FITTING SEASONAL STOCHASTIC MODELS TO INFLOWS OF BEKHME RESERVOIR

Dr. Salah Tawfeek Ali
College of Engineering, Babylon University

ABSTRACT:

In this study time series analysis is applied to records of mean inflow to Bekhme reservoir in the north of Iraq for the period from water year 1933/1934 to water year 2001/2002. Three Box-Jenkins seasonal multiplicative models are fitted to this series. These are the $(1, 1, 0)x(1, 1, 0)_{12}$, $(2, 1, 0)x(1, 1, 0)_{12}$, and $(0, 1, 1)x(0, 1, 1)_{12}$ models. The unconditional sum of squares method is used to estimate the parameters of the models and to compute the sum of squared errors for each of the fitted model. It is found that the model which corresponds to the minimum sum of squared errors is the $(0, 1, 1)x(0, 1, 1)_{12}$ model with moving average parameters θ= 0.378 and Θ= 0.953. The adequacy of this model is checked by plotting the normalized cumulative periodogram which does not indicate nonrandomness of the residuals. The 95 and 99 percent confidence regions for the model parameters θ and Θ are shown graphically. Forecasts of monthly inflow for the period from October, 2002, to September, 2006 are graphically compared with observed inflows for the same period and since agreement is very good adequacy of the selected model is confirmed.

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

الخلاصة :

تم في هذه الدراسة تطبيق تحليل السلاسل الزمنية على التصاريف الشهرية الداخلة الى خزان بخمة في شمـال العراق للفترة من بداية السنة المائية 1933/1934 وحتى نهاية السنة المائية 2001/2002، حيث تمت مطابقة ثلاثة أنواع من المـوديل التصـادفي المعروف بموديـل (Box and Jenkins) الفصلي وهـي المـوديل $(1, 1, 0)x(1, 1, 0)_{12}$ والموديـل $(2, 1, 0)x(1, 1, 0)_{12}$ والموديل $(0, 1, 1)x(0, 1, 1)_{12}$. أن نتائـج أستخدام طريقة مجموع المربعات غير المشروطة لتقدير معاليم الموديلات أظهـرت بأن مجمـوع مربعـات الأخطـاء لـلمـوديـل $(0, 1, 1)x(0, 1, 1)_{12}$ بمعـلمي المعـدل الـحركـي θ=0.378 و Θ=0.954 أقل من الموديلين الآخرين، وكما أن الفحص المعروف بفحص مخطط الذبذبه لم يبين وجود عدم عشوائية فـي الباقيات لهذا الموديل. لقد تم أيضا توليد السلسلة المستقبليةحسب الموديل وللفترة مـن تشرين الأول / 2002 ولغاية أيلول / 2006 وعند مقارنتها مع القيم الفعلية المسجلة وجد تطابقا جيدا مما يؤكد ملائمة الموديل.

1. INTRODUCTION

Review of past time series studies in Iraq , e.g., Abdul-Razaq, 1984, Mahmood, 2000, reveals that a series is decomposed into three components–a "trend," a "seasonal component," and a "random component". The trend is fitted by a polynomial and the seasonal component by a Fourier series.

An autoregressive (abbreviated AR), moving average (abbreviated MA), mixed autoregressive moving average (abbreviated ARMA) is then fitted to the random component of the series. Other authors , e.g., Shaker, 1986, Abed , 2007, apply autoregressive integrated moving average (ARIMA) models to deseasonalised series. In this study we fit other types of models which are the seasonal models called Box-Jenkins models to inflows of Bekhme reservoir. Application of these seasonal models to data from Iraq is not found in the literature.

2. STOCHASTIC PROCESSES

A model which describes the probability structure of a sequence of observations is called a "stochastic process". A time series of N successive observations X (= X_1, X_2,…..,X_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 (AR), moving average (MA), and mixed autoregressive moving average processes (ARMA). Another class of stochastic processes which are nonstationary and nonseasonal processes are the autoregressive integrated moving average (ARIMA) models.

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 :

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Z_t = \theta_q(\beta)\Theta_Q(B^s)a_t \qquad \text{.....(1)}$$

or

$$(1-\sum_{j=1}^{p}\phi_j B^j)(1-\sum_{j=1}^{P}\Phi_j B^j)\nabla^d\nabla_s^D Z_t = (1-\sum_{j=1}^{q}\theta_j B^j)(1-\sum_{j=1}^{Q}\Theta_j B^{sj})a_t \qquad \text{.....(2)}$$

where

∇ is backward difference operator such that

$$\nabla Z_t = Z_t - Z_{t-1},$$

and

$$\nabla_s Z_t = Z_t - Z_{t-s},$$

$$Z_t = \ln X_t,$$

X_t is observed value,

B is backward shift operator such that

$BZ_t = Z_{t-1}$, and

$$B^s Z_t = Z_{t-s},$$

s is the number of seasons (= 12 for monthly data), and

a_t is a random drawing 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)x(P, D, Q)$_s$.
Now, if p=P=0, d=D=1, q=Q=1, and s=12, Equation 2 becomes:

$$\nabla\nabla_{12} Z_t = (1-\theta B)(1-\Theta B^{12})a_t \qquad (4)$$

or

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

where

$W_t = \nabla\nabla_{12} Z_t$

This is a seasonal multiplicative model which is called (0, 1, 1)x(0, 1, 1)$_{12}$ Box-Jenkins model .

If p=P=1, d=D=1, q=Q=0, and s=12, Equation 2 becomes:

$$(1-\phi B)(1-\Phi B^{12})W_t = a_t \qquad (5)$$

Equation 5 is a seasonal multiplicative model which is known as Box- Jenkins seasonal model of order (1 ,1, 0)(1, 1, 0)$_{12}$.
If p=2, P=1, d=D=1, q=Q=0, and s=12, then it is obtained from Equation 2 that:

$$(1-\sum_{j=1}^{2}\phi_j B^j)(1-\Phi B)W_t = a_t \qquad (6)$$

or

$$(1-\phi_1 B-\phi_2 B^2)(1-\Phi B^{12})W_t = a_t \qquad (7)$$

This is known as (2, 1, 0)(1, 1, 0)$_{12}$ Box-Jenkins model .

The main objective of the present study is to fit Box-Jenkins models to the seasonal time series of monthly inflow to Bekhme reservoir in the north of Iraq.

3. INFLOW TO BEKHME RESERVOIR

The Bekhme dam is located at about 7 km upstream of Bekhme village and approximately 2 km downstream of the confluence with the tributary Ruwandus river which flows from the left-bank side as shown in Figure 1.

In Figure 2 we plot the means of inflows to Bekhme reservoir for the entire period of record of 70 years from Oct.,1933 to Sep.,2002. To show the periodic behavior of this series only a small part of it corresponding to the period from water year 1933/1934 to water year 1937/1938 is given as Figure 3 which, as expected, shows a marked seasonal pattern since mean inflow is at its highest always in the spring months, i,e., similarities occur after 12 months period.

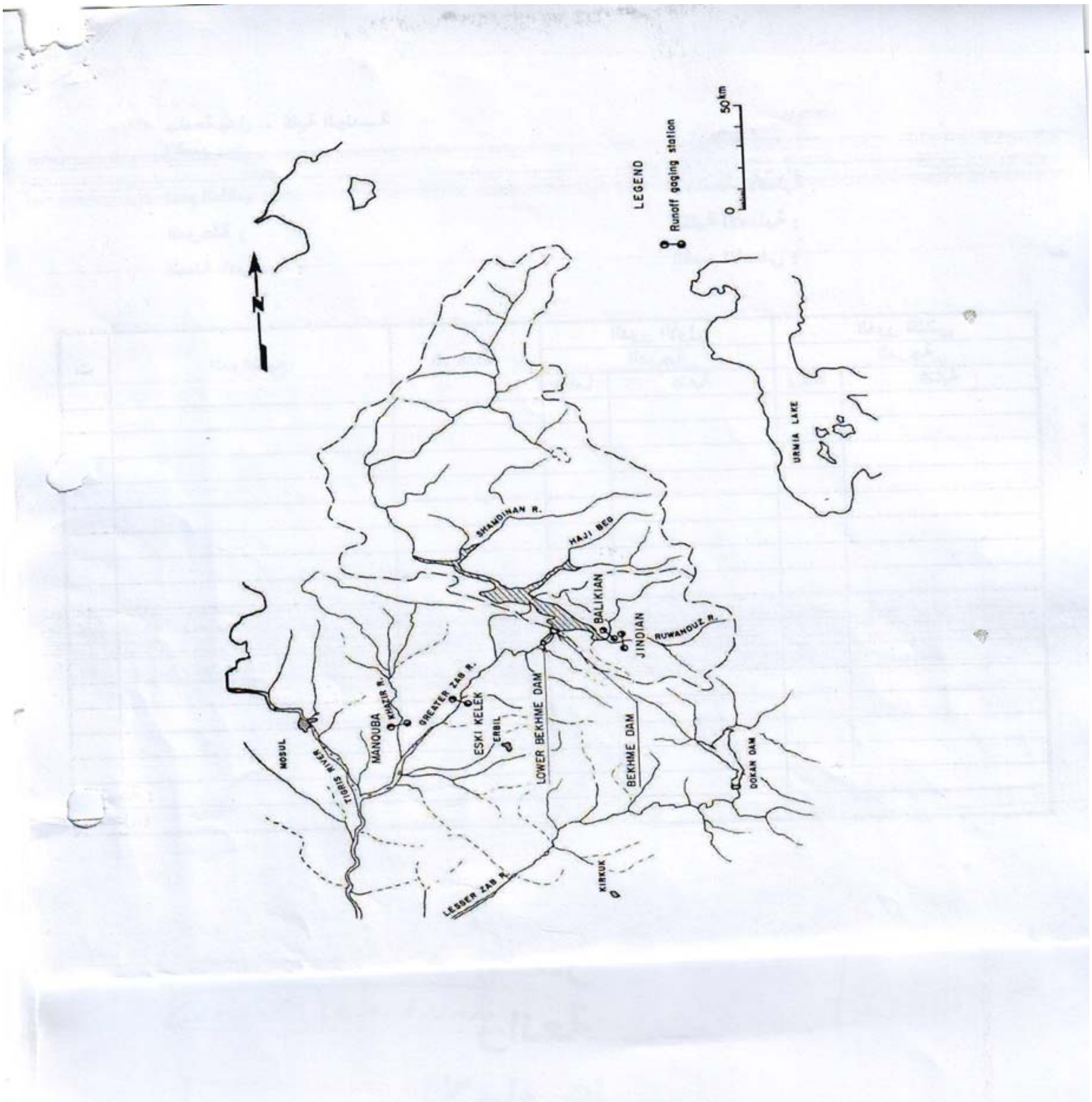

Figure 1: Site Map of Bekhme Reservoir (Iraqi Ministry of Irrigation, 1986)

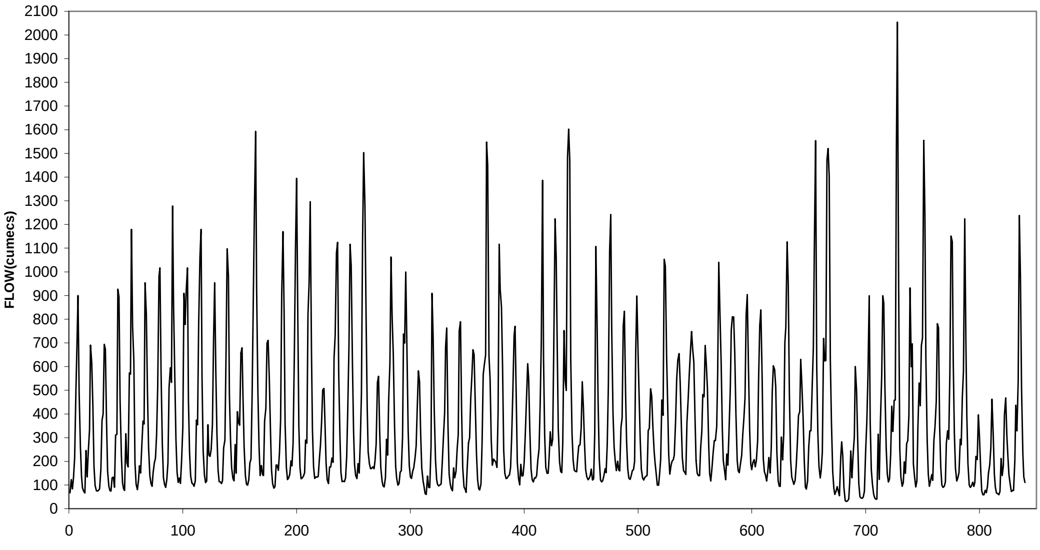

Figure 2:Monthly Inflow t0 Bekhme Reservoir for the Period (Oct,1933-Sep,2002).

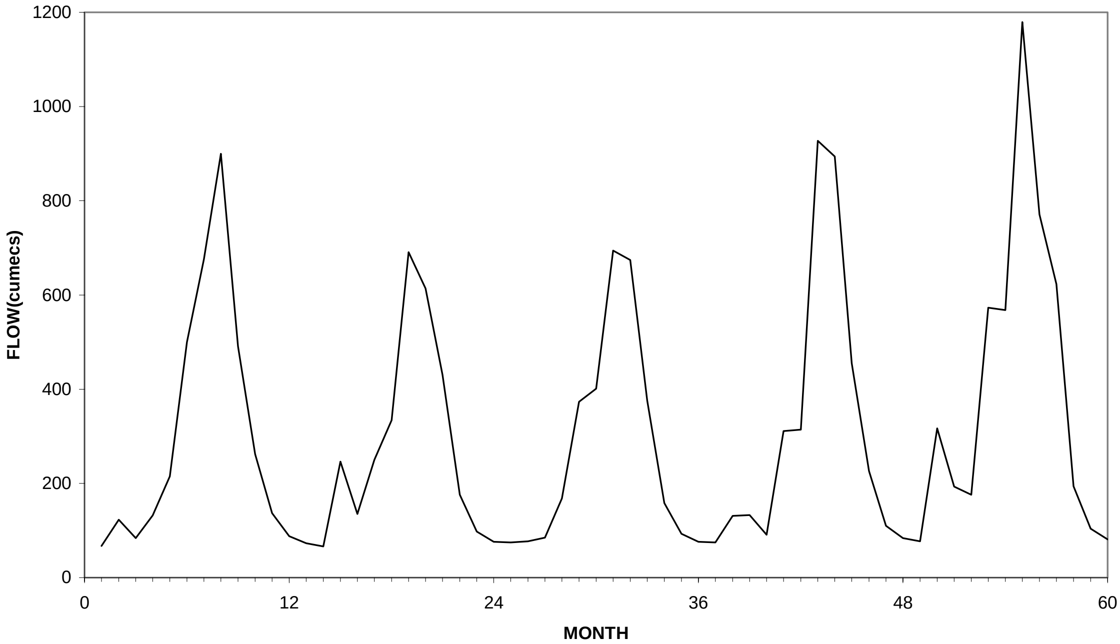

Figure 3: Monthly Inflow to Bekhme Reservoir for the period (Oct,1933-Sep,1937).

4. FITTING SEASONAL BOX-JENKINS MODELS

Relating of a Box-jenkins seasonal model to data is achieved by three stage iterative procedure based on identification, estimation of model parameters, and diagnostic checking of the estimated parameters (Young, 1974).

4. 1 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 orders 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 functions are plotted to guess the orders of p,q,P, and Q.

The estimated autocorrelation and autocorrelation functions 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. This behavior resembles that associated with autoregressive process(Chatfield, 1975). Therefore, we apply the autoregressive seasonal models $(1, 1, 0)x(1, 1, 0)_{12}$ and $(2, 1, 0)x(1, 1, 0)_{12}$. However, the estimated values of the lag 11 correlation coefficient r_{11} (=0.146) is very close to the estimated value of the lag 13 correlation coefficient r_{13}(=0.150) and to the product of the estimated values of lag 1 and lag 12 correlation coefficients r_1r_{12}\{= (-0.286\}x(-0.499)= 0.143\}. These results appear to indicate a moving average process (

Box and Jenkins, 1976). Therefore, in addition to the autoregressive seasonal models the moving average seasonal model $(0,1,1)x(0,1,1)_{12}$ is also fitted..

4. 2 ESTIMATION OF PARAMETERS

The unconditional sum of squares method is used to estimate the parameters of the models. Trial values for the parameters of the $(0, 1, 1)x(0, 1, 1)_{12}$ model, i.e., θ and Θ in Equation 3 are assumed and the sum of squares, S, is computed. Computation of S is repeated with different values of θ and Θ until minimum sum of squares is obtained. To illustrate this procedure results are given in Table 1 for $\theta = 0.378$ and $\Theta = 0.953$ corresponding to the minimum sum of squares of 70.67. Assuming $e_t = 0$ for t= 828, 829, ….., 840, $a_t = 0$ for t= -13, -14, …, -25, and $e_t = 0$ for t= 0, -1, …., -12 computation in this table starts from the bottom of it using either the forward or backward forms of the model which are given below as Equations 8 and 9 respectively:

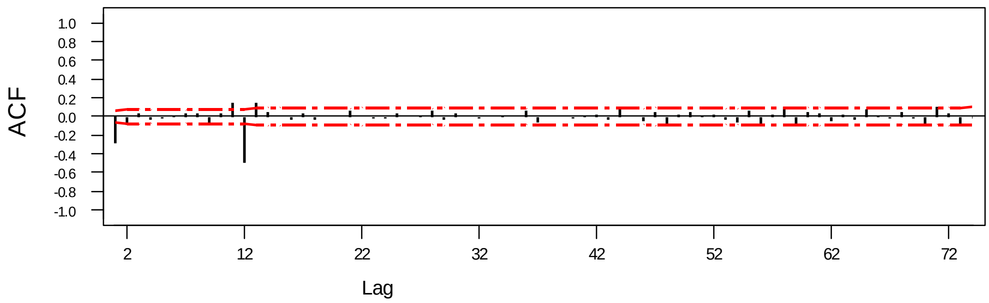

Figure 4:Autocorrelation Function (ACF) of W.

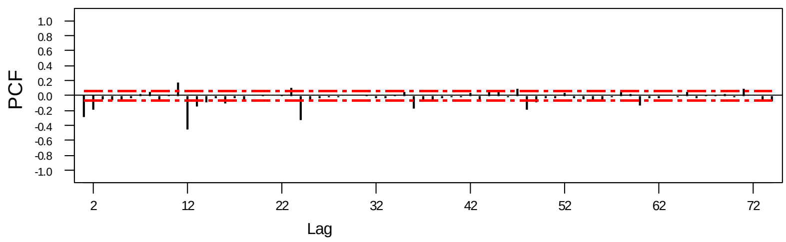

Figure 5: Partial Autocorrelation Function (PACF) of W.

`

Table 1: Computation of sum of squares (S) for the $(0, 1, 1)x(0, 1, 1)_{12}$ model with θ= 0.378 and Θ= 0.953

t	X(t)	Z(t)=ln X(t)	W(t)	e(t)	a(t)	a(t)^2
-12	67	4.204693	-0.05312	0	-0.05312	0.002821
-11	123	4.812184	0.323697	0	0.303619	0.092185
-10	84	4.430817	-0.57967	0	-0.4649	0.216136
-9	132	4.882802	0.175361	0	-0.00037	1.39E-07
-8	215	5.370638	0.052108	0	0.051967	0.002701
-7	500	6.214608	0.448716	0	0.46836	0.219361
-6	676	6.516193	-0.24914	0	-0.0721	0.005199
-5	900	6.802395	0.311501	0	0.284246	0.080796
-4	491	6.196444	-0.07895	0	0.028491	0.000812
-3	262	5.568345	0.08075	0	0.09152	0.008376
-2	137	4.919981	-0.07111	0	-0.03652	0.001333
-1	88	4.477337	-0.16903	0	-0.18284	0.033429
0	73	4.290459	-0.13988	0	-0.25961	0.067396
1	66	4.189655	-0.7083	-0.14745	-0.49795	0.247949
814	104	4.644391	0.006711	-0.25705	0.329498	0.108569
815	72	4.276666	-0.33695	-0.13148	-0.01783	0.000318
816	77	4.343805	0.148485	0.10999	0.017169	0.000295
817	77	4.343805	-0.19913	-0.10184	-0.28362	0.080439
818	203	5.313206	-0.11052	0.257382	0.493696	0.243736
819	437	6.079933	1.181671	0.973284	0.835669	0.698342
820	329	5.796058	-0.53519	-0.55129	-0.33482	0.112105
821	540	6.291569	-0.29295	-0.04259	-0.03152	0.000994
822	1238	7.121252	0.662629	0.662319	0.296769	0.088072
823	977	6.884487	0.17191	-0.00082	0.01205	0.000145
824	504	6.222576	-0.2974	-0.45696	-0.12739	0.016228
825	234	5.455321	-0.36179	-0.42212	-0.1187	0.014091
826	135	4.905275	-0.22462	-0.15962	-0.06964	0.00485
827	111	4.70953	0.17198	0.17198	0.009963	9.93E-05

SS= 70.6746

$$[e_t]=[W_t]+\theta[e_{e+1}]+\Theta[e_{t+12}]-\theta\Theta[e_{t+13}]......................................(8)$$

$$[a_t]=[W_t]+\theta[a_{t-1}]+\Theta[a_{t-12}]-\theta\Theta[a_{t-13}]......................................(9)$$

The results of application of the unconditional sum of squares method to estimate the parameters for the other two models are as follows:

Model	Model parameters	Sum of squared errors
$(1, 1, 0)x(1, 1, 0)_{12}$	Ø= -0.281 Φ= -0.504	S= 100.57
$(2, 1, 0)x(1, 1, 0)_{12}$	$Ø_1$= -0.33 $Ø_2$= -0.17 Φ= -0.5	S= 97.79

Hence, according to these results it is found that the $(0, 1, 1)x(0, 1, 1)_{12}$ model agrees with the historical series better than the other two models. However, a diagnostic test should be performed before a final decision on the most appropriate model is made.

4. 2 DIAGNOSTIC CHECKING

The periodogram $I(f_i)$ of a time series of residuals a_t, t= 1, 2, ……., n is defined as (Chatfield and Prathero, 1973):

$$I(f_i)=\frac{2}{n}[(\sum_{t=1}^{n}a_t Cos2\pi f_i t)^2+(\sum_{t=1}^{n}a_t Sin2\pi f_i t)^2],i=1,2,......,(\frac{n-1}{2})......(10)$$

where $f_i=(i/n)$ is the frequency.

The normalized cumulative periodogram which is an effective means for detection of periodic nonrandomness is given by Barnlett(Chatfield, 1975) as:

$$C(f_i)=\frac{\sum_{i=1}^{j}I(f_i)}{ns^2}..(11)$$

where s^2 is an estimate of the variance σ_a^2.

Deviations of the cumulative periodogram from that expected if the estimated values of the residuals were purly random, i.e., white noise, is assessed from a $C(f_i)$ versus f_i plot.Thus, limit lines are placed about the theoretical line corresponding to white noise such that if the residual series were white noise, the $C(f_i)$ would deviate from the straight line sufficiently to cross these limits only with a stated probability. The limit lines are drawn at distances $\pm K_\xi\sqrt{[(n-1)/2]}$ above and below the theoretical line. For 5% probability limits, i.e., ξ= 0.05, K_ξ is approximately 1.36.

The normalized cumulative periodogram plot of the residuals from the $(0, 1, 1)x(0, 1, 1)_{12}$ model fitted to the inflow series with θ= 0.378, and Θ= 0.953 is shown in Figure 6 which fails to indicate any significant departure from the assumed model.

4. 3 CONFIDENCE REGIONS

The (1-ξ) confidence region for the model parameters is bounded by the contour on the sum of squares surface for which (Chatfield and Prathero, 1973)

$$S(\beta)=S\hat{\beta})\{1+\frac{\chi_\xi^2(K)}{n}\}..(12)$$

where

$(\beta) = model\ parameters,$

$(\hat{\beta}) = estimates\ of\ model\ parameters$

$\chi^2_{\xi}(K) = chi - squared\ value\ corresponding\ to\ \chi^2\ proportion\ of\ the\ distribution$
$with\ K\ degrees\ of\ freedom,$

$K = number\ of\ estimated\ parameters$

Equation 12 is used to determine the 95 and 99 percent confidence regions for the model parameters which are shown in Figure 6.

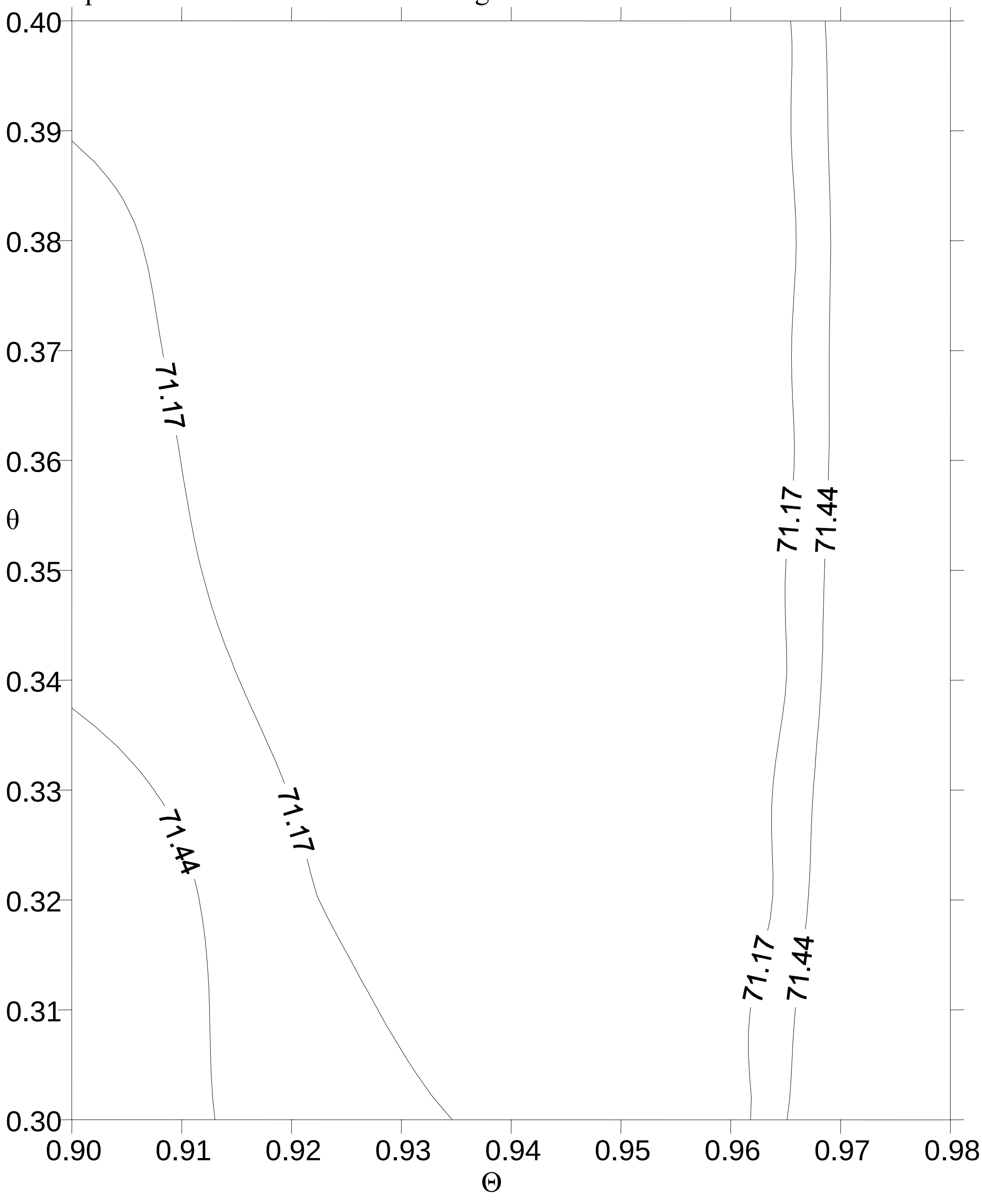

Figure 6: The sum of squares surface showing 95 and 99 percent regions of the parameters of the $(0, 1, 1)x(0, 1, 1)_{12}$ model. The 71.17 represents the boundary of the 95% confidence region while the 71.44 represents the the boundary of the 99% confidence region.

5. COMPARISON OF FORECASTED AND OBSERVED INFLOWS

At time t+ℓ the (0, 1, 1)x(0, 1, 1)$_{12}$ model which is fitted to the inflow series, i.e., Equation 4 with θ=0.378, and Θ=0.953, gives:

$$Z_{t+\ell}=Z_{t+\ell-1}+Z_{t+\ell-12}-Z_{t+\ell-13}+a_{t+\ell}-0.378a_{t+\ell-1}-0.953a_{t+\ell-12}+0.36a_{t+\ell-13} \quad \ldots\ldots\ldots\ldots (13)$$

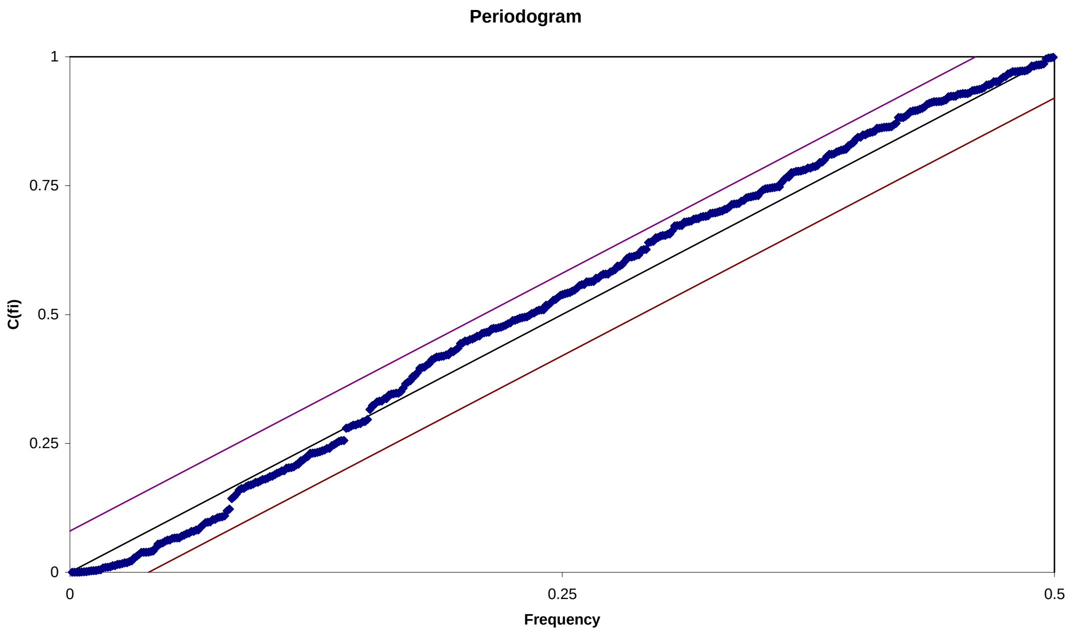

Figure 7: Cumulative Periodogram of Residuals from (0, 1, 1)x(0, 1, 1)$_{12}$ Model.

This equation is used to obtain forecasts with the unknown Z's being replaced by their already forecasted values and unknown a's by zeroes (Box and Jenkins, 1976). Hence, the forecasts so obtained for lead times up to 48 months, all made at the origin September 2002 are shown in Figure 7. The corresponding observed values are also shown in this figure and since agreement between observed and forecasted values is very good, it is confirmed that the model is adequate.

6. CONCLUSIONS

The (October 1933-September 2006) time series of monthly inflow to Bekhme reservoir is a periodic series and, therefore, the stochastic model which represents it is a seasonal one. Three of such models are fitted to the series and it is found that sum of squared errors of the (0, 1, 1)x(0, 1, 1)$_{12}$ model with moving average parameters of θ= 0.378 and Θ= 0.953 is less than the other two (1, 1, 0)x(1, 1, 0)$_{12}$ and (2, 1, 0)x(1, 1, 0)$_{12}$ models. The diagnostic checking of cumulative periodogram does not indicate significant departure of the series from the selected model. Forecasts using the model for the period from October, 2002 to September 2006 agrees well with the observed values. The 95 and 99% confidence regions forthe estimated parameters of the selected $(0, 1, 1)x(0, 1, 1)_{12}$ model is also determined.

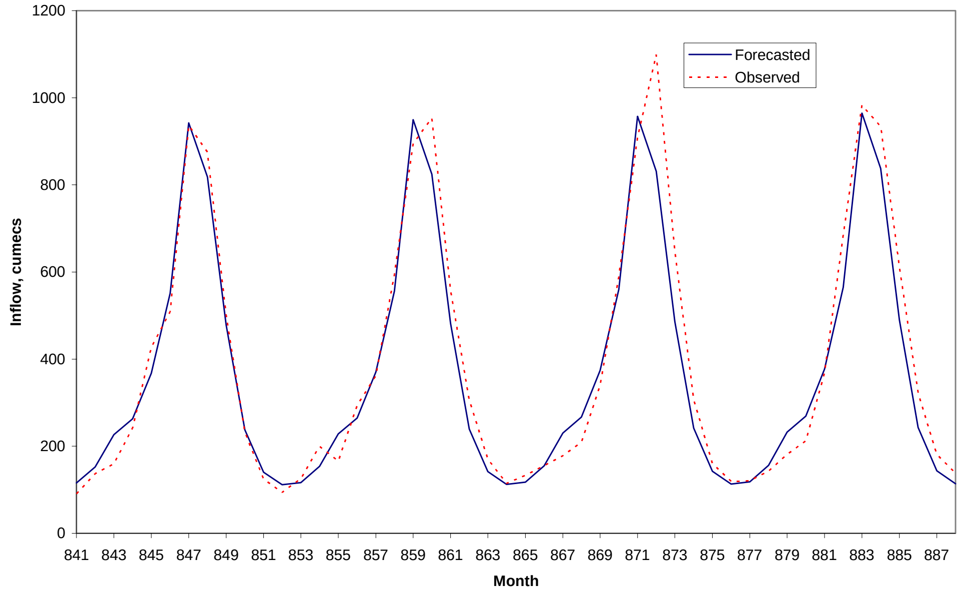

Figure 8: Comparison of Forecasted and Observed Inflows(OCt. 2002-Sep.2006).

REFERENCES

1. Abdul Razaq, T. T. (1984) " Multivariate stochastic generation of daily rainfall," MSc Thesis, University of Baghdad.
2. Abed, Z. A. (2007) " Stochastic models of some properties of water in the middle Euphrates region of Iraq,"MSc Thesis, Babylon University.
3. Box, G. E. P., and Jenkins, G. M. (1976) Time series analysis, Holden-Day Inc., San fransisco.
4. Chatfield, C. (1975) The analysis of time series, Chapman and Hall, London.
5. Chatfield, C., and Prathero, D. L. (1973) "Box-Jenkins seasonal forecasting," J. R. Statist. Soc. A, 136.
6. Iraqi Ministry of Irrigation (1986) " Planning report on Bekhme dam project."
7. Mahmood, H. S. (2000) " Time series analysis of hydrochemical pollutant," MSc Thesis, Babylon University.
8. Shaker, R. H. (1986) " Stochastic analysis of daily streamflow of Tigris River," MSc Thesis, University of Baghdad.
9. Young, P. (1974) "Recursive approach to time series analysis, Bull. I. M. A., 10.

NOTATIONS

ACF	autocorrelation function
B	backward shift operator
$C(f_i)$	cumulative periodogram
D,d	degree of differencing
$I(f_i)$	periodogram
PACF	partial autocorrelation function
S	sum of squared errors
s	number of seasons
s^2	variance
X	observed value
Z	lnX
Ø,Φ	autoregressive parameters
θ,Θ	moving average parameters
∇	Backward difference operator