant Months for Water Quality Variables based on Tukey Test

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
TDS		√	√		√		√	√	√			
BOD	√		√	√					√			√
FC	√	√	√	√	√	√	√	√	√	√	√	√
TSS			√		√		√					√
TN	√	√	√	√	√	√	√	√	√	√	√	√
TP	√	√	√			√	√	√	√	√	√	

Conclusions And Recommendations:

Gharbia drainage catchment area (GDA) is one of the biggest drainage area in the Nile Delta and valley. Most of GDA drainage water is reused officially by the Ministry of Water Resources and Irrigation; by mixing its drainage water with the fresh water with designed mixing ratio to cope with the irrigation requirements. Also, the drainage water is used unofficially for irrigation and fish farms. Keeping the water quality of GDA in a good condition is taking seriously by the government and the public.

The high rate of population growth and insufficient sanitation impose huge threats to the water resources, especially on the non-conventional ones. Results of factor analysis showed that five loading factors are representing the water quality variables of GDA and accounted for 74.7% of the total variance. The first varifactor, explained 17.7% of the total variance which is clearly dominated by the nitrogen nutrients. The second varifactor; explained 16.4% of the total variance and it is clearly dominated by the organic load effects. The third varifactor explain 14.3% of the total variance and dominated by the agricultural drainage water. The fourth varifactor explain 13.3% of the total variance and dominated by fecal contamination. The last factor explains 12.9% of the total variance and dominated by agricultural runoff and domestic pollution.

ANOVA analysis indicates that the seasonal variations of FC, BOD, Nitrate, Ammonia, TSS, TDS and TN are statistically significant. Based on Tukey test, Most of water quality variables have different significant values in months January, February, March, May, June, July, August, and September.

It is concluded that five water quality; TDS, BOD, FC, TN, TSS; are used to represent the water quality of GDA out of fifteen water quality parameters. These water quality variables can be measured from four to eight times per year instead of twelve times. The secured budget due to the reduction in the water quality monitoring parameters and sampling frequency can be used for water quality monitoring network enhancement and resources development. It is recommended to intensify the measurement of the water quality variables in months January, March, July, and September. These months have the most significant different values than the other months. In addition, it is recommend reassessing the water quality monitoring network of GDA after five years to ensure the sustainability of the results.

ACKNOWLEDGMENT

The research is supported by the National Water Research Center (NWRC). Special thanks to Drainage Research Institute (DRI) and Central Laboratory of Environmental Quality Monitoring (CLEQM) at NWRC for providing the field data and laboratory analysis.

REFERENCES

1. MWRI, 2013. National Strategy for Water Resources Development and Management in Egypt till 2050, Cairo, Egypt.
2. CCME, 2015. Canadian Council of Ministers of the Environment, Guidance Manual for Optimizing Water Quality Monitoring Program Design, Canada.
3. Abdel Rashid A., 2002. Water Quality control in open drains and its effect on soil properties. PhD. Thesis, faculty of Agriculture, Ain Shams University, Egypt.
4. WMO, 2013. Planning of Water –Quality Monitoring system, WMO-No.1113. World Meteorological Organization, Geneva, Switzerland.
5. APHA, AWWA, and WEF, 2012. Standard Methods for the Examination of Water and Wastewater, 22nd edition, American Public Health Association, American Water Works Association and Water Pollution Control Federation, Washington, DC., USA.
6. Singh, K.P., A. Malik and S. Sinha, 2005. Water Quality Assessment and Apportionment of Pollution Sources of Gomati River(India) Using multivariate statistical techniques – a Case Study. *Analtica Chimica. Acta*, 538: 355-374.
7. Kowalkowski, T., R. Zbytniewski, J. Szpejna and B. Buszewski, 2006. Application of Chemometrics in River Water Classification. *Journal of Water Research*, 40(4): 744-752. Lohani, M.B., A. Singh, D.C. Rupainwar and D.N. Dhar, 2008. Seasonal Variations of Heavy Metal
8. Boyacioglu, H. and H. Boyacioglu, 2006. Water Pollution Sources Assessment by Multivariate Statistical Methods in the Tahtali Basin, Turkey." *Journal of Environmental Geology*, 54(2): 275-282.
9. Shrestha, S. and F. Kazama, 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environ. Model. Software*, 22(4): 464-475.
10. DRI, 2011. Re-Assessment of the Egyptian National Water Quality Monitoring Network (NWQMN), Final Report, National Water Research Center, Cairo, Egypt.
11. Vega, M., R. Pardo, E. Barrado and L. Deban, 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.*, 32(12): 3581-3592.