Dirasat, Engineering Sciences, Volume **32**, No. **2**, 2005

A Statistical Approach for the Analysis and Forecasting of Water Quantity and Quality in King Talal Reservoir

*Ahmad Bani Hani, Ahmad Jamrah and Sulieman Tarawneh **

ABSTRACT

This study concentrated on presenting statistical methods for the use in studying and projecting quantity and quality variables in Zarka River, which is a main supplier of King Talal Reservoir (KTR). The quantity variable used was the flow, and the quality variables were (TSS, BOD5, COD, T-P, and T-N). The data collected for each variable covered a period of 156 months from the year 1988 till the end of the year 2000. The procedure used in analyzing the six variables includes auto, cross and vertical distance correlation at the point of direct discharge into KTR, which receives water from the mixing of Zarka River and the effluent of Samra wastewater treatment plant. Deterministic and stochastic forecasting of the six variables was used in finding the best model to be used in projection. The study results indicate that the exponential growth method yielded the least percentage of mean error for predicting TSS and T-P. The ARIMA model yielded the least percentage of mean error for BOD5 forecasting. None of the models satisfied the 10 percentage of mean error for BOD5 forecasting. Based on the percentage of mean error, the COD was best described by the quadratic method, the T-N was best predicted by the linear method, and the auto regression method succeeded most in predicting the flow.

KEYWORDS: water quality, King Talal Reservoir, deterministic forecasting, stochastic forecasting.

INTRODUCTION

Water consumption categories in Jordan are mainly divided into three types. These include domestic, industrial, and agricultural water consumption. These categories constitute 25%, 5%, and 70% of the total water consumption, respectively. It should be noted that the annual agricultural water consumption in 1995 was about 639.7 MCM (RSS, 1998).

Water is the main factor in agricultural production, and the main factor for extension in land reclamation. King Talal Reservoir (KTR) was constructed on the Zarka river in 1977 to collect and utilize surface water for irrigation. The capacity of the reservoir was 56 MCM, and was increased to 86 MCM in 1988 (Salameh, 1996). The height of the dam is 108m, and the catchment area covers 3157 km^2 with a population of 2.439 million in the year 1996. The reservoir, together with King Abdullah Canal, irrigates about 10,000 donums of agricultural farms (RSS, 1988-2001).

The inlet of water to KTR is mainly from Zarka Stream and Ramemen Wadi. The influent water to the reservoir is a mixture of rainfall and spring water mixed with domestic and industrial waste (WAJ, 1998). The percent of water drained from Al Zarka River to the lake of the dam is about 90%, while the contribution from Rumaimeen Valley is about 10%. The population of the catchment area is about half of the kingdom's population, which is distributed in Amman, Zarka, Sweileh, Rusaifeh, Baka'a and Jerash with several villages surrounding these cities (RSS, 1988-2001).

The difference in quality and quantity of water in KTR mandates the implementation of a continuous monitoring program for quantity and quality. Monitoring activities should aim at investigating whether the effluent from KTR is within the limits of the irrigation standards. Forecasting the water quantity and quality should help in establishing some precautional measures, which should assist in resolving the anticipated problems.

The aim of this research is to study the water quality and quantity in KTR. Trends of influent rate and quality

* Department of Civil Engineering, Faculty of Engineering and Technology, University of Jordan (1&2); and Department of Civil Engineering, Mu'tah University, Al-Karak, Jordan. Received on 29/5/2005 and Accepted for Publication on 8/11/2005.

will be established and analyzed as a function of time. A time series model will then be derived to represent the trends and to help making future projects. The water quality parameters that will be investigated and forecasted are the Total Suspended Solids (TSS), the Biochemical Oxygen Demand (BOD_5), the Chemical Oxygen Demand (COD), the Total Phosphorous (T-P), and the Total Nitrogen (T-N). The TSS is the most important physical characteristic of wastewater, which is composed of floating matter, settleable matter, and colloidal matter. Analytically, the total suspended solid content is defined as all matter that remains as residue upon filtration and then drying at 103 to 105 ^{o}C. The BOD_5 is a biochemical property, and the most widely used parameter of organic pollution applied to both wastewater and surface water. Determination of BOD_5 involves the measurements of the dissolved oxygen used by microorganisms in the biochemical oxidation of organic matter. The COD is a chemical property used to measure the organic and inorganic pollutants in wastewater. Finally, the total Nitrogen and Phosphorus are chemical characteristics and are essential nutrients to the growth of the protista and plants. Inspite of the fact that they are chemical characteristics, they are necessary to evaluate the treatability of wastewater by biological processes (Metcalf and Eddy, 2003).

These five water quality variables were analyzed and forecasted using several statistical methods. Any statistical analysis of data has to be based upon some assumed probability model for those data (Green and Margerison, 1977). If a series has shown some trend or persistent pattern in its variations for a long period of time in the past, it will be sensible to assume that such patterns or regularities will continue in the future (Chao, 1974). But one should take into consideration that forecasts may contain some errors. The magnitude of the forecasting errors experienced will depend upon the forecasting system used (Montogomery and Johnson, 1976).

Forecasting of environmental quality parameters can be justified because data is frequently expensive to accumulate, and correlations between the water constituents may help the filling of missing data or the identification of outlier data. Additionally, forecasting of quality parameters may prove useful in the prevention of any deterioration in environmental quality, because early warnings may provide the opportunity for controlling the problem at a lower cost before the problem magnifies (Bean and Rover, 1998).

METHODOLOGY

Relevant data concerning the flow and quality parameters of water entering KTR were obtained from the Ministry of Water and Irrigation. The data collected cover a period of 13 years starting January 1988 and ending December 2000. The collected data were reported on a monthly basis. The thirteen-year period was chosen to make sure that the collected data reflect the monthly and seasonal variations. The flow and quality data were measured in Zarka River, at the point of direct discharge into KTR, which receives water from the mixing of Zarka River and the effluent of Samra wastewater treatment plant. In addition to the flow rate, the data includes selected physical and chemical water quality parameters such as TSS, BOD_5, COD, T-P, T-N.

Analysis of the data started with a graphic presentation in a scatter diagram. The graphic presentation of statistical data gives a pictorial effect, makes the data easy to understand and grasp, and shows any prevailing trend that may be present and the direction in which the trend may change (Pillai and Bagavalthi, 1997).

The next step in the analysis of data was the determination of outliers. The time series plot contains fluctuations that can be identified, explained, and measured by simplifying assumptions. The fluctuations that appear in the plot are due to four basic types of variations: secular trend, seasonal, cyclic and irregular variations (Sakakini, 2001). The variation in the data will lead to outliers. In general, an outlier is an observation that is far from the rest of the data.

The missing data were determined and estimated by averaging the data of the same month. This was followed by a plot of a scatter diagram between each of the flow and quality parameters versus time. A box diagram was then plotted using “Minitab 13” for the original data and residuals reflecting the seasonal period. The outliers are shown in those two box diagrams as points. The outliers were then determined in the two cases; and the original data were checked to confirm whether they were actual outliers or not. The actual outliers were not changed; and others, which were attributed to human errors and negligence, were adjusted to the average monthly value. A new scatter diagram was plotted for the new adjusted data.

After adjusting the data, normality tests were carried out on the data using four methods:

1) Weibull distribution model histogram, which was determined by calculating the average monthly value for each of the flow and quality parameters. A plot of the Weibull Distribution Model Histogram was obtained using "Minitab 13" to determine whether the histogram is skewed to left or right.
2) Coefficient of Variation (COV), which was determined by dividing the data into four quarters, then finding the coefficient of variation for each quarter. The value of the coefficient of variation was compared with the number 1, and it was concluded that a value of less than $|1|$ is not skewed; and that otherwise the values are skewed either left or right.
3) The transformed Kurtosis Coefficient, where data were also divided into four quarters, then the transformed Kurtosis Coefficient was found for each quarter. The value of the Kurtosis Coefficient was compared with the number $|1|$, and it was concluded that a value of less than $|1|$ results in what is called Mesokurtic data and Leptokurtic data for the positive values, and Platykutric data for the negative values, and
4) The Shapiro-Wilk test, where the data were divided into four quarters. The Shapiro-Wilk value was found for each quarter and was compared with a standard value. Values larger than the standard one indicate normally distributed data, while values smaller than the standard one indicate skewed data. It should be noted that ARIMA's model variables can only be calculated for normal data, and that skewed data should undergo a lognormal transformation and should be normalized before the calculation of ARIMA's model parameters.

Data were then used for the determination of ARIMA's parameters; which are Autoregressive (AR), Moving Average (MA), and integrated (I) parameter. The Autoregressive (AR) value, expressed by the item "p", was determined through a plot of the autocorrelation function for each of the flow and quality parameters using Minitab13 software, and taking into consideration that the value of p should not exceed unity in surface water forecasting. This is due to the fact that in small rivers (such as Zarka River) the water characteristics do not need more than a few days to dilute, so that the correlation does not exceed one month (Viessman and Lewis, 1996). The MA value, expressed by the item "q", was determined by drawing the moving average graphs with different q values for each of the flow and quality parameters using Minitab13 software. The correct value of q was determined when the graph showed that the trend was minimized; and when the following graphs were approximately having the same trend as the previous one. The value of I, expressed by the item "d", was determined by plotting four figures using Minitab13 software. The figures were plotted for the original data, detrended data, seasonally adjusted data, and seasonally adjusted and detrended data. Differentiation was carried out when these four figures show significant difference; whereby a diagnostic model diagram for the ARIMA model was drawn using Minitab13 software with an I value of 0 or 2. This is due to the fact that there are only two (summer and winter) seasons in Jordan. It was then concluded that the diagnostic model that has less residual has the correct I value that should be used.

Forecasting of future values was carried out after finding the ARIMA parameters. The method of forecasting was divided into two parts; the deterministic and the stochastic forecasting. The deterministic forecasting is subdivided into four different methods; which are linear regression, quadratic regression, exponential growth regression, and single exponential smoothing models. The acceptance of a model was determined based on whether the model has an error that is less than or equal to 10% of the real data. Additionally, three methods were used in stochastic forecasting; AR, MA; and the ARIMA model. A model was also accepted for stochastic forecasting when an error of less than or equal to 10% is achieved. The results of the error in each model were reported as the percentage of error; which were calculated based on the difference in mean between the real data and the forecasted one. The stochastic forecasting model was concluded to be appropriate if it satisfies the maximum accepted error, otherwise the deterministic model with the lowest error was concluded to be appropriate.

RESULTS AND DISCUSSION

The results of the study regarding analysis and forecasting of all studied water quantity and quality are too extensive to be presented here. The following is a demonstration of the study outcome for Total Phosphorus (T-P). A similar approach was employed to obtain results for other parameters covered by this study.

Detection of Missing Data and Outliers

The T-P data collected from the Ministry of Water and Irrigation did not contain any missing data, so the second step is to find the outliers. Data were plotted in a scatter diagram as shown in Figure (1) so that outliers will be clearly observed. These data, which contain 156 observations from January 1988 till December 2000, were noted to have approximately two outliers in the months of March 1992 and December 1999. Investigation of these outliers revealed that the rainfall in these months was relatively high; resulting in substantial dilution of phosphorous in Zarka river. As a result, these two data points were assumed to be outliers and were adjusted in a way similar to that of the missing data where the average monthly value was used.

Outliers for the seasonal trends for the original and the residual data are shown in Figure (2). It can be clearly seen from the charts presented in the figure that there are two outliers in both the original data and the residual data in the seasonal condition. Figure (2) also shows the variation in the data for the same month, and that the variation was the highest in December, and was the lowest in October. Another two outliers were found in the seasonal drawings for the months of March 1992 and December 1999. These two outliers were observed in the original data, so no adjustment was carried out.

After adjusting the outliers, the new adjusted data were plotted and shown in Figure (3). The figure shows that some outliers remain in the data due to the fact that these original data will eventually influence the results of statistical treatment of data. However, comparison of Figures (1) and (3) shows that the original data and the adjusted data have the same trend. This indicates that the effect of the outliers on the data is negligible.

Normality of Data

Normality of data was checked out using four different approaches; Weibull's distribution model, coefficient of variance, Kurtosis coefficient, and Shapiro-Wilk test. Lognormal transformation of data was carried out whenever needed to ensure that the modeled data are normally distributed.

Weibull's Distribution Model Histogram

Data were transformed to the average monthly values for the Total Phosphorous (T-P) variable. This resulted in a total of twelve data points. The Weibull's distribution histogram, which indicates the normality of seasonal data, was plotted for these twelve data points as shown in Figure (4). It can be observed from Figure (4) that the data of T-P shows reasonable normality with slight skewness to the left and bulked to the right, and that the overall indication of the graph is normal.

Coefficient of Variation (COV), Preliminary Test:

A preliminary test was carried out whereby the data were divided into four quarters; each quarter consisted of 39 data points. Table (1) provides the value of the mean, variance, standard deviation, and the coefficient of variation for the T-P variable.

$$COV = \frac{Std.Dev.}{Mean}$$

It can be observed from the table that the value of the coefficient of variation for each quarter is less than unity, indicating that each quarter of the data is slightly skewed (either to the right or left). Accordingly, all of the data of the T-P variable have less skewness than each of the four quarters individually. As a result, it can be concluded that skewness of this variable can be ignored for further analysis.

Transformed Kurtosis Coefficient (peakedness), Vertical Test

The data were divided into four quarters as shown in Table (2), which provides the values of the transformed Kurtosis coefficient for each quarter and along with needed statistical indicators (mean, variance, standard error, and the value of K) for the calculation of the transformed Kurtosis coefficient. It can be observed from Table (2) that the data in the first, second, and third quarters were normally distributed (mesokurtic), and that the fourth quarter was fairly leptokurtic. As a result, it can be concluded that the total data of this variable can be assumed normally distributed (mesokurtic) for further analysis.

Shapiro-Wilk Test

The data were divided into four quarters and the Shapiro–Wilk statistical values were calculated for each quarter. These values were compared with the five-percent critical value for a sample size of 20. The data was considered to show evidence of normality when the Shapiro-Wilk test was greater than the five-percent critical value. The results of this analysis are shown in Table (3).

It can be clearly observed from Table (3) that the data in the first three quarters suggest the prevelance of normality of data, while the data of the fourth quarter suggest prevelance of a slight non–normality. Overall, the data of the T-P variable show an acceptable indication of normality according to Shapiro-Wilk test.

Order of AR

Analyses and projections of water quality parameters in water bodies similar to King Talal Dam may require that the value of AR, which is expressed by the term p, shall not be more than unity. This is due to the fact that the autocorrelation for a particle of T-P does not last more than one month before it gets analyzed (Viessman and Lewis, 1996). Figure (5) shows a plot of the autocorrelation function for the total phosphorous versus time lag. It clearly indicates that the value of Autoregressive or time lag (AR or p) is 4. The Figure also shows a cyclic behavior where trends are repeated every twelve months. However, the value of AR will be assumed to be unity for further analysis of the T-P variable. An AR value of unity will indicate that T-P concentrations in any month are related only to the previous month.

Order of MA

Determination of the order of the q involves plots of the variable of interest versus its moving average with p. Figure (6) shows plots of T-P concentration in mg/L versus time. This figure is intended to show the change between the real data of the variable T-P and its moving average with different lengths of p. The figure shows a p–length of up to six. Investigation of the figure indicates that a p-length value of 4 is appropriate. Accordingly, the order of the MA was concluded to be MA(4) for further analysis.

Order of (I)

The last coefficient of ARIMA's parameters is I, which is generally expressed by the term d. It should be noted that determination of the order of I requires that the data be differentiated when there is a trend or shift or seasonality in the data. Figure (7) shows a plot of T-P variable in mg/L versus time showing the component analysis of original data, detrended data, seasonally adjusted data, and seasonally adjusted and detrended data for the determination of the order of I. The figure shows clearly that there is a significant difference between the original and detrended data, while seasonality has very little effect on the data. Accordingly, the detrended effect should be taken into consideration in a further analysis of the data.

Water quality and quantity data in Jordan can be influenced by two seasons; summer and winter. Accordingly, the order of the integrated model can be assigned a d-value of zero when these data do not show any influence due to seasonality; and a d-value of 2 should be assigned when the influence of seasonality becomes apparent in the data. Figures (8 and 9) provide ARIMA model diagnostics for ARIMA (1,0,4) and (1,2,4). It is shown in the two graphs that the residual in Figure (9) is less than that in Figure (8). As a result, the coefficients of ARIMA that will be used are (1,2,4).

Forecasting Analysis

Two general methods; the deterministic and the stochastic, were used in the forecasting of the flow and the quality parameters. The collected data were divided into two parts; the first part consisted of 90% of the data and was used for analysis and prediction of future data. The second part consisted of the remaining 10% of the data, and was compared to the predicted data in terms of both value and mean. The stochastic method of forecasting is preferable over the deterministic method of forecasting if the error in both is less than 10%.

Deterministic Forecasting

The deterministic approach of forecasting utilizes regression, which involves using a regression or a least-square equation to predict a dependent variable from the values of an independent variable. The type of the least-square regression equation used is determined by the type of relationship that prevails between the two (dependent and independent) variables. It is a general practice to fit several types of lines to the data and choose the one that result in the best fit.

Figures (10, 11, and 12) show respectively, linear regression, quadratic regression, and exponential growth regression fit lines for the T-P variable. These lines were applied for 90% of the original data. The figures also show the projections of regression lines to forecast the last 10% of the data.

It can be observed from Figure (10, 11, and 12) that these three types of regression resulted in an increasing relationship. Table (4) shows the predicted values for the last 10% of data as compared to the actual data for these three regression relations.

Comparison of the actual values with the predicted ones indicates a prediction error of 13.9%, 18.6%, and 10.1% for the three methods of forecasting. This indicates that these regression methods did not satisfy the error requirement for the forecasting of T-P.

The single exponential smoothing trend model is another deterministic approach of forecasting that can only be used in its basic approach for non-seasonal time series showing no trend. When original data show trend or seasonality, then it is generally recommended to eliminate this trend prior to applying the single exponential smoothing regression. The regression of the additive single exponential smoothing trend model is shown in Figure (13).

Table (5) shows the exponential smoothing prediction for the next 10% of the data. The predicted upper and lower values along with the actual data are shown. The forecasted value of total phosphorous was taken as the average between the upper and lower values.

Comparison of the actual values with the predicted ones indicates that the simple exponential smoothing model resulted in a prediction error of 30.7%. This indicates that this method did not satisfy the error requirement for the forecasting of T-P.

Stochastic Forecasting

The ARIMA model is a stochastic forecasting model that has a classical and a linear time series analysis dealing with a non-stationary and random behavior. It handles these processes within the framework of the classical time series analysis through forming the differences in order to get a stationary process. An ARIMA process is made up of sums of autoregressive and moving-average components, and the integration part.

The results of stochastic forecasting of T-P using AR(1) are shown in Table (6). The table shows the AR(1) prediction for the last 10% of data along with the original data. The forecasted data are reported as the average of the upper and lower values.

Comparison of the actual values with the predicted ones indicates that the AR(1) resulted in a prediction error of 55.6%. This indicates that this method by itself did not satisfy the error requirement for the forecasting of T-P.

Figures (14) shows the MA(4) for the T-P variable. This line was applied for 90% of the original data. The figures also show the projections of regression lines, MA(4), to forecast the last 10% of the data.

In addition, results of the stochastic forecasting of T-P using MA(4) are shown in Table (7). The table shows the MA(4) prediction for the last 10% of data along with the original data. The forecasted data are reported as the average of the upper and lower values.

Comparison of the actual values with the predicted ones indicates that the moving average model of order four; MA(4), resulted in a prediction error of 8.5%. This indicates that this method satisfied the error requirement for the forecasting of T-P.

The results of stochastic forecasting of T-P using autoregressive integrated moving average model of the order ARIMA (1,2,4) are shown in Table (8). The table shows the ARIMA(1,2,4) prediction for the last 10% of data along with the original data. The forecasted data are reported as the average of the upper and lower values.

Comparison of the actual values with the predicted ones indicates that the autoregressive integrated moving average model of order ARIMA(1,2,4) resulted in a prediction error of 23.2%. This indicates that this method did not satisfy the error requirement for the forecasting of total phosphorous.

The results of the percentage of mean error for all the variables analyzed in this study are summarized in Table (9), which provides a summary of the models used in the prediction and the percentage of mean error for each variable.

Finally, it should be noted that the accuracy of forecasting carried out using time series analysis increases with the available data as this leads to emphasizing the trend and seasonal effect. Additionally, it is vital that enough information be available about any strange reading. This will facilitate the recognition of outliers and justify their presence, and consequently, result in a decrease in the randomness of data.

CONCLUSIONS

Original data regarding the flow and some water quality parameters included some missing data. The T-P data did not include any missing values. Data on the TSS variable in Zarka River was missing in December 1999. Samra's effluent had nine missing data points for the different variables and all were in the year 2000. Missing data were estimated by taking the average of the data for the same months.

The Weibull distribution was used in plotting the histogram of the normality test for each variable because of its simplicity and due to the fact that it does not have a 100% probability. The Kurtosis and Shapiro–Wilk tests evidently resulted in almost identical results in most data quarters. This indicates that the Shapiro-Wilk test is a good method for testing the normality in the vertical direction.

The five variables (TSS, BOD_5, COD, T-P, and T-N) had an approximately normal distribution, while the flow data showed some abnormality due to the variation in data with a number of abnormal observations (9 in total). These abnormal observations were kept and used for further analyses due to the fact that they were real data. A lognormal test was carried out on the flow data and showed a significant improvement in the normality of the data. The Weibull distribution model was skewed to the right in the distribution of flow data, the coefficient of variation was more than one, and the Kurtosis coefficient reached a value of 20.

The percentage of mean error was calculated after analyzing the data and forecasting 10% of the original data. For the TSS variable, it was shown that the least percentage of mean error was in the exponential growth method, which is a deterministic forecasting method, with an error of 1.7%. The least percentage of mean error in the stochastic models for TSS variable was found to be 5.4% using AR(1). The other stochastic model that yielded acceptable forecasting results for TSS variable was the ARIMA (1,0,4), with a mean error of 8.2%.

None of the forecasting methods resulted in an error of less than 10% for the prediction of the BOD_5. However, the least percentage of mean error for BOD_5 variable was 16.1%, and resulted from the use of ARMA (1,3) model. Similarly, none of the stochastic models for COD variable satisfied the 10% error. Additionally, in the deterministic modeling for COD variable, the quadratic method resulted in an error of 3.8% constituting the least percentage of mean error.

For the T-P variable, the only method that satisfied the 10% error was the moving average MA(4), which resulted in a percentage of mean error that equals to 8.5%. In forecasting the T-N variable, many forecasting methods satisfied the 10% of the mean error, and the method that resulted in the least error was the linear method with a mean error of 3.3%. However, the best model to be used in forecasting the T-N variable is ARIMA(1,2,5), which gave 4.8% of mean error.

Forecasting of the flow variable was carried out for log (Q) variable. Many models have satisfied the 10% mean error. However, it was not possible to estimate the model ARIMA (1,2,6) due to the fact that the computer software (Minitab13) can only carry out this estimation when the coefficients of ARIMA model (p,d,q) are less than 5. However, the AR(1) resulted in an error of 4.9%; constituting the least for the forecasting of log(Q). It should be mentioned that the ARIMA model has satisfied the forecasting model for most of the variables. The least amount of mean error resulting from ARIMA model was when calculating the percentage of mean error for ARIMA(1,2,5), which was 4.8%.

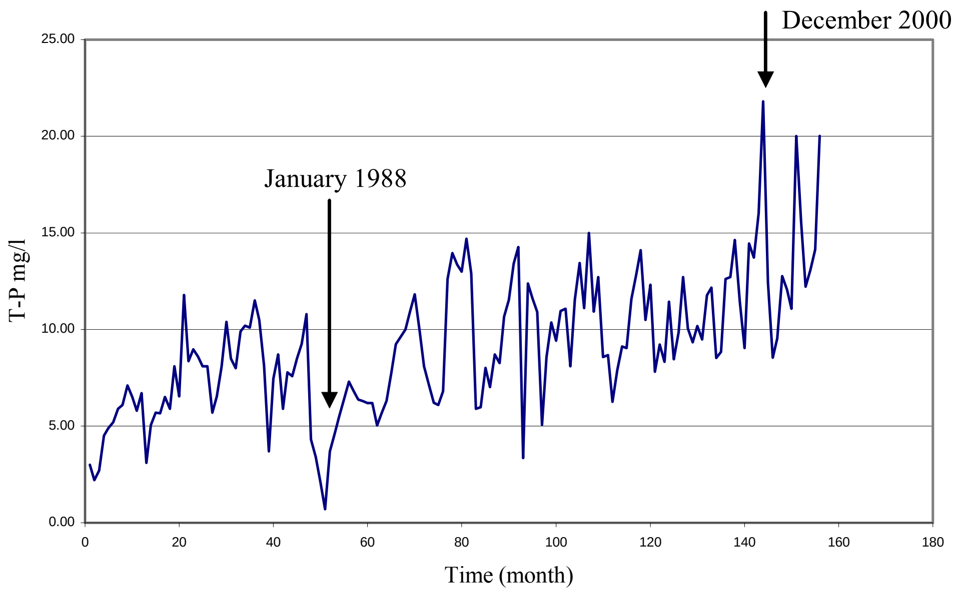


Figure (1): A plot of original data of concentration (T-P) versus time. Concentrations values are in mg/L.

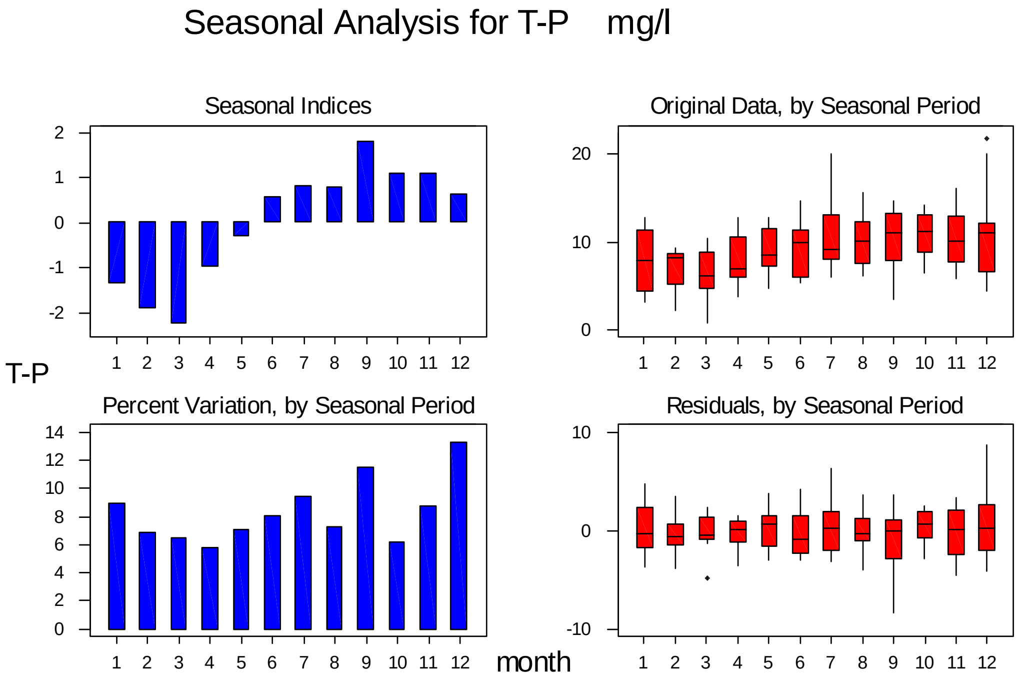


Figure (2): Seasonal analyses for T-P concentration.

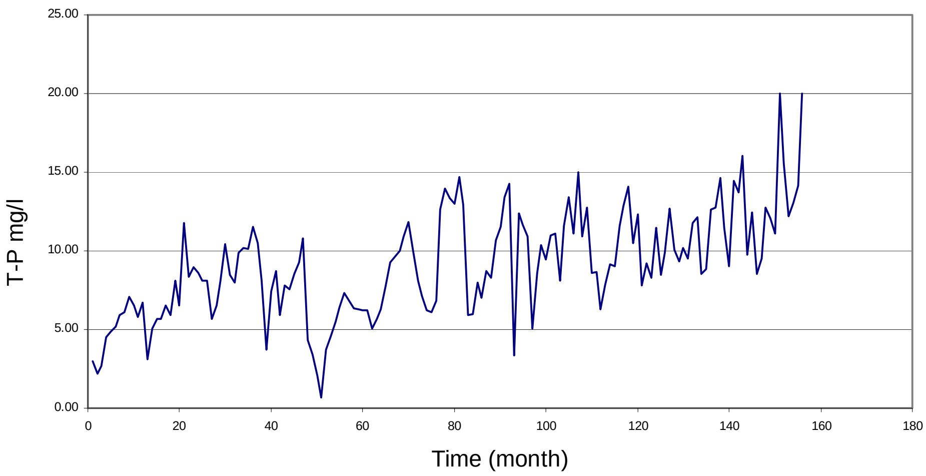


Figure (3): A plot of the adjusted data of T-P concentration versus time. Concentrations values are in mg/L.

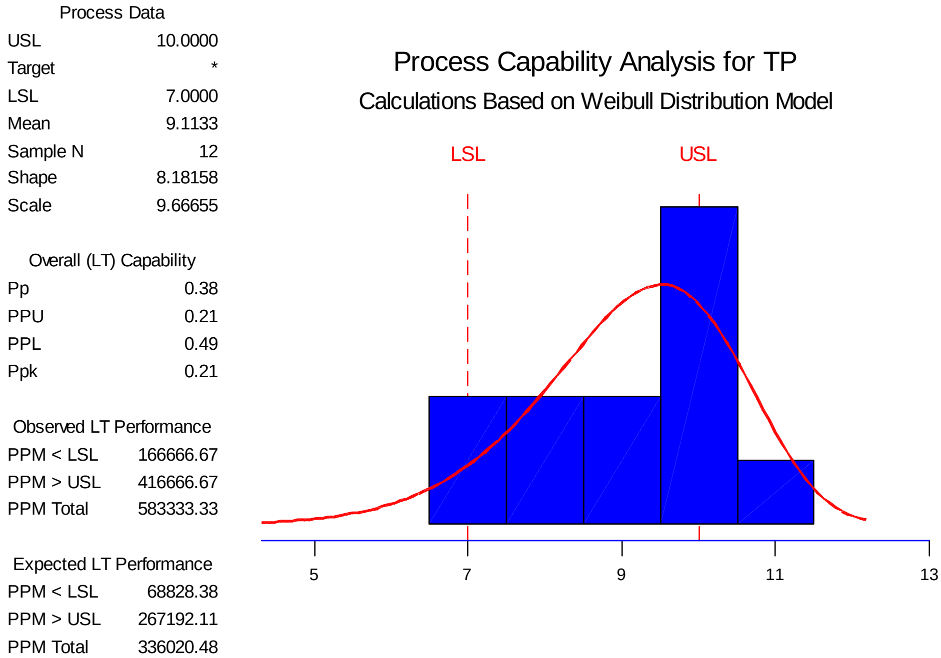


Figure (4): Weibull distribution model histogram for the T-P variable.

Table (1): The coefficient of variation for theT-P variable.

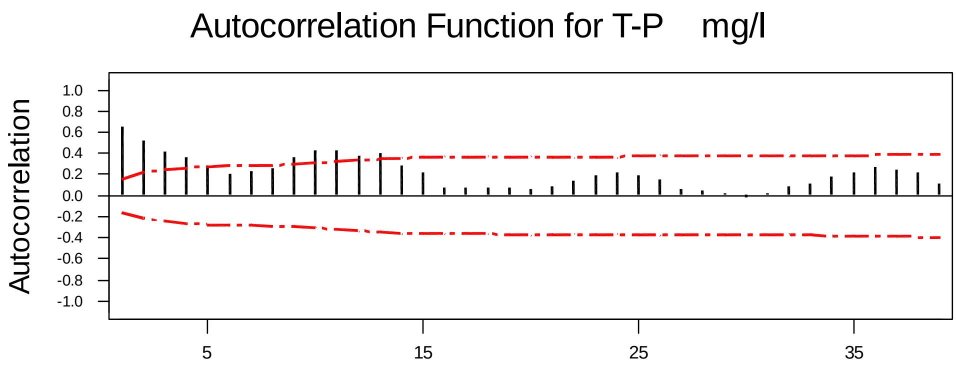
	MEAN	**VARIANCE**	**ST. DEV. (S)**	**C.O.V.**
T-P (1st Quarter)	7.0	5.9	2.4	0.3
T-P (2nd Quarter)	7.3	7.6	2.8	0.4
T-P (3rd Quarter)	10.2	7.9	2.8	0.3
T-P (4th Quarter)	12.1	10.8	3.3	0.3

Table (2): The transformed Kurtosis Coefficient for the T-P variable.

	MEAN	VARIANCE	ST. DEV. (S)	K	Kurtosis Coeff. C'_K
T-P (1st Quarter)	7.0	5.9	2.4	93.3	-0.3
T-P (2nd Quarter)	7.3	7.6	2.8	208.2	0.6
T-P (3rd Quarter)	10.2	7.9	2.8	174.7	-0.2
T-P (4th Quarter)	12.1	10.8	3.3	555.7	1.8

Table (3): The Kurtosis Coefficient for the T-P variable.

	Weighted sum of extreme observation (b)	Standard Deviation (S)	Shapiro – Wilk statistics values (W)	Shapiro – Wilk critical value
T-P (1st Quarter)	14.78	2.43	0.971	0.939
T-P (2nd Quarter)	16.83	2.75	0.984	0.939
T-P (3rd Quarter)	17.09	2.82	0.968	0.939
T-P (4th Quarter)	19.12	3.28	0.891	0.939



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.65	8.13	67.40	13	0.40	2.28	333.06	25	0.19	1.02	385.62	37	0.24	1.23	437.44
2	0.52	4.81	111.18	14	0.28	1.54	346.38	26	0.15	0.78	389.78	38	0.22	1.10	447.39
3	0.41	3.32	138.49	15	0.21	1.17	354.31	27	0.06	0.31	390.43	39	0.11	0.54	449.89
4	0.36	2.72	159.52	16	0.07	0.40	355.27	28	0.05	0.27	390.93				
5	0.28	2.00	172.05	17	0.08	0.42	356.35	29	0.02	0.09	390.98				
6	0.20	1.41	178.67	18	0.07	0.37	357.20	30	-0.01	-0.08	391.03				
7	0.23	1.58	187.25	19	0.08	0.42	358.26	31	0.02	0.13	391.15				
8	0.25	1.73	197.90	20	0.05	0.30	358.81	32	0.09	0.48	392.83				
9	0.37	2.47	220.49	21	0.09	0.49	360.29	33	0.11	0.57	395.17				
10	0.43	2.79	251.93	22	0.14	0.76	363.93	34	0.17	0.90	401.24				
11	0.42	2.61	282.31	23	0.19	1.00	370.33	35	0.22	1.13	410.90				
12	0.37	2.20	305.97	24	0.21	1.13	378.66	36	0.26	1.36	425.32				

Figure (5): Autocorrelation function for the T-P variable.

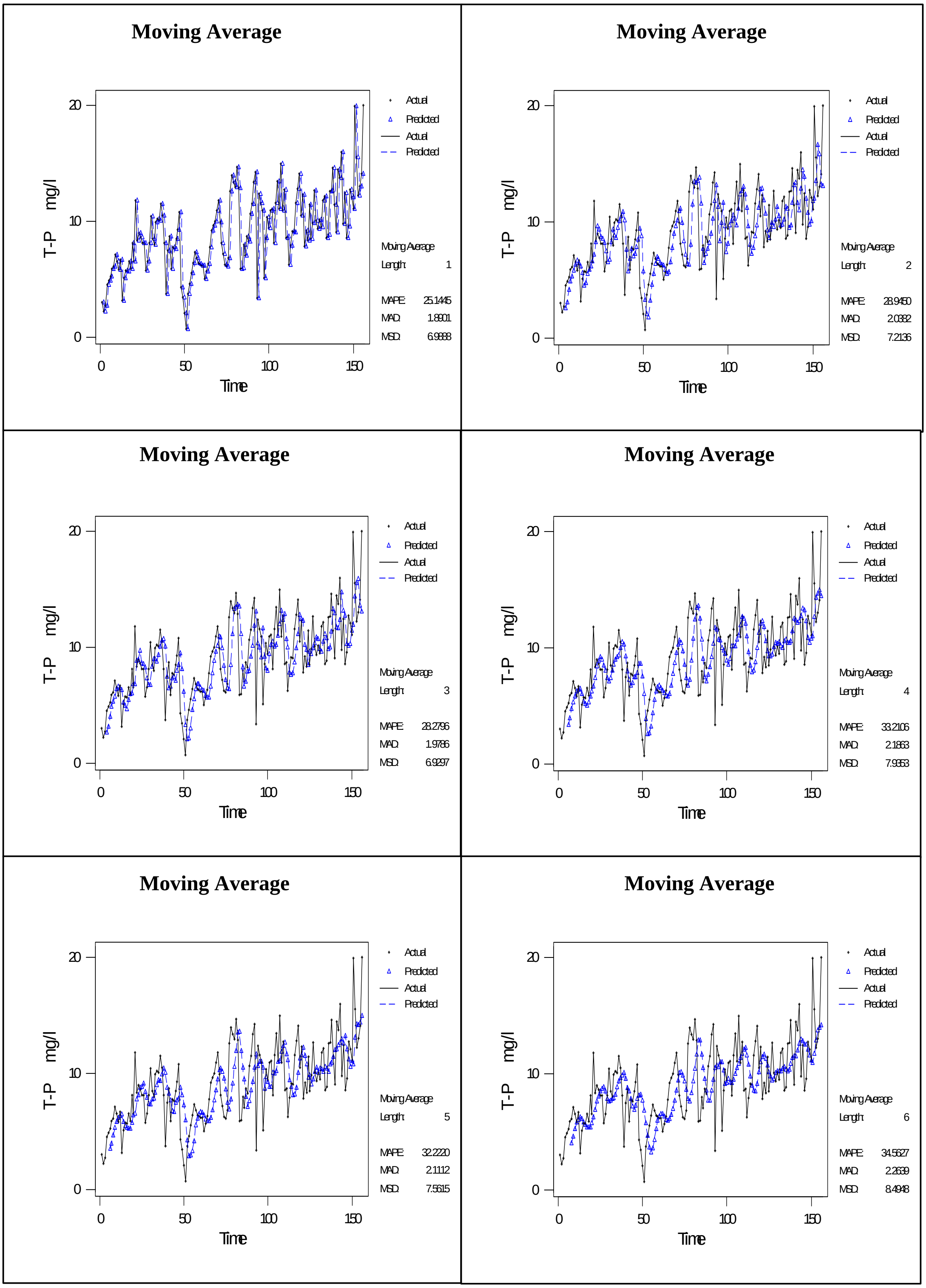


Figure (6): Plots of the moving average of variable T-P with different values of p.

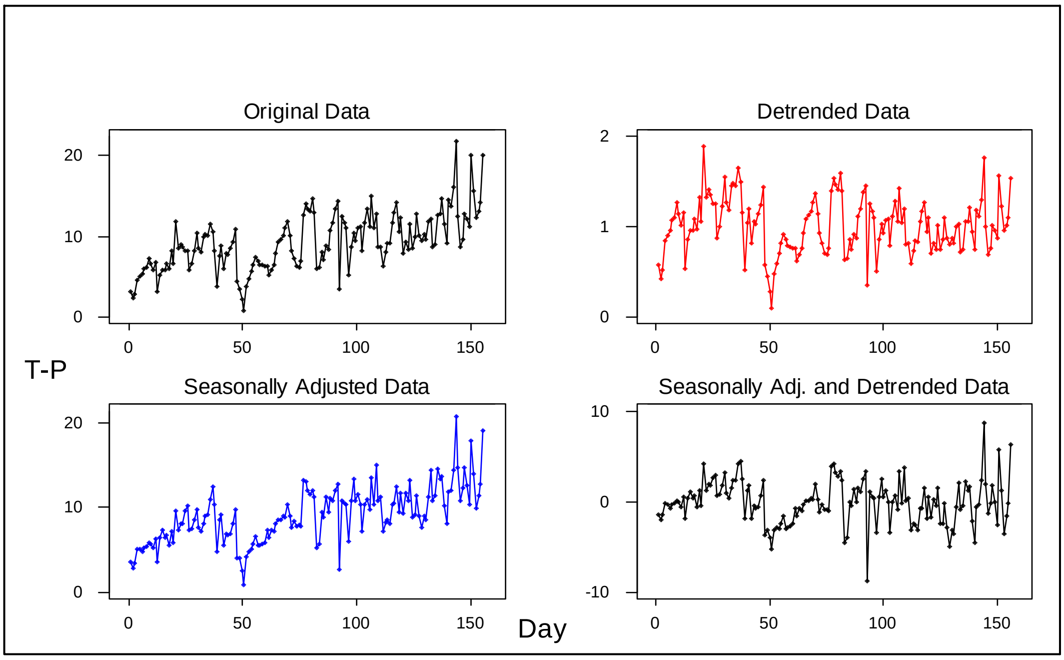


Figure (7): Plots of T-P variable in mg/L versus time showing the component analysis of original data, detrended data, seasonally adjusted data, and seasonally adjusted and detrended data for the determination of the order of I.

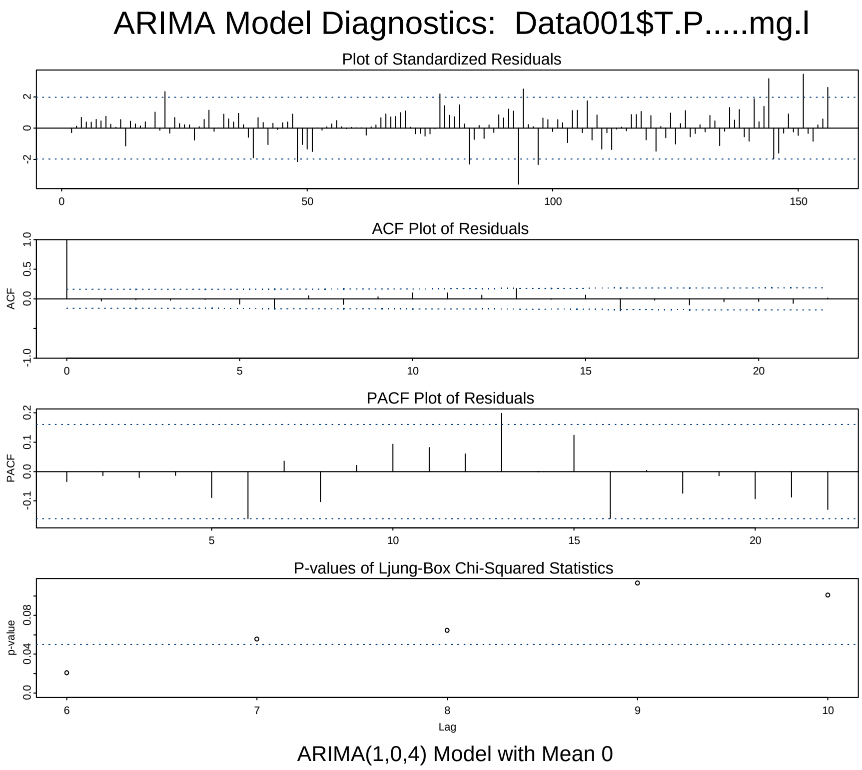


Figure (8): ARIMA (1,0,4) model diagnostic for the T-P variable.

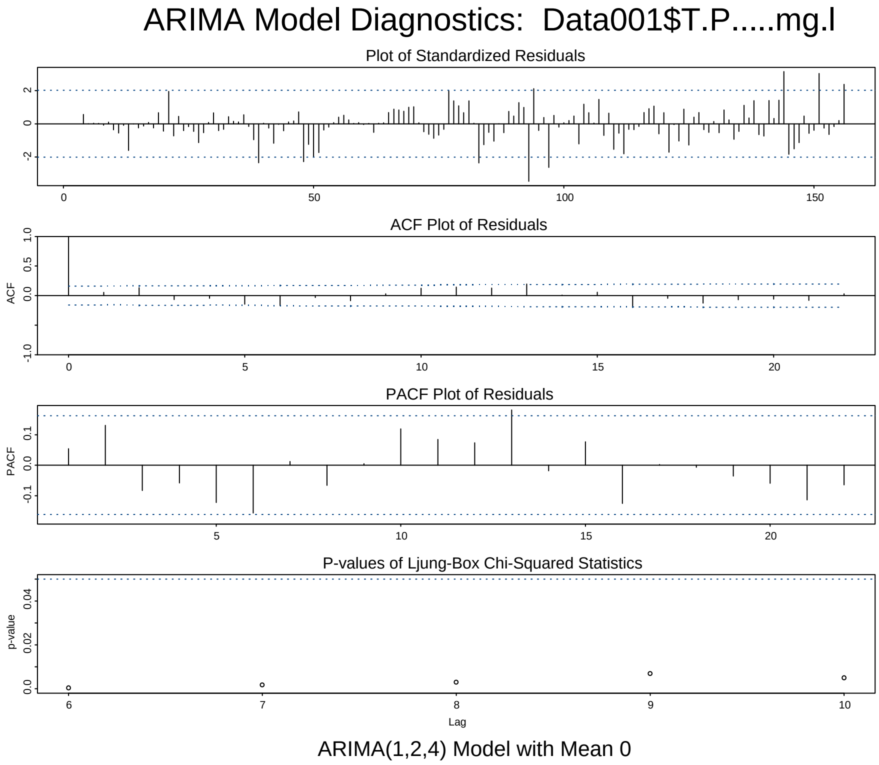


Figure (9): ARIMA (1,2,4) model diagnostic for the T-P variable.

Table (4): Comparison of actual T-P data to forecast outcome using linear, quadratic, and exponential growth regression lines.

Row	Period (months)	Linear Forecast (mg/L)	Quadratic Forecast (mg/L)	Exponential Growth Forecast (mg/L)	Actual Data (mg/L)
1	141	11.51	11.16	11.72	14.45
2	142	11.55	11.19	11.79	13.73
3	143	11.59	11.21	11.85	16.03
4	144	11.63	11.24	11.92	9.78
5	145	11.67	11.27	11.99	12.43
6	146	11.72	11.29	12.05	8.55
7	147	11.76	11.32	12.12	9.53
8	148	11.80	11.34	12.19	12.76
9	149	11.84	11.37	12.26	12.07
10	150	11.88	11.39	12.33	11.08
11	151	11.92	11.42	12.39	19.99
12	152	11.97	11.44	12.46	15.57
13	153	12.01	11.47	12.53	12.22
14	154	12.05	11.49	12.60	13.07
15	155	12.09	11.52	12.67	14.14
16	156	12.13	11.54	12.75	20.02

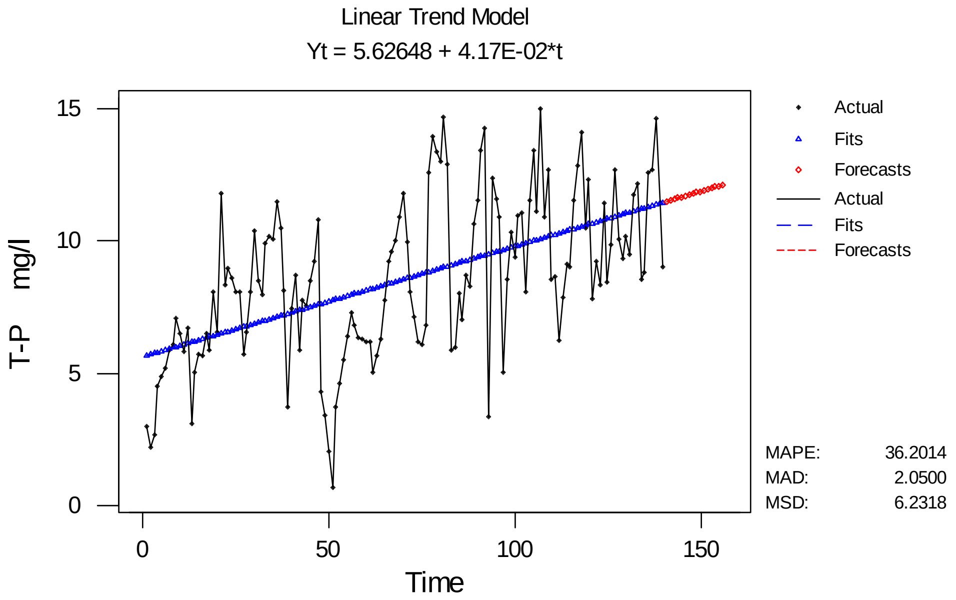


Figure (10): A plot of original data of T-P variable along with linear regression line for 90% of data and forecasting for the remaining 10% of data.

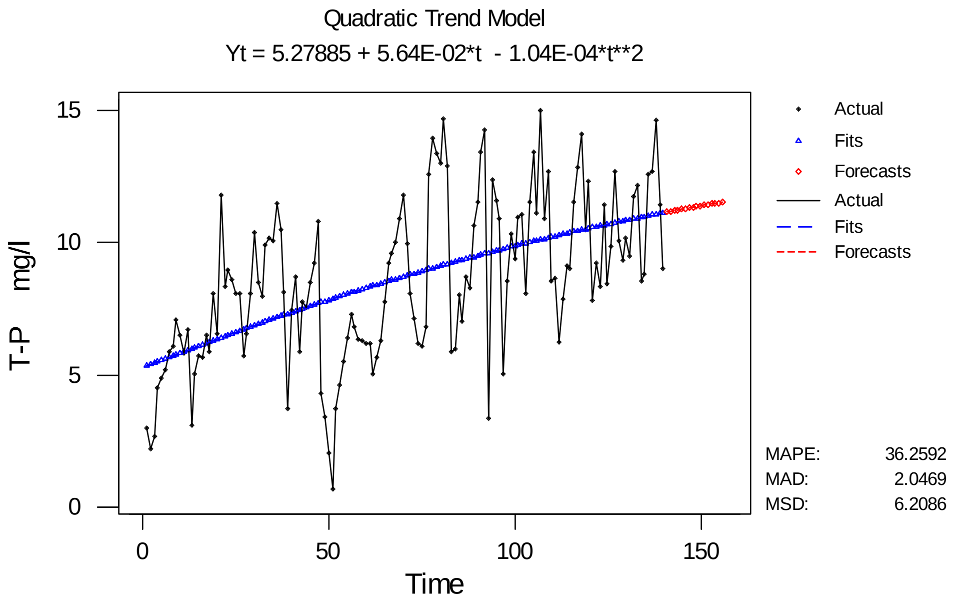


Figure (11): A plot of original data of T-P variable along with quadratic regression line for 90% of data and forecasting for the remaining 10% of data.

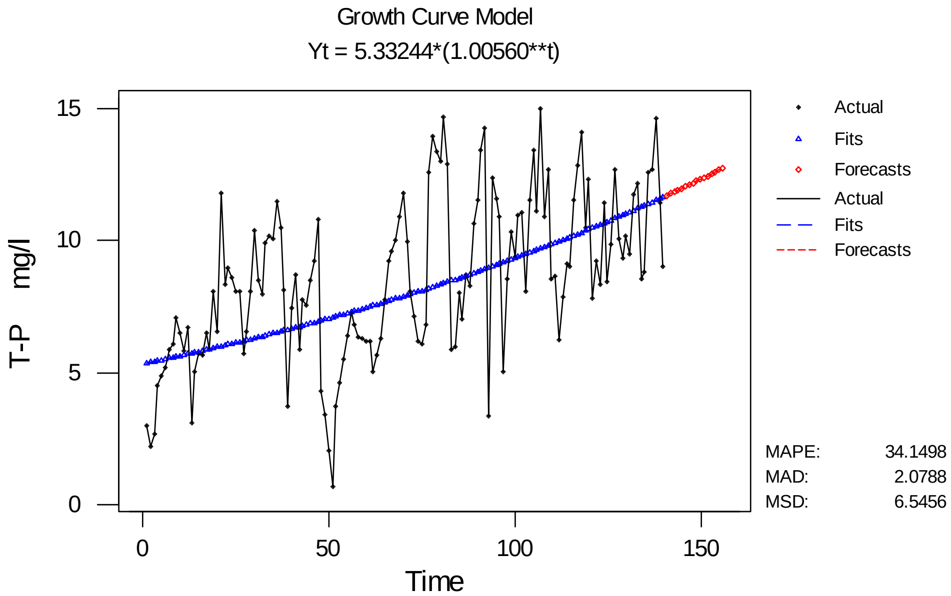


Figure (12): A plot of original data of T-P variable along with growth curve regression line for 90% of data and forecasting for the remaining 10% of data.

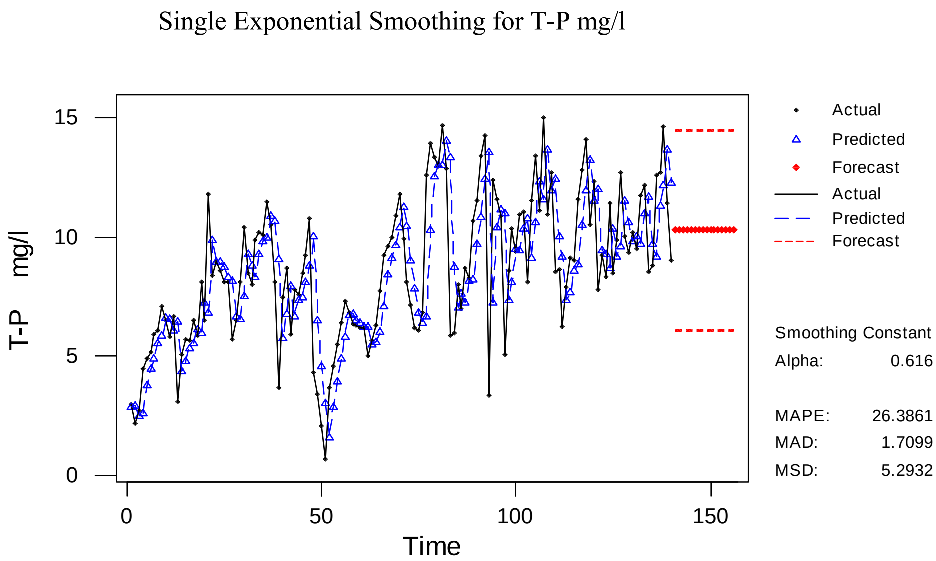


Figure (13): A plot of original data of T-P variable along with single exponential smoothing regression line for 90% of data and forecasting for the remaining 10% of data.

Table (5): Comparison of actual T-P data to forecast outcome using single exponential smoothing.

Row	Period (month)	Forecast mg/l	Lower mg/l	Upper mg/l	Actual mg/l
1	141	10.30	6.11	14.49	14.45
2	142	10.30	6.11	14.49	13.73
3	143	10.30	6.11	14.49	16.03
4	144	10.30	6.11	14.49	9.78
5	145	10.30	6.11	14.49	12.43
6	146	10.30	6.11	14.49	8.55
7	147	10.30	6.11	14.49	9.53
8	148	10.30	6.11	14.49	12.76
9	149	10.30	6.11	14.49	12.07
10	150	10.30	6.11	14.49	11.08
11	151	10.30	6.11	14.49	19.99
12	152	10.30	6.11	14.49	15.57
13	153	10.30	6.11	14.49	12.22
14	154	10.30	6.11	14.49	13.07
15	155	10.30	6.11	14.49	14.14
16	156	10.30	6.11	14.49	20.02

Table (6): Comparison of actual T-P data to forecast outcome using autoregressive model of order one; AR(1).

Row	Period (month)	Forecast mg/l	Lower mg/l	Upper mg/l	Actual mg/l
1	141	8.86	4.47	12.25	14.45
2	142	8.74	3.44	14.05	13.73
3	143	8.66	2.99	14.34	16.03
4	144	8.61	2.77	14.45	9.78
5	145	8.57	2.66	14.49	12.43
6	146	8.55	2.60	14.49	8.55
7	147	8.53	2.56	14.49	9.53
8	148	8.52	2.55	14.49	12.76
9	149	8.51	2.53	14.48	12.07
10	150	8.50	2.53	14.48	11.08
11	151	8.50	2.52	14.48	19.99
12	152	8.50	2.52	14.47	15.57
13	153	8.50	2.52	14.47	12.22
14	154	8.49	2.52	14.47	13.07
15	155	8.49	2.52	14.47	14.14
16	156	8.49	2.52	14.47	20.02

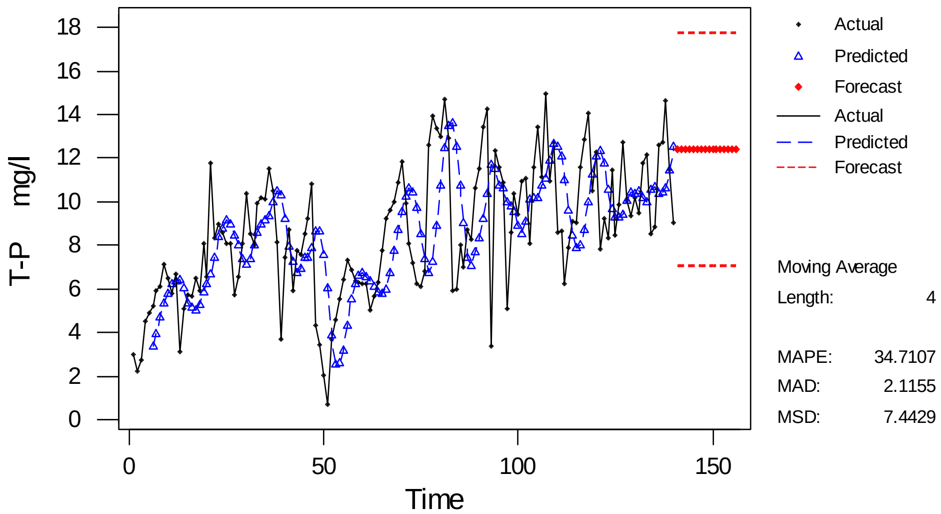


Figure (14): A plot of original data of T-P variable along with the moving average regression fit of order 4; MA(4) regression line for 90% of data and forecasting for the remaining 10% of data.

Table (7): Comparison of actual T-P data to forecast outcome using moving average model of order four; MA(4).

Row	Period (month)	Forecast mg/l	Lower mg/l	Upper mg/l	Actual mg/l
1	141	12.41	7.06	17.76	14.45
2	142	12.41	7.06	17.76	13.73
3	143	12.41	7.06	17.76	16.03
4	144	12.41	7.06	17.76	9.78
5	145	12.41	7.06	17.76	12.43
6	146	12.41	7.06	17.76	8.55
7	147	12.41	7.06	17.76	9.53
8	148	12.41	7.06	17.76	12.76
9	149	12.41	7.06	17.76	12.07
10	150	12.41	7.06	17.76	11.08
11	151	12.41	7.06	17.76	19.99
12	152	12.41	7.06	17.76	15.57
13	153	12.41	7.06	17.76	12.22
14	154	12.41	7.06	17.76	13.07
15	155	12.41	7.06	17.76	14.14
16	156	12.41	7.06	17.76	20.02

Table (8): Comparison of actual T-P data to forecast outcome using autoregressive integrated moving average model of order ARIMA(1,2,4).

Row	Period (month)	Forecast mg/l	Lower mg/l	Upper mg/l	Actual mg/l
1	141	9.45	4.78	14.13	14.45
2	142	9.85	3.98	15.71	13.73
3	143	10.05	3.05	17.05	16.03
4	144	9.98	2.26	17.69	9.78
5	145	10.06	1.51	18.60	12.43
6	146	10.05	0.81	19.30	8.55
7	147	10.10	0.13	20.06	9.53
8	148	10.38	0.00	20.75	12.76
9	149	10.72	0.00	21.44	12.07
10	150	11.06	0.00	22.11	11.08
11	151	11.39	0.00	22.77	19.99
12	152	11.71	0.00	23.41	15.57
13	153	12.03	0.00	24.05	12.22
14	154	12.34	0.00	24.68	13.07
15	155	12.66	0.00	25.31	14.14
16	156	12.96	0.00	25.92	20.02

Table (9): Percentage of error of each model for forecasting of variable.

Model	Percentage of Mean Error (%)					
	TSS	BOD_5	COD	T-P	T-N	Q
Linear Method	5.0	38.2	19.5	13.9	3.3	9.2
Quadratic Method	8.6	37.9	3.8	18.6	7.2	15.3
Exponential Growth Method	1.7	46.4	23.8	10.1	3.7	7.9
Simple Exponential Smoothing	12.4	28.5	7.1	30.7	13.8	35.6
Auto Regression, AR	5.4	44.7	47.4	55.6	30.3	4.9
Moving Average, MA	11.1	33.9	21.3	8.50	6.7	9.0
ARIMA	8.2	16.1	20.7	23.2	4.8	--

REFERENCES

Green, J.R, and Magerison, D. 1978. *Statistical Treatment of Experimental Data*. New York-USA.

Lincoin Chao, L. 1974. *Statistic Methods and Analyses*, 2nd edition. McGraw-Hill Kogukusha, LTD, California-USA.

McBean, E. A., and Rovers, F. A. 1998. *Statistical Procedures for Analysis of Environmental Monitoring Data and Risk Assessment.* Prentice-Hall, New Jersey-USA.

Metcalf and Eddy. 2003. *Wastewater Engineering: Treatment, Disposal, and Reuse,* 4th edition. McGraw-Hill, Inc.

Ministry of Water and Irrigation. 1998. Water Authority of Jordan, Operational Manual, Amman- Jordan.

Montgomery, D., and Johnson L. 1976. *Forecasting and Time Series Analysis*. McGraw-Hill Book Company, Japan.

Pillai, R.S.N. and Bagavathi, V. 1997. *Statistics: Theory and Practice*. 9th edition. Ram Nagar, New Delhi, India.

Royal Scientific Society. 1988-2001. Monitoring the Water in King Talal Dam. Amman-Jordan. RSS Reports

Royal Scientific Society. 1998. National Project for Assessment of Water Quality in Jordan. RSS Report.

Sakakini A. E. 2001. Forecasting the Characteristics of Jordanian Domestic Wastewater Using Time Series, Master Thesis. University of Jordan. Amman-Jordan.

Salameh, Elias. 1996. Water Quality Degradation in Jordan: Impact on Environment, Economy and Future Generations Resources Base. Jordan.

Warren Viessman, Jr. and Gray Lewis, L. 1996. *Introduction to Hydrology*, 4th edition. Addison-Wesley Educational Publishers, Inc., USA.

طريقة احصائية للتحليل والتنبؤ بكمية ونوعية المياه الداخلة إلى سد الملك طلال

*أحمد بني هاني، وأحمد الجمرة، وسليمان الطراونة**

ملخـص

ركزت هذه الدراسة على استعراض طرق احصائية لاستعمالها في التحليل والتنبؤ بكمية ونوعية المياه في نهر الزرقاء الذي يعتبر الرافد الرئيس لسد الملك طلال. ان المتغير الذي تم استخدامه للتعبيرعن كمية المياه في هذا البحث هو معدل التدفق. ايضا، لقد تم استخدام المتغيرات التالية للتعبير عن نوعية المياه الداخلة الى السد: المواد الصلبة العالقة الكلية، الاكسجين المستهلك عضويا، الاكسجين المستهلك كيميائيا، الفسفور الكلي، والنيتروجين الكلي. ان البيانات التي تم جمعها لكل متغير قد تم تسجيلها خلال مدة 156 شهرا، وذلك ابتداء من العام 1988 وحتى نهاية العام 2000. لقد تم تحليل المتغيرات الستة المتعلقة بكمية ونوعية المياه عند مدخل السد، الذي يستقبل المياه الناتجة عن اختلاط مياه نهر الزرقاء مع المياه المعالجة الخارجة من محطة الخربة السمراء. ولقد تم استخدام الطرق التالية في معالجة البيانات: علاقة المتغير بنفسه (autocorrelation)، وعلاقة المتغير مع متغير اخر عند نقطة معينة (cross-correlation)، وعلاقة المتغير مع نفسه ولكن بعد مسافة افقية معينة (distance correlation). بالاضافة الى ذلك، فقد تم استخدام التنبؤ الحتمي (deterministic) والاحتمالي (stochastic) للمتغيرات الستة من اجل ايجاد النموذج الافضل للاختطاط. بينت نتائج الدراسة ان نموذج النمو الاسي (exponential growth) قد حقق اقل نسبة في خطا المتوسط الحسابي من بين النماذج الاخرى وذلك للتنبؤ بالمواد الصلبة العالقة الكلية والفسفور الكلي، في حين ان نموذج ARIMA حقق اقل نسبة في خطأ المتوسط الحسابي من بين النماذج الاخرى وذلك للتنبؤ بالاكسجين المستهلك عضويا. لقد بينت الدراسة ايضا انه بالاعتماد على اقل نسبة في خطأ المتوسط الحسابي، فان الدالة التربيعية حققت افضل النتائج للتنبؤ بالاكسجين المستهلك كيميائيا، في حين ان الدالة الخطية تنبات بالنيتروجين الكلي بشكل افضل، وان الاختطاط الذاتي (auto-regression) مثل معدل التدفق بشكل افضل.

الكلمات الدالة: نوعية المياه، سد الملك طلال، التنبؤ الحتمي، التنبؤ الاحتمالي.

* قسم الهندسة المدنية، كلية الهندسة والتكنولوجيا، الجامعـة الأردنيـة، عمان، الأردن (1، 2)؛ وقسم الهندسـة المدنيـة، جامعة مؤتة، الكرك، عمان. تاريخ استلام البحث 2005/5/29، وتاريخ قبوله 2005/11/8.