Fitting ARIMA Models for Forecasting to Inflow of Dokan Reservoir

Ruqaya K.M. Al-Masudi
College of Engrg., Babylon Univ., Babil, Iraq

Abstract

In this study Box-Jenkins seasonal model is applied to records of mean flow to Dokan reservoir in the north of Iraq for the period from water year 1953/1954 to water year 2004/2005.Seven Box-Jenkins seasonal multiplicative models are fitted to this series. These are $(1,1,1)\times(1,1,1)_{12}$, $(1,1,1)\times(0,1,1)_{12}$, $(1,1,2)\times(0,1,1)_{12}$,$(0,1,1)\times(0,1,1)_{12}$,$(2,1,0)\times(1,1,0)_{12}$,$(1,1,0)\times(1,1,0)_{12}$,$(1,1,1)\times(0,1,0)_{12}$ models. The unconditional sum of squares method is used to estimate the parameters of the models and to compute the sum of square errors for each of the fitted model. It is found that the model which corresponds to the minimum sum of squared errors is the $(1,1,1)\times(1,1,1)_{12}$ model with parameters $\theta=0.899$,$\Phi=0.60$,$\Theta=0.96$, and $\Phi=0.03$. Portmanteau lack of fit test is used as diagnostic checking. Forecasts of monthly inflow for the period from October, 2005, to September,2007 are graphically compared with observed inflow for the same period and since agreement is very good adequacy of the selected model is confirmed.

الخلاصة

في هذه الدراسة تم تطبيق طريقة بوكس–جنكينز (Box-Jenkins) للنموذج الفصلي (model) للجريان الشهري إلى سد دوكان في شمال العراق للفترة من بداية السنة المائية 1955/1954 وحتى نهاية السنة المائية 2004-2005. حيث تمت مطابقة سبعة أنواع من النموذج التصادفي الفصلي وهي النموذج $(1,1,1)\times(1,1,1)_{12}$ والنموذج $(1,1,1)\times(0,1,1)_{12}$ والنموذج $(1,1,2)\times(0,1,1)_{12}$ والنموذج $(0,1,1)\times(0,1,1)_{12}$ والنموذج $(2,1,0)\times(1,1,0)_{12}$ والنموذج $(1,1,0)\times(1,1,0)_{12}$ والنموذج $(1,1,1)\times(0,1,0)_{12}$. إن نتائج استخدام طريقة مجموع المربعات غير المشروطة لتقدير معالم النماذج أظهرت بان مجموع مربعات الأخطاء للنموذج $(1,1,1)\times(1,1,1)_{12}$ بمعالم $\theta=0.899$ $\phi=0.600$ $\Theta=0.960$ و $\Phi=0.030$ اقل من باقي النماذج، كما وان الفحص المعروف بفحص مخطط الذبذبة لم يبين وجود عشوائية في الباقيات لهذا النموذج. لقد تم أيضا توليد السلسلة المستقبلية حسب النموذج وللفترة من تشرين الأول 2005 ولغاية أيلول 2007 وعند مقارنتها مع القيم الفعلية المسجلة وجد تطابقا جيدا مما يؤكد ملائمة النموذج.

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

Introduction

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.. **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. These are $(1,1,0)\times(1,1,0)_{12}$, $(2,1,0)\times(1,1,0)_{12}$,$(0,1,1)\times(0,1,1)_{12}$.It is found that the model which corresponding to the minimum sum of squared error is the ,$(0,1,1)\times(0,1,1)_{12}$.

In this study we fit other types of models which are the seasonal models called Box-Jenkins models to inflow of Dokan reservoir. Application of these seasonal models to data from Iraq is not found in the literature.

Stochastic Processes

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 :

$$\phi_p(\beta)\Phi_P(\beta^s)W_t = \theta_q(\beta)\Theta_Q(\beta^s)a_t \quad \ldots\ldots(1)$$

or

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

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 $(p,d,q)\times(P,D,Q)_s$. For example, if p=P=1, d=D=1, and q=Q=1, equation 2 becomes:

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

This model is called Box-Jenkins seasonal multiplicative model or Box-Jenkins seasonal moving average model of order (1, 1, 1) × $(1, 1, 1)_{12}$.By the same way we obtained the equations of the other models and the results are listed in table below.

Model	Model Equation
$(1,1,1)\times(1,1,1)_{12}$	$(1-\phi\beta)(1-\Phi\beta^{12})W_t = (1-\theta\beta)(1-\Theta\beta^{12})a_t$
$(1,1,1)\times(0,1,1)_{12}$	$(1-\phi\beta)W_t = (1-\theta\beta)(1-\Theta\beta^{12})a_t$
$(1,1,2)\times(0,1,1)_{12}$	$(1-\phi\beta)W_t = (1-\theta_1\beta-\theta_2\beta^2)(1-\Theta\beta^{12})a_t$
$(0,1,1)\times(0,1,1)_{12}$	$W_t = (1-\theta\beta)(1-\Theta\beta^{12})a_t$
$(2,1,0)\times(1,1,0)_{12}$	$(1-\phi_1\beta-\phi_2\beta^2)(1-\Phi\beta^{12})W_t = a_t$
$(1,1,0)\times(1,1,0)_{12}$	$(1-\phi\beta)(1-\Phi\beta^{12})W_t = a_t$
$(1,1,1)\times(0,1,0)_{12}$	$(1-\phi\beta)W_t = (1-\theta\beta)a_t$

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

Inflow to Dokan Reservoir

The Dokan dam is located at about 60km from the north west of Al-Sulaimania town and at about 300 km from Kerkok governorate as shown in figure 1).The main purposes of the Dokan project are to store and regulate the abundant water of the Lesser Zab River, a tributary of the Tigres River, by creating a large scale reservoir, to supply irrigation water required in the area downstream of the dam, and to control discharges downstream by impounding and regulating floods.

In addition to the abovementioned purposes, the discharge and head obtained by the dam are to be utilized for power generation for effective use of hydraulic energy, thereby making this a multi-purpose, for irrigation, flood control, and power generation.

In figure 2 we plot the means of inflows to Dokan reservoir for the entire period of record of 48 years from Oct.,1953 to Sep.,2004.To show the periodic behavior of this series only a small part of it corresponding to the period from water year 1953/1954 to water year 1957/1958 is given in figure 3 which, as expected, shows a marked seasonal pattern.

Fitting Box-Jenkins Models to Dokan 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 4 and 5 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 $(1,1,1)\times(1,1,1)_{12}$ model ,i.e. ,θ ,ф ,Θ ,and Φ in Equation 3 are assumed and the sum of squares ,SS , is computed. Computation of SS is repeated with different values of θ ,ф ,Θ ,and Φ until minimum sum of squares is obtained. To illustrate this procedure results are given in Table 1 for θ=0.899,Φ=0.60,Θ=0.96, and Φ=0.03 corresponding to the minimum sum of squares of 78.40.

The number of observations is N=624 , n=N-d-SD =624-1-12= 611 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 ,ф ,Θ ,and Φ vary from -1 to 1. We choose θ =ф= Θ= Φ=-1 as trial values and sum of square (SS) is computed. The procedure is repeated with different values of θ ,ф , Θ , and Φ until we obtain minimum sum of squares. Table 1 shows calculation corresponding to specific assumed values of θ ,ф , Θ , and Φ. Then, the model may be written in either forward form as shown in the following equation(**Box and Jenkins, 1976**):

$$[a_t]=[W_t]-\phi[W_{t-1}]-\Phi[W_{t-12}]+\phi\Phi[W_{t-13}]+\theta[a_{t-1}]+\Theta[a_{t-12}]-\theta\Theta[a_{t-13}] \quad \text{.............}(4)$$

or backward form as follows:

$$[e_t]=[W_t]-\phi[W_{t+1}]-\Phi[W_{t+12}]+\phi\Phi[W_{t+13}]+\theta[e_{t+1}]+\Theta[e_{t+12}]-\theta\Theta[e_{t+13}] \quad \text{.........}(5)$$

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 611,.The beginning of calculation is done by equation 5,i.e. when t=611, to compute the $[e_t]$'s (column 6) for t=611, 610,…..,1 by setting $[e_t]$ =0 for t=612 , 613,and so on .Then the $[W_t]$ series for t= 0,-1, -2,…., -12 is computed by equation (5) by setting $[e_t]$ =0 for t=0, -1,…., -12. After that $[a_t]$'s (in fourth column) are computed by equation 4 for t= -12 ,-11,…..,-611 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:

$$SS = \sum_{t=-12}^{t=611} a_t^2 \quad \text{..}(6)$$

Table 1 shows that the (SS=78.40 when ф=0.60, Φ=0.03,θ=0.899, and Θ=0.96).Therefore, the model equation is:

$$W_t(1-0.60\beta)(1-0.03\beta^{12})=(1-0.899\beta)(1-0.96\beta^{12})a_t \quad \text{.............................}(7)$$

The results of application of the unconditional sum of squares method to estimate the parameters for the all models are listed in table below:

Model	Parameters						SS
	θ_1	θ_2	ϕ_1	ϕ_2	Θ	Φ	
$(1,1,1)\times(1,1,1)_{12}$	0.899	-	0.600	-	0.960	0.030	78.40
$(1,1,1)\times(0,1,1)_{12}$	0.870	-	0.569	-	0.955	-	78.54
$(1,1,2)\times(0,1,1)_{12}$	-0.380	0.390	-0.758	-	0.910	-	79
$(0,1,1)\times(0,1,1)_{12}$	0.380	-	-	-	0.926	-	80
$(2,1,0)\times(1,1,0)_{12}$	-	-	-0.370	-0.220	-	-0.500	106
$(1,1,0)\times(1,1,0)_{12}$	-	-	-0.322	-	-	-0.501	111
$(1,1,1)\times(0,1,0)_{12}$	0.501	-	0.122	-	-	-	140

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 Lak 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 :

$$Q = \left(N - d - DS\right)\sum_{k=1}^{M} r_k^2 a_t \quad \text{...} \quad (8)$$

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**),SARIMA 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= (157.79) less the χ^2- table (188.01) with 152 degree of freedom (M-P-p-Q-q) at 97.5% confidence limits where M=156 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 (±2/√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 2005 to 2007 by applying the following equation (**Box and Jenkins, 1976**):

$$\hat{Z}_t(\ell)=Z_{t+\ell}=Z_{t+\ell-1}+Z_{t+\ell-12}-Z_{t+\ell-13}+a_{t+\ell}-\theta a_{t+\ell-1}-\Theta a_{t+\ell-12}+\theta\Theta a_{t+\ell-13}+\phi W_{t+\ell-1}+\Phi W_{t+\ell-12}-\phi\Phi W_{t+\ell-13}\text{........}(9)$$

where t is the origin time (t=624) and ℓ is the lead time (=1,2,…36).After obtaining the forecasted series (Z_t for t=625,626,627….660) ,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.

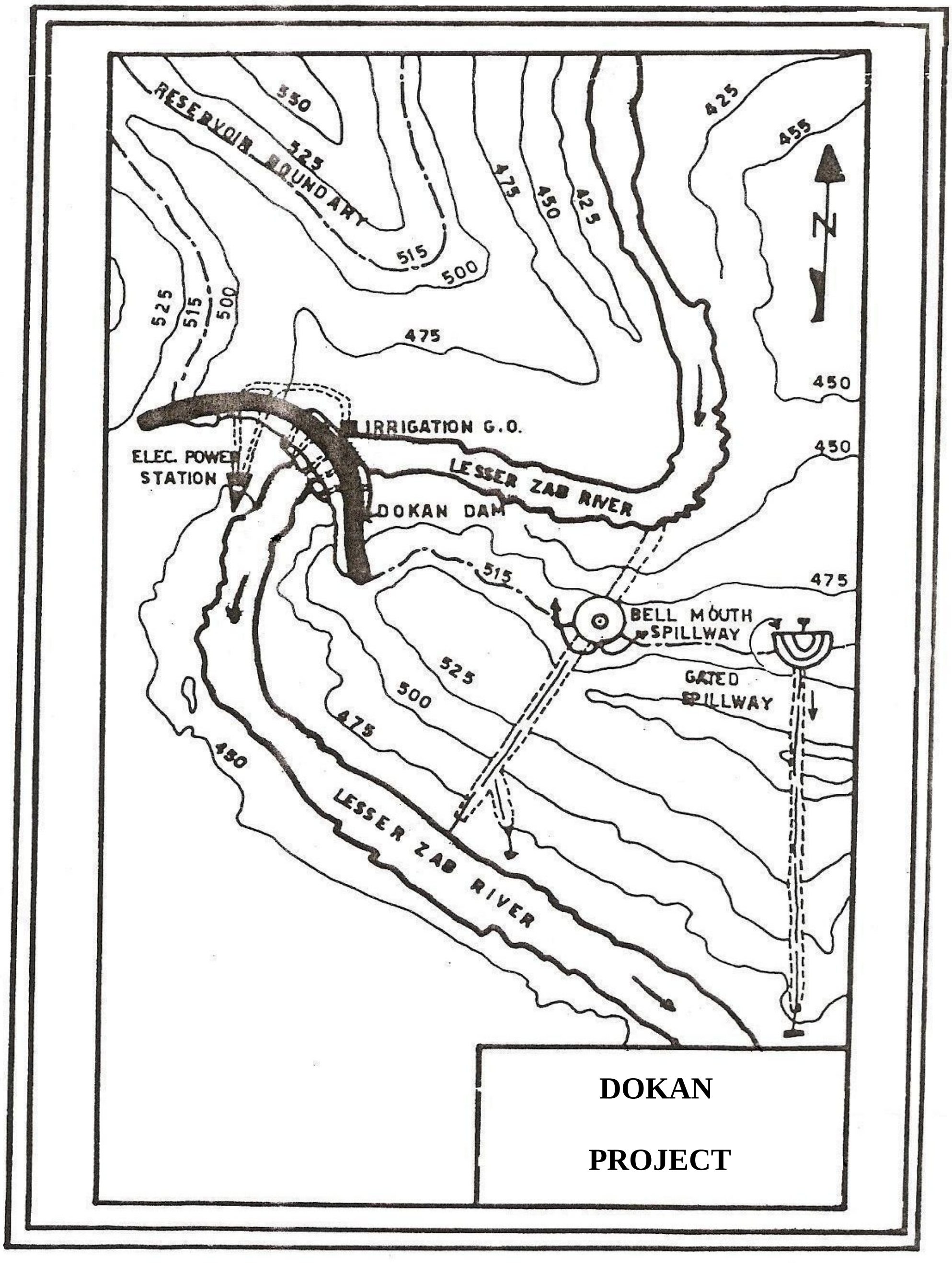

Figure 1: Sit Map of Dokan Reservoir.
(Planning Report on Dokan Dam Project,2007)

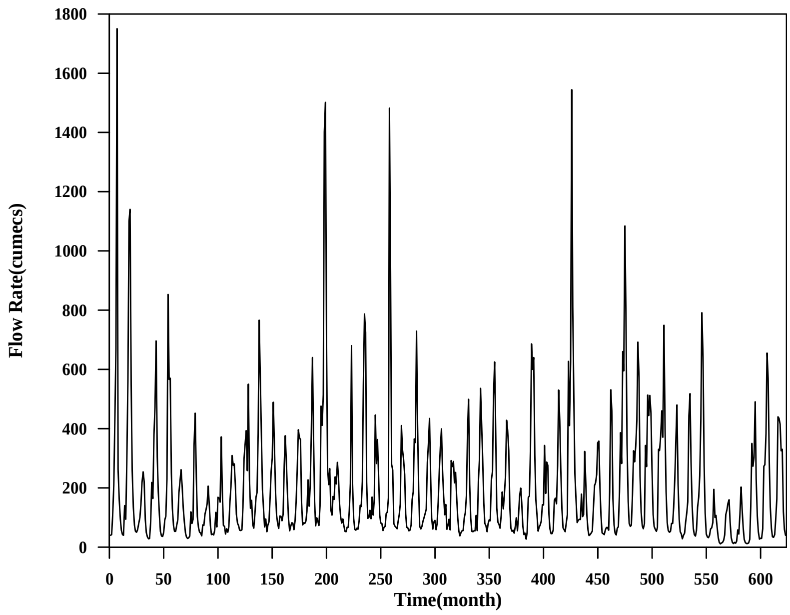

Figure 2: Monthly Inflow to Dokan Reservoir for the Period (Oct,1953-Sep.,2004).

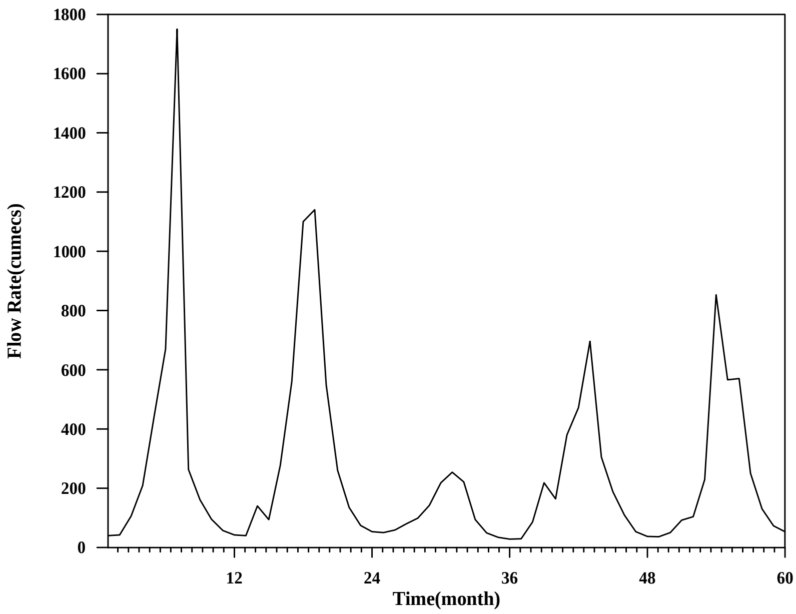

Figure 3: Monthly Inflow to Dokan Reservoir for the Period (Oct.,1953-Sep.,1957).

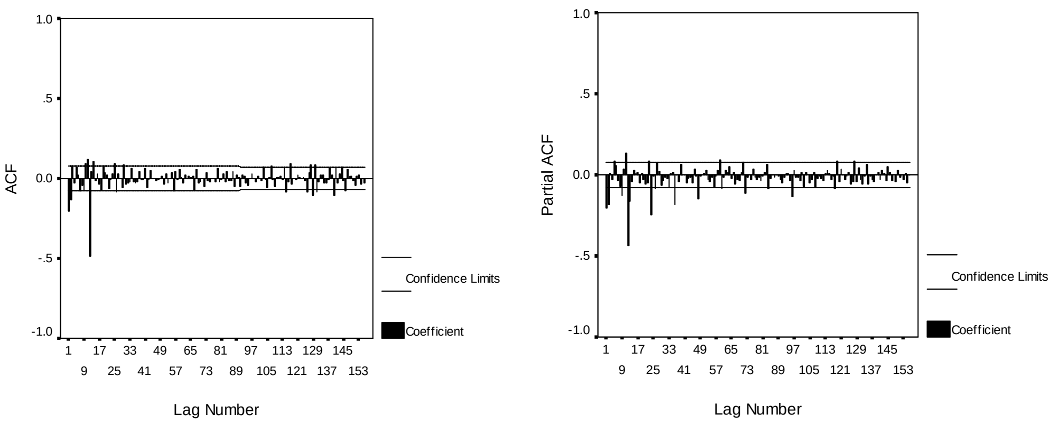

Figure 4: ACF and PACF of the series.

Table 1: Sum of squares(SS) computation for (1,1,1)×(1,1,1)$_{12}$ model with ϕ=0.60, Φ=0.03, θ=0.899, and Θ=0.96.

t(month)	X_t	Z_t=lnX_t	a_t	W_t	e_t	a^2
-12	40	3.688879	-0.09813	-0.098135	0	0.00963
-11	42	3.73767	-0.43232	-0.40298	0	0.186902
-10	106	4.663439	0.357215	0.50408	0	0.127602
-9	209	5.342334	0.321455	0.30277	0	0.103333
-8	441	6.089045	0.436095	0.32877	0	0.190179
-7	670	6.507278	0.230599	0.03581	0	0.053176
-6	1750	7.467371	1.027137	0.84131	0	1.05501
-5	263	5.572154	-0.97608	-1.3947	0	0.952741
-4	161	5.081404	0.231822	0.27251	0	0.053742
-3	95	4.553877	0.099736	0.054832	0	0.009947
-2	57	4.043051	-0.15116	-0.20793	0	0.02285
-1	42	3.73767	-0.14414	-0.13300	0	0.020778
0	40	3.688879	-0.11905	0.0212	0	0.014173
1	140	4.941642	0.763737	1.20397	0.169214	0.583294
↓	↓	↓	↓	↑	↑	↓
599	33	3.496508	-0.08598	-0.16420	-0.06401	0.007393
600	42	3.73767	0.073038	0.27506	0.18368	0.005335
601	93	4.532599	0.254692	0.03509	-0.07823	0.064868
602	163	5.09375	0.115457	-0.921187	-0.74086	0.01333
603	440	6.086775	0.828667	0.97487	0.851225	0.686689
604	435	6.075346	-0.14911	-0.33184	-0.35902	0.022235
605	415	6.028279	-0.29523	-0.58367	-0.41977	0.087159
606	325	5.783825	-0.33382	-0.1037	0.113101	0.111433
607	330	5.799093	0.459118	0.6893	0.701179	0.21079
608	120	4.787492	-0.16698	-0.2544	-0.15652	0.027883
609	60	4.094345	-0.08299	0.0764	0.159818	0.006887
610	40	3.688879	-0.15382	0.18232	0.214503	0.023661
611	42	3.73767	0.040858	0.10763	0.107631	0.001669
					0	SS=78.40

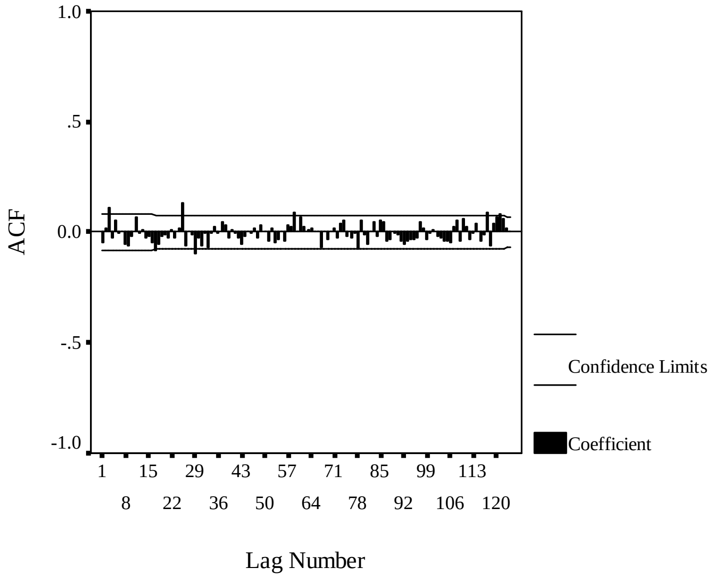

Fig.(5): Autocorrelogram of residual series parameter.

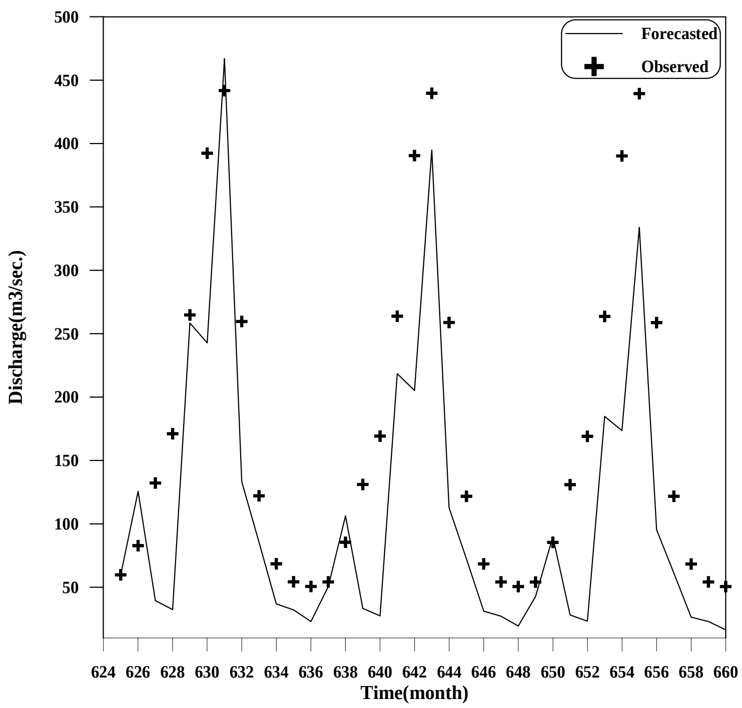

Figure 6: Comparison of Forecasted and Observed Inflow (Oct.2005-Sep.2007).

Conclusions

The(October 1953-September 2004) time series of monthly inflow to Dokan reservoir is a periodic series and , therefore, the stochastic model which represents it is a seasonal one. Seven of such models are fitted to the series and it is found that sum of squared errors of the $(1,1,1)\times(1,1,1)_{12}$ model with autoregressive moving average parameters of θ=0.899,Φ=0.60,Θ=0.96, and Φ=0.03 is less than the other$(1,1,1)\times(0,1,1)_{12}$,$(1,1,2)\times(0,1,1)_{12}$,$(0,1,1)\times(0,1,1)_{12}$,$(2,1,0)\times(1,1,0)_{12}$,$(1,1,0)\times(1,1,0)_{12}$, $(1,1,1)\times(0,1,0)_{12}$ models. The diagnostic checking show that model is adequate. Forecasts using the model for the period from October, 2005 to September 2007 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 Jornal for Mechanical and Material Engineering , Special Issue(A), College of Engineering, University of Babylon.

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.**",McGraw-Hill.

Ministry of water resources.,(2007):" **Planning report on Dokan dam Project".**

Notations

ACF	autocorrelation function
D,d	degrees of differencing
PACF	partial autocorrelation function
SS	sum of square errors
s	number of seasons
X	observed value
Z	lnX
ϕ ,Φ	autoregressive parameters
θ ,Θ	moving average parameters
β	backward shift operator
∇	backward difference operator