Forecasting by Box-Jenkins (ARIMA) Models to Inflow of Haditha Dam

Zahra Abd Saleh
College of Eng., Babylon Univ., Babil, Iraq

Abstract

Box-Jenkins seasonal model is applied in this study to records of mean flow to Haditha reservoir in the middle west of Iraq for period from water year 1999/2000 to water year 2008/2009 . Two types of model $(0,1,1) \times (0,1,1)_{12}$ and $(0,1,2) \times (0,1,1)_{12}$ are suggested, and the selected model is the one which give minimum sum of squares (SS). The unconditional sum of squares is used to estimate the model parameters. It is found that the model which corresponds to the minimum sum of squared errors is the $(0,1,2) \times (0,1,1)_{12}$ model with parameters $\theta_1 = 0.368$, $\theta_2 = 0.321$ and $\Theta = 0.910$. Port Manteau Lack of fit test and Residual Autocorrelation Function (RACF) test are applied as diagnostic checking. Forecasts of monthly inflow for the period from October ,2009, to September,2011, are compared with observed inflow for the same period and since agreement is very good adequacy of the selected model is confirmed.

الخلاصة

تم تطبيق طريقة بوكس–جنكينز(Box-Jenkins) في هذه الدراسة للنموذج الفصلي (model) للجريان الشهري إلى سد حديثة في منتصف غرب العراق للفترة من بداية السنة المائية 2000/1999 وحتى نهاية السنة المائية 2009/2008. حيث تمت مطابقة نوعين من النماذج التصادفية الفصلية وهي النموذج $(0,1,1)_{12}\times(0,1,1)$ والنموذج $(0,1,1)_{12}\times(0,1,2)$ واختير النموذج الذي يعطي اقل مجموع مربعات أخطاء. إن نتائج استخدام طريقة مجموع المربعات غير المشروطة لتقدير معالم النماذج أظهرت بان مجموع مربعات الأخطاء للنموذج $(0,1,1)_{12} \times (0,1,2)$ بمعالم $\theta_1 = 0.368$ ، $\theta_2 = 0.321$ و $\Theta = 0.910$ هو الاقل. كما وان فحص مخطط الذبذبة وفحص Port Manteau Lack لم يبينا وجود عشوائية في الباقيات لهذا النموذج. ولقد تم أيضا توليد السلسلة المستقبلية حسب النموذج وللفترة من تشرين الأول 2009 ولغاية أيلول ٢٠١١ وعند مقارنتها مع القيم الفعلية المسجلة وجد تطابقا جيدا مما يؤكد ملائمة النموذج.

KEY WORDS: Box-Jenkins; Stochastic Model; Time Series; Seasonality.

Introduction

Stochastic models are widely used in forecasting flow, water quality parameters, rainfall and other hydrologic phenomena. Time series analysis approach has been used to develop parametric models for (daily, weekly, monthly, …) data by many researchers like (Mahloch, 1974), (Stidinger, 1981), (Cooper and Wood, 1982) and (Salas and Abedlmohsen, 1993).

A model which describes the probability structure of a sequence of observations is called a "stochastic process ". A time series of N successive observations $z = (z_1 ,z_2 ,..., z_n)$ is regarded as a sample realization, from an infinite population of such samples, which could have been generated by the process. An important class of stochastic processes is the stationary processes. They are assumed to be in a specific form of statistical equilibrium and in particular vary about a fixed mean. Particular stationary stochastic processes of value in modeling time series are the autoregressive (abbreviated AR), moving average (abbreviated MA), and mixed autoregressive moving average processes (abbreviated ARMA). Another class of stochastic processes which is non stationary processes like autoregressive integrated moving average (abbreviated ARIMA) models (Box and Jenkins, 1976). **Al-Suhaili (1986)** used singlesite AR(1), autoregressive integrated moving average(ARIMA (1,0,1)) and (matalas model) for four Tigris river flow stations. **Mahmood (2000)** applied AR(1), AR(2) and ARIMA(1,1,1) models in analyzing monthly time series of water quality data on the Euphrates river at Kufa city. **Abed (2007)** applied Box-Jenkins seasonal multiplicative model of order $(0, 1, 1) \times (0, 1, 1)_{12}$. to monthly records of some physical and chemical properties of river water in Babylon, Najaf, and Diwaniya governorates. **Al-Ta'ee (2009)** applied ARIMA models to records of rainfall and evaporation at Babylon governorate. **Ali (2009)** fitted three Box-Jenkins seasonal multiplicative models to monthly inflow to bekhem reservoir. **Al-Masudi (2011)** fitted seven seasonal multiplicative models to monthly inflow of Dokan reservoir.

Stochastic Methods

Box and Jenkins (1976) have generalized the autoregressive integrated moving average ARIMA (p,d,q) model to deal with seasonality and define a general multiplicative seasonal model in the form :

$$\phi_p(\beta)\Phi_P(\beta^s)W_t = \theta_q(\beta)\Theta_Q(\beta^s)a_t \qquad \text{..............................(1)}$$

or

$$\left(1-\sum_{j=1}^{p}\phi_j\beta^j\right)\left(1-\sum_{j=1}^{P}\Phi_j\beta^{sj}\right)W_t = \left(1-\sum_{j=1}^{q}\theta_j\beta^j\right)\left(1-\sum_{j=1}^{Q}\Theta_j\beta^{sj}\right)a_t \qquad \text{.........(2)}$$

Where:
ϕ ,Φ=autoregressive parameters
θ ,Θ=moving average parameters

β=backward shift operator such that:

$\beta Z_t=Z_{t-1}$,and $\beta^s Z_t=Z_{t-s}$,

$Z_t=\ln X_t$,

X_t=observed inflow (m^3/sec.),

$W_t = \nabla^d \nabla_{12}^D Z_t$,and

∇ =difference operators such that:

$\nabla^d Z_t = Z_t - Z_{t-d}$,

$\nabla_{12}^D Z_t = Z_t - Z_{t-12D}$,

D,d=degrees of differencing,
s= number of seasons (=12 for monthly data), and a_t's are drawings from a distribution of zero mean and constant variance ,i.e., white noise.

This model ,i.e., Equation 1, is said to be of order (p,d,q)×(P,D,Q)$_s$. For example, if p=P=0, d=D=1, and q=Q=1, equation 2 becomes:

$$W_t = (1-\theta\beta)(1-\Theta\beta^{12})a_t \qquad \text{....................................(3)}$$

This model is called Box-Jenkins seasonal multiplicative model or Box-Jenkins seasonal moving average model of order (0, 1, 1) × $(0, 1, 1)_{12}$.By the same way the equation of the model (0, 1, 2) × $(0, 1, 1)_{12}$ is become:

$$W_t = (1 - \theta_1 \beta - \theta_2 \beta^2)(1 - \Theta \beta^{12}) a_t \qquad \text{..........................(4)}$$

The main objective of the present study is to apply the Box and Jenkins model to monthly in flow to Haditha reservoir in middle west of Iraq.

Description of Haditha Dam

Haditha dam is a multi-purpose hydro-development designed to control the Euphrates River flow in the interests of irrigation, electric power generation and for partial accumulation of extreme Euphrates River inflows into Haditha reservoir.

Haditha dam was constructed on the Euphrates River in the Middle West of Iraq 7km upstream from Haditha town. In 1988 the project was completed. The project generates (660 Mw) of electrical power a side from performing its flood control function. Central and southern parts of Iraq get the benefit of irrigation water from its reservoir. Figure (1) shows a general layout of the project (**Hydroprojekt**, 1988).

The means of inflows to Haditha reservoir for the entire period of record of 10 years from Oct.,1999 to Sep.,2008 are plotted as shown in figure (2). The periodic behavior of this series is show a marked seasonal pattern.

Fitting Box-Jenkins Models to Haditha Inflow

The seasonal ARIMA (p,d,q)×(P,D,Q) model –which is called "seasonal multiplicative autoregressive integrated moving average model "or Box – Jenkins seasonal model –is to fitted to a time series by using a three-stages procedure. These stages are model identification, estimation of model parameters and diagnostic checking of the estimated parameters.

First of all the observations are plotted against time this will show up important features such as trend, seasonality, discontinuities and outliers. Figures (2) show that there is little trend ,high fluctuation, and seasonal variation for monthly inflow.

The procedure of fitting is summarized by the following steps(**Box and Jenkins, 1976**):

1. Transform data using natural log transformation which was found the most appropriate.
2. Removing trend component by using the first order differencing.
3. Removing the seasonal variation by using the first order seasonal differencing.
4. Model identification by plotting ACF and PACF of monthly observations.

Identification of Representative Models

By identification it is meant the use of the data, and of any information on how the series was generated, to suggest a subclass of models from the general Box-Jenkins family , i.e., Equation 1 for further examination. In other word, identification provides clues about the choice of the order of p ,d ,q ,P ,D ,and Q. However, in practice the degrees of differencing d and D are assumed 1 while autocorrelation and partial autocorrelation function are plotted to guess the order of p ,q ,P ,and Q.

The estimation autocorrelation and partial autocorrelation function as shown in figures (3) and (4) respectively are characterized by correlations and autocorrelations which alternate in sign and which tend to damp out with increasing lags.

Estimation of the Model Parameters

The unconditional sum of square method is used to estimate the model parameters.

Trial values for the parameters of the $(0,1,2)\times(0,1,1)_{12}$ model ,i.e. θ_1 , θ_2 and Θ in Equation (4) are assumed and the sum of squares ,SS , is computed. Computation of SS is repeated with different values of θ_1 , θ_2 and Θ until minimum sum of squares is obtained. To illustrate this procedure results are given in Table 1 for θ_1=0.368, θ_2=0.321 and Θ=0.910 corresponding to the minimum sum of squares of 12.017439.

The number of observations is N=120 , n=N-d-SD =120-1-12= 107 is the number of dependent stochastic component (W_t), d is the degree of simple difference (d=1), D is the degree of seasonal difference (D=1), and S is the length of periodic cycle (S=12).

The values of both θ_1 , θ_2 and Θ vary from -1 to 1. We choose $\theta_1 = \theta_2 = \Theta$ -1 as trial values and sum of square (SS) is computed. The procedure is repeated with different values of θ_1 , θ_2 and Θ until we obtain minimum sum of squares. Table (1) shows calculation corresponding to specific assumed values of θ_1 , θ_2 and Θ. Then, the model may be written in either forward form as shown in the following equation(**Box and Jenkins, 1976**):

$$[a_t]=[W_t]+\theta_1[a_{t-1}]+\theta_2[a_{t-2}]+\Theta[a_{t-12}]-\theta_1\Theta[a_{t-13}]-\theta_2\Theta[a_{t-14}]$$

……..……..…..(5)

or backward form as follows:

$$[e_t]=[W_t]+\theta_1[e_{t+1}]+\theta_2[e_{t+2}]+\Theta[e_{t+12}]-\theta_1\Theta[e_{t+13}]-\theta_2\Theta[e_{t+14}]$$

………..…....(6)

Hence, it is convenient to use a numbering system so that the first observation in the X_t series (the second column in table 1) has a subscript -12, the last observation has a subscript 107,.The beginning of calculation is done by equation (6),i.e. when t=107, to compute the $[e_t]$'s (column 6) for t=107, 106,…..,1 by setting $[e_t]$ =0 for t=108 , 109,and so on .Then the $[W_t]$ series for t= 0,-1, -2,…., -12 is computed by equation (6) by setting $[e_t]$ =0 for t=0, -1,…., -12. After that $[a_t]$'s (in fourth column) are computed by equation (5) for t= -12 ,-11,…..,107 by setting $[a_t]$'s=0 for t=-13,-14,and so on. Hence ,the sum of squares(SS) is computed by the following equation:

$$SS=\sum_{t=-12}^{t=107} a_t^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

Table (1) shows that the (SS=12.017439 when θ_1=0.368, θ_2=0.321 and Θ=0.910).Therefore, the model equation is:

$$W_t=(1-0.368\beta-0.321\beta^2)(1-0.910\beta^{12})a_t \quad \ldots\ldots\ldots\ldots (8)$$

The result of application of the unconditional sum of squares method to estimate the parameters for the other model $(0, 1, 1) \times (0, 1, 1)_{12}$ is (SS=13.45575 when θ_1=0.488 and θ_2=0.902).

Diagnostic Check

After estimating the model parameters, the diagnostic checking is applied to see if the model is adequate or not. Therefore the following statistical tests are used:

1. Port Manteau Lack of fit Test

Portmanteau lack of fit test is used for this purpose. It is a test of the residual independency and uses the Q-statistic defined as :

$$Q = (N - d - DS)\sum_{k=1}^{M} r_k^2 a_t \quad \text{……………………………………………} (9)$$

where $r_k(a_t)$is the autocorrelation coefficient of the residual (a_t) at lag k ,and M is the maximum lag considered (about N/4) **(Chatfield,1982**), ARIMA model is considered adequate if $Q<\chi^2_{\alpha,(M-np)}$ where α is the level of significant , np is the number of model parameters, and the expression (M-np) represents the degree of freedom.

The results of this test indicate that model is adequate because the calculated Q= (36.4005) less the χ^2- table (43.192) with 27 degree of freedom (M-P-p-Q-q) at 97.5% confidence limits where M=30 for monthly data.

2. Residual autocorrelation Function (RACF) Test

The second test is the independency of the resulting (a_t) series, the correlogram of this series are computed for lag (M=N/5) are shown in figure (5) . The figure shown that the most of computed lags lie inside the tolerance interval ($\pm 2/\sqrt{N}$, at 95% confidence limits). Hence, the suggested model can be considered as appropriate model because of its capability of removing the dependency from data.

Forecasting

Forecasted monthly data are computed for 3 years a head of original data for the period from 2009 to 2011 by applying the following equation (**Box and Jenkins, 1976**):

$$\hat{Z}_t(\ell) = Z_{t+\ell} = Z_{t+\ell-1} + Z_{t+\ell-12} - Z_{t+\ell-13} + a_{t+\ell} - \theta_1 a_{t+\ell-1} - \theta_2 a_{t+l-2} - \Theta a_{t+\ell-12} + \theta_1 \Theta a_{t+\ell-13} + \theta_2 \Theta a_{t+l-14} \text{…..}(10)$$

where t is the origin time (t=120) and ℓ is the lead time (=1,2,…36).After obtaining the forecasted series (Z_t for t=121,122,123….156) ,then the final series (X_t) is determined by reversing (ln) transformation .Figure (6) show the forecasted series for these data. The corresponding observed values are also shown in the figure and since agreement between observed and forecasted values is very good, it is confirmed that the model is adequate.

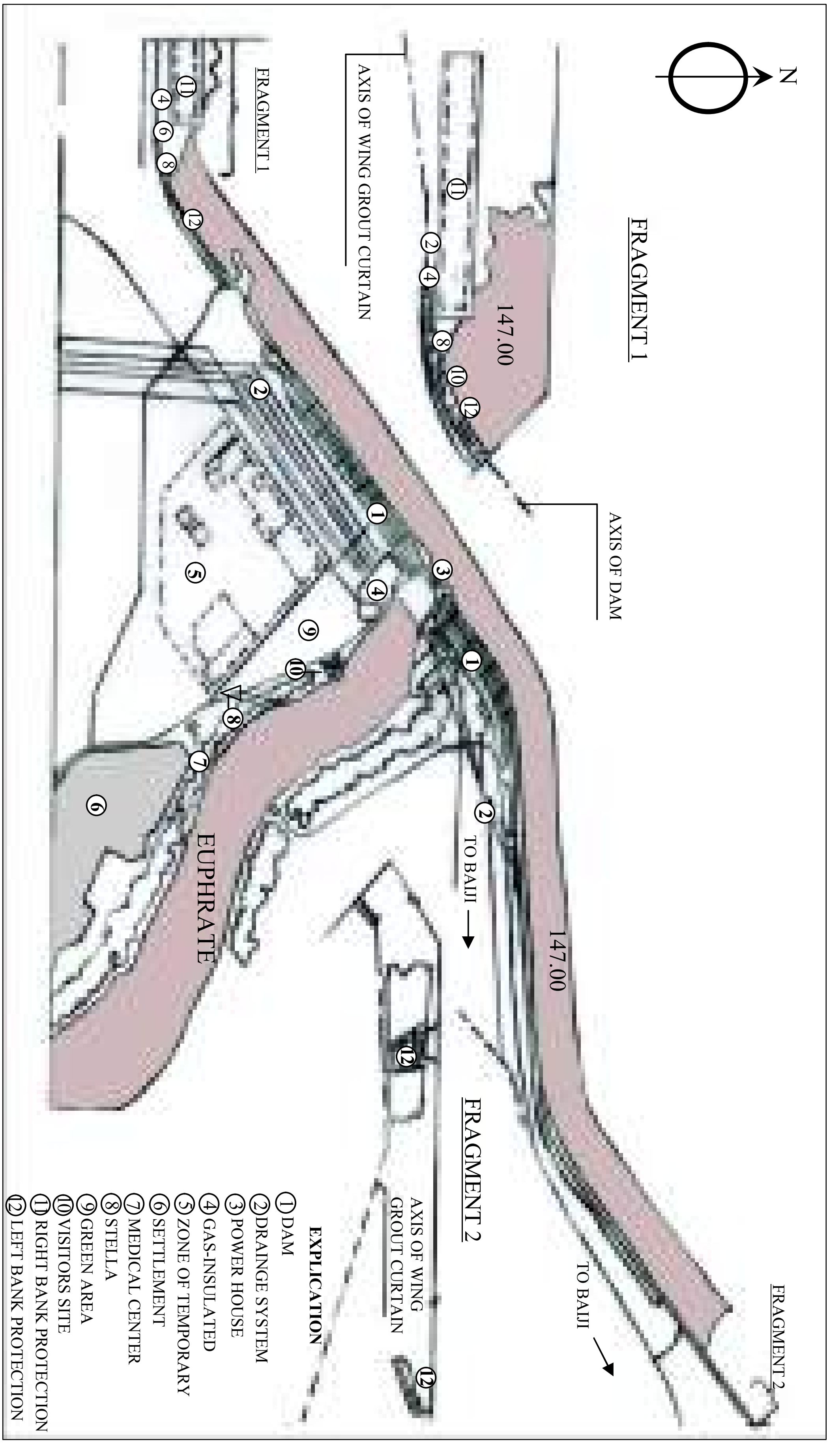
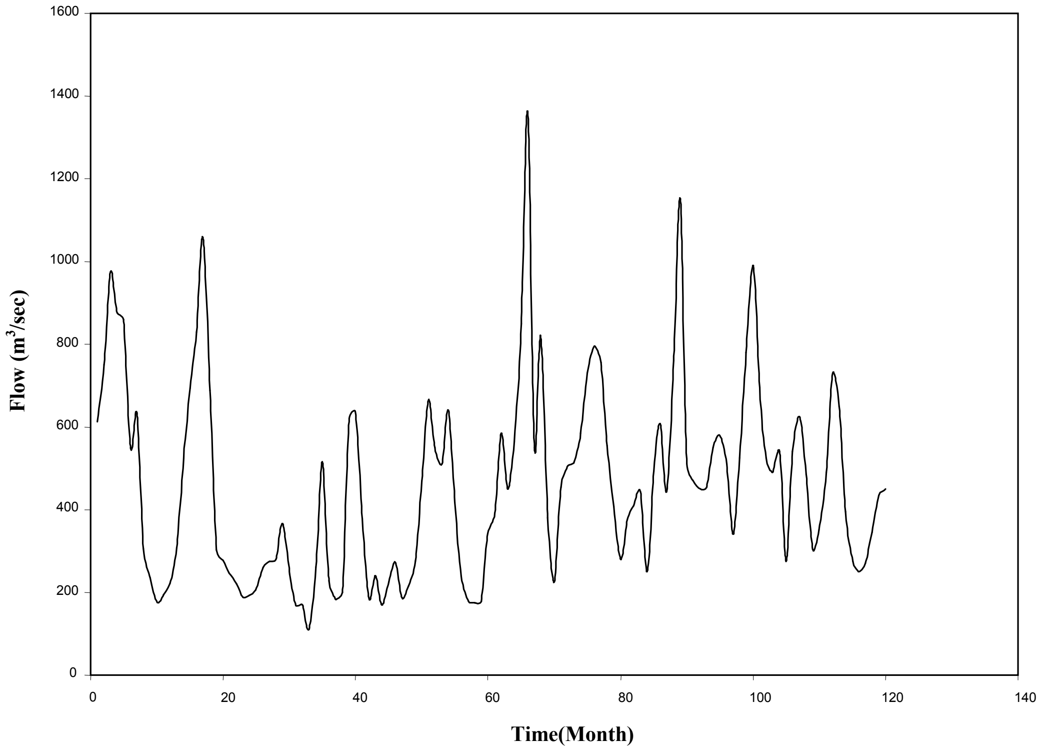
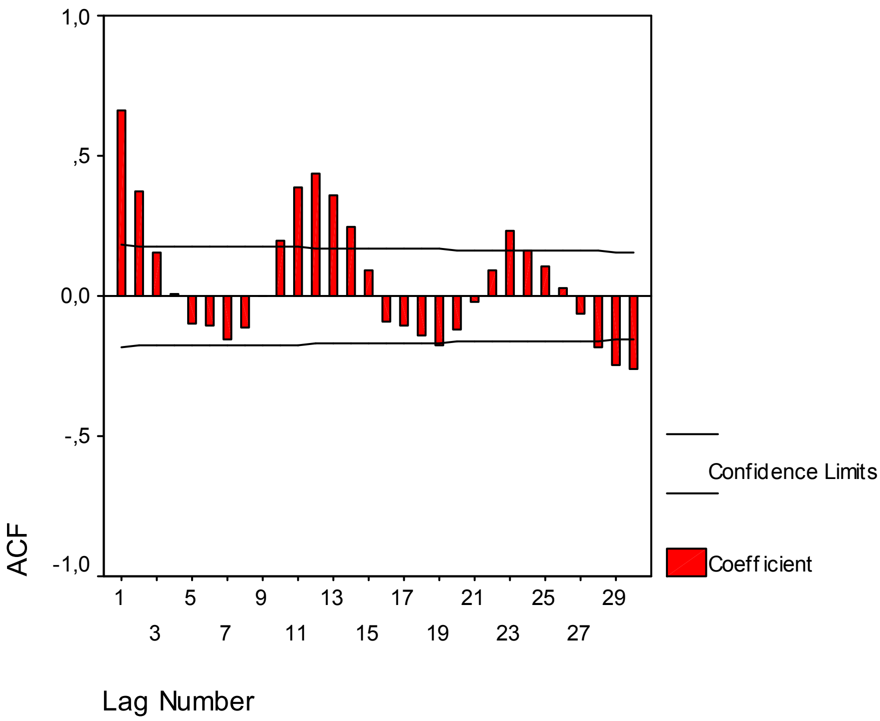


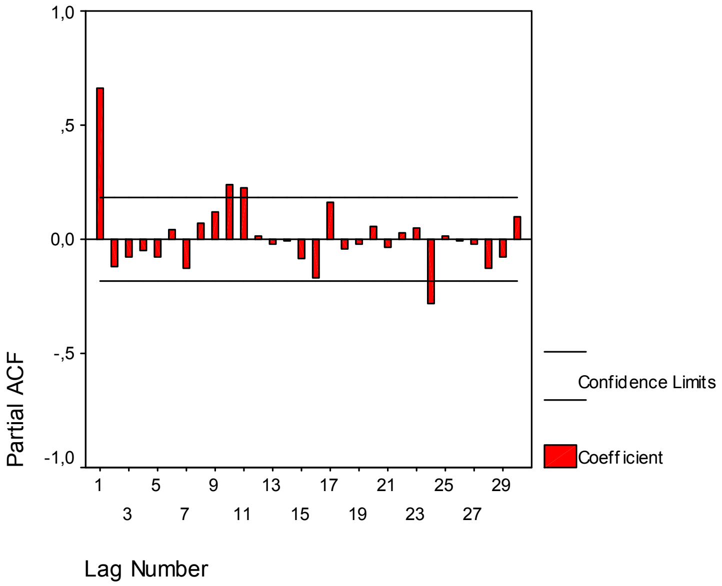
Figure (1): Schematic general layout for Haditha Dam(**Hydroprojekt**, 1988)



Figure(2): Monthly Observed Inflow to Haditha Reservoir from Oct., 1999 to Sep., 2008(**Ministry of water resources**, **2011**).



Figure(3): Autocorrelation Function for the Series.

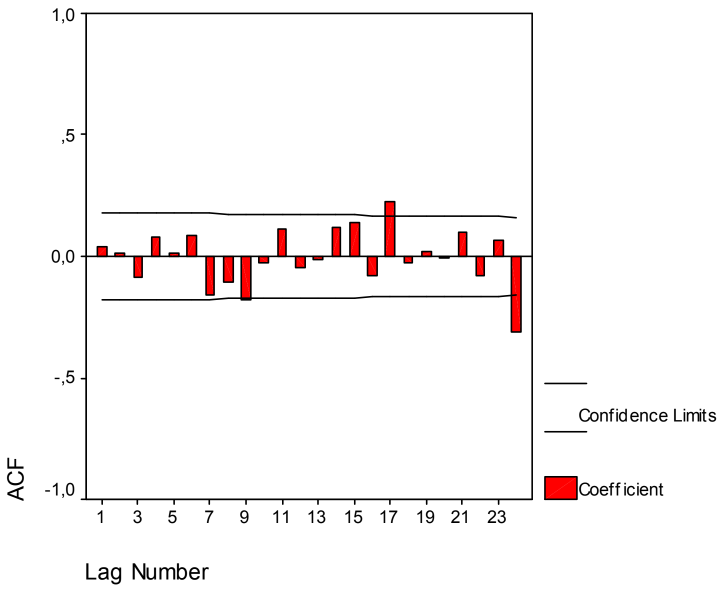


Figure(4): Partial Autocorrelation Function for the Series.

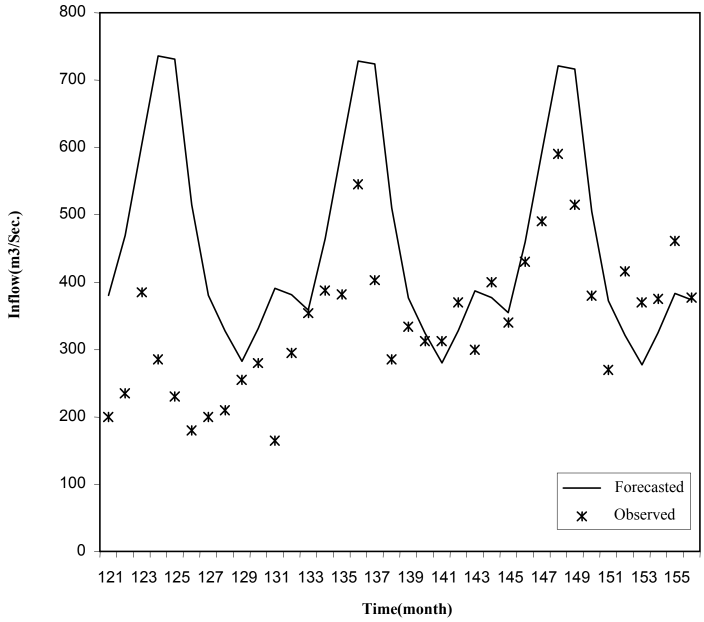
Table(1): Calculation of sum of squares for model $(0,1,2)\times(0,1,1)_{12}$ with $\theta_1=0.368$, $\theta_2=0.321$ and $\Theta=0.910$

t(month)	X_t	$Z_t=\ln X_t$	a_t	W_t	e_t	a^2
-12	613	6.418365	0.147125154	0.147125154	0	0.021645811
-11	750	6.620073	0.02111979	-0.033022267	0	0.000446046
-10	975	6.882437	0.025621828	-0.029377429	0	0.000656478
-9	875	6.774224	-0.311103331	-0.327311616	0	0.096785283
-8	860	6.756932	-0.096896619	0.0093648	0	0.009388955
-7	550	6.309918	-0.190599069	-0.055076944	0	0.036328005
-6	631	6.447306	0.356195435	0.457439707	0	0.126875188
-5	305	5.720312	-0.491782175	-0.561679793	0	0.241849707
-4	235	5.459586	-0.239649338	-0.173012233	0	0.057431805
-3	176	5.170484	-0.693148427	-0.447095393	0	0.480454742
-2	194	5.267858	-0.408382614	-0.076376555	0	0.166776359
-1	222	5.402677	-0.297875905	0.074909542	0	0.088730055
0	307	5.726848	0.213592195	0.320417457	0	0.045621626
1	525	6.263398	0.287775744	0.334842244	0.2577119	0.082814879
2	694	6.542472	0.164440542	0.016709433	0.2082181	0.027040692

⋮	⋮	⋮	⋮	⋮	⋮	⋮
94	625	6.437752	0.051772936	0.093072685	0.1281152	0.002680437
95	485	6.184149	-0.057300307	-0.124985381	0.0059274	0.003283325
96	305	5.720312	-0.458288778	-0.058372006	-0.1403493	0.210028604
97	350	5.857933	-0.326316428	-0.296831267	-0.2463165	0.106482411
98	490	6.194405	-0.176390174	-0.05942342	0.0270005	0.031113494
99	730	6.593045	0.053639214	0.16022812	0.1264132	0.002877165
100	650	6.476972	-0.157862645	0.237720926	0.1243111	0.024920615
101	395	5.978886	-0.202958577	-0.217573015	-0.2478549	0.041192184
102	275	5.616771	-0.189685545	-0.293121796	-0.0691564	0.035980606
103	250	5.521461	-0.075526775	-0.192473928	-0.0150540	0.005704294
104	270	5.598422	0.157176952	0.751759083	0.7149698	0.024704594
105	345	5.843544	0.127650858	-0.429675584	-0.2669439	0.016294742
106	435	6.075346	0.171812733	0.085619104	0.1914207	0.029519615
107	450	6.109248	0.168568394	0.28750431	0.2875043	0.028415304
						12.01743926



Figure(5): Autocorrelogram of Residual Series Parameter.



Figure(6): Comparison of Forecasted and Observed Inflow (Oct. 2009- Sep.2011).

Conclusions

The time series(October 1999-September 2008) of monthly inflow to Haditha reservoir is a periodic series and , therefore, the stochastic model which represents it is a seasonal one. The models(0,1,1) × $(0,1,1)_{12}$ and (0,1,2) × $(0,1,1)_{12}$ are fitted to the series and it is found that sum of squared errors of the (0,1,2)×$(0,1,1)_{12}$ model with autoregressive moving average parameters of θ_1 =0.368 , θ_2 =0.321 and Θ =0.910 is less than the other model. The diagnostic checking show that model is adequate. Forecasts using the model for the period from October, 2009 to September 2011 agrees well with the observed values.

References

Abed, Z.A. Al-Ridah,(2007):"**Stochastic Models of Some Properties of water in the middle of Euphrates Region in Iraq**", M. Sc. Thesis ,College of Engineering, University of Babylon.

Ali, S.T.,(2009):"**Fitting Seasonal Stochastic Models to Inflow of Bekhme Reservoir**", The Iraqi Journal for Mechanical and Material Engineering , Special Issue(A), College of Engineering, University of Babylon.

Al-Masudi, R. K.M.,(2011): "**Fitting ARIMA Models for Forecasting to inflow of Dokan Reservoir**", Journal of Babylon University, Vol.19, No.4.

Al-Ta'ee, M.H.A.,(2009):" **Analysis of Records of Rainfall and Evaporation in Babylon"**, M. Sc. Thesis, College of Engineering, University of Baghdad.

Al-Suhaili, R.H.,(1986):"**Stochastic Analysis of Daily Streamflow of Tigris River**", M. Sc. Thesis , College of Engineering, University of Baghdad.

Box, G.E.P. and Jenkins, G.M.,(1976):"**Time Series Analysis; Forecasting and Control.** ",Holden Day , San Francisco, California.

Chatfield, C.,(1982):"**The Analysis of Time Series; An Introduction**", Chapman and Hill, 2nd Edition.

Chow, V.T., Maidment, D.R., and Mays, L.W.,(1988):"**Applied Hydrology**", Mc Graw-Hill.

Cooper, D. M., Wood,(1982): "**Parameter Estimation of Multiple Input-Output Time Series Models: Application to Rainfall-Runoff Process**", Journal of Water Resource Research, Vol.18, No.5, Pages 1352-1364.

Hydroprojekt,(1988):"**Haditha project on the Euphrates River, operation and maintenance manual for earth-fill dam**", Book1, No.950-10-76, Moscow.

Mahloch, J. L.,(1974): "**Multivariate Techniques for Water Quality Analysis**", Journal of Environmental Engineering Division, Vol.100, No.EE5.

Mahmood, H.S.,(2000):"**Time Series Analysis of Hydrochemical Pollutant: Application of Box-Jenkins Models on Euphrates River at Kufa**", M. Sc. Thesis College of Engineering, University of Babylon.

Ministry of water resources,(2011):" **Report on Haditha dam Project"**.Salas, J. D. and Abdulmohsen, M. W.,(1993):"**Initializing for Generation Single-Site and Multisite Low-Order Periodic Autoregressive and Moving Average Process**", Journal of Water Resource Research, Vol.29, No.6, Pages 1771-1776.

Stedinger, J. R.,(1981):"**Estimation Correlations in Multivariate Stream Flow Models**", Journal of Water Resource Research, Vol.17, No.1.