TREND DETECTION OF DRAINAGE WATER QUALITY IN EGYPT

Bahaa M. KHALIL Rasha M.S. El KHOLY & Shaden T. ABDEL GAWAD*

Abstract

Water quality has been the principal limiting factor to water availability. The assessment of short and long-term water quality changes is a challenging problem. During the last two decades, there has been an increasing demand for monitoring water quality of many water bodies by regular measurements of various water quality parameters. The result has been the gradual accumulation of reliable water quality records and the examination of these data for trends (Hirsch et al.*, 1991). Without such information of the trend detection of the water bodies, effective water quality management remains impossible.*

The goal of this research is to identify water quality trends in Egyptian drains. The proposed analysis aims at determining how and to what degree several water quality parameters are changing, and characterizing the function and response of the drains to seasonal variability besides the correlation of load-discharge relationships and concentration-discharge relationship where the load-discharge relationships showed better correlation than that of concentration-discharge relationships.

Although several parameters are examined, particular emphasis is given herein to ascertaining trends in nutrients, organic matter and physical parameters. An examination of a vital drainage catchement in the eastern region of Egypt's Nile Delta is conducted to describe the short-term trends. The data of five water quality variables (NO_3, P, BOD, COD, TSS) and the discharge monitored on a monthly basis for the period August 1997-December 2002 were selected for this analysis.

This study examines the time series of monthly values of water quality parameters and the discharge using statistical methods and the existence of trends and thus presents the evaluation of the best-fitted trend models. Trends are detected using the regression analysis of the variables involved. Due to the wide variation over time in the statistical tests for nutrients, organic matter and physical parameters, the trend varied as for BOD, COD, TSS and NO_3 concentrations it was downwards following the quadratic equation while the concentration of P showed no trend.

1 INTRODUCTION

During the last two decades, there has been an increasing demand for monitoring water quality of many water bodies by regular measurements of various water quality parameters. According to Liebetrau (1979), some of the necessities of water quality monitoring are the following: 1) to provide a system-wide synopsis of water quality, 2) to monitor short and long-range trends in selected water quality parameters, 3) to detect actual or potential water quality problems and 4) to enforce standards.

---

* Research Assistant, Researcher and Vice Chairperson. National Water Research Center. Ministry of Water Resources and Irrigation.
National Water Research Center (NWRC) Administration Building, El Qanater, 13621, Cairo, Egypt

More frequent sampling generally helps up to a point; however, data collected much more frequently than monthly lose independence, which adversely affects the statistical tests for trend. This paper examines statistically: (1) the time series of monthly values of water quality parameters and the discharge at one of the vital drainage outfalls in the eastern region of Egypt's Nile Delta, (2) the existence of trends and the evaluation of the best fitted trend models and (3) the relationships between concentration and loads of water quality parameters and the discharge.

Testing water quality data for trend over a period of time has received considerable attention recently. The interest in methods of water quality trend arises for two reasons. The first is the intrinsic interest in the question of changing water quality arising out of the environmental concern and activity. The second reason is that only recently has there been a substantial amount of data that is amenable to such an analysis.

Trend analysis determines whether the measured values of a water quality variable increase or decrease during a time period. In statistical terms, has the probability distribution from which they arise changed over time or not. It would be useful to describe the amount or rate of that change, in terms of changes in some central value of the distribution such as mean or median (Hirsch *et al.*, 1982).

2 THE STUDY SITE

Bahr Hadus drain outfall discharges annually about 1 bcm of agricultural drainage water into El Salam Canal (Figure 1). El Salam Canal Project is considered as a strategic national development project for land reclamation of 620,000 acres in Sinai Governorate.

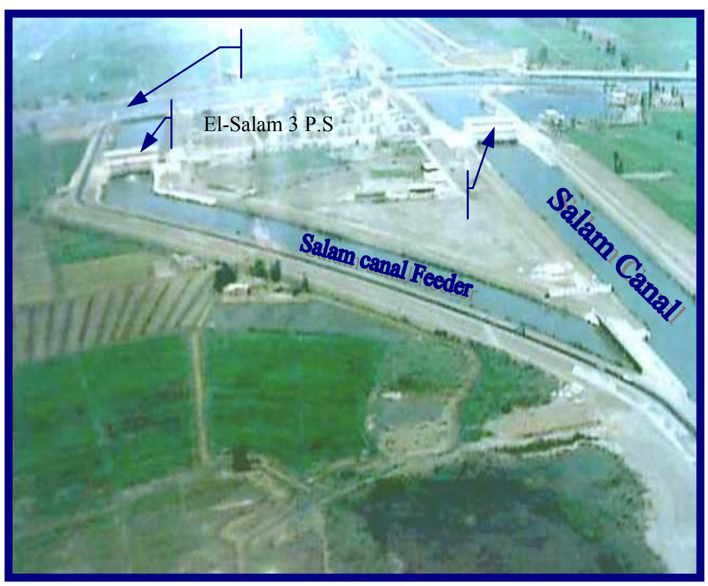


Figure 1. Satellite image showing the area of Bahr Hadus outfall

Since the catchment area of Bahr Hadus drain is located in a highly polluted area, the drain system is susceptible to pollution from legal and illegal dumping of domestic and industrial wastewater. Most of the water received by the drain is from agricultural diffuse sources. Although the domestic diffuse sources are only 4% of the total discharge, it contributes 94% of the organic load received by the drain, expressed as BOD (DRI, 2000).

3 METHODOLOGY OF ANALYSIS

In planning control and management program of streams, the statistical and trend analysis as well as the relations between concentrations or loads and discharge are important steps for understanding the behavior and the variation of water quality parameters and stream flow (Antonopoulos *et al*., 2001).

Water quality data do not usually follow convenient probability distributions such as the well-known normal and lognormal distributions on which many classical statistical methods are based (Lettenmaier *et al.*, 1991).

Frequency histograms are used to determine how well data fit a theoretical distribution. This could be achieved by comparing visually histograms of measured values to the density curve of normal distributions. To check normality, the Shapiro-Wilk W-statistic was used. The W-statistic has values ranging from 0 to 1; small values for W are significant and indicate non-normality (Shapiro and Wilk 1965). For samples less than 200 the Shapiro-Wilk test should be used (SAS 1985). For larger samples, the Kolmogorov-Smirnov test should be used.

The Shapiro-Wilk test, proposed in 1965, calculates a *W* statistic that tests whether a random sample, $x_1, x_2, .. x_n$ comes from (specifically) a normal distribution . Small values of *W* are evidence of departure from normality and percentage points for the *W* statistic. The *W* statistic is calculated as follows:

$$W = \frac{\left[\sum_{i-1}^{n} a_i x_{(i)}\right]^2}{\sum_{i-1}^{n}\left(x_i - \bar{x}\right)^2}$$

where the $x_{(i)}$ are the ordered sample values ($x_{(1)}$ is the smallest) and the a_i are constants generated from the means, variances and co-variances of the order statistics of a sample of size *n* from a normal distribution.

For evaluating the concentration-discharge and load-discharge relationships, the method of least squares for the pairs of monthly measured values of each variable with the discharge was used to determine the constants of these models. For this analysis, the linear (C_{ij}=a+bQ_j), the power (C_{ij}=aQj^b), the exponential (C_{ij}=a exp(bQ_j)) and the logarithmic (C_{ij}=a+b ln(Q_j)) models were used.

4 RESULTS AND DISCUSSION

Figure 2 shows the time series of monthly measured values of Q, BOD, COD, TSS, NO_3 and P at Hadus drain outfall from August 1997 to December 2002. In figure 2 it is shown that many water quality parameters show striking seasonal variations, only the nitrates show relatively uniform seasonal variation during the time even though it gives the highest coefficient of variation. Some of the variations depend on the discharge and some on the seasonality. The statistical measures of time series of water quality variables used in the study analysis are given in Table 1.

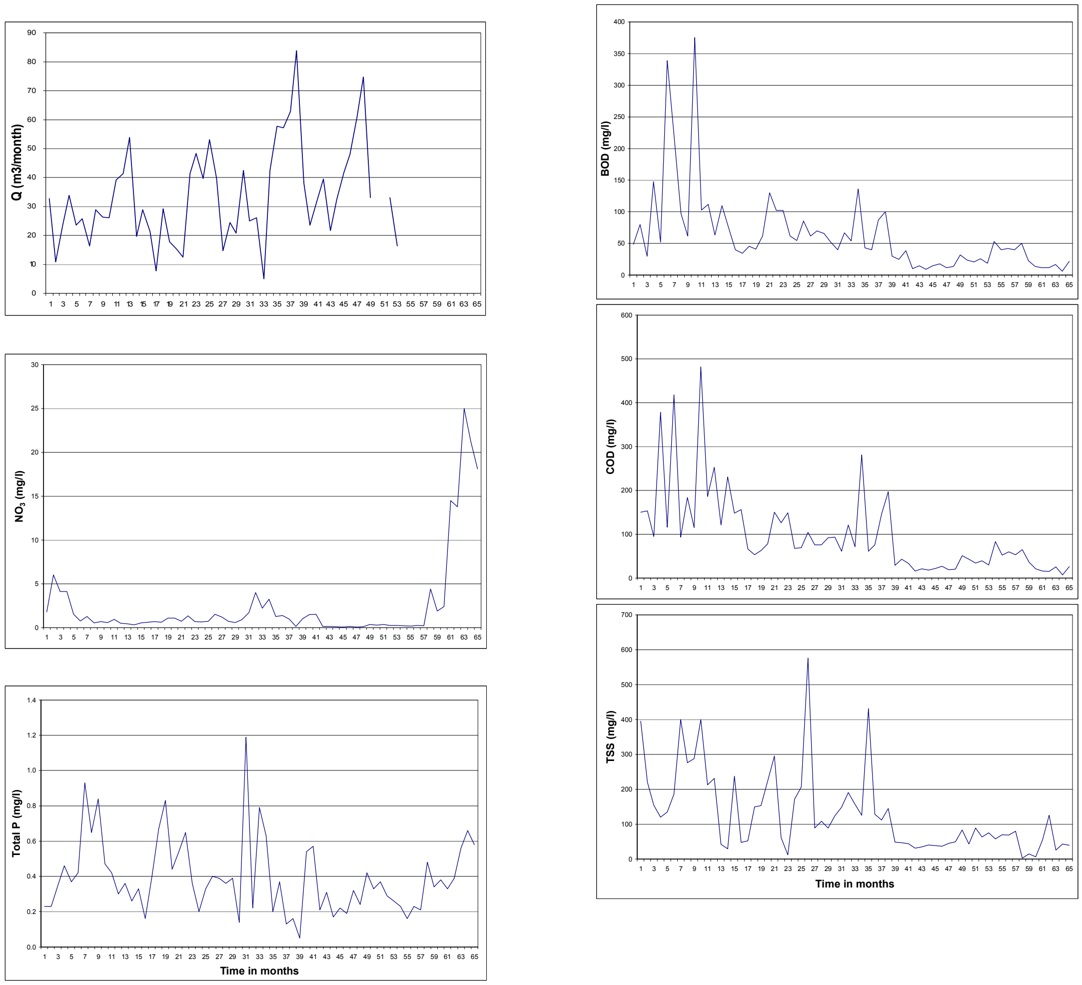


Figure 2. Time Series of water quality parameters and discharge at Bahr Hadus outfall (August 1997 – December 2002)

Table 1. Statistical parameters of the time series of monthly values of water quality parameters and discharge of Bahr Hadus drain outfall

Variable	Sample Size	Mean	Median	Maximum	Minimum	Range	STDV	CV
Q m3/sec	51	33.64	31.90	83.75	5.05	78.71	16.75	0.50
BOD mg/l	65	63.47	45.40	375.20	6.00	369.20	66.00	1.04
COD mg/l	65	99.35	69.20	482.00	7.00	475.00	95.33	0.96
TSS mg/l	65	130.67	89.00	576.00	3.00	573.00	117.60	0.90
NO3 mg/l	60	1.13	0.71	6.01	0.06	5.95	1.24	1.09
P mg/l	65	0.39	0.36	1.19	0.05	1.14	0.21	0.54

STDV = standard deviation CV = coefficient of variation

From figure 2 and table 1, there is a variation in concentrations and discharge. There are different reasons for these variations, some of them depend on seasonality and discharge; the magnitude of the fluctuation in discharge is much greater than those of concentrations. The ratio of the highest to the lowest concentrations is very large for total suspended solids (192:1) followed by nitrates (100:1), chemical oxygen demand (69:1), biological oxygen demand (62:1) and phosphorus (23:1) while the ratio for discharge was about (16:1).

A very useful and concise graphical display for summarizing the distribution of a data set is the boxplot (Helsel and Hirsch, 1992). Box plots provide visual summaries of: 1) The center of the data (the median = the centerline of the box), 2) The variation or spread (interquartile range = the box height), 3) The skewness (quartile skew = the relative size of box halves) and 4) The presence or absence of unusual values (outliers and extreme values). Boxplots are even more useful in comparing these attributes between several data sets.

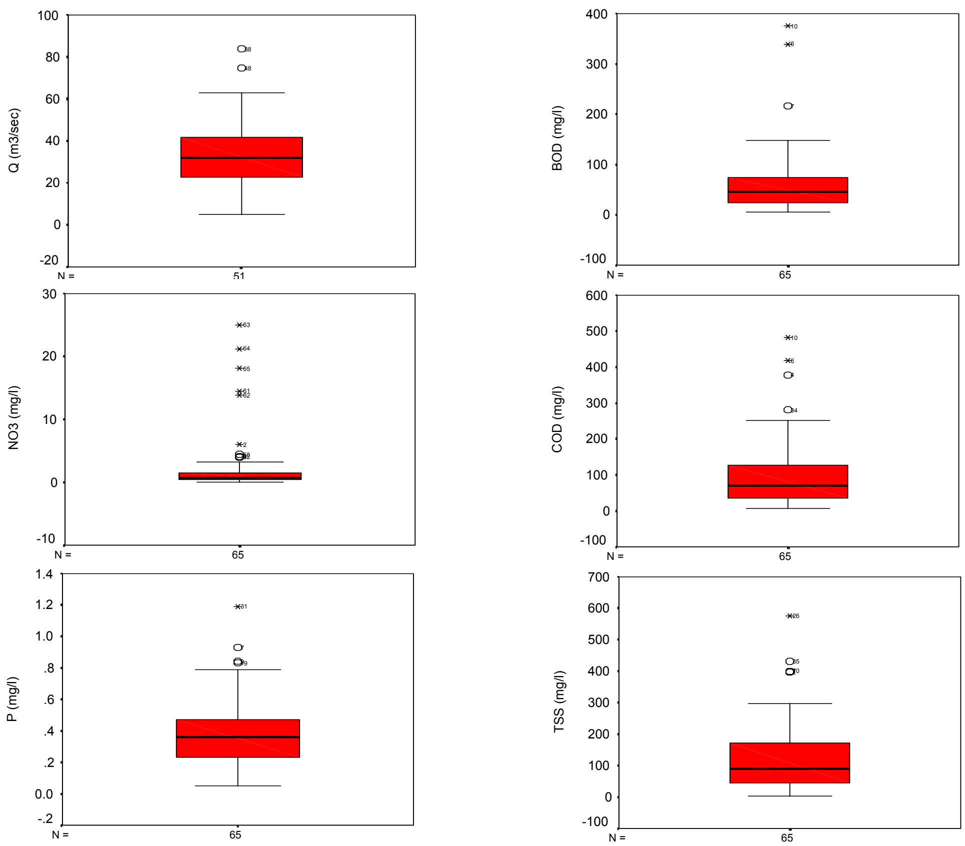


Figure 3. Boxplots of water quality parameters and discharge at Bahr Hadus outfall

Figure 3 shows the box and whiskers plots of water quality parameters and discharge at Bahr Hadus outfall. The plots show that the data of the variables BOD, COD, TSS, NO_3, and P depart from a normal distribution not only in skewness, but also by the number of outliers and the extreme values which are unexpected and might be due to non-suitable measurements or handling the water samples. The data of discharge is approaching normality with only two outliers.

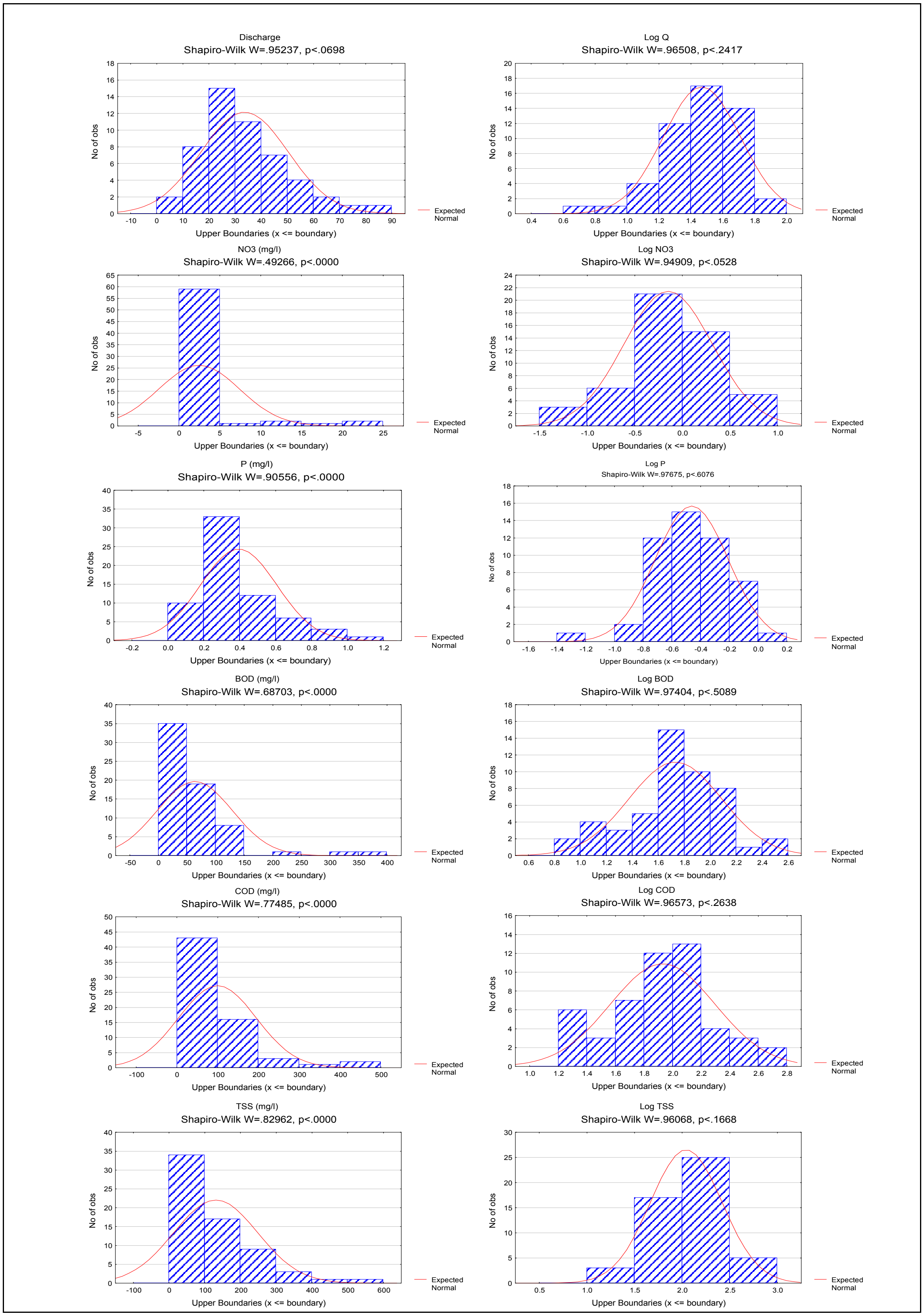


Figure 4. Frequency histograms of water quality parameters data and the log-transformed data

The Shapiro-Wilk W-statistic was used to check normality, where normality has not been exactly met the transformation recommended was the log-transformation as shown in figure 4. The values of W and P are presented in table 2 for the data before and after the transformation. Only discharge values did not need any transformation.

Table 2. The Shapiro-Wilk W-statistic and P values for water quality parameters before and after the log-transformation.

Variable	W	P	W (Transformed)	P (Transformed)
Q m^3/month	0.952	0.069	0.965	0.242
BOD mg/l	0.687	0.000	0.974	0.509
COD mg/l	0.775	0.000	0.966	0.264
TSS mg/l	0.829	0.000	0.961	0.167
NO_3 mg/l	0.493	0.000	0.949	0.052
P mg/l	0.905	0.000	0.977	0.607

It is not considered necessary to use nonparametric approaches simply because the assumption of normality has been exactly met after-transformation.

5 TREND ANALYSIS FOR WATER QUALITY PARAMETERS

The selection of the best-fitted model was based on the values of the MAPE, MAD, and MSD. MAPE, or Mean Absolute Percentage Error, measures the accuracy of fitted time series values. It expresses accuracy as a percentage. MAD, which stands for Mean Absolute Deviation, measures the accuracy of fitted time series values. It expresses accuracy in the same units as the data, which helps conceptualize the amount of error. MSD stands for Mean Squared Deviation. It is very similar to MSE, mean squared error, a commonly-used measure of accuracy of fitted time series values. Because MSD is always computed using the same denominator, n, regardless of the model, it is easy to compare MSD values across models. For all three measures of accuracy, the smaller the value the better the fit of the model. These statistics were used to compare the fits of the different methods. The values of the statistical tests and their graphs, along with the observed values of the data time series are presented in Figure 5, while the models and values of statistical tests are given in table 3.

Table 3 Trend models with the values of the statistical tests for the goodness-of-fit.

Variable	Model	MAPE	MAD	MSD
Q	$22.0045 + 0.536888*T - 0.00263*T^2$	56.311	12.175	239.635
Log NO_3	$0.292777 - 0.02334*T + 0.000198*T^2$	279.577	0.346	0.181
Log P	$-0.29 - 0.0112*T + 0.000136*T^2$	82.8552	0.1762	0.0524
Log BOD	$2.03524 - 0.00795*T - 0.0000899*T^2$	13.6426	0.2067	0.0682
Log COD	$2.33101 - 0.0138*T - 0.0000311*T^2$	11.9692	0.2041	0.0625
Log TSS	$2.32623 - 0.00793*T - 0.0000869*T^2$	17.6513	0.2631	0.1178

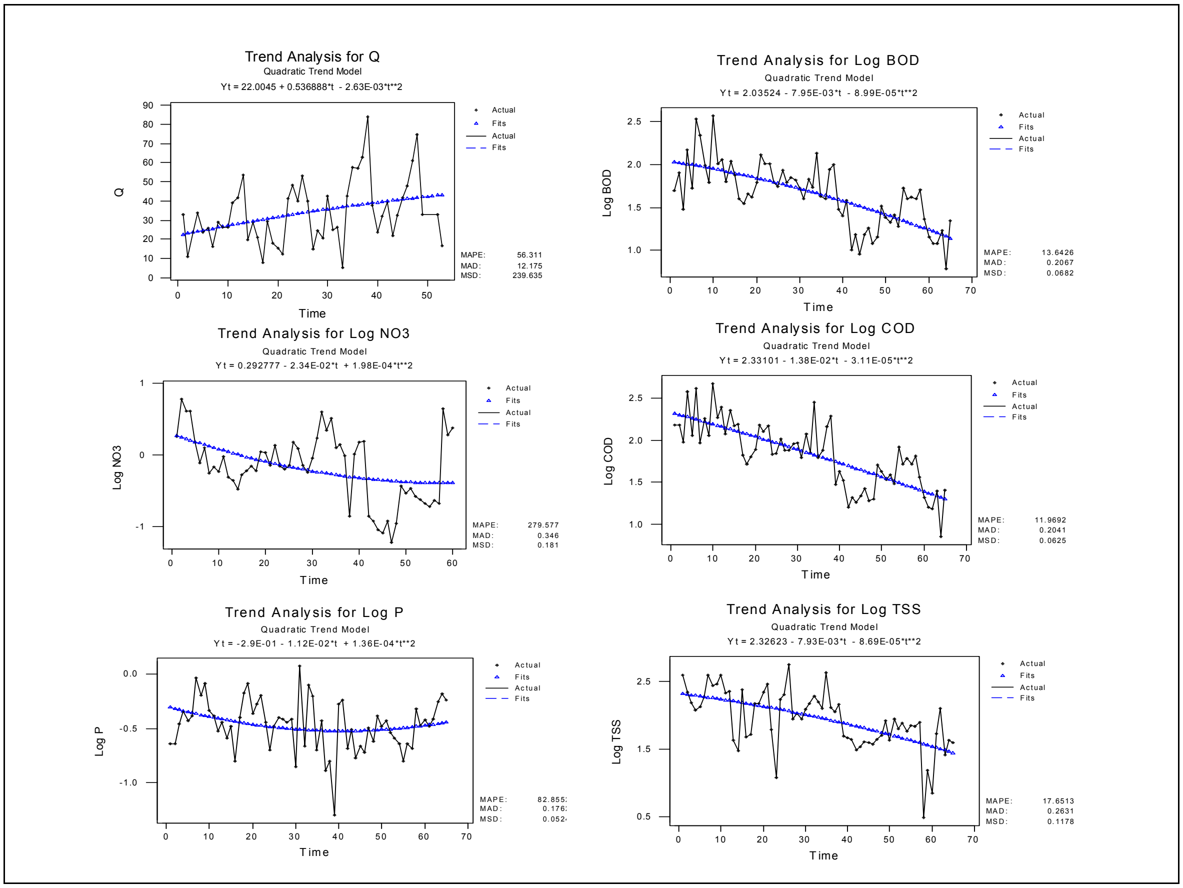


Figure 5. Best fitted trend model of monthly measured values of water quality parameters

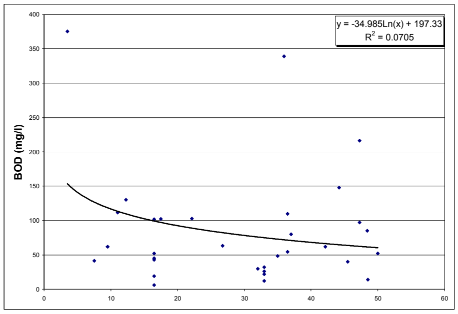
It is clear from the previous trend analysis that: a) the second order quadratic equation describes the trend of the data time series better, b) the trends of BOD, COD and TSS are steeply downwards while the trend of NO_3 is mildly downwards, c) the concentration of total P have no trends, while discharge shows upward trend.

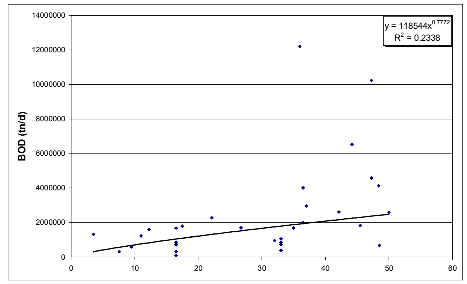
6 CONCENTRATION/LOAD-DISCHARGE RELATIONSHIPS

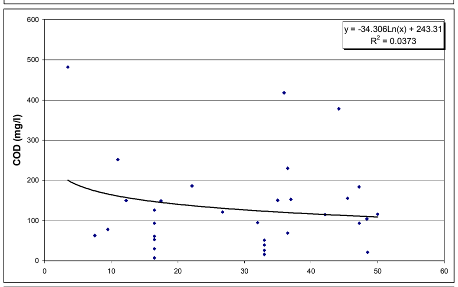
The results of simple regression applied between each water quality variable's concentration (dependant variable) and the discharge (independent variable) as well as between each water quality variable's load (dependant variable) and the discharge (independent variable), monitored at Bahr-Hadus drain outlet are given in table 4. In this table the values of parameters a and b are given for linear ($C_{ij}=a+bQ_j$), power ($C_{ij}=aQj^b$), exponential ($C_{ij}=a\ \exp(bQ_j)$) and logarithmic ($C_{ij}=a+b\ \ln(Q_j)$) models as well as the correlation coefficients are given. Figure 6 shows the best fitted to the data regression models.

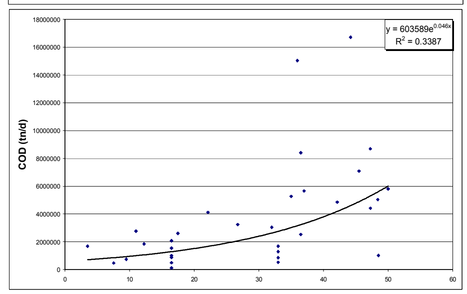
Table 4. Concentration-discharge and Load-discharge relationship

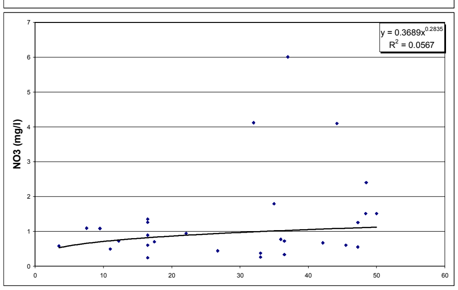
Variable	**Concentration-discharge**				**Load-discharge**			
	Equation	*a*	*b*	*r*	*Equation*	*a*	*b*	*r*
BOD	Log	197.33	-34.98	0.0705	Power	10.242	0.7772	0.2338
COD	Log	243.31	-34.31	0.0373	Exp	52.15	0.046	0.3387
NO_3	Power	0.3699	0.2835	0.0567	Power	0.0319	1.2835	0.5518
P	Power	0.5005	-0.075	0.0107	Exp	0.2451	0.0424	0.6226
TSS	Power	272.64	-0.257	0.0249	Linear	-133.4	20.282	0.2886

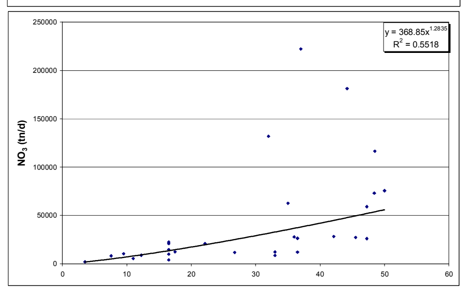

BOD (mg/l)
y = -34.985Ln(x) + 197.33
R² = 0.0705


BOD (tn/d)
y = 118544x^0.7772
R² = 0.2338


COD (mg/l)
y = -34.306Ln(x) + 243.31
R² = 0.0373


COD (tn/d)
y = 603589e^0.046x
R² = 0.3387


NO3 (mg/l)
y = 0.3689x^0.2835
R² = 0.0567


NO3 (tn/d)
y = 368.85x^1.2835
R² = 0.5518

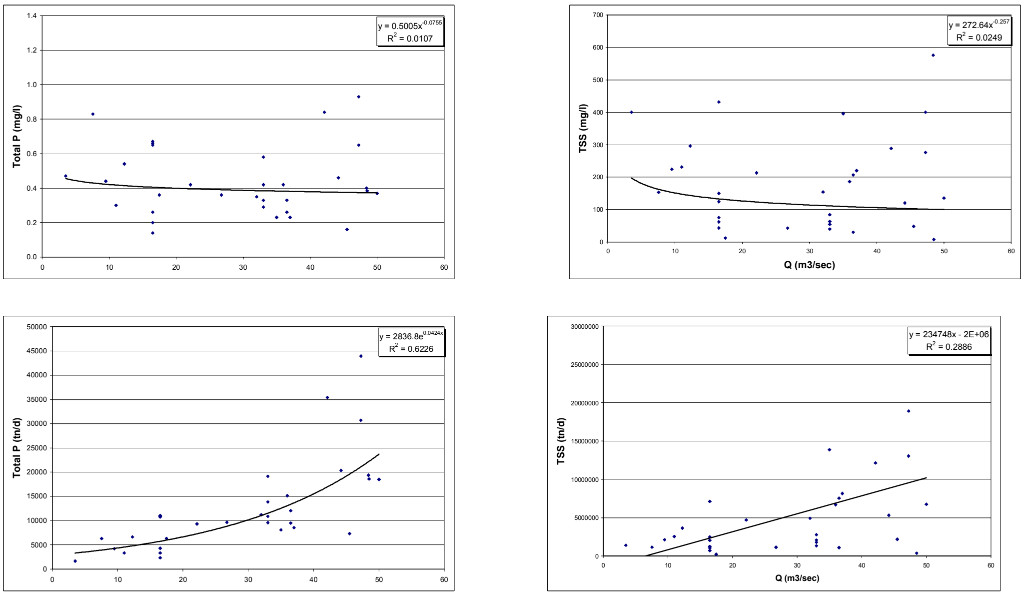


Figure 6. The best fitted model of water quality parameters concentrations and loads against discharge

From the values of table 4 and the graphs of figure 6 it is clear that: a) the logarithmic and the power models describe better the concentration-discharge relationships, b) the exponential and the power models describe better the load-discharge relationships, c) the load-discharge relationships show better correlation than that of concentration-discharge relationships even though both correlation relations are not significant.

7 CONCLUSIONS

Many factors, such as different sources of contaminants, seasonal cycles, precipitation, and natural variability affect measured water quality. As a consequence, it often takes many years of regular water quality data collection to statistically detect a trend that is, small, gradual changes.

Water quality trend assessment serves primarily as a warning system for change. This can be extremely useful for policy evaluation, but it must be emphasized that definitive conclusions on water quality trends may require years of sampling. Ultimately, if a trend is identified, additional scientific assessment is often essential to understand the implications of the trends and to identify effective corrective actions.

The load-discharge relationships show better correlation than that of concentration-discharge relationships even though both relations are not significant, extra data could lead to better correlation.

The trends of BOD, COD, TSS and NO_3 are downwards following the order quadratic equation while the discharge shows upward trend, only the concentration of total P has no trend.

REFERENCES

Antonopoulos, V.Z., Papamichail, D.M. and Mitsiou, K.A., 2001, *Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece*, Hydrology & Earth System Sciences, pp 679-691.

Drainage Research Institute, 2000, *Drain Pollution Sources Study for the Delta and Fayoum*, Technical report no. **56**.

Helsel, D.R. and Hirsch, R.M., 1992. *Statistical Methods in Water Resources*. Elsevier, The Netherlands.

Hirch, R.M., Slack, J.R. and Smith, R.A., 1982. *Techniques of trend analysis for monthly water quality data*, Water Resources Research **18**, pp 107-121.

Lettenmaier, R.P., Hooper, E.R., Wagoner, C., and Fans, K.B., 1991. *Trends in stream quality in continental United States*, 1978-1987, Water Resources Research **27**, pp 327-339.

Liebertrau, A.M, 1979. *Water quality sampling: some statistical considerations*, Water Resources Research **15**, pp 1717-1728.

SAS Institute Inc. 1985, *SAS user's guide: Statistics*. Cary. NC 25711-8000. 956 p.

Shapiro, S. S. and Wilk, M. B. (1965). "*An analysis of variance test for normality (complete samples)*", Biometrika, 52, 3 and 4, pages 591-611.