Journal of Scientific Research and Development 2 (14): 33-37, 2015
Available online at www.jsrad.org
ISSN 1115-7569

Feasibility study of water quality trend in long term time series for Langat River

Yusof N.A.U *, Mat M.Z, Khalid N

Faculty of Civil Engineering, University Technology MARA, Malaysia

Abstract: The main river basin that supply water to two third of the state of Selangor is Langat River Basin. The rapid development of the area and increase in population has put pressure on the river ecosystem. Particularly obvious is river pollution and water deficit. Protecting the ecosystem health of the river will require an integrated approach. Analyzing surface water quality parameters and prediction of variation in future is a principal step in water quality management. Various techniques can be applied to analysis and prediction among which time series model including Box – Jenkins is one of the suitable tools. The Box-Jenkins method for the time series analysis was applied to model the water quality data recorded in the Langat River, for biological oxygen demand, dissolved oxygen and ammoniacal nitrogen. This paper examines the feasibility of using time series analysis to detect long term water quality trend and the existence of trends and the evaluation of the best fitted trend models for Langat River from years 2002 to 2006.

Key words: ARMA model; Water quality parameters; Long term time series; Langat River; Box-Jenkins; Model

1. Introduction

Streams and river are important freshwater resources to the country because of high water consumption due to the development of the country. About 98 % of freshwater from river and stream are used in Malaysia and the balance is from groundwater. Pollutant discharge from the industrial and municipal wastewater, agricultural land run – off, slaughter house and other pollutant are discharge into river as carries of this pollutant. When there is climate change, there is surface run – off known as seasonal phenomenon which is largely affected the basin by the climate. Where industrial and municipal discharge or sewage is continuously pollute the river. The activities of land use in upper streams of the river have affect the quality of water in Malaysia and it not will be improve or getting better and it's important to have a good water quality because as the nature of river in Malaysia is very long.

During last decades, measurement of water quality parameters is an important tool for monitoring water quality due to increasing demand. The water quality for water resources is a serious subject of ongoing concern. Long term water changes assessment becoming a challenging problem (Vassilis et al., 2001). The result has been the gradual accumulation for examination of these data for long term trends and it reliable for long term water quality records. The main river basin that supply water to two third of the state of Selangor is Langat River Basin. Due to changes of undeveloped area to develop area has contribute to changes of direct runoff and discharges into Langat River. The urbanization will increase the impervious and previous area, which is a main factor that been identified in increases of pollution loading as well as increases of direct runoff volume (Hafizan et al., 2010). For the prediction of water quality, recently more researches have become interested in the application of the time series models to detect any changes occur to the water quality of the river. Time series analysis objective is to forecasting and find the changes model. Box – Jenkins method is one of the techniques to forecast the time series behavior. The statisticians George Box and Gwilynn Jenkins introduced Box – Jenkins methodology and named the methodology as Box – Jenkins. Box- Jenkins methodology where use in order to make forecasting, autoregressive moving average ARMA or ARIMA models was applied to find the best fit of a time series to past values of this time series.

Several studies to predict the river water quality through random methods in time series analysis were conducted by different researchers. Time series approach for analyzing water resources were first applied by Thomann (1967) who studied variation of temperature by the time and dissolved oxygen level for the Delaware Estuary cited from Asadollahfardi, Rahbar and Fatemiaghda. Dordoi (2009) predicted boron concentration in 5 hydrology stations of Buick run – down River in Turkey using ARIMA and SARIMA models for a period of 36 months. Ragavan and Fernadez (2006) predicted long – term of water quality of some rivers that were selected randomly by using ARIMA seasonal adjustment parameters model in SAS

* Corresponding Author.

software cited from (Mina and Mohammadreza, 2014).

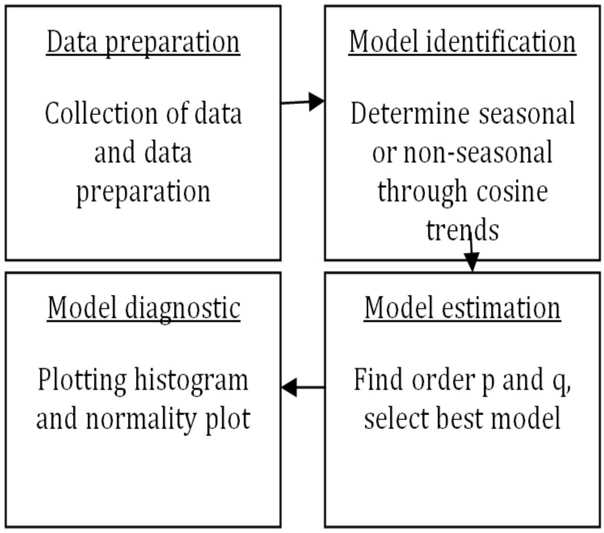
The Box-Jenkins method for the time series analysis was applied to model the water quality data recorded in the Langat River, for biological oxygen demand, dissolved oxygen and ammoniacal nitrogen. This paper examines: (1) the feasibility of using time series analysis to detect long term water quality trend (2) the existence of trends and the evaluation of the best fitted trend models.

2. Problem statement

There are various sources of pollution that polluted Langat River such as the main pollution is due to the discharge from the industries about 58 %, 28 % from the treatment plant that discharges the domestic sewage, projects under construction contribute about 12 % and pig farm contribute about only 2 %. Pollution increase in Selangor and problem with deficit water in the Basin has increase the pollution of the Langat River. The WQI for Langat River in 2000 is 36 to 89 which made the river to be classified as overall polluted river. The WQI is still class 1 at the upstream of the Hulu Langat area before Lui River which is low polluted area, but WQI has reach to class III and IV at the downstream of Langat River after the tributaries Batang Benar River and Balak River (Lee et al., 2006). Trend analysis has proved to be useful tool for effective water resources planning, design and management since trend detection of land use and hydrological variables such as discharge, direct runoff and precipitation provides a useful information on the possibility of change tendency of the variables in the future (Hamilton et al., 2001; Yue and Wang, 2004) cited from (Hafizan et al., 2010).

3. Research methodology

Study area for this research is Langat River which is located in the southeast of Selangor State. For Langat River, six stations were referred for data which are station 2815603, 2817641, 2918606, 2917642, 3118647 and 2814602. Three parameters involves in this research which are biological oxygen demand, dissolved oxygen and ammoniacal nitrogen. Length of data used for this research is from years 2002 until 2006. For the prediction of water quality, recently more researches have become interested in the application of the time series models to detect any changes occur to the water quality of the river. Time series approach for analyzing water resources were first applied by Thomann (1967) who studied variation of temperature by the time and dissolved oxygen level for the Delaware Estuary cited from Asadollahfardi, Rahbar and Fatemiaghda. The Box-Jenkins method for the time series analysis was applied to model the water quality data recorded in the Langat River, for biological oxygen demand, dissolved oxygen and ammoniacal nitrogen. The three stages involves in Box- Jenkins methodology is the model identification, model estimation and model diagnostic. Identification process is important to determine if the series is seasonality or not. In the model estimation is the step for estimating the value of parameters in models which is to estimate the parameter from autoregressive and parameter from the moving average model. At the model diagnostic stage, the values of the model parameter will be choice to best fit the model.



4. Test results and discussion

Raw data that been collected from DID were analyse. Six stations were selected for Langat River which is station number 2918606, 2917642, 3118647, 2814602, 2815603 and 2817641. The parameters for this study are DO, BOD and NH_3. Data used for this study is from 2002 until 2006. In each month, data for monitoring the station have 6 data. For each month average of the data will be used for these studies. The average data for the parameters will be used to detect long term water quality trends.

4.1. Model identification

From the regression analysis, data can be determined as seasonal or non-seasonal. For example Fig. 1 shows the regression analysis of the DO for Langat River. In the regression check, two conditions must be follows to determine whether the data is seasonal or non-seasonal which is the P value must have small value and the R^2 also have low percentage. For example Fig. 1 shows the DO parameter data for Langat River. From the regression analysis, the P value is 0.169 and the R^2 value is 19.2%. From the analysis, DO parameter is a non-seasonal because fulfill the requirements. Fig. 2 shows regression analysis of BOD for Langat River. From the analysis, the BOD data show that the P value is 0.378 and the R^2 value is 4.0%. Which the BOD regression analysis have fulfill all the requirements of non-seasonal. Fig. 3 shows the regression analysis of NH_3 for Langat River. The P value is 0.780 and the R^2 is 7.8%.

```
Regression Analysis: C1 versus C3, C4

The regression equation is
C1 = 5.15 + 0.570 C3 - 0.258 C4

Predictor      Coef  SE Coef      T      P
Constant     5.1460   0.1307  39.37  0.000
C3           0.5700   0.1848   3.08  0.003
C4          -0.2581   0.1848  -1.40  0.169

S = 0.922922   R-Sq = 19.2%   R-Sq(adj) = 15.8%
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5
1	4.460	-0.49382	0.86603	0.50000	
2	3.790	-1.44670	0.50000	0.86602	
3	2.650	-1.27022	0.00000	1.00000	
4	4.640	0.32840	-0.50000	0.86603	
5	4.210	-0.45779	-0.86602	0.50000	
6	4.580	-0.76459	-1.00000	0.00000	
7	2.430	-1.85732	-0.86603	-0.50000	
8	4.100	-0.13676	-0.50000	-0.86602	
9	4.610	0.43602	-0.00000	-1.00000	
10	5.490	-0.25008	0.50000	-0.86603	

Fig. 1: Shows the regression analysis of the DO

```
Regression Analysis: C1 versus C3, C4

The regression equation is
C1 = 5.70 - 0.632 C3 + 0.511 C4

Predictor      Coef  SE Coef      T      P
Constant     5.6990   0.4062  14.03  0.000
C3          -0.6318   0.5743  -1.10  0.277
C4           0.5114   0.5743   0.89  0.378

S = 2.86780   R-Sq = 4.0%   R-Sq(adj) = 0.0%
```

Worksheet 1 ***

↓	C1	C2	C3	C4
1	12.170	1	0.86603	0.50000
2	11.560	2	0.50000	0.86602
3	11.330	3	0.00000	1.00000
4	3.220	4	-0.50000	0.86603
5	5.830	5	-0.86602	0.50000
6	6.110	6	-1.00000	0.00000
7	8.670	7	-0.86603	-0.50000
8	5.430	8	-0.50000	-0.86602
9	6.000	9	-0.00000	-1.00000
10	3.670	10	0.50000	-0.86603

Fig. 2: Show the regression analysis of BOD

```
Regression Analysis: C1 versus C3, C4

The regression equation is
C1 = 1.83 - 0.358 C3 + 0.051 C4

Predictor      Coef  SE Coef      T      P
Constant     1.8337   0.1276  14.38  0.000
C3          -0.3576   0.1804  -1.98  0.053
C4           0.0506   0.1804   0.28  0.780

S = 0.900635   R-Sq = 7.8%   R-Sq(adj) = 3.9%
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5
1	2.490	0.75478	0.86603	0.50000	
2	3.120	0.77906	0.50000	0.86602	
3	3.850	1.68161	0.00000	1.00000	
4	0.980	-1.29905	-0.50000	0.86603	
5	1.720	0.40718	-0.86602	0.50000	
6	1.990	0.97924	-1.00000	0.00000	
7	3.960	0.73039	-0.86603	-0.50000	
8	1.720	0.00584	-0.50000	-0.86602	
9	1.490	-0.16899	-0.00000	-1.00000	
10	1.110	-0.37885	0.50000	-0.86603	

Fig. 3: Shows the regression analysis of the NH_3

4.2 Model estimation, selection of best model and model diagnostic.

In order to estimated parameters of the model, each parameter will be test from ARMA (1, 0, 0) until (5, 0, 5) to select the best model for the residual. The best model is selected based on the P value. The highest P value is selected. The P value must greater than 0.1. Plotting histogram and normality plot to test the p and q value for the highest p value. The histogram should be normally distributed and normality plot should be straight line. Fig. 4 show the best model for DO of Langat River, which is ARMA (4, 0, 3) with highest P value is 0.218. The histogram show in Fig. 5 for the ARMA (4, 0, 3) is normally distributed and the normally plot show in Fig. 6 is straight line. The best model for BOD is ARMA (4, 0, 2) with the P value is 0.722 show in Fig. 7. The histogram show in Fig. 8 is normally distributed and the normality plot show in Fig. 9 is straight line. The best model for NH_3 is ARMA (1, 0, 2) with P value is 0.558 show in Fig. 10.Fig. 11 show the histogram is normally distributed and the normality plot is straight line show in Fig. 12. DO, BOD and NH_3 fit the ARMA to detect the long term water quality of Langat River.

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12      24      36      48
Chi-Square     1.0    12.2    21.6    46.7
DF               4      16      28      40
P-Value      0.902   0.733   0.802   0.218

MTB >
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5	C6
1	4.460	-0.19402	0.86603	0.50000		
2	3.790	-1.00652	0.50000	0.86602		
3	2.650	-1.32232	0.00000	1.00000		
4	4.640	0.48715	-0.50000	0.86603		
5	4.210	-0.44089	-0.86602	0.50000		
6	4.580	-0.05425	-1.00000	0.00000		
7	2.430	-1.47983	-0.86603	-0.50000		
8	4.100	0.59669	-0.50000	-0.86602		
9	4.610	0.43128	-0.00000	-1.00000		
10	5.490	0.12760	0.50000	-0.86603		

Fig. 4: Show the best model for DO

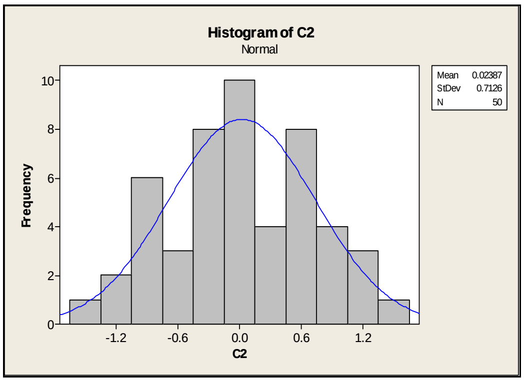


Fig. 5: Show the histograms of DO

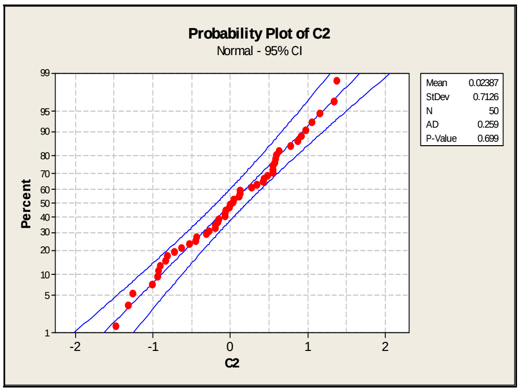


Fig. 6: Show normality plot for DO

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12      24      36      48
Chi-Square     2.9    20.1    33.1    50.6
DF               5      17      29      41
P-Value      0.722   0.269   0.274   0.144

MTB > Histogram C2;
SUBC>   Bar;
SUBC>   Distribution;
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5
1	12.170	3.91550	0.86603	0.50000	
2	11.560	3.29245	0.50000	0.86602	
3	11.330	1.28350	0.00000	1.00000	
4	3.220	-4.16972	-0.50000	0.86603	
5	5.830	-0.97528	-0.86602	0.50000	
6	6.110	-0.27507	-1.00000	0.00000	
7	8.670	-0.14826	-0.86603	-0.50000	
8	5.430	-0.26390	-0.50000	-0.86602	
9	6.000	-1.20011	-0.00000	-1.00000	
10	3.670	-1.03194	0.50000	-0.86603	

Fig. 7: Show the best model for BOD

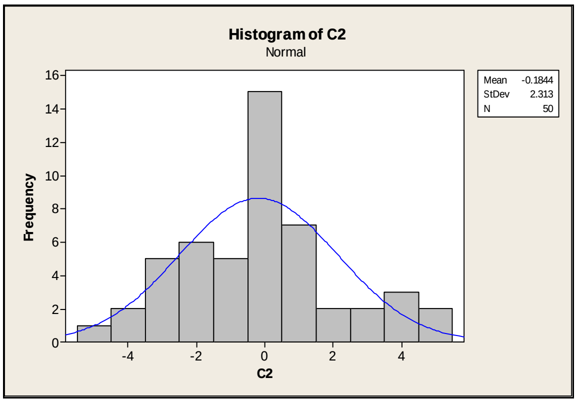


Fig. 8: Show the histogram of BOD

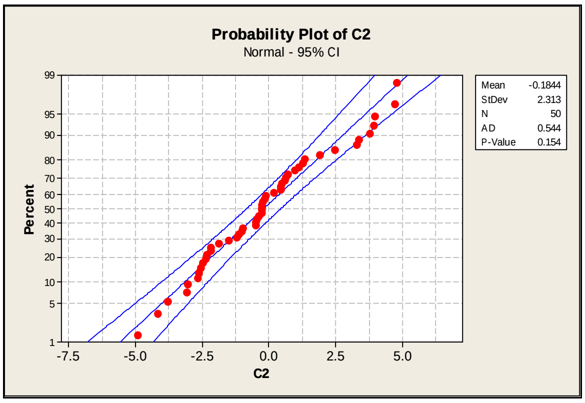


Fig. 9: Show normality plot for BOD

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            12     24     36     48
Chi-Square    6.8   22.4   38.9   65.6
DF              8     20     32     44
P-Value     0.558  0.321  0.187  0.019

MTB > |
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5
1	2.490	0.39206	0.86603	0.50000	
2	3.120	1.38806	0.50000	0.86602	
3	3.850	1.19475	0.00000	1.00000	
4	0.980	-1.11387	-0.50000	0.86603	
5	1.720	0.09636	-0.86602	0.50000	
6	1.990	0.42225	-1.00000	0.00000	
7	3.960	1.66719	-0.86603	-0.50000	
8	1.720	-0.50229	-0.50000	-0.86602	
9	1.490	-0.46494	-0.00000	-1.00000	
10	1.110	-0.16727	0.50000	-0.86603	

Fig. 10: Show the best model for NH3

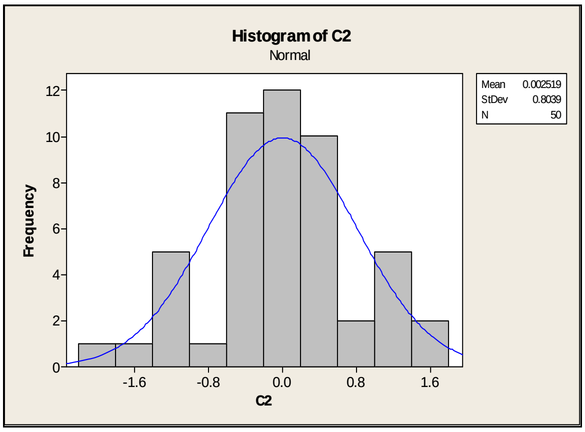


Fig. 11: Show the histogram of NH3

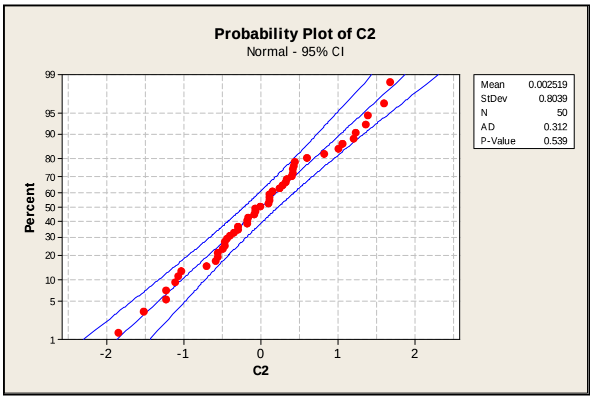


Fig. 12: Show normality plot of NH3

4.3. Count error

The error for each parameter is calculated using the Mean Absolute Percentage Error. Fig. 13 show the forecast value for DO of Langat River. Fig. 14 show the forecast value for BOD and Fig. 15 show the forecast value for NH_3. The forecast value will be used to calculate the error. The error percentage must less than 20%. The percentage error for D0 is 1.4 %, for BOD is 4.4 % and for NH_3 is 1.2 %.

```
                        Limits
Period  Forecast    Lower    Upper  Actual
    51   5.94934  4.43967  7.45902
    52   6.56552  4.83022  8.30082
    53   6.39725  4.59631  8.19820
    54   6.63762  4.79830  8.47694
    55   5.59685  3.66709  7.52660

MTB >
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5
1	4.460	-0.19402	0.86603	0.50000	5.949342
2	3.790	-1.00652	0.50000	0.86602	6.56552
3	2.650	-1.32232	0.00000	1.00000	6.39725
4	4.640	0.48715	-0.50000	0.86603	6.63762
5	4.210	-0.44089	-0.86602	0.50000	5.59685
6	4.580	-0.05425	-1.00000	0.00000	
7	2.430	-1.47983	-0.86603	-0.50000	
8	4.100	0.59669	-0.50000	-0.86602	
9	4.610	0.43128	-0.00000	-1.00000	
10	5.490	0.12760	0.50000	-0.86603	

Fig. 13: Show forecast value for DO

```
Forecasts from period 50

                       95 Percent
                         Limits
Period  Forecast    Lower    Upper  Actual
    51    5.9403   1.5392  10.3414
    52    2.7491  -2.0798   7.5780
    53    4.0446  -0.8274   8.9166
    54    4.1051  -0.8033   9.0134
    55    7.3716   2.1627  12.5804

MTB >
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5
1	12.170	3.91550	0.86603	0.50000	5.94028
2	11.560	3.29245	0.50000	0.86602	2.74908
3	11.330	1.28350	0.00000	1.00000	4.04458
4	3.220	-4.16972	-0.50000	0.86603	4.10507
5	5.830	-0.97528	-0.86602	0.50000	7.37157
6	6.110	-0.27507	-1.00000	0.00000	
7	8.670	-0.14826	-0.86603	-0.50000	
8	5.430	-0.26390	-0.50000	-0.86602	
9	6.000	-1.20011	-0.00000	-1.00000	
10	3.670	-1.03194	0.50000	-0.86603	

Fig. 14: Show forecast value for BOD

```
Forecasts from period 50

                    95 Percent Limits
Period  Forecast     Lower     Upper  Actual
    51   2.15493   0.52846   3.78139
    52   1.53546  -0.22223   3.29314
    53   2.10943   0.35092   3.86793
    54   1.53564  -0.22369   3.29496
    55   2.10925   0.34911   3.86939

MTB >
```

Worksheet 1 ***

↓	C1	C2	C3	C4	C5
1	2.490	0.39206	0.86603	0.50000	2.154927
2	3.120	1.38806	0.50000	0.86602	1.53546
3	3.850	1.19475	0.00000	1.00000	2.10943
4	0.980	-1.11387	-0.50000	0.86603	1.53564
5	1.720	0.09636	-0.86602	0.50000	2.10925
6	1.990	0.42225	-1.00000	0.00000	
7	3.960	1.66719	-0.86603	-0.50000	
8	1.720	-0.50229	-0.50000	-0.86602	
9	1.490	-0.46494	-0.00000	-1.00000	
10	1.110	-0.16727	0.50000	-0.86603	

Fig. 15: Show forecast value for NH3

5. Conclusion

In the regression check, two conditions must be follows to determine whether the data is seasonal or non-seasonal which is the P value must have small value and the R^2 also have low percentage. The best ARMA model for Langat River of DO is ARMA (4, 0, 3), BOD is (4, 0, 2) and NH_3 is (1, 0, 2). The best model for each parameter selected based on the highest P value and the histogram must normally distribute and normality plot is straight line.

The error for each parameter is calculated using the Mean Absolute Percentage Error. For checking the error, the five last data in year 2006 is used and forecast the data. The forecast value will be used to calculate the error. The error percentage must less than 20%. The percentage error for Langat River of D0 is 1.4 %, for BOD is 4.4 % and for NH_3 is 1.2 %. The time series analysis is suitable to detect long term water quality trend based on the parameter that fit the ARMA model using the Minitab software. Time series analysis is suitable method to detect long term water quality of parameter because it cannot be explained or modeled by normal analytical procedures where water quality variables change continually through time and arise from dynamic processes.

References

G. Asadollahfardi, M. Rahbar and M. Fatemiaghda. Application of Time Series Models to Predict Water Quality of Upstream and Downstream

Hafizan, J., Sharifuddin, M.Z., Ahmad, Z.A., Mohd, K.Y., Mohd, A.A.S., and Mazlin, M. (2010). Hydrological Trend Analysis Due to Land Use Changes at Langat River Basin. UPM, UKM and UM.

Komornik,J.,Komornikova,M.,Mesiar,R.,Szokeova,D.,Szolgay,J.(2006). Comparison of forecasting performance of nonlinear models of hydrological time series. J.Physics and Chemistry of the Earth. 31,1127-1145

Lee, Y.H, Md., P.A, Chai, S.Y, Mazlin., M, Rohani, A. (2006). Development of possible indicators for sewage pollution for the assessment of Langat River ecosystem health, Vol 10, No 1. UKM.

Mason C.F. and Parr B.L.9 (2003). Long-term trends in water quality and their impact on macroinvertebrate assemblages in eutrophic lowland rivers, Water Resources.

Marques,C.A.F,Ferreira,J.A,Rocha,A.,Castanheira,J.M, MeloGencolves,P.,Yaz,N.,Dias,J.M.,(2006).Singular spectrum analysis and forecasting of hydrological time series.J.Physics and Chemistry of the Earth.31,1172-1179.

Mina, R., Mohammadreza, K., Using Arima time series model in forecasting the trend of changes in qualitative parameters of Serfidrud River. Islamic Azad University, Tonekabon Branch and University of Guilan.

Robinsoos, G.M., (2009). Time Series Analysis. Elseveir, 285-293.

Rojas,I.,Valenzuela,O.,Rojas,F.,Guillen,A.,Herrera,L.J.,Pomares,H.,Marques,L.,Pasadas,M.,2008.Soft-computing techniques and ARMA model for time series prediction. Neurocomputing. 71,519-537.

Sengul,S.,Can,I.,2009.Stochastic modelling of mean monthly flows of Karasu River, in Turkey.J.Water and Environment.25,31-39.

Vassilis, Z.A., Dimitris, M.P., and Konstantina, A.M. (2001). Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece, 5,(4), 679-691 Greece.

Wu,C.L.,K.W.,2010.Data – driven models for monthly streamflows time series prediction.J.Engineering Applications of Artificial Intelligence.