Journal of Applied Sciences 6 (9): 1929-1934 2006
ISSN 1812-5654

Time Series Analysis for Hydrological Features: Applications of Box-Jenkins Models to Euphrates River

Naba Shakir Hadi
Department of Civil Engineering and Architecture, Wuhan University of Technology, Mail Box 169, West Mafang Shan Campus, #205 Luoshi Road, Wuchang 430070, People's Republic of China

Abstract: Water plays a significant role at the present time and receives more concern in future due to its necessity for the development of the life of people, because water is considered as the most important natural resources and depends on it different activities. This study presents the statistical models via the application of the theories of a time series to water variables of Thalweg of the Euphrates River between Al-Qadissiya Dam and Abu-Ghraib stream. Measurement were made and SPSS 7.5 Software was used for the analysis and the best model was determined by checking the explained variation.

Key words: Time series analysis, box Jenkins, Euphrates river

INTRODUCTION

The Thalweg of the cross section of the Euphrates river between Al-Qadisya Dam and Abu-Graib stream were recorded. Box-Jenkins method is used to estimate the Auto Correlation Factor (ACF) and Partial Auto Correlation Factor (PACF) for the notices above by using the SPSS software (Hussain, 2000). This system accounts the final estimates of the autocorrelation and the autoregressive and moving average in addition to the greatest log likelihood, AIC and SBC and the total of square and variance and residual difference and checking the whiteness after achieving the account of the ACF of the residuals by using Chi-square test and at the end we chose the best model to the ARIMA from the possible alternatives which gives the best check of variations for the results depending on the statistical tests (Al-Samawi and Abas, 1994).

ANALYSIS

It considers observed data across time periods and statistic methodology which deal with analysing these previews are called time series analysis (Muhsin, 1995), this analysis is distinguished among others by giving it a significant in keeping on previews sequences in time series (Pandit and Wu, 1983).

METHOD

Box-Jenkins method was used, this method is considered one of the common prediction methods because it is a general method use for dealing with all kinds of stationary and non stationary time series (Box and Jenkins, 1976). The theoretical basis of the method is explained below;

- Linear Filter model

$$X_1 = \mu + \sum_{j=0}^{\infty} \psi_i \in_{t-j} \tag{1}$$

Ψ_i: Values of coefficient weights
μ : Average of variable series

- Autoregressive Model (AR)

$$X_t = \xi + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots.. + \varphi_p X_{t-p} + \in_t \tag{2}$$

X_t: Values of observations of time series during the period (t-p)t
ξ : Time series average.
ϵ_t: Random error combination that follows natural distribution with zero average and contrast is (σ^2_t)

- Moving Average Model (MA)

$$X_t = \mu + \in_t - \theta_1 \in_{t-1} - \theta_2 \in_{t-2} - \ldots.. - \theta_q \in_{t-q} \tag{3}$$

X_t: The value of real observation in (t)
μ : Time series mean
$\theta_1, \theta_2, \ldots, \theta_q$: Features of the model
$\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q}$: Random changes

- Autoregressive Moving Average Model (ARMA)

$$X_t = \xi + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p} \\ - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - ..... - \theta_q \epsilon_{t-q} + \epsilon_t \quad (4)$$

X_t: Values of real observation for time series in (t)
ξ : Time series mean
$\varphi_1, \varphi_2, ..., \varphi_p$: Coefficient of autoregressive limitations
$\theta_1, \theta_2, ..., \theta_q$: Features of moving average limitations
$\epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-q}$: Values of random changes

Box and Jenkins (1976) presented a general way to indicate on model ARIMA the best for any time series then using this model for purposes of prediction.
This way is summarized with the following steps:-

- Model Identification (Coulbeck and Orr, 1988).
- Estimation (Box and Jenkins, 1976).
- Diagnostic checking (Hanke and Reitsch, 1981).

Selection of parsimonious model: The best degree of the model which has the smallest number of variables is used to give the best possible results. We account well the variable (a) standard error test (Donovan and Thomas, 1983). (b) variance ratio test (c) Akaike information criteria, which is used by Srinivas Edwin (1982).

EUPHRATES THALWEG MODEL

The Thalweg variable consists of (52) observations in the year 1999 between Al-Qadisya dam and Abu-Graib stream from the Euphrates. The Fig. 1 shows the drawing of the thalweg in the return of the distance which illustrates the most important series characteristics. The clear characteristics such as the direction, the discontinuities will be clear if it is founded in the series so the drawings illustrates the chosen test of stable series and the model definition.

The Fig. 1 shows wild observations or they are outliers which don't appear organized with other information. The outlier observations form wild observations. That needs to their smooth to their expected value before the analysis of the series. The thalweg series has three outliers the smoothed of these wild observation is it taking equated to the average of the two observation which occurs before and after which is shown in Table 1.

Figure 2 represents relationship between standard deviation and mean transformation test using pearson method the correlation coefficient is found equal to (0.77).

Common logarithm function is used to transform the data to the normal distribution (Fig. 3) the correlation coefficient between the standard deviations and the means of the thalweg is reduced to (-0.016). This means that fluctuations of the data are reduced.

Table 1: Thalweg outliers and their smooth points

Distance index (km)	Outlier (m)	Last point (m)	Next point (m)	Smoothed point (m)
340	46.00	51.77	54.65	53.21
410	35.70	43.85	39.60	41.73
485	28.00	35.88	34.30	35.09

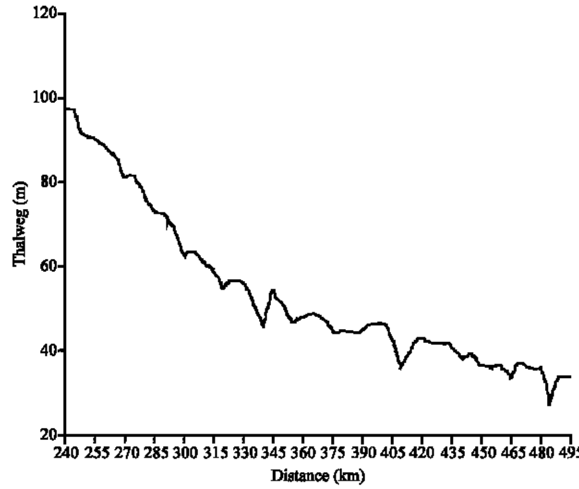

Fig. 1: Thalweg time series

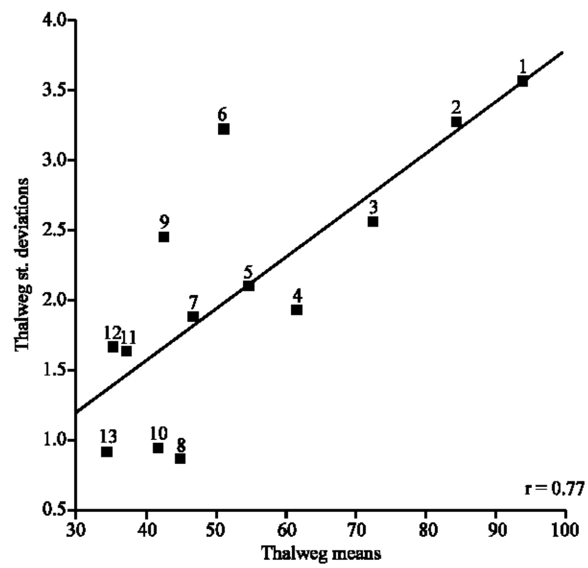

Fig. 2: Standard deviations vs means of thalweg

Figure 4 represents (ACF) autocorrelation function for the variables the extend of values decreases gradually till it reach to zero.

The partial autocorrelation function (PACF) which is illustrated in the Fig. 5 is accounted. The cut is after the first lag so the addition is to the first real value. This value is bigger than the level of confidence limit error. From these Fig. 4 and 5 the model ARIMA(1,0,0) or we can define it $AR_{(1)}$.

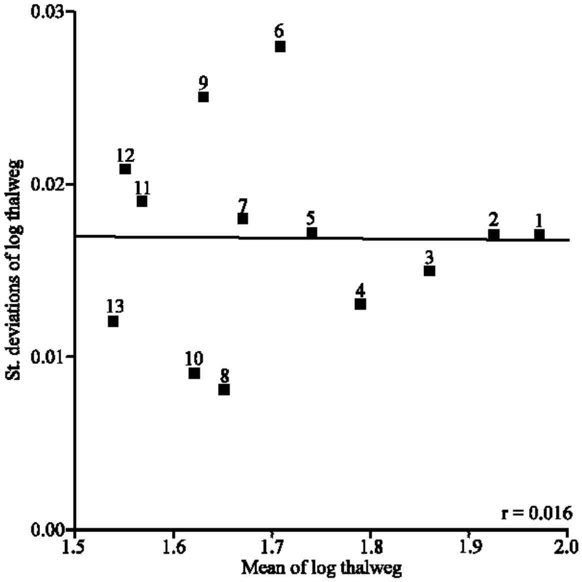

Fig. 3: Standard deviations vs. means of thalweg common log

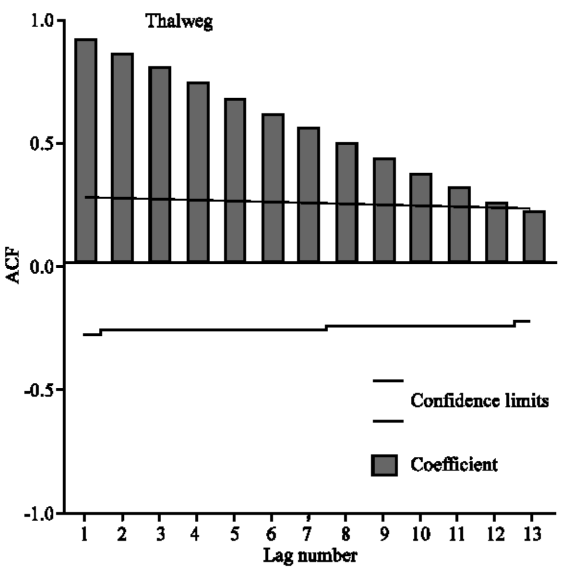

Fig. 4: Auto correlation function of thalweg

Figure 6 shows the model ($AR_{(1)}$)all the variables of the model are new, because the values is more than of double of standard error.

The model equation:

$$X_t - 0.99X_{t-1} = 4.03 + \epsilon_t \quad (5)$$

The statistic tests:

LLH = 76, AIC = -148, SBC = -144

Because of the value (ϕ_1 = 0.99) is nearer to one it is possible to take first differencing.

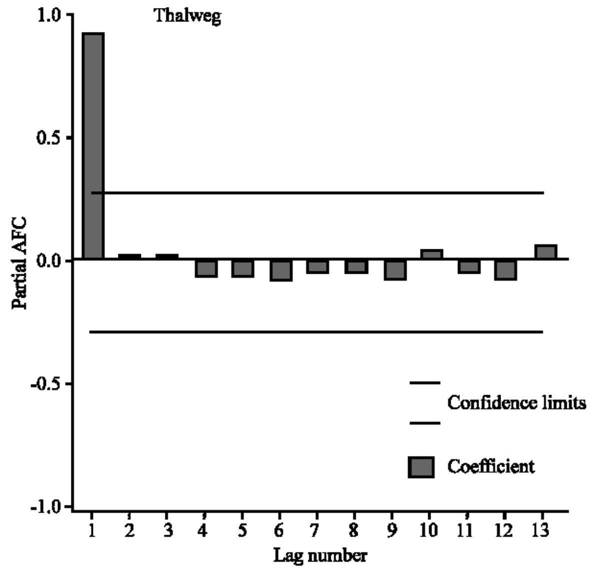

Fig. 5: Partial auto correlation function of thalweg

MODEL: MOD_2
Model Description:
Variable: THALWEG
Regressors: NONE
Non-seasonal differencing: 0
No seasonal component in model.
Parameters:
AR1 ———— < value originating form estimation >
CONSTANT ———— < value originating form estimation >
Analysis will be applied to the natural logarithm of the data.
95.00 percent confidence intervals will be generated.
Split group number: 1 Series lenght: 52
No missing data.
Melard' s algorithm will be used for estimation.
_ Conclusion of estimation phase.
Estimation terminated at iteration number 4 becaus:
All parameter estimates changed by less than .001
FINAL PARAMETERS:
Number of residuals 52
Standard error .05479328
Log likeihood 76.024727
AIC -148.04945
SBC -144.14697
Analysis of Variance:

	DF	Adj. Sum of Squares	Residual Variance
Residuals	50	.16337475	.003300230

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
ARI	.9938515	.00936712	106.10006	.000000
CONSTANT	4.0305786	.46002307	8.76169	.000000

Fig. 6: Estimation of thalweg model $AR_{(1)}$

The (ACF) illustrates for a model to the lowest level of the bottom thalweg after taking the first difference in Fig. 7 which has the real value to three of Autocorrelation these values are bigger than the confidence limit error at lag (1), lag (9) and lag (10).

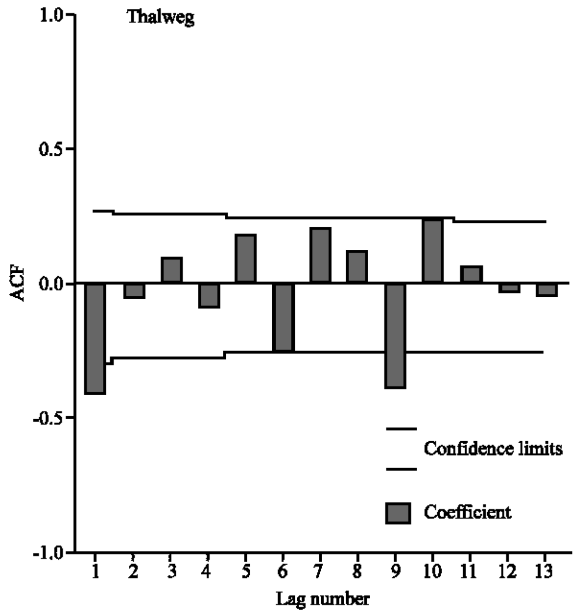

Fig. 7: Auto correlation function of thalweg difference (1)

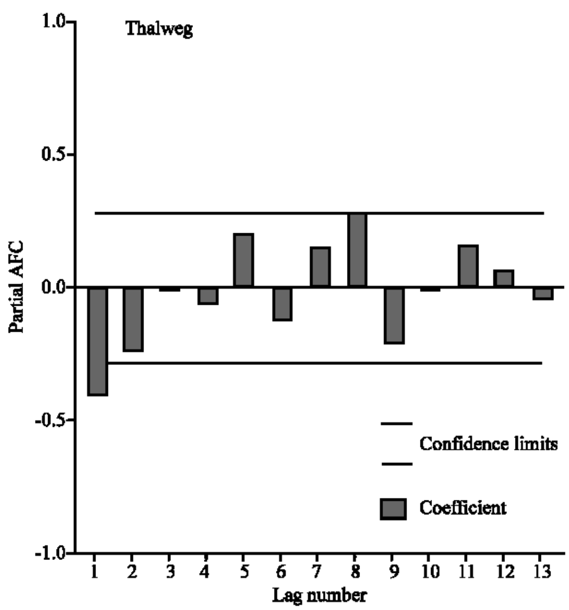

Fig. 8: Partial auto correlation function of thalweg difference (1)

Figure 8 shows the partial Auto correlation function of thalweg after taking difference (1), which has real values for two of partial auto correlation these values are bigger than confidence limit error at lag (1), lag (8).

From the two Fig. 7 and 8 we can know the model ARIMA (2, 1, 3).

Figure 9 represents estimation of Thalweg model ARIMA (2, 1, 3).

The model equation:

MODEL: MOD_6
Model Description:
Variable: THALWEG
Regressors: NONE
Non-seasonal differencing: 1
No seasonal component in model.
Parameters:

AR1	————	< value originating form estimation >
AR2	————	< value originating form estimation >
MA1	————	< value originating form estimation >
MA2	————	< value originating form estimation >
MA3	————	< value originating form estimation >
CONSTANT	————	< value originating form estimation >

Analysis will be applied to the natural logarithm of the data.
95.00 percent confidence intervals will be generated.
Split group number: 1 Series lenght: 52
No missing data.
Melard' s algorithm will be used for estimation.
Conclusion of estimation phase.
Estimation terminated at iteration number 5 becaus:
All parameter estimates changed by less than .001
FINAL PARAMETERS:

Number of residuals	51
Standard error	.04525121
Log likeihood	87.734247
AIC	-163.46849
SBC	-151.87754

Analysis of Variance:

DF	Adj. Sum of Squares	Residual Variance
Residuals 45	.09530565	.00204767

Variables in the Model:

	B	SEB	T-RATIO	APPROX. PROB.
ARI	-.00307892	.2161156	-.0142466	.98869618
AR2	.83581609	.2092170	3.9949712	.00023692
MA1	.49729621	2.5413523	.1956817	.84574023
MA2	-.58970974	3.7781296	.2390046	.81218717
MA3	-.58970974	1.4833445	-.3975541	.69283847
CONSTANT	-.02095664	.0063411	-3.3048769	.00187058

Fig. 9: Estimation of thalweg model ARIMA (2,1,3)

$$X_t + 0.003X_{t-1} - 0.836X_{t-2} = 0.02 + \in_t - 0.497\in_{t-1} - 0.903\in_{t-2} + 0.589\in_{t-3} \quad (6)$$

the model didn't achieve the standard error test because the variable of the first $AR_{(1)}$ is (-0.003) and the variable of MA1, MA2, MA3 are (0.5), (0.9), (0.59) less than the double confidence limit (1.48), (3.78), (2.54), (0.22). So the model ARIMA (1,1,0) is identified to adjust the failure in the model.

Figure 10 shows estimation of thalweg model ARIMA (1,1,0)

The model equation is :

$$X_t + 0.39X_{t-1} = -0.02 + \in_t \quad (7)$$

the statistic to test are :

LLH = 85, AIC = 166, SBC = -163

The model variables of the two models are equal so we can not use variance ratio test. From the statistical test

```
MODEL:      MOD_7
Model     Description:
Variable:    THALWEG
Regressors:  NONE
Non-seasonal differencing: 1
No seasonal component in model.
Parameters:
AR1           ____________   < value originating form estimation >
CONSTANT  ____________   < value originating form estimation >
Analysis will be applied to the natural logarithm of the data.
95.00 percent confidence intervals will be generated.
Split group number:  1  Series lenght:  52
No missing data.
Melard' s algorithm will be used for estimation.
Conclusion of estimation phase.
Estimation terminated at iteration number 2 becaus:
   Sum of squares decreased y less than .001 percent.
FINAL PARAMETERS:
Number of residuals  51
Standard error          .04636721
Log likeihood           85.185544
AIC                           -166.37109
SBC                           -162.50744
            Analysis of Variance:
                DF   Adj. Sum of Squares    Residual Variance
Residuals  49                 .10568092            .00214992
Variables in the Model:
                        B              SEB         T-RATIO    APPROX. PROB.
ARI          -.38661402   .13111194   -2.9487325        .00487823
CONSTANT  -.02108492   .00470823   -4.4783114        .00004506
```

Fig. 10: Estimation of thalweg model ARIMA (1,1,0)

```
MODEL:    MOD_9.
Variable:   ERR_3         Missing cases: 1       Valid cases: 51
Autocorrelations:   ERR_3   Error for THALWEG form ARIMA_7 LN
C
        Auto-Stand.
Lag   Corr.   Err.   -1   -.75   -.5  -.25   0   .25   .5   .75   1
                      +----+----+----+----+----+----+----+----+
  1   -.100   .136                      .   **I    .
  2   -.195   .135                      .****I     .
  3    .901   .133                      .     I**  .
  4    .035   .132                      .     I*   .
  5    .119   .130                      .     I**  .
  6   -.142   .129                      .  ***I    .
  7    .241   .128                      .     I*****
  8    .113   .126                      .     I**  .
  9   -.336   .125                    **.****I     .
 10    .179   .123                      .     I****.
 11    .184   .122                      .     I****.
 12   -.017   .120                      .     *    .
 13   -.100   .119                      .   **I    .
Plot Symbols:          Autocorrelations  *       Two Standard Error
Limits.
Total cases:    52       Computable first lags: 50
```

Fig. 11: Auto correlation function of thalweg residuals model ARIMA (1,1,0)

to the Table 2 the model ARIMA is the best chosen model because it has the largest value of LLH and the smallest values of AIC, SBC.

Figure 11 shows the accuracy of auto correlation function (AIC) of residual model ARIMA (1,1,0) FOR (lag 13) it is obvious that one of the auto correlation has bigger value than the double confidence limit error at lag (9). by using diagnostic process by finding the total squares for the values (ACF) and multiplying the total with the final number of the observation (N = 51), (Q) portmanteau coefficient was equal to (17.94) which is less than chi-square (χ^2 = 19.68) at (11) degree of freedom and 5% of confidence limit error level (χ^2>Q) and more than the auto correlation which doesn't have real values and

Table 2: Comparison between Thalweg models $AR_{(1)}$ and ARIMA (1,1,0)

	AR(1)	ARIMA (1,1,0)
Observation	52	51
Model variables	2	2
Variance	330	332
LLH	76	85
AIC	-148	-166
SBC	-144	-163

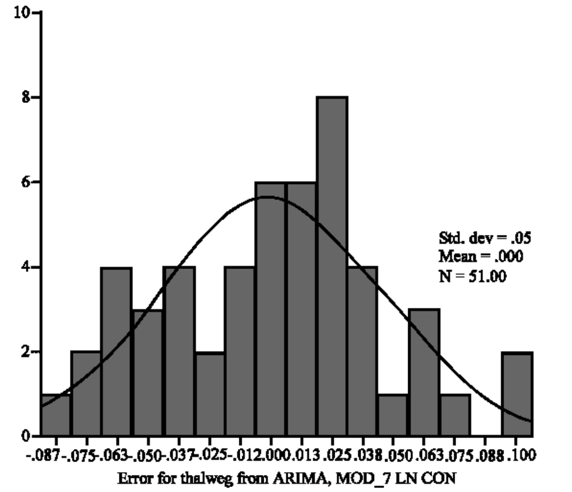

Fig. 12: Thalweg residuals histogram model ARIMA (1,1,0)

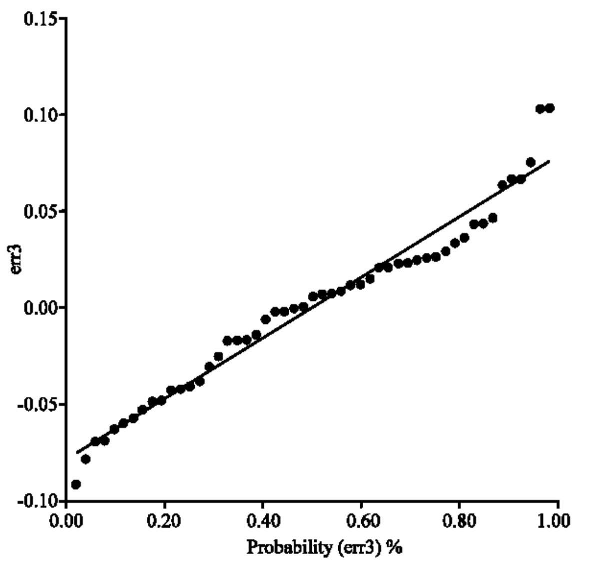

Fig. 13: Cumulative probability plot of a residuals form thalweg ARIMA (1,1,0)

Table 3: Chi-square test

Lower-limit	Upper-limit	Observed frequency	Expected frequency	Chi-square
At or below	-0.03586	14	11.03	0.784523
-0.03586	-0.00802	6	11.04	2.181618
-0.00802	0.01982	12	11.93	0.000298
0.01982	0.04766	13	9	1.410671
Above 0.04766		6	8	0.355936

so the checking of ARIMA (1,1,0) model validity is done concerning the residuals are whiteness.

Figure 12 shows the thalweg residuals histogram model ARIMA (1,1,0). While Fig. 13 shows cumulative probability plot of a residuals from thalweg ARIMA (1,1,0).

Table 3 shows Chi-square test for the better application to common model.

CONCLUSIONS

- The flexibility of Box-Jenkins method that flexibility comes of the numbering of the models of this method widely so there aspace to choose the suitable model to the series under research with bigger freedom.
- There are irregular values, which are substituted by the observations which are surrounded to it, because the deviations can not be completed or they have irregular values.
- From the variables drawings to the following study thalweg to the Euphrates between the Qadisya Dam abd Abu-Grabe stream.
- The thalweg model represented by the first order autoregressive model $AR_{(1)}$ this refers that the memory of the variables, which are measured according.
- To the distance is short and the deviations, extences to one distance only the lowest level of the bottom thalweg has the smallest value of the memory (-0.394) which refer that the periodic value to the error component to this variables is large.

In this research the formation of stochastic vocabularies which suits with the standard common distribution in addition the suitability test to these deviations for the common standard distribution by using probability plot for the common standard distribution which probability values spread to the deviations on the straight line whenever these points on the straight line that will a referring on the weakness of this suitability. The probability plot to models residual shows that the residual deviations to the model the thalweg is the closest for the common standard distribution. This is affirmed the value of Chi-square χ^2 which is accounted for the residual model ($\chi^2_{(2)}$ = 4.73305) it shows confidence at the level of (0.0938063) function as in Table 3.

REFERENCES

Al-Samawi, A.A. and K.A. Abas, 1994. A two-site model on the generation of missing monthly river flow data: A case study on Iraqi rivers. J. Eng. Technol., Vol. 13, No. 2.

Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis, Forecasting and Control. Holden-Day.

Coulbeck, B. and C. Orr, 1988. Computer Applications in Water Supply: Vol. 1 System Analysis and Simulation. Research Studies Press Ltd.

Donovan, O. and M. Thomas, 1983. Short Terms Forecasting: An Introduction to the Box-Jenkins Approach, John Wiley and Sons. New York.

Hanke, J.E. and A.G. Reitsch, 1981. Business Forecasting. Allyn and Bacon, Inc.

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.

Muhsin, A.O., 1995. Time series models for treatment of surface waters: A case study of the Tigris River in Baghdad. M.Sc. Thesis, University of Technology, Iraq.

Pandit, S.M. and S. Wu, 1983. Time Series Analysis with Applications. John Wiley and Sons. New York.

Srinivas, G.R. and W.Q. Edwin, 1982. Stochastic ARIMA Models for Monthly Stream Flows. A Research Edited in Time Series Methods in Hydrosciences. By A.H. El-Sharawi. Proceedings of an International Conference at Canada Center for Inland Waters, Burlington, Ontario, Canada, pp: 474-486.