STOCHASTIC MODELS OF SOME PROPERTIES OF WASTE WATER IN THE MAAMERA SEWAGE TREATMENT PLANT

Dr. Nabaa Shakir Hadi

nabaalshimmri@yahoo.com

College of Engineering, University of Babylon, Babylon, Iraq

Received 30 November 2014   Accepted 3 February 2015

ABSTRACT

The records of influents concentration for both BOD_5 and TSS of Maamera sewage treatment plant which were chosen of this study are very important parameters, They play an important part in the planing and management of the national water resources. Most of these records have periods of missing data of the influent BOD_5 and TSS. In this study a model for generating missing monthly concentrations influent BOD_5 and TSS. Data are introduced. Initially univariate models using the Box-Jenkins approach were fitted to the logarithmically transformed series. Both transformed series were found to be generated by a random process using sampling theory were considered to be white noise. Ordinary regression analysis was performed. No significant correlation between influent BOD_5 and TSS concentration were found.

KEY-WORDS: Maamera sewage treatment plant, Statistical analysis, stochastic models.

نماذج عشوائية لبعض خصائص فضلات المياه في محطة معالجة مياه المجاري في المعيميرة

الخلاصة

تم في هذا البحث دراسة الخصائص لمياه الفضلات الداخلة لمحطة معالجة المعميرة لأهم المتغيرات الأساسية لنوعية المياه وهي المطلب الحيوي للأوكسجين والمواد الصلبة العالقة, حيث أنها تلعب دورا هاما في التخطيط وإدارة الموارد المائية. معظم هذه البيانات فيها فترات لبيانات مفقودة للمطلب الحيوي للأوكسجين والمواد الصلبة العالقة. تم دراسة موديل بإدخال البيانات الشهرية المفقودة لتركيز المطلب الحيوي للأوكسجين والمواد الصلبة العالقة. موديل احادي المتغير باستخدام طريقة بوكس جينكينز يلائم التحويلات للسلاسل الزمنية. في كل التحويلات للسلاسل يوجد متغير عشوائي بأستخدام نموذج نظري يعبر عنه . وباستخدام تحليل الانحدار تبين عدم وجود ارتباط بين التراكيز الداخلة للمطلب الحيوي للأوكسجين والمواد الصلبة العالقة.

1. INTRODUCTION

Engineers who take up the task of analyzing flows, BOD_5, etc. into stream for the purposes of design and planning are often confronted with the problem of working with records having a sequences of missing data. In this study, the readings of influent of BOD_5 and influent of Total suspended solid of the wastewater from Maamera sewage treatment plant were considered for the analysis in time series.

The infilling of missing values in hydrological data involves the use of statistical procedures of data generation one of such methods is the use of univariate linear stochastic models[Al-Samawi,1986]. Box and Jenkins (univariate models) techniques which were used to determine the appropriate model. These represent the structure of the time series Then tests of these values by the chi-square goodness of fit test and K.S test were performed to check the normality of the data[Hussain,2000]. Many wastewater treatment plants were built within Hilla city. Al-Maamera sewage treatment plant is one of these plants and has begun to operate in 1982. the plant works with an activated sludge system which biologically treats compounds of carbon and nitrogen in raw wastewaters. Maamera sewage treatment plant serve 50000 populations and the treatment facility is a conventional activated sludge system with an average wastewater inflow of 12000 m^3/day. The sewerage system is designed to accommodate the industrial wastewater, as well as domestic effluent. The treated wastewater in the plant is then being discharged to Shatt Al-Hilla River. A full outline of the plant units is shown in Fig. 1.

The objectives of the study is to investigate and analyzed the applicability of such stochastic models to the influent of BOD_5 and TSS. in the wastewater of the city of Hilla during the years, 2008 to 2013.

2. METHODS

In the present study, certain data have been collected yearly by the Mayoralty of Hilla from the influent in Maamera sewage treatment plant. Major water quality parameters were selected for this study; biochemical oxygen demand (BOD_5), Total suspended solids (TSS) over a period of six consecutive years.

3. THEORY

The mean of every monthly readings of influent BOD_5 and TSS. The parameters that the study depends on the first must be known so that of the time series and its components could be construct.

3.1 Definitions time series model

A time series is defined as a set of observations that measure the variation in time of some aspect of a phenomenon, such as the rate of the dissolved oxygen in the stream and the total suspended solids, or the sediment load in a channel [kottegoda, 1980].

3.2 Components of time series:

3.2.1 Trend:

Trend is a steady and regular movement in a time series through which the values are on average either increasing or decreasing.

3.2.2 Periodicity

This represents a regular or oscillatory form of variations such as seasonal effect which clearly evident in closely spaced data. In general, the periodic component in a time series can be represented through a system of sin functions after the trend component, if it exists, has been estimated and removed [kottegoda, 1980].

3.3 Time series model

If a high degree of dependency between sequential observations exists, then forecasting technique which express this dependency and which will generally produce superior results can be applied. These techniques which are presented by Box and Jenkins are called Box – Jenkins model.
These techniques are used to identify the appropriate model, other variables and estimate the parameters of the stochastic models.
In general; the model are formulated so that the current value of a variable is the weighted sum of past values and a random values which represents the unknown.

3.4 Parameters of the model:

3.4.1 Autocorrelation function(ACF)

For series,which are not random, there will be dependency between sequential observations. A useful tool to measure this effect is the autocorrelation function which may be defined as:

$$\rho(k) = \frac{E(X_t - U_x)(X_{t+k} - U_x)}{\sqrt{E(X_t - U_x)^2 . E(X_{t+k} - U_x)^2}} \qquad (1)$$

The autocorrelation function has the following properties:
$\rho(0) = 1$
$|\rho(k)| \leq 1 \, for \, all \, k \neq 0$
$And \; \rho(k) = \rho(-k)$
For an observed time series X_t of length N, the autocorrelation function of lag k can be estimated from

$$r(k) = \frac{\sum_{t=1}^{N-K}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{N}(X_t - X)^2} \qquad (2)$$

Where:

$$\bar{X} = \frac{1}{N}\sum_{t=1}^{N} X_t \qquad (3)$$

3.4.2 Partial autocorrelation function(PACF)

The partial autocorrelation function at lag k is the correlation between X_t and X_{t+k} with the effects of the intervening observations(X_{t+1} ,X_{t+2} ,…….,X_{t+k-1}) removed. [Montgomery and Johnson, 1976].
Notationally, we shall refer to the K^{th} partial autocorrelation coefficient as $\emptyset_{k,k}$
The set of parameter $\emptyset_{1,1}, \emptyset_{2,2}, \emptyset_{3,3}$, which are the last coefficients of the autoregressive models of order 1,2,3,… respectively represent the partial autocorrelation coefficient. A plot of $\emptyset_{k,k}$ versus the lag K is called the sample partial autocorrelation function.
In general, the partial autocorrelation $\emptyset_{P,P}$ is the autocorrelation remaining in the series after fitting a model of order (P-1) and removing the liner dependence. The partial autocorrelation function(PACF) is an important tool in determining the order of the model if the serial correlation function suggests that the process could be approximated by a linear autoregressive model.
As a general rule,we would assume a partial autocorrelation coefficient to be zero if the absolute value of its estimate is less than twice its standard error [Kottegoda,1980].

3.4.3 Autoregressive processes(AR)

The autoregressive processes means that the current observations X_t is "regressed" on previous realizations $X_{t-1}, X_{t-2}, \ldots X_{t-p}$ of the same time series[Montgomery and Johnson,1976]. The autoregressive model AR(P) takes the form.

$$X_t = \emptyset_{P,1} X_{t-1} + \emptyset_{P,2} X_{t-2} + \cdots \emptyset_{P,P} X_{t-P} + a_t = \sum_{i=1}^{P} \emptyset_{P,1} X_{t-1} + \emptyset_t \quad (4)$$

Where $\emptyset_{P,i} i = 1,2,3, \ldots, P$ are the autoregressive parameters or weights and (a_t) is a white noise process or residuals, the model in eauation(4) is called an autoregressive process of order P, abbreviated AR(P).
Also, the model called a linear autoregressive model, in which the current value of a variable is equated to the weighted sum of a (P) number of past values. A variant (at) that is completely random, the word linear merely signifies that the current value is dependent additively upon the past values and not for example, on their squares or square roots [Kottegode, 1980].

4. RESULTS AND DISCUSSION

4.1 The data

The data used in this study are the average of each months for the six-year (2008 – 2013) period for both of influent of BOD_5 and TSS into Maamera Sewage Treatment Plant.

4.2 Time plot

2.1 Both of the two Figs (2&3) show that the behavior of the original time series for both influents BOD_5 and TSS these Fig show:

a. The maximum value for influent BOD_5 was (214mg/l) in April 2012. While the minimum value for influent BOD_5 was (75mg/l) in Novmaber 2013.
b. The maximum value for influent TSS was (301mg/l) in March 2009. And the minimum value for influent TSS was (93mg/l) in March 2010.

2.2 From Figs (4&5) it was noted that, the standard deviation for every year was directly perpotional to the mean in that year. It is noted in the beginning, the standard deviation was low and so was the mean while during the last year the standard deviation became higher with the mean.
All these indicated that a logarithmic transformation of the data was needed to stabilize the variance and to make multiplicative effects additive.

4.3 Transformation

After adjusting the outlier observation the logarithms for the original time – serieses were taken and are plotted as shown in the Figs (6&7) for both influents BOD_5 and TSS these Fig show:

a. The standard deviation become constant with the increase of the mean.
b. The variation patterns during every year for these series are similar to the variation patterns of the original series.

The values for both influent BOD_5 and influent TSS are shown in Table1.

4.4 Autocorrelation

From Figs (8&9) for influent BOD_5 and influent TSS respectively, the autocorrelation function of the series have no trend and seasonality. since the autocorrelation function have the ability of all

lags are not significant and also,the function have no seasonal cycles[Hipel et al.,1977b], hence the time series has no deterministic for stochastic component
From Figs (10&11) for influent BOD_5 and influent TSS. It can be that, show the partial autocorrelation functions for two series with confidence limits of (95%).
from these four Figs, it can be seen that all autocorrelation coefficient will be within the confidence limits (95%). Hence , it can be said that the two series were (serially independent).

4.5 Test of Normality

The test is carried out by two ways:

4.5.1 Chi-Square Test

The Chi –Squared statistic depends on specifying the number of histogram classes into which the data will be grouped, and there is no rule that gives the correct number to use [Vose, 2010]. The Chi –Squared test statistic is computed from the relationship:

$$x^2 = \frac{\sum_{i=1}^{k}(O_i - E_i)^2}{E_i} \qquad (5)$$

Where O_i is the observed and E_i is the expected number of observation in the ith class interval(based on the probability distribution being tested). The expected numbers are calculated by multiplying the expected relative frequency by the total number of observation[Barkotulla et al.,2009].The chi square test parameters are shown in Tables (2) and(3) for influent BOD_5 and TSS respectively.

From Table (2) it is seen that, the values of $x^2 = 7.8508$ for influent BOD_5 and all the expected frequencies were be larger than or equal to 5[Crof,1979]. The chi-square value is found to be (0.25). This value is within the acceptable region for the normally distributed and that it is white noise series as shown in Fig(12).

For influent of TSS, the values of $x^2 = 3.5747$ and all expected frequencies were greater than(5) as shown in Table (3). The chi-square value was (0.75). This value is within the acceptable region for the normally distributed and that it is white noise series as shown in Fig(13).

4.5.2 Kolmogorov-Smirnov Test

Kolmogorov-Smirnov(K-S)goodness of fit test is based on a statistic that measures the deviation of the observed cumulative histogram from the hypothesized cumulative distribution function [Soong, 2004].By using this test, the significant level for influent BOD_5 was (0.441), and for influent TSS.was (0.642) as show in Table (4).
From all this it can be concluded that the series are white noise and have normal distribution as was obtained from(Chi-square test).

4.6 Regression Analysis

The study of regression had done on the three relationships the first relation was between influent of BOD_5 and TSS. The data of this relation can be seen from Table (5) and the plot of this relation is shown in Fig(14).
Second trial was carried out between the influent BOD5 and the transformed values of TSS, as it seen in Table(6) and Fig(15).
Third trial had performed out between the transformed function of influent BOD_5 and the transformed values of TSS, as it seen in Table(7) and Fig(16).

From these three relations, it can be seen that there is no physical relation exist between influent BOD_5 and TSS. The values of R^2 for this test were (0.019),(0.028), and (0.017),respectively. These values were too low to say that the model was adequate for prediction.

4.7 One-Step-ahead-Forecast

The forecasting of the sample for influent BOD_5 and TSS, during the period of recording the data (2008 to 2013) is depend upon the sampling theory. From the theory of sampling is the estimate of both BOD_5 and TSS. Can be found by the following expression:

$$\bar{X} \pm 2Se \quad (6)$$

Where:
$\bar{X}$: is the mean for the influent BOD_5 and TSS.
Se: is the standard error for the mean influent of BOD_5 and TSS.
Then for influent BOD_5 the forecasting value is $(168.19\ ,241.71\ )$ mg/l while for the influent TSS. It is $(247.78\ ,353.32)$mg/l.

5. CONCLUSIONS

The following conclusions are drawn from this study:
1. The need for the logarithmic transformation of both influent BOD_5 and TSS concentrations data indicates that the two parameters which generate data are non linear in nature.
2. The deterministic component of data of both influent BOD_5 and TSS.
3. The time series of both transformed influent BOD_5 and TSS is white noise series without residual series.
4. The seasonal effect is not present, so if the time series tales values more than 72 value may be the seasonal effect appear.
5. Box-Jenkins models are not applicable here because the randomness of the data.
6. The forecasting values are derive from the sampling method are tabulated these forecasting values(no each case an interval estimate is given) should be up dated to monitor the values of $\bar{X}$and S_e for each variable(BOD_5 and TSS).
7. Relationship between influent BOD_5 and TSS concentration:
An attempt was made to relate the influent TSS concentration, which is usually easy to measure, with the influent BOD_5 which is takes lengther time to determine.
The range of possible mathematical relationships covered in this analysis are as follow:

(i)The simple linear form,

$$BOD_5 = a + b\,TSS \quad (7)$$

(ii)The inverse form,

$$BOD_5 = \acute{a} + \acute{b} ln\,TSS \quad (8)$$

(iii)The semi inverse form,

$$lnBOD_5 = \acute{\acute{a}} + \acute{\acute{b}}\,ln\,TSS \quad (9)$$

Figs (14),(15) and (16) show the following

No visual relationship between influent BOD_5 and TSS, exists according to the mathematical formulations as given in equations(7),(8) and(9). This finding is supported by the results of the statistical regressions which are tabulated in Table(5),(6) and (7). In all mathematical formulations, the slope coefficients b, $\acute{b}, and\ \acute{\acute{b}}$ were found to be insignificant, thus supporting the findings that no physical relations between influent BOD_5 and influent TSS.
Hence, the best model which represent the variability of the influent BOD_5 is given by the log-normal distribution. Similarly, influent TSS. A concentration may be modeled in the same manner.

REFERENCES

[1] Al-Samawi, A. A. (1986). "The generation of missing river flow data", Arab Gul,J.of scientific Research. Vol.4, No.2. PP.761-774.

[2] Barkotulla M.A.B., Rahman M.S., and Rahman M.M.,(2009). "Characterization and frequency analysis of consecutive days maximum rainfall at Boalia, Rajshahi and Bangladesh", India, Journal of Development and Agricultural Economics Vol. 1(5). PP.121-126.

[3] Chatfield, D. (1979). "Statistics for Technolgy", Champaman and Hall,2nd edition.

[4] Croft. D. (1979)."Applied statistics for management studies", Macdonald and Evans, 2nd edition.

[5] Hipel, K, W., Mcleod, A. I. and Lennox, W. C. (1977). "Advances in Box-Jenkins modeling 2. Applications", Water Resources Research, Vol.13,No.3.PP.577-586.

[6] Hussain, S. M., (2000). "Time series analysis for hydrochemical pollutants: Applications of Box-Jenkins models to Euphrates river at kufa". M.Sc. Thesis, University of Babylon.

[7] Kottegode, N. T., (1980). "Stochastic Water Resources Technology", the macmillan press Ltd.

[8] Maidment, D. R. and Parzen, E. (1984). "Cascade model of monthly municipal water use", Water Resources Research, Vol. 20. No. 1, PP. 15-23.

[9] Soong T.T., (2004). "Fundamentals of probability and statistics for engineers", John Wiley & Sons Ltd, State University of New York at Buffalo, New York, USA.

[10]Vose, David.,(2010). "Fitting distributions to data and why you are probably doing it wrong", 15 Feb 2010, www.vosesoftware.com.

Table (1): Descriptives			
Parameter	Mean	Standard Deviation	Standard Error
BOD_5	4.9703	0.2552	0.03095
TSS	5.2185	0.2194	0.02661

Tables (2): Chi-Square Test for the influent BOD_5				
Lower limit	Upper limit	Observed frequency	Expected frequency	Chi-square
At or below	4.55	6	5	0.1403
4.55	4.67	6	7	0.1428
4.67	4.79	4	8	2.0075
4.79	4.91	8	10	0.4036
4.91	5.03	13	10	0.8410
5.03	5.15	10	10	0.0022
5.15	5.21	9	6	1.3787
5.21	5.27	4	7	1.2857
Above 5.27		8	5	1.6490
Chi-square=7.8508 with 6 dif, Sig.level=0.25				

Tables (3): Chi-Square Test for the influent TSS				
Lower limit	Upper limit	Observed frequency	Expected frequency	Chi-square
At or below	4.92	4	5	0.2568
4.92	5.05	9	7	0.4787
5.05	5.13	8	6	0.5565
5.13	5.21	9	13	1.3097
5.21	5.29	14	13	0.0769
5.29	5.37	10	8	0.4199
5.37	5.45	5	6	0.2150
5.45	5.58	5	5	0.0044
Above 5.58		4	5	0.2568
Chi-square= 3.5747 with 6 dif, Sig.level=0.75				

Table (4): The values of Kolmogorov-Smirnov Test for all the influents and with confidence level equal 95%				
Parameters	Estimated KOLMOGOROV Statistics DPLVS	Estimated KOLMOGOROV Statistics DPLVS	Estimated Overall statistics DN	Approximate significance level
BOD_5	0.105	0.085	0.105	0.441
TSS	0.090	0.087	0.090	0.642

Table (5): Regression Analysis-Linear Model y=a+bx					
Dependent variable:BOD5			Independent variable:TSS		
Parameter	Estimate	Standard error	T-value	Probability level	
Intercept Slope	125.613 0.121	20.906 0.108	6.008 1.119	0.000 0.267	
Analysis & Variance					
Source	Sum of square	D.f.	Mean square	F-ratio	Probability
Model Error	1567.889 82599.097	1 66	1567.889 1251.501	1.253	0.267
Total(correlation)= 84166.985 D.f.= 67 Correlation coeffication=0.136 Standard Error Estimate=35.377 R-squared=0.019					

Table (6): Regression Analysis-Linear Model y=a\`+b\`lnx					
Dependent variable:BOD5			Independent variable: lnTSS		
Parameter	Estimate	Standard error	T-value	Probability level	
Intercept Slope	8.541 26.822	102.398 19.605	0.083 1.368	0.934 0.176	
Analysis & Variance					
Source	Sum of square	D.f.	Mean square	F-ratio	Probability
Model Error	2321.257 81845.728	1 66	2321.257 1240.087	1.872	0.176
Total(correlation)= 84166.985 D.f.= 67 Correlation coeffication=0.166 Standard Error Estimate=35.215 R-squared=0.028					

Table (7): Regression Analysis-Linear Model lny=a\`\`+b\`\`lnx					
Dependent variable:lnBOD5			Independent variable: lnTSS		
Parameter	Estimate	Standard error	T-value	Probability level	
Intercept Slope	4.169 0.154	0.741 0.142	5.625 1.082	0.000 0.283	
Analysis & Variance					
Source	Sum of square	D.f.	Mean square	F-ratio	Probability
Model Error	0.076 4.288	1 66	0.076 0.065	1.171	0.283
Total(correlation)= 4.364 D.f.= 67 Correlation coeffication=0.132 Standard Error Estimate=0.25488 R-squared=0.017					

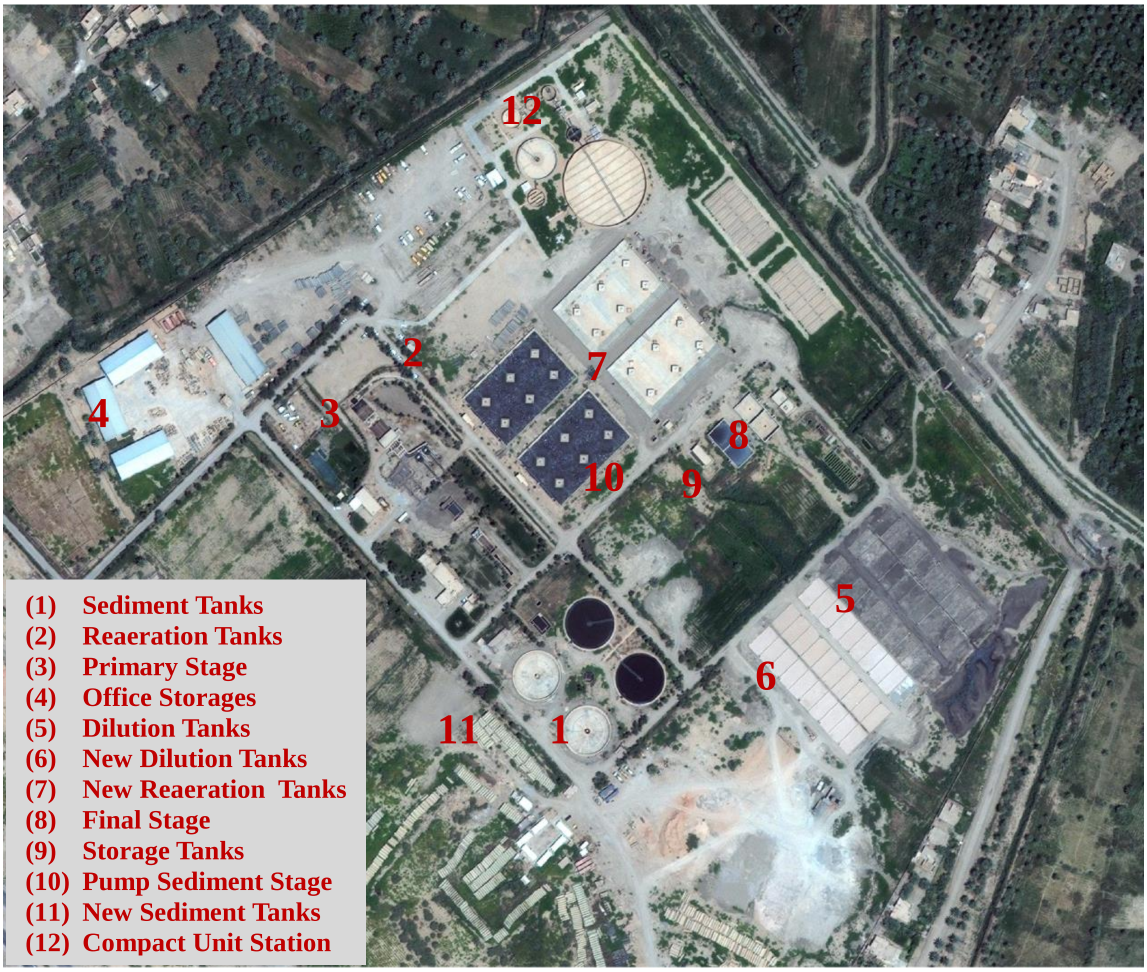

Figure (1): Image map of Maamera sewage treatment plant, Hilla (Al-Maamera project office, 2012).

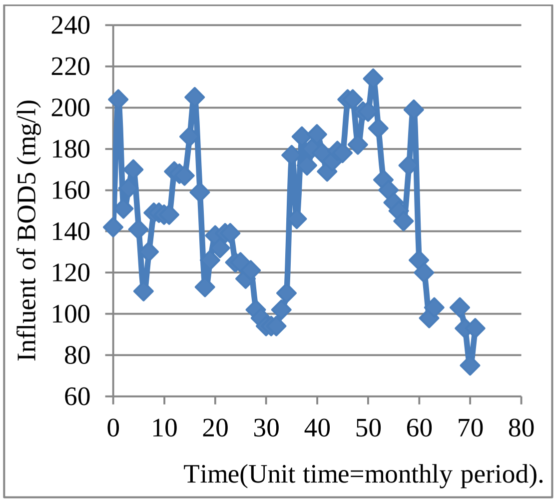

Figure (2): Time series of influent BOD_5 of Al-Maamera S.T.P.

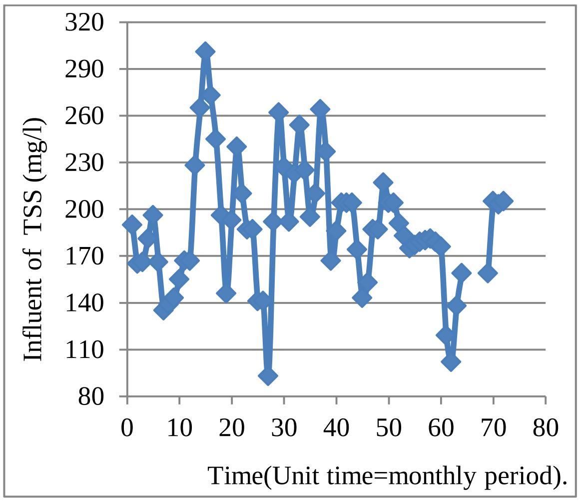

Figure (3): Time series of influent TSS of Al-Maamera S.T.P.

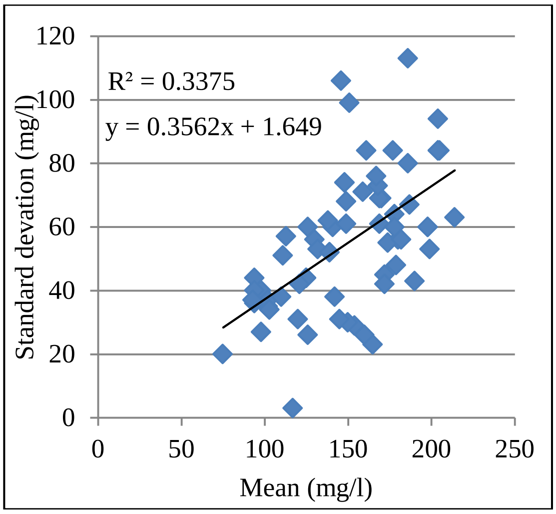

Figure (4): Standard devation of influent of BOD_5 Vs. Mean.

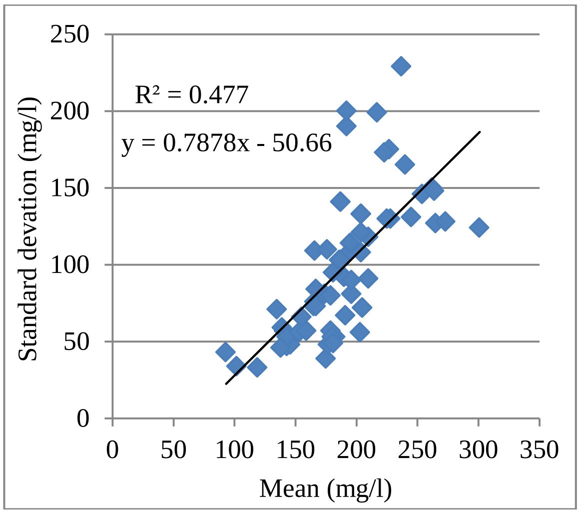

Figure (5): Standard devation of influent of TSS Vs. Mean.

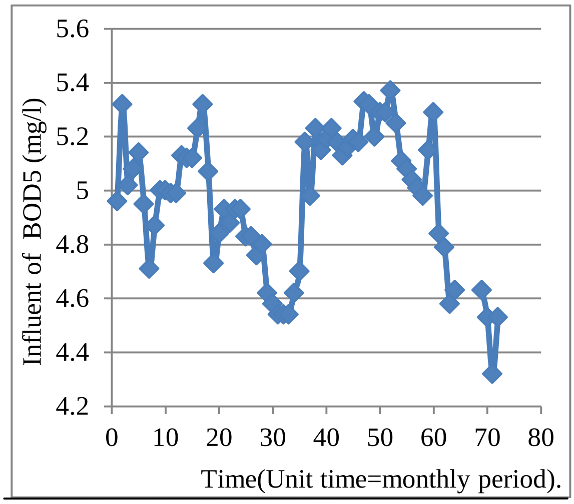

Figure (6): Transformed of Infl. BOD_5 series

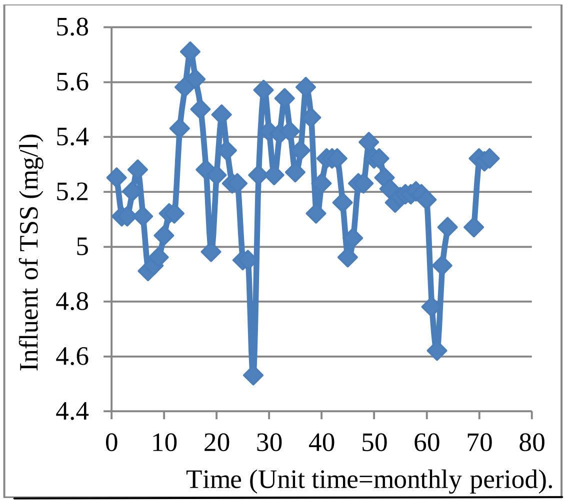

Figure (7): Transformed of Infl. TSS series

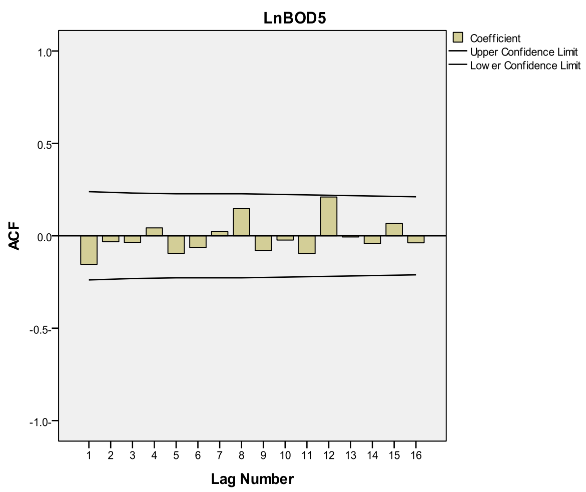

Figure (8): Autocorrelations for influent BOD_5 series

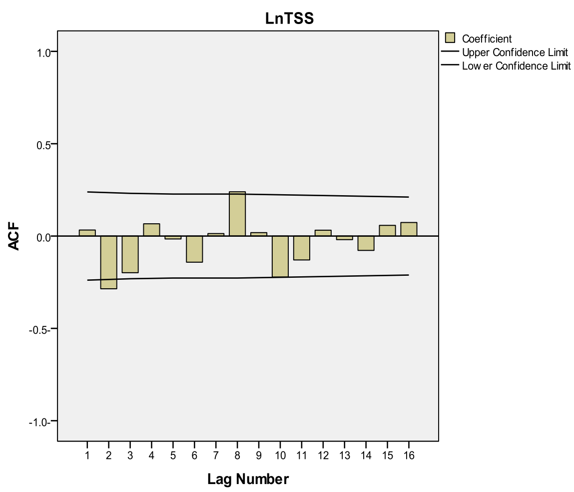

Figure (9): Autocorrelations for influent TSS series

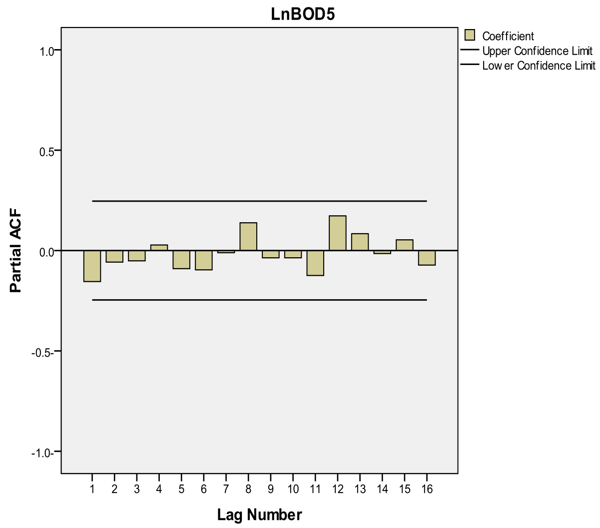

Figure (10): Partial autocorrelations for influent BOD_5 series

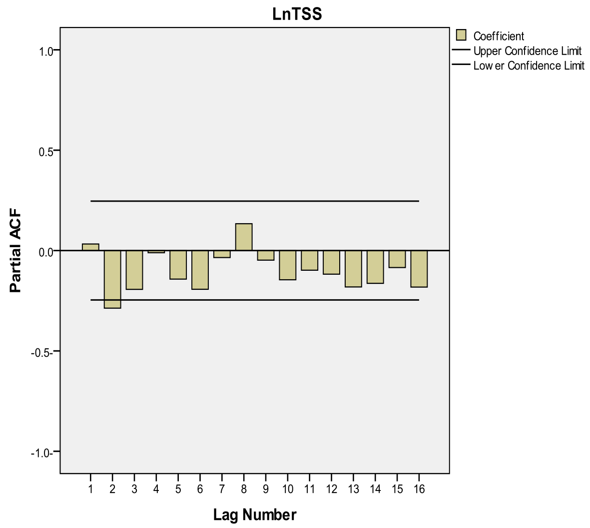

Figure (11): Partial autocorrelations for influent TSS series

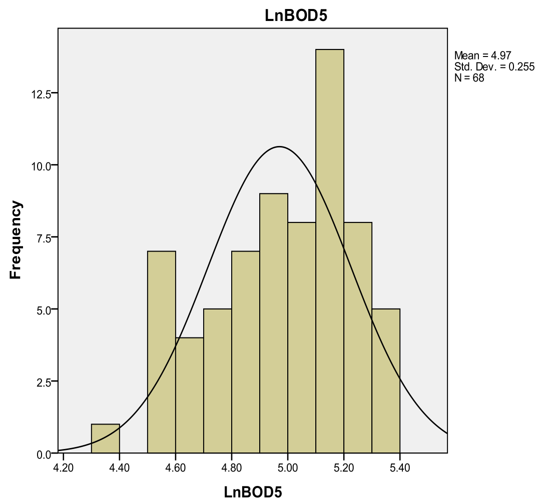

Figure (12): Frequency Histogram for transformed series of influent BOD_5

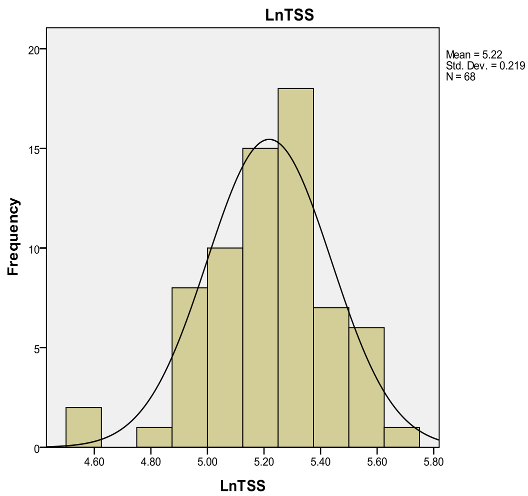

Figure (13): Frequency Histogram for transformed series of influent TSS

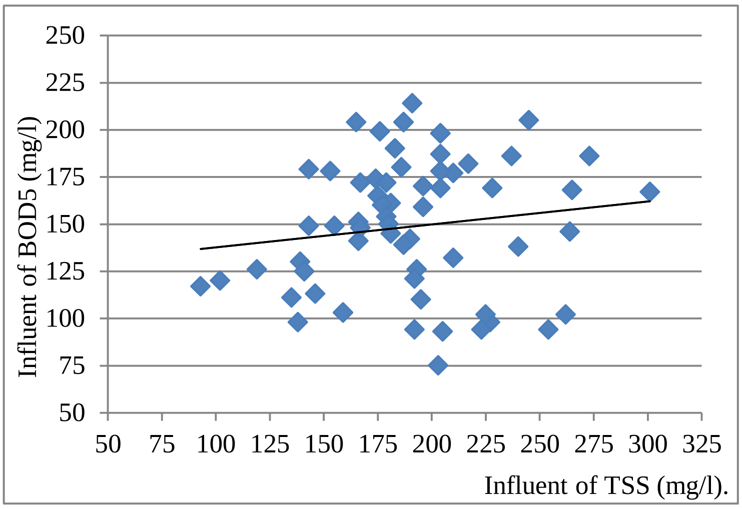

Figure (14): Regression of influent of BOD5 on the influent of TSS.

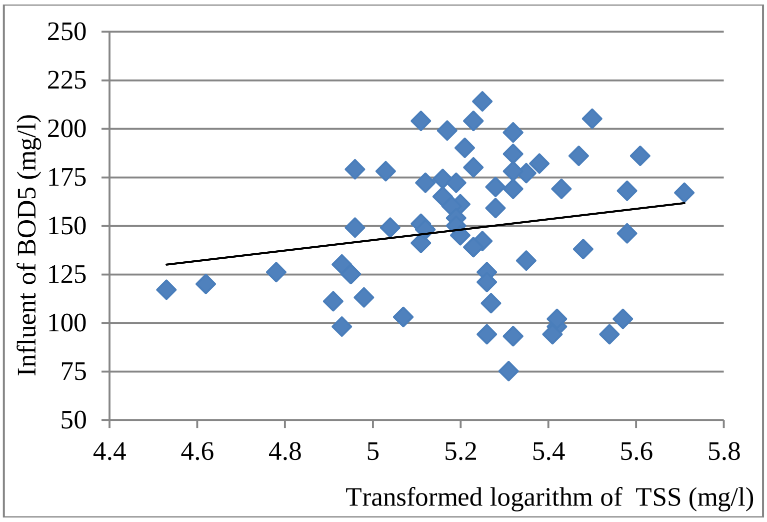

Figure (15): Regression of influent of BOD_5 on the logarithm transformed influent of TSS.

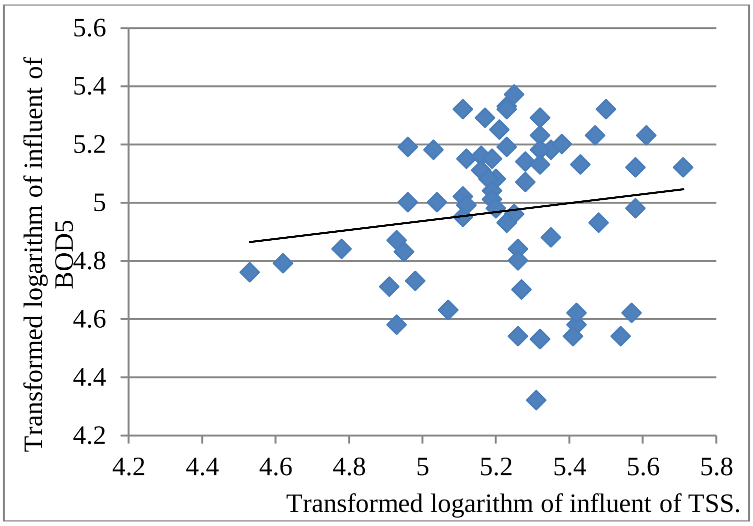

Figure (16): Regression of Transformed logarithm of influent of BOD_5 on the logarithm transformed influent of TSS.