**Iraqi Journal of Statistical Science (14) 2008**
**p.p. [1-20]**

# A Proposed Technique for the Problem of Selecting the best Forecasting Model in Time Series: A Case Study

**Thafer R. Muttar***

## ABSTRACT :

Forecasting is considered as one of the essential goals regarding time series analysis, and forecasting accuracy decreases the risk possibility regarding decision making. Whereat the best model to represent the time series data, might not be the same model used for forecasting. The forecasting evaluation criterion used such as RMSE, MAPE, and MAE provides almost different results concerning one time series, which confuse the researcher to select the best model for forecasting. Therefore this research deals with the problems of criterion differences results, that affects the evaluation performance to select the best model, and providing a statistical manner that employs the forecasting criterion results that are mentioned as a weighted mean for each model of ARIMA which is considered as a candidate model with the least weighted mean that provides the best forecasting performance. This has been applied on the monthly time series for the water of the Tigris River (M-cu-m) that enters Mosul City for the period 1963-1995. Meanwhile Box-Jenkins model shows SARIMA $(1,1,2) * (3,1,1)_{12}$ Very encouraging forecasting results depending on the suggested manner compared with the rest of models, meanwhile the best model for representing the data is SARIMA $(1,1,2) * (0,1,1)_{12}$.

## اسلوب مقترح لمسالة اختيار افضل نموذج تكهن في السلاسل الزمنية: حالة دراسية

### الملخص

يعد التكهن أحد الأهداف الأساسية المتعلقة بتحليل السلاسل الزمنيـة ،وان دقـة التكهن تقلل المخاطر لدى متخذ القرار. ان النموذج الأفضل لتمثيل بيانات السلسـلة الزمنية لايكون غالبا النموذج المستخدم للتكهن نفسه، وبما أن معايير تقيـيم اداء

*Asst. Prof.- Dept. of Statistics and Informations-College of Computers Sciences and Mathematics

Received: 3/ 7 /2008 ____________________Accepted: 21/ 9 / 2008

التكهن المستخدمة في السلاسل الزمنية مثـلMAE, MAPE, RMSE تقـدم غالبا نتائج مختلفة للسلسلة الزمنية الواحدة باختلاف النماذج فأن هـذا ممـا يربـك الباحث في اختيار النموذج الأفضل للتكهن ، لذا فان هذا البحث يتعامل مع اشكالية تناقض نتائج معايير تقييم اداء التكهن في ترشيحها للنموذج الأفضل ويهـدف الـى تقديم اسلوب احصائي يعمل على توظيف نتائج معايير التكهن المـذكورة كوسـط موزون لكل نموذج من نماذج ARIMA التي يتم ترشيحها ومن ثم اختيار النموذج الذي يمتلك أقل وسط موزون بوصفه يقدم افضل اداء للتكهن . لقد تم التطبيق على السلسلة الزمنية الشهرية لمنسوب مياه نهر دجلة (مليون متر مكعب) الداخلة الـى مدينة الموصل للفترة من 1963- 1995 وقد اظهـر نمـوذج بـوكس جنكنـز $SARIMA\ (1,1,2) * (3,1,1)_{12}$ نتائج مشجعة جدا في التكهن اعتمـادا علـى الاسلوب المقترح مقارنة ببقية النماذج ، في حين ان النمـوذج الأفضـل لتمثيـل البيانات هو $SARIMA(1,1,2) * (0,1,1)_{12}$.

## 1. Introduction

Forecasting a time series is a common problem in many domains of science (electricity, hydrology, etc.), and has been addressed for a long time by statisticians ( Lendasse et al.,2000).

Forecasting a hydrologic time series has been one of the most complicated tasks owing to the wide range of data , the uncertainties in the parameters influencing the time series and also due to the non availability of adequate data ,thus the development and use of stochastic models of hydrological phenomena play an important role in water resources engineering, including their use to forecast river flows. The choice of the right model for a given hydrological series is an important aspect of the modeling process (Mujumdar & Nagesh Kumar , 1990;Kumar et.al.2004).Thus, forecasting will be essentially part in time series.

The challenge of predicting future values of a time series spans a variety of disciplines. The multiplicity of techniques developed to make predictions manifests a heritage from biology, computer sciences, economics, engineering, mathematics, physics, statistics, and other areas. These methods find applications in such diverse data sets as animal populations, equity market prices, disease control, meteorological measurements, astronomic observations, and others (Pomares & Rojas, ,2004).

## 2.Box – Jenkins Analysis

Traditionally, time series analysis is defined as a branch of statistics that generally deals with the structural dependencies between the observation data of random phenomena and the related parameters. The observed phenomena are indexed by time as the only parameter; therefore, the name time series is used (Ajoy & Dobrivoje ,2005).

The primary objective of time series modeling is to study techniques and measures for drawing inferences from past data. It accounts for the fact that data points taken through time may have an underlying structure (such as autocorrelation, trend or seasonal variation) and this structure will persist over time. The approach consists of establishing mathematical models to represent the data set. Then, the models can be employed to describe and analyze the sample data, and make forecasts for the future. The main advantage of time series models is that they can handle any persistent patterns in data (Abdullah & Tayfur ,2004).
In statistics, two basic mathematical system models are used:
*deterministic models, mathematically viewed as analytical models represented by deterministic relations like $x_t=f(t)$ or the succession of values in a time series is usually influenced by some external (or exogenous) information. If this information is not known, only the past values of the series itself can be used to build a model, i.e. a mathematical function of the form recurrence equations like

$$x_{t+1} = f_\theta(x_t, x_{t-1}, \ldots) \qquad ...(1)$$

where an unknown new value $x_{t+1}$ is estimated from the known current and past values of x . The parameters $\theta$ of the model $f_\theta$ are chosen according to the information available, i.e. to all known values of x ; this step if called learning or fitting, most widely known prediction tools use linear models $f_\theta$, sometimes other information is available, in this case, it is a good idea to use this external information in the model, usually in the form

$$x_{t+1} = f_\theta(x_t, x_{t-1}, \ldots, x_{t-N+1}, y_t^1, y_t^2, \ldots, y_t^p) \quad \ldots(2)$$

where the values at time t of p external (exogenous) variables are used in the model ( Lendasse et al.,2000).
*stochastic models, statistically viewed as functions of random variables, mathematical models used of time series analysis are generally (regression models, time-domain models such as transfer functions models and state-space models, frequency models ) ( Lendasse et al.,2000;Ajoy & Dobrivoje ,2005).

The Box-Jenkins variant of the ARMA model is predestinated for applications to nonstationary time series that become stationary after their differencing. Differencing is an operation by which a new time series is built by taking the successive differences of successive values, such as x(t) – x(t-1) along the nonstationary time series pattern. In the acronym ARIMA, the letter I stands for integrated. The widely accepted convention for defining the structure of ARIMA models is ARIMA(p, q, d), where p stands for the number of autoregressive parameters, q is the number of moving-average parameters, and *d* is the number of differencing passes (Ajoy & Dobrivoje ,2005).

The basic concepts necessary to study the time dependent dynamics of random phenomena are called random signals, stochastic processes, or random time series. The emphasis here is on random dynamics which are stationary, that is governed by underlying statistical mechanisms that do not change in time (Wojbor,2006).A time series is said to be stationary, if its statistical properties remain constant over time, i.e., its mean is independent of time and its autocorrelation function is independent of time for each lag. The autocorrelation function provides valuable information about how much successive values in a time series depends on each other. It can be thought of an indication of change in one observation if there is a change in the other. In addition, it plays an important role in forecasting future values based on the present and past values. Box and Jenkins (1976) provides a methodology for fitting a model to an empirical data set. The systematic approach identifies a class of models appropriate for empirical data sequence at hand and estimates its parameters. A general class of Box and Jenkins models include ARIMA models that can model a large class of autocorrelation functions (Box and Jenkins 1976, Brockwell & Davis 2002),The model is a combination of auto regressive (AR) and moving-average (MA) models for differenced data. An AR model is simply a regression of the current observation to the previous ones. Formally, if $\{x_t\}$, t = 0, 1, 2, . . . are the values of observations recorded at time t , then

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_p x_{t-p} + z_t \quad \text{...(3)}$$

is called an AR process of order p where $z_t$ is a white noise process with mean 0 and variance $\sigma^2$ and $\varphi$'s are finite weight parameters. On the other hand, an *MA model* is a regression of the current value against the previous white noise, i.e.,

$$x_t = z_t + \theta_1 z_{t-1} + \ldots + \theta_q z_{t-q} \quad \text{...(4)}$$

where $\theta$'s are constants. Then, $\{x_t\}$ is an ARMA(p,q) process,if $\{x_t\}$ is stationary and if for every t,

$$x_t - \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_p x_{t-p} = z_t + \theta_1 z_{t-1} + \ldots + \theta_q z_{t-q} \quad \text{...(5)}$$

The process $\{x_t\}$ is said to be an ARMA(p,q) process with mean $\mu$, if $\{x_t - \mu\}$, deviations from the mean, is an ARMA(p,q) process. Finally, the integrated ARMA model, ARIMA(p,d,q), is an ARMA(p,q) model to the d times differenced data. Differencing is a tool in order to remove trend and seasonality from the empirical data. (Abdullah & Tayfur,2004).

There are three steps in ARMA modeling: (Jack & John,1997)

1. Check the series for stationarity , and, if necessary, transform the series to induce stationarity.
2. From the autocorrelation properties of the transformed series choose a few ARMA specifications for estimation and testing in order to arrive at a preferred specification with white noise residuals.
3. Calculate forecast over a relevant time horizon from the preferred specification.

Last chatfield(1995) provide the difference between terms predication and forecasting ," where some authors using terms 'prediction' and 'forecasting' interchangeably , but some authors do not, for example Brown (1963) uses 'prediction' to describe subjective methods and 'forecasting' to describe objective methods, whereas Brass (1974) uses 'forecast' to mean any kind of looking into the future, and 'prediction' to denote a systematic procedure for doing so. Prediction is closely related to control problems in many situation".

## 3. Model Selection

The problem of model selection is an important one in time series analysis as there are infinitely many possible models and the choice of a wrong model may result in a costly decision (Mujumdar & Kumar, 1990).

Several criteria proposed for selecting time series models such as

(A) the Akaike information criterion (AIC), $nLn(S/n)+2p$;

(B) the Bayesian information criterion (BIC),$nLn(S/n)+p+pLn(n)$;

where p denotes the number of parameters fitted in the model and n denotes the number of effective observations used in fitting the model ,

the residuals sum of squares ,S, can only become smaller and the residual standard deviations $\hat{\sigma}_a$ will tend to become smaller as a model is made 'larger'. Thus the minimization of a criterion such as the AIC or BIC is more satisfactory for choosing a 'best' model from candidate models having different numbers of parameters. Strictly speaking (a) and (b) above are approximations to the variable part of the AIC and BIC respectively. In both cases the first term is a measure of (Lack of) fit and the remainder is a penalty term to prevent overfitting. The BIC penalizes extra parameters more severely than the AIC dose, leading to ' smaller' models. Several similar criterion have been proposed including alternative closely related Bayesian criterion which depend on different priors on model size. In particular Schwarz's Bayesian criterion (SBC) has the penalty term pln(n) rather than p+pLn(n). (see Chatfield and Faraway,1998)

Mujumdar and Nagesh Kumar (1990) gives flaws in the AIC rule. Firstly, the AIC has no optimal property, i.e. it does not minimize the average value of any criterion function. Secondly, the AIC rule is not consistent, i.e. the probability that the decision rule will choose a wrong model does not go to zero even when the number of observations tends to infinity. Also, Pena (2001) displays the problem with AIC=1-2(Log maximum likelihood)+2(number of parameters) is that tends to overestimate the number of parameters, even asymptotically, where in the BIC=-2(Log maximum likelihood)+(Log n)(number of parameters) criterion, the penalty for introducing new parameters is greater than AIC, so the BIC tends to select simpler models than those chosen by AIC. The difference between both criterion can be very large if n is large.

Chatfield and Faraway (1998) recommended to use bias-corrected version of the AIC ($AIC_c$), as recommended by Brockwell and Davis (1993), which is obtained by adding 2(p+1)(p+2)/(n-p-2) to the AIC .This makes little difference for small values of p but, for larger values of p , penalizes extra parameters (much) more severely than the AIC.
used only BIC and ($AIC_c$) criterions for selecting best model in this paper.

## 4. Forecasting Performance

To evaluate the forecast performance of model (see Liu,2006) ,it is common to reserve a small portion of the data at the end of a time series solely for forecast comparison. The data used for such a purpose are referred to as hold out sample or post-sample , and in principle are not used in model identification or estimation when evaluating forecast performance. A number of criteria are available for the computation of forecast performance, including Root Mean Squared error (RMSE) , Mean Absolute Percent Error(MAPE), and Mean Absolute Error( MAE or (MAD) as defined below.

Root Mean Squared Error

$$\text{RMSE}=\sqrt{\frac{1}{m}\sum_{t=1}^{m}(x_t-\hat{x}_t)^2} \qquad ...(6)$$

Mean Absolute Percent Error

$$\text{MAPE}=\left[\frac{1}{m}\sum_{t=1}^{m}\frac{|x_t-\hat{x}_t|}{x_t}\right]\times 100\% \qquad ...(7)$$

Mean Absolute Error ( also known as mean absolute deviation)

$$\text{MAE}=\left[\frac{1}{m}\sum_{t=1}^{m}|x_t-\hat{x}_t|\right] \qquad ...(8)$$

Where $x_t$ is the actual observation at time t , $\hat{x}_t$ is the forecast value of $x_t$ based on a particular model or method , and m is the total number of observations in the post-sample period. Typically m is small in comparison to the total length of a time series.

The definition of RMSE is similar to that of the residual standard error $\hat{\sigma}_a$ of the estimated model. Assuming that the estimated model is representative of the forecasting period, the post-sample RMSE should be consonant with the residual standard error. When the post-sample RMSE is much larger than $\hat{\sigma}_a$, it signifies that the model may not be appropriate for forecasting. The RMSE statistic is a measure expressed in the same scale as $x_t$ making it directly comparable to $x_t$. Because of this property, care should be taken when trying to compare the post-sample RMSE of different time series.

The MAPE criterion is easy to understand and convenient for communication. It has the nice property of being scale independent. Therefore, MAPE is often used when there is a need to compare forecast across different time series. Even though it is frequently used, it has a number of pitfalls. Since the forecast error is divided by the actual observation $x_t$ in the computation of MAPE, a larger $x_t$ will automatically decrease MAPE suggesting better forecast performance and a smaller $x_t$ will automatically increase MAPE suggesting poorer forecast performance , regardless of the model . this implies that MAPE is not a good criterion to use for time series such as the S& P500 daily return( due to the fact that some $x_t$'s are small and some are zero or negative), the monthly mortgage rate ( due to downward drift of the series),and the monthly airline passenger series( due to high variability in different months of a year). It may make sense if it is used to evaluate the forecast performance of the U.S. population series or time series with similar nature. In this case , however , we would expect that the accuracy of a forecasting model or method would improve over the year simply due to the increase of $x_t$ . Close attention must be paid when MAPE is used as a criterion for forecasting comparison.

To avoid the pitfalls of MAPE, the simpler MAE criterion shown above may be considered . Brown (1962) shows that the MAE is approximately 1.25 RMSE. Therefore, MAE and RMSE are related measures. The MAE criterion is useful whenever the loss associated with an error increases linearly. The RMSE criterion, on the other hand, is appropriate when the loss associated with an error increases at a quadratic rate or proportionately to $(x_t - \hat{x}_t)^2$.

It is obvious that any researcher does not depend completely upon any criterion from criterions above ,so we present in the next section mathematical manner depend on employ the results of three criterions to determine the model which has best forecast performance, and here we monition this model will not necessary it is the same model to representation the model..

## 5. Proposal Manner

As we mentioned in the pervious section, the criterions of evaluation the forecast performance (RMSE,MAPE and MAE) suffer problems and constraints at uses. The decision making to select best forecast performance upon one criterion may not lead to this target since every criterion indicates the selection different model, so we present proposal statistical manner which depends on the employ of the results of RMSE,MAPE and MAE criterions in weighted mean calculate to each model in candidate models. The idea of a weighted mean was inspired from Program Evaluation and Review Technique (PERT) which deals with optimistic, most likely and pessimistic time in network analysis/operation research(Hillier and Lieberman(2001)). This technique gives only four weights to most likely time ,here we give four weights in combination form to each criterion to arrive at a model which present best forecast performance that has minimum weighted mean , as follows:

$$\text{weighted mean} = (RMSE + 4\,MAPE + MAP)/6 \quad \ldots(9)$$

$$= (4RMSE + MAPE + MAP)/6 \quad \ldots(10)$$

$$= (RMSE + MAPE + 4MAP)/6 \quad \ldots(11)$$

This form of weighted mean will be expected achieve the same result . Vandaele (1983) indicates that the best model for representing the data may give poor forecasting, the same remark is given by Chatfield and Faraway(1998) about models in neural networks, therefore, the best model in performance forecasting upon weighted mean is not necessarily the same to represent the data .

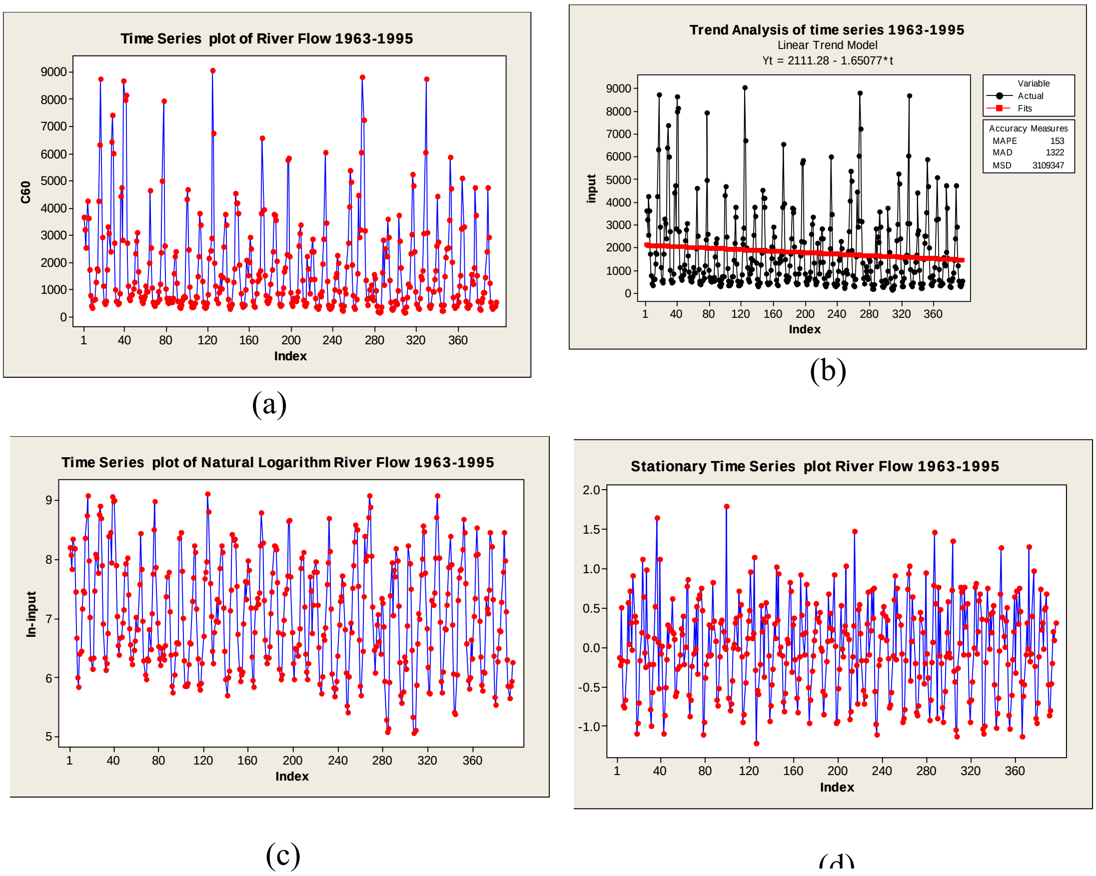
## 6. Application

The time series used in this paper is the monthly totals flow of Tigris River which entered Mosul City for the period 1963-1995(Suleiman (2008)). We used the data from the period 1963-1994 to estimate the model and the months of 1995, Table 1 below gives the monthly mean and standard deviation of Tigris River where it enter Mosul City for the period 1963-1994.

Table 1. Results for monthly mean and standard deviation of Tigris River where it enter Mosul City for the period 1963-1994

| month | Jan. | Feb. | Mar. | Apr. | May. | Jun. | Jul. | Aug. | Sep. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean (M-cu-m) | 1541 | 1875 | 3217 | 4784 | 4182 | 1690 | 723.2 | 421.9 | 357.2 | 503.4 | 873 | 1403 |
| S.D. | 966 | 868 | 1531 | 1882 | 2162 | 733 | 272.9 | 140.5 | 106.5 | 151.5 | 439.4 | 1014 |

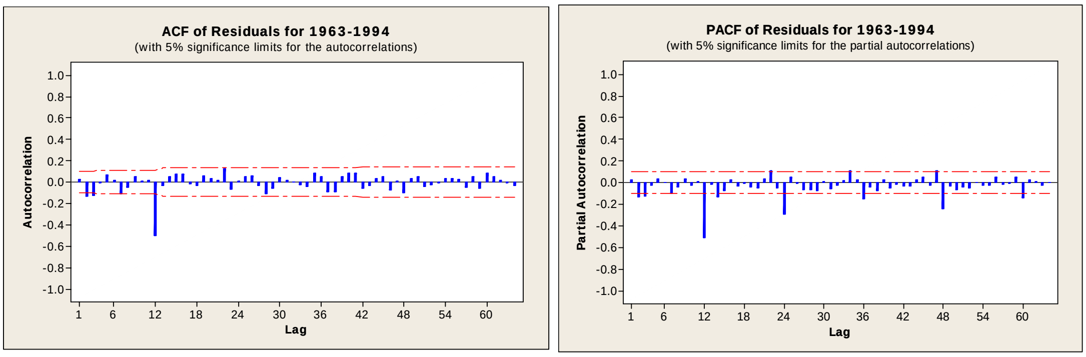
Fig.1 shows that the data have a downward trend together with seasonal behavior which gives the so-called multiplicative seasonality and this will reflex the complex nature of the hydrological series, the behavior of this series will generally be like this series that analyses of air line passengers by Chatfield and Faraway(1998), the standard Box-Jenkins analysis that satisfactory transformation to this type of seasonality takes natural logarithms to transform the behavior of seasonality from multiplicative to additive, also we take the sequentially and seasonality difference of order one to make the series stationary .




(a) (b) (c) (d)

Fig.1.Flow Tigris River when it enter to Mosul City from January 1963 to December 1995 :(a) raw data ;(b)Trend analysis ;(c) natural logarithms ;(d)stationary.

Through Fig.2. that shows sample autocorrelation function and sample partial autocorrelation function of stationary series we expect the seasonal autoregressive integrated moving average (SARIMA) model, of order $(3,1,2)\times(5,1,1)_{12}$ to be fitted as initially chosen.




(a) (b)

Fig .2. Stationary time series of Tigris River Flow: (a) ACF of the series ; (b) PACF of the series

The model diagnostic check includes checking the model sensitivity to the characteristics of the input data, for this Box and Jenkins proposed the overfitting procedure , which in starting with a high order model, if the previous model is already overfitted, order models, the dimensions of which are reduced and repeatedly checked against overfitting (Ajoy & Dobrivoje,2005). Following this manner to checking overfitting, Table 2. shows eleven candidate models of the seasonal autoregressive integrated moving average with Schwarz Bayesian criterion (SBC), also bias corrected version of the Akaike information criterion which denotes ($AIC_c$)

Table 2. Results for various SARIMA models

| | model | SBC | $AIC_c$ |
|---|---|---|---|
| 1 | $(3,1,2)\times(5,1,1)_{12}$ | 277.06636 | 231.09154 |
| 2 | $(2,1,2)\times(5,1,1)_{12}$ | 271.17036 | 228.96365 |
| 3 | $(1,1,2)\times(5,1,1)_{12}$ | 265.77808 | 227.35144 |
| 4 | $(0,1,2)\times(5,1,1)_{12}$ | 282.20356 | 247.56886 |
| 5 | $(1,1,2)\times(4,1,1)_{12}$ | 261.16785 | 226.53341 |
| 6 | $(1,1,2)\times(3,1,1)_{12}$ | 288.61568 | 257.78467 |
| 7 | $(1,1,2)\times(2,1,1)_{12}$ | 251.68925 | 224.67363 |
| 8 | $(1,1,2)\times(1,1,1)_{12}$ | 245.89864 | 222.70997 |
| 9 | $(0,1,2)\times(1,1,1)_{12}$ | 263.19018 | 243.83994 |
| 10 | $(0,1,2)\times(0,1,1)_{12}$ | 258.13623 | 242.6358 |
| 11 | $(1,1,2)\times(0,1,1)_{12}$ | 240.30127* | 220.95103* |

(*)minimum value

From the above table we choose model SARIMA $(1,1,2)\times(0,1,1)_{12}$ to represent the data of series depending on SBC, $AIC_c$ criterions and access the diagnostic checking (ACF and PACF of the residuals and Box-peries test) as shown below

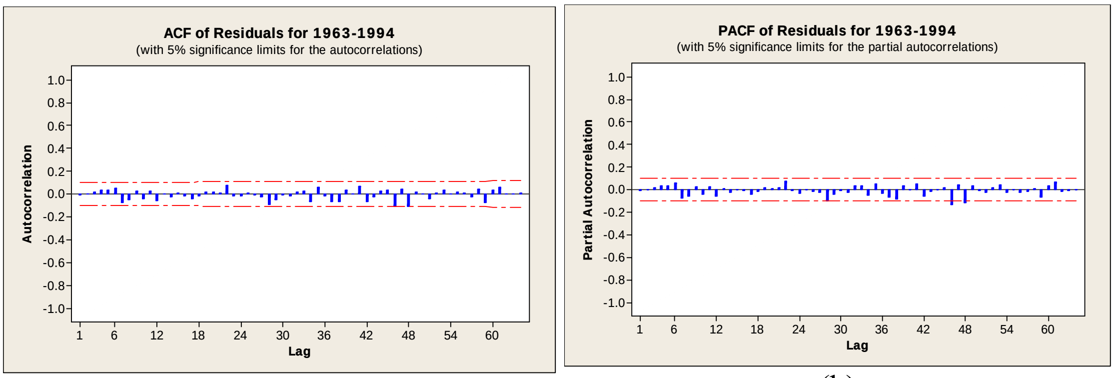
Final Estimates of Parameters

| Type | Coef | SE Coef | t-value | P-value |
|---|---|---|---|---|
| AR 1 | 0.6337 | 0.0530 | 11.95 | 0.000 |
| MA1 | 0.8364 | 0.0563 | 14.85 | 0.000 |
| MA 2 | 0.1134 | 0.0484 | 2.34 | 0.020 |
| SMA 12 | 0.9114 | 0.0307 | 29.69 | 0.000 |
| Constant | 0.00007 | 0.0002376 | 0.31 | 0.759 |

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

| Lag | 12 | 24 | 36 | 48 |
|---|---|---|---|---|
| Chi-Square | 8.1 | 11.9 | 20.4 | 40.7 |
| DF | 7 | 19 | 31 | 43 |
| P-Value | 0.327 | 0.890 | 0.926 | 0.573 |

we show that term constant is insignificant and can be removed from the model.




(a) (b)

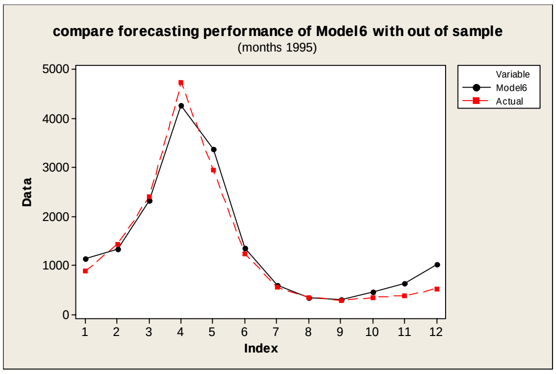
Fig .3.Diagnostic checking of the SARIMA $(1,1,2)\times(0,1,1)_{12}$ model: (a) ACF of the model residual ; (b) PACF of the model residual.

To choose best model for forecasting, Table 3 shows the result of performance criterions RMSE,MAPE and MAP by using out of sample months of 1995 year to eleven candidate models ,it is obviously the model selection upon any criterion is different from criterion to another. Upon RMSE criterion the best model performance to forecast is SARIMA $(1,1,2)\times(3,1,1)_{12}$, while MAPE criterion gives SARIMA $(0,1,2)\times(5,1,1)_{12}$ and last, MAP gives SARIMA $(0,1,2)\times(0,1,1)_{12}$, although these three models do not accesse the diagnostic checking, this result agrees with Mujumdar and Nagesh Kumar (1990) that best model for representating the data and best model for forecasting are often not the same. Also we show that SARIMA $(1,1,2)\times(3,1,1)_{12}$ model presents best coefficient of determination .

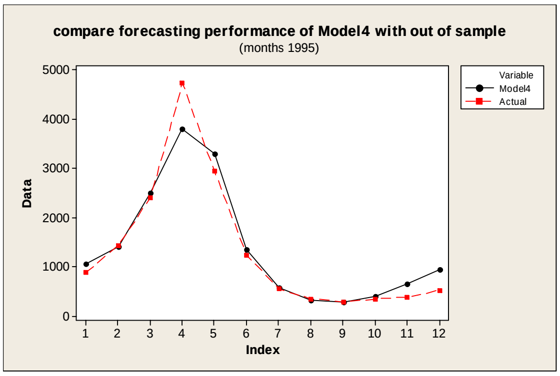
Table 3. Results for performance forecast to eleven candidate models

| | model | RMSE | MAPE | MAP | $R^2$ |
|---|---|---|---|---|---|
| 1 | $(3,1,2)\times(5,1,1)_{12}$ | 370.777 | 32.7033 | 281.108 | 0.943 |
| 2 | $(2,1,2)\times(5,1,1)_{12}$ | 371.331 | 32.707 | 281.014 | 0.943 |
| 3 | $(1,1,2)\times(5,1,1)_{12}$ | 370.257 | 32.4936 | 278.373 | 0.943 |
| 4 | $(0,1,2)\times(5,1,1)_{12}$ | 328.019 | 20.1644* | 206.173 | 0.951 |
| 5 | $(1,1,2)\times(4,1,1)_{12}$ | 340.293 | 29.6569 | 253.635 | 0.949 |
| 6 | $(1,1,2)\times(3,1,1)_{12}$ | 259.114* | 22.9657 | 194.163 | 0.968** |
| 7 | $(1,1,2)\times(2,1,1)_{12}$ | 342.744 | 32.2316 | 259.713 | 0.951 |
| 8 | $(1,1,2)\times(1,1,1)_{12}$ | 347.207 | 32.435 | 261.585 | 0.949 |
| 9 | $(0,1,2)\times(1,1,1)_{12}$ | 347.69 | 21.6013 | 192.728 | 0.953 |
| 10 | $(0,1,2)\times(0,1,1)_{12}$ | 352.695 | 20.811 | 188.594* | 0.955 |
| 11 | $(1,1,2)\times(0,1,1)_{12}$ | 321.784 | 28.1287 | 231.846 | 0.955 |

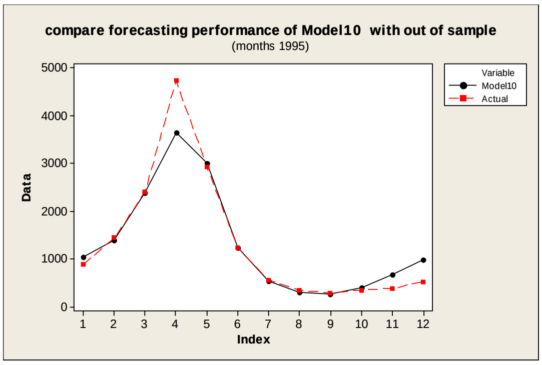
(*)minimum values to criterion : (**) maximum value of $R^2$




(a)




(b)




(c)

Fig. 4. performance of forecasting comparable with monthly data of 1995 year : (a)SARIMA $(1,1,2)\times(3,1,1)_{12}$; (b) SARIMA $(0,1,2)\times(5,1,1)_{12}$; (c) SARIMA $(0,1,2)\times(0,1,1)_{12}$.

From table 3. and Fig.4. the researcher found there is problem to select best performance of forecasting through three models, although fig.4. shows that performance of the model $(1,1,2)\times(3,1,1)_{12}$ will seem better. To solve this problem we try to use proposal statistical manner by calculating weighted mean to results of three criterions to any model as demonstrated in section 5. Table 4. shows that minimum weighted mean is with model $(1,1,2)\times(3,1,1)_{12}$ in all three forms of weighted mean, this is result consistent with the result in fig.4.,

Table 4. shows the weighted mean of three criterions of performance forecast of eleven candidate models.

| | model | weighted mean from eq.(9) | weighted mean from eq.(10) | weighted mean from eq.(11) |
|---|---|---|---|---|
| 1 | $(3,1,2)\times(5,1,1)_{12}$ | 130.45 | 299.487 | 254.652 |
| 2 | $(2,1,2)\times(5,1,1)_{12}$ | 130.529 | 299.841 | 254.682 |
| 3 | $(1,1,2)\times(5,1,1)_{12}$ | 129.767 | 298.649 | 252.707 |
| 4 | $(0,1,2)\times(5,1,1)_{12}$ | 102.475 | 256.402 | 195.479 |
| 5 | $(1,1,2)\times(4,1,1)_{12}$ | 118.759 | 274.077 | 230.748 |
| 6 | $(1,1,2)\times(3,1,1)_{12}$ | 90.857* | 208.931* | 176.455* |
| 7 | $(1,1,2)\times(2,1,1)_{12}$ | 121.897 | 277.154 | 235.638 |
| 8 | $(1,1,2)\times(1,1,1)_{12}$ | 123.089 | 280.475 | 237.663 |
| 9 | $(0,1,2)\times(1,1,1)_{12}$ | 104.471 | 267.515 | 190.034 |
| 10 | $(0,1,2)\times(0,1,1)_{12}$ | 104.089 | 270.031 | 187.98 |
| 11 | $(1,1,2)\times(0,1,1)_{12}$ | 111.024 | 257.852 | 212.883 |

(*)minimum values to weighted mean.

Finally, we note from fig.4 that there is a exist difference between three values (Jan.,Nov.and Dec.)of year 1995 and forecasting values, we can understand this difference if we look at table 5. below

Table 5. the difference between values of Jan.,Nov.and Dec. months of year 1995 and mean, standard deviation of the period 1963-1994 for these months.

| month | Mean period1963-1994 | Std. period1963-1994 | values1995 |
|---|---|---|---|
| Jan. | 1541.0 | 966.0 | 884.0 |
| Nov. | 873.0 | 439.4 | 381.0 |
| Dec. | 1403.0 | 1014.0 | 522.0 |

It is obvious exist decrease of water enter mosul city in Jan.,Nov.and Dec.,1995

## CONCLUSIONS

1- The suggested statistical manner proved its competence in limiting the best model according to forecasting performance that coincide with the determinant of coefficient criterion. There hence getting rid off the difference criterion problem for RMSE, MAPE, and MAE that limits the best model.

2- Changing the given weights for each criterion of the three criterions within the weighted mean in the suggested manner did not serve the change for reaching the best model from the side of forecasting performance, which leads to the flexibility and qualified suggested manner.

3- Box-Jenkins model is considered the best for representing the data not always is the best for providing a better forecasting, the application side approves it in the analysis the Tigris River water flow it appeared that the best model for representing the data is SARIMA $(1,1,2) * (0,1,1)_{12}$ while the best forecasting offered model is SARIMA $(1,1,2) * (3,1,1)_{12}$

4- The capability of Box-Jenkins model to deal with the hydrology data from the point that has the ability to provide very promising forecasting in spite of the unsure information effecting the time series of the river water flow.

5- The employment of long range in time hydrology data provides a larger opportunity for limiting Box-Jenkins model that can provide new forecasting.

6- Box -Jenkins manner can provide an opportunity to skip the overfitting problem through information criterions in a better way than other manners like neural networks.

7- Box-Jenkins manner skips the problem of falling for forecasting that are not acceptable like the manner used in neural networks that results from over training.

Lastly, we recommended using the suggested weighted mean for reaching the model that provides best forecasting in other forecasting manners like neural networks and fuzzy models .

**References**

1-Abdullah,S. Karaman and Tayfur,Altok (2004). An Experimental Study on Forecasting Using Tes Processes. Proceedings of the 2004 winter simulation conference , eds, lngalls R.G.,Rossetti M.D.,Smith J.S. and Peters B.A., PP.437-442.

2-Ajoy, K.Palit and Dobrivoje,Popovic (2005). Computational Intelligence in Time Series Forecasting – Theory and Engineering Applications. Springer – Verlag London Limited.

3-Chatfield, Chris (1995). The Analysis of Time Series – An Introduction. 5th .ed. Chapman &Hall/CRC, New york.

4-Chatfield, Chris and Faraway, Julian (1998). Time Series Forecasting with Neural Networks:A Comparative Study Using the Airline Data. Applied Statistics ,volume 47, Issue 2,pp.231-250.

5-Hillier, Frederick S. and Lieberman, Gerald J.(2001). Introduction to Operations Research.7th ed. Mc Graw- Hill Companies ,New York,U.S.A.

6-Lendasse,A.,DE BODT,E.,Wertz,V. and Verleysen.M.(2000).Non-Linear financial time series forecasting – Application to the Bel 20 stock market index. European Journal of Economic and Social Systems 14 N°1,pp.81-91.

7-Liu,L.-M.(2006). Time Series Analysis and Forecasting. 2nd ed. Scientific Computing Associates Crop.Illinois,U.S.A.

8-Jack, Johnston & John , Dinardo (1997). Econometric Methods.Mc Graw- Hill – A Division of the Mc Graw- Hill Companies.

9-Mujumdar,P.P., and Nagesh Kumar,D. (1990).Stochastic models of streamflow – Some case studies. Hydrol.Sci.J.35,pp.395-410.

10-Nagesh Kumar, D.,Sathish,T. and Srinivasa,R.(2004). River Flow Forecasting Using Recurrent Neural Networks . J.Water Resources Management 18:pp.143-161.

11-Pena,D.(2001). Prediction and Model Selection: A Course in Time Series Analysis (Edited by Pena,D , Tiao G. and Tsay R.S.) John Wiley &Sons, Inc.Newyork.

12-Pomares, H. and Rojas,I.(2004). Soft – computing techniques for time series forecasting . Proceedings- European Symposium on Artificial Neural Networks,ISBN 2-930307-04-8,PP.93-102.

13-Suleiman, Intisar Ibrahim (2008).Forecasting by Time Series Models and Feed Forward Artificial Neural Network with Application – Comparative Study. Unpublished M.Sc thesis ,College of Computer Sciences and Mathematics, University of Mosul, Iraq.

14-Vandaele, Walter (1983). Applied Time Series and Box-Jenkins Models. Academic Press.

15-Wojbor,A.Woyczyn'ski (2006). A First Course in Statistics for Signal Analysis. Birkhauser Boston.