American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-02, Issue-06, pp-49-54
www.ajer.us

Research Paper Open Access

Stochastic Modelling of Shiroro River Stream flow Process

Musa, J. J.
Department of Agricultural & Bioresources Engineering, School of Engineering & Engineering Technology, Federal University of Technology, P M B 65, Minna, Nigeria

Abstract: - Economists, social scientists and engineers provide insights into the drivers of anthropogenic climate change and the options for adaptation and mitigation, and yet other scientists, including geographers and biologists, study the impacts of climate change. This project concentrates mainly on the discharge from the Shiroro River. A stochastic approach is presented for modeling a time series by an Autoregressive Moving Average model (ARMA). The development and use of a stochastic stream flow model involves some basic steps such as obtain stream flow record and other information, Selecting models that best describes the marginal probability distribution of flows. The flow discharge of about 22 years (1990-2011) was gotten from the Meteorological Station at Shiroro and analyzed with three different models namely; Autoregressive (AR) model, Autoregressive Moving Average (ARMA) model and Autoregressive Integrated Moving Average (ARIMA) model. The initial model identification is done by using the autocorrelation function (ACF) and partial autocorrelation function (PACF). Based on the model analysis and evaluations, proper predictions for the effective usage of the flow from the river for farming activities and generation of power for both industrial and domestic us were made. It also highlights some recommendations to be made to utilize the possible potentials of the river effectively.

Keywords: - *ARMA, ARIMA, AR, Climate, water, stream flow*

I. INTRODUCTION

There had not been serious attention given to the depleting ozone layer, global warming and climate change until about four decades ago when it became obvious that anthropogenic damage to the earth's stratospheric ozone layer will lead to an increase in solar ultraviolet (UV) radiation reaching the earth's surface, with a consequent adverse impact, (Ghanbarpour, *et. al*., 2010). Climate change is a complex and comprehensive process that can only be understood on the basis of the combined insights from various scientific disciplines (Saremi, *et. al*., 2011). Natural scientists contribute to an improved understanding by looking at issues like the global energy balance, the carbon cycle and changes in atmospheric composition (Gangyan, *et. al*., 2002). At the same time, economists, social scientists and engineers provide insights into the drivers of anthropogenic climate change and the options for adaptation and mitigation, and yet other scientists, including geographers and biologists, study the impacts of climate change (Szilagyi, et. al., 2006; Sharif, et. al., 2007; Krishna, et. al., 2011). They also stated that a key factor of interaction is the availability of water. Water is needed for agriculture, energy production, residential water demand and industry and will be influenced by climate change. These impacts could, certainly locally, be so strong that they would influence the human activities sufficiently to create feedbacks.

Water resources play a crucial role in the economic development of Nigeria. Due to the increasing population growth and resulting demands on limited water resources, an efficient management of exiting water resources needs to be put in place for further use rather than building new facilities to meet the challenge. In the water management communities, it is well known that to combat water shortage issues, maximizing water management efficiency based on stream flow forecasting is crucial.

In design, the hydrologist is most often required to estimate the magnitude of river flow for an ensuing period of hours, days, months or possibly longer. The time sequence of flows during critical periods can be of considerable importance. The operation of a reservoir is necessarily based on anticipated flows into the reservoir and at key points downstream. Reliable flow forecasts are particularly important in the case of multipurpose reservoirs, and they are indispensable to the operation of flood mitigation reservoir systems (Cigizoglu, 2003; Valipour, 2012).

Generation of synthetic sequences of daily hydroclimatic variables like stream flow is often used for efficient short-term and long-term planning, management and assessment of complex water resources systems. Downscaling methods are an important component of the hydrologist's tool kit for generating such flow traces, which should be statistically indistinguishable from the observations (Rajagopalan *et al*., 2010). On the other hand, nonparametric methods require only a limited set of assumptions about the structure of the data, and they may therefore be preferable when *a priori* postulations required for parametric models are not valid (Higgins, 2004).

According to Otache *et al*., (2011), the principal aim of time series analysis is to describe the history of movement in time of some variables such as the rate flow in a dam at a particular size. Time series modeling for either data generation or forecasting of hydrologic variables is an important step in planning and operational analysis of water resource systems.

A stochastic approach is presented for modeling a time series by an Autoregressive Moving Average model (ARMA). This enforces stationarity on the autoregressive parameters and in inevitability on the moving average parameter, thus taking into account the uncertainty about the correct model by averaging the parameter estimates. Several stochastic models have been proposed for modeling hydrological time series and generating synthetic stream flows. These include ARMA models, disaggregation models, models based on the concept of pattern recognition. Most of the time-series modeling procedures fall within the framework of multivariate ARMA models (Otache and Bakir, 2008). Generally, AR models and Autoregressive Integrated Moving Average (ARIMA) models have an important place in the stochastic modeling of hydrologic data. Such models are of value in handling what might be described as the short-run problem; that of modeling the seasonal variability in a stochastic flow series.

The Box-Jenkins methodology, commonly known as the ARIMA model, has already been widely used in a number of related areas such as economic time series forecasting, ecological and weather prediction, medical monitoring, traffic flow prediction, and also physical activity recognition. Generally, the application of ARIMA models is mostly focused on predicting a single univariate time series (Halim, *et al*., 2007). Box and Jenkins (1976) stated that the ARIMA modeling aims at constructing the most appropriate model to fit observed data. Several types of ARIMA modeling methods and their derivatives could be used in the modeling seasonal time series, such as monthly stream flow time series. They are seasonal ARIMA, periodic ARIMA and deseasonalized ARMA model. The deseasonalized ARMA type of modeling strategy was adopted in this study due to its simplicity and effectiveness of modeling. The general form of ARIMA model is expressed as (Vandaele, 1983, Otache, *et. al*., 2011):

$$\varphi(B)y_t = \theta(B)a_t \qquad 1$$

Where

$y_t = (1 - B)^d Y_t$ - Stationary series after differencing

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots \ldots - \varphi_p B^p$ - Non-seasonal autoregressive polynomial

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots \ldots \ldots - \theta_q B^q$ - Non-seasonal moving average polynomial

a_t = white noise process

Y_t = dependent variable

B is the backward shift operator defined as $BX_t = X_{t-1}$

Examination of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) provides a thorough basis for analyzing the system behavior under time dependence, and will suggest the appropriate parameters to include in the model. The Box and Jenkins (1976) three-stage standard modeling procedure (identification, estimation, and diagnostic check) can be used to develop ARIMA models.

The objective of this study is aimed at analyzing the flow discharge from Shiroro River between 1990-2011 using AR, ARMA and ARIMA Models.

II. MATERIALS AND METHODOLOGY

The study area is located on latitudes 9^0 55^1 and 10^0 00^1N and longitudes 6^0 40^1 and 6^0 45^1E with its elevation ranging between 274 and 305 m. The Shirroro hydro-electricity dam has it source of water supply form river Kaduna. This study makes us of the inflow and outflow information/data form the dam.

The development and use of a stochastic stream flow model involves some basic steps such as obtain stream flow record and other information, Selecting models that best describes the marginal probability distribution of flows in different sections and estimate the models parameters, selecting an appropriate model of the spatial and temporal dependence of the stream flows, verifying the computer implementation of the model, and validating the model for water resources system information.

The ARMA model basically includes the AR, and the seasonal ARIMA models (Vandaele, 1983, Otache, 2011). Box and Jenkins (1976) give the paradigm for fitting ARMA models as

1. Model identification:-Determination of the ARM model orders.
2. Estimation of model parameters:-The unknown parameters in equation 1 are estimated.
3. Diagnostic and Criticism:-the residuals are used to validate the model and interval suggests potential alternative models which may be better.

These steps are repeated until a satisfactory model is found. To enhance the understanding of these paradigms, a brief discussion of the steps is imperative here.

Model Identification

The initial model identification is done by using the autocorrelation function (ACF) and partial autocorrelation function (PACF). Despite this, an alternative procedure for selecting the model order is by using a penalized log likelihood measure. One of the popular measures is the Akaike's information criterion (AIC). This is defined as;

$$AIC(k) = 2\log ML + 2k \qquad 2$$

Where ML is the maximum likelihood and K is the number of independently adjusted parameters within the model.

The best model is the one with the lowest AIC value for ARMA (*p*, *q*) models, *k* = *p* + *q*, and the AIC value can be calculated as;

$$AIC\ (p,q) = N\log(\delta_t^2) + 2\ (p+q) \qquad 3$$

Where, δ_t^2 is the variance of the innovation process.

III. RESULTS AND DISCUSSION

The analysis was carried out using the MatLab 2009 statistical package. The results are presented in the Autocorrelation Functions (ACF) and Partial autocorrelation function (PACF) graphs to show the iterated variables in a simplified form. The ACF was initially carried on the available data. The stationarity condition here implies that the mean and the variance of the process were constant. The autocovariances model developed is stated in equation 4 below

$$r_k = cov(Z_t, Z_{t-k}) = E\ (Z_t - N)(Z_{t-k} - N) \qquad 4$$

While that of the autocorrelation is obtained as

$$p_k = \frac{cov\,(Z_t, Z_{t-k})}{[v(Z_t)\,.v(Z_{t-k})]^{1/2}} \qquad 5$$

k depends on the lag or time deference since these conditions apply only to the first and second-order or weak stationarity. The autocorrelations p_k are independent of the scale of time series which is considered as a function of k and thus referred to as the autocorrelation function (ACF) or correlogrm. Since $r_k = r - k$ $\left[r_k = \right.$ *cov Zt,Zt−k=cov Zt−k,Zt=cov Zt,Zt+k=r−k* and *pk=p−k*. It is important to note that only the positive half of the ACF is usually considered. Figures 1 and 2 below shows the ACF for the standardized monthly and unstandardized flow of Shirorro River respectively.

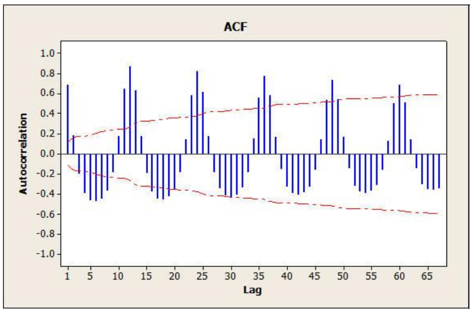


Fig. 1 Autocorrelation for standardized monthly flow

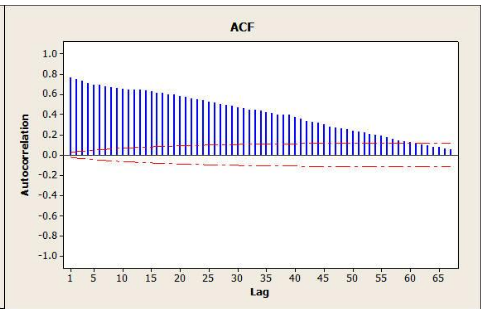


Fig. 2 Autocorrelation for daily flow (unstandardized)

The positive section of the graph shows that there exists seasonality effect on the monthly flow of the river. This further implies that carrying out ACF alone won't satisfy to build our model, thus the need for PACF. The ACF plays a major role in modeling the dependencies among observations, since it characterizes together with the process mean $E(Z_t)$ and variance $r_o = v(Z_t)$, the stationary stochastic process.

The estimate of p_k is given by the lag k sample autocorrelation in equation 6

$$\vartheta_{kk} = \frac{r_k - \sum_{j=1}^{k=1} \vartheta_{k-1,}k_{k-j}}{1 - \sum_{j=1}^{k=1} \vartheta_{k-1,j} r_j} \qquad 6$$

J=0,1,2------------k-1

K=0,1,2------------k-1

For uncorrelated observations the variance of r_k is given by

$$V(r_k) = \frac{1}{n} \qquad 7$$

$$\vartheta_{k,j} = \vartheta_{k-1,j} - \vartheta_{kk}\, \vartheta_{k-1,k-j} \qquad 8$$

The plots of the PACF for the standardized monthly and daily flow of the Shiroro River are given below in the figure 3 and 4 respectively. It was noticed that the PACF is of a particular form. The autocorrelations decrease as the lag k increase indicating that observations closer together are more correlated than the ones far apart. For Ø > 0 the autocorrelations decay geometrically to zero, and for Ø < 0 the autocorrelations decay in an oscillatory pattern.

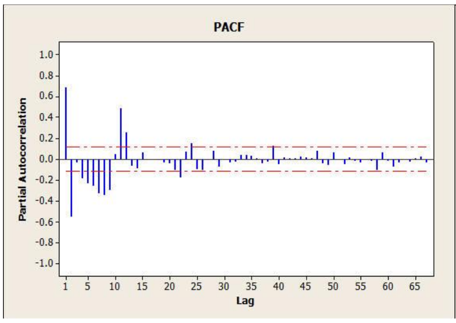


Fig. 3 Partial autocorrelation for standardized monthly flow

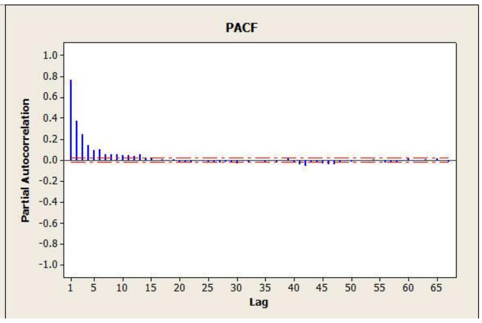


Fig. 4 Partial autocorrelation for daily flow (unstandardized)

The standardized data showed some degree of seasonality and thus gave room for the gap in the data points thereby standardizing the data brought which about a uniform decay in our data and the plot is as given in figure 4. From the graph, it was seen that our model is of lag 1 which gave us the basis for selecting our AR (p) model.

The ACF of the residuals in our data reveals addition structure in the data that the regression did not capture. Instead, the introduction of correlation as a phenomenon that leads to proposing the AR and ARMA models. Adding nonstationary models to the mixed leads to the ARIMA model. Adding nonstationary models to the mix leads to the ARMA models popularized by Box and Jenkins (1970).

AR models are based on the idea that the current value of the series, x_t, can be explained as a function of past values, x_{t-1}, x_{t-2},--------, x_{t-p}, where p determines the number of steps into the past needed to forecast the current value.

An AR model of order p, is presented in equation 9

$$X_t = \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + ---- + \emptyset_p x_{t-p} + w_t \qquad 9$$

Where x_t is stationary and $\emptyset_1$, $\emptyset_2$,----------- $\emptyset_p$ are constants ($\emptyset_p \neq 0$). Assuming w_t is a Gaussian white noise series with mean zero and variance δ^2_w, μ= mean of x_t

$$X_t - \mu = \emptyset_1(x_{t-1} - \mu) + \emptyset_2(x_{t-2} - \mu) + ---- + \emptyset_p(x_{t-p} - \mu) + w_t \qquad 10$$

Or

$$X_t = \alpha + \emptyset_1 x_{t-2} + \emptyset_2 x_{t-2} + ---- + \emptyset_p x_{t-p} + w_t \qquad 11$$

Where

α= μ (1-Ø,-------------$Ø_p$)

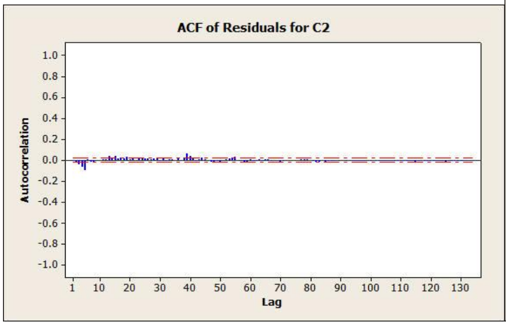
Equation 11 above have some technical difficulties, because of the regressors, x_{t-1},---- ,x_{t-p}, which are random components, A useful form follows by using the backshift operator to write the AR (p) model, (1) as

$$(1 - \emptyset_1 B - \emptyset_2 B^2 ----- \emptyset_p B^p)x_t = w_t \qquad 12$$

Or

$$\emptyset(B)x_t = w_t \qquad 13$$

Figure 5 shows the ACF of residual daily flow (AR) and figure 6 shows the PACF of residual for daily flow (AR). From the graphs, the value for ACF and PACF in terms of lag is seen to be not significant since it both shows very low lag value, thus, lag 1 was chosen.



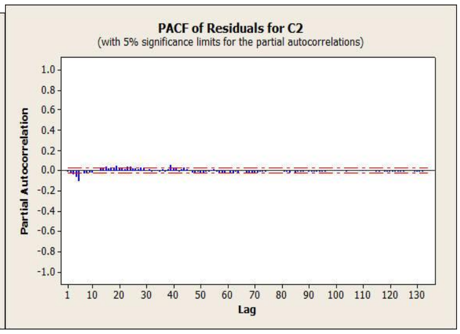


Fig. 5: Autocorrelation functions of residuals for daily flow (AR) Fig. 6: Partial autocorrelation functions of residuals for daily flow (AR)

Proceeding with the general development of autoregressive moving average and mixed ARMA models.

A time series of the form $\{x_t; t = 0, \pm1, \pm2, ----3\}$ is ARMA (p, q) if it is stationary and

$$X_t = \emptyset_1 x_{t-1} + ---- + \emptyset_p x_{t-p} + w_t + \theta_1 w_{t-1} + --- + \theta_q w_{t-p} \quad 14$$

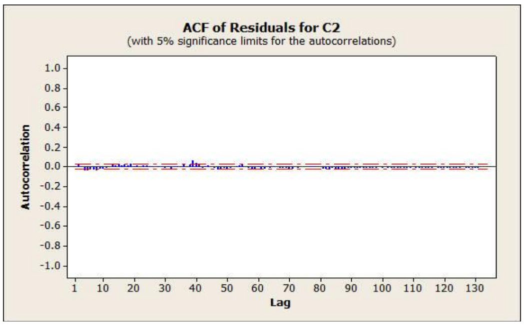
With $\O_p \neq 0$, $\Theta_p \neq 0$, and $\delta^2_w > 0$. The parameters p and q are called the autoregressive and the moving average order respectively. If x_t has a nonzero mean μ, we set $\alpha = \mu\,(1-\O_1---------- \O_p)$ and write the model as

$$X_t = \alpha + \emptyset_1 x_{t-1} + --- + \emptyset_p x_{t-1} + w_t + \theta_1 w_{t-1} + --- + \theta_q w_{t-q} \quad 15$$

To aid the investigation of ARAM models it will be useful to write them using the AR operator, and the MA operator. The ARMA (p, q) model can be written in concise form as

$$\emptyset(B) x_t = \theta(B) w_t \quad 16$$

ARMA models were carried out on the flow data as well and it was observed that the lag values obtained for both ACF and PACF were of very low ranges. This is indicated in the figure below.



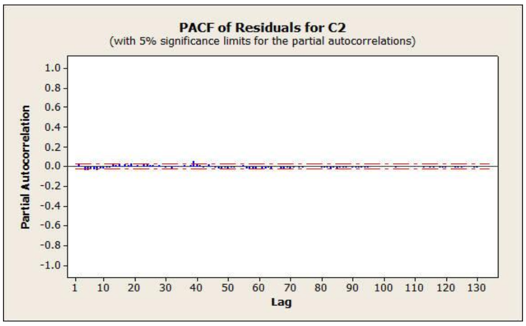
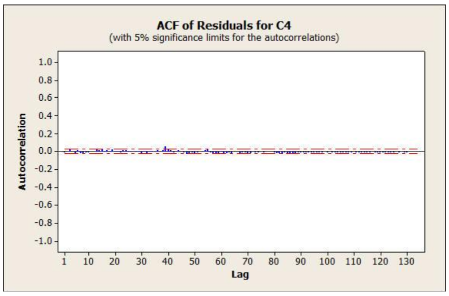


Fig. 7 Autocorrelation functions of residuals for daily flow (ARMA) Fig. 8 Partial autocorrelation functions of residuals for daily flow (ARMA)

It is seen also from the figures 7 and 8 above as analyzed with ARMA model, that the ACF and PACF has very low lag values. Thus, lag 1 is also selected as the best option for the model. Including an integrated domain we have an ARIMA model figure which further show the relations in our river flow. This is shown in the close relations in the data points from figure 9 and 10 respectively.

Since the ACF and PACF shows some seasonality in the ARIMA model due to the closeness in their data points, thus, ARIMA model is similar to both AR and the ARMA model, but only has a difference in its components. The ARIMA plots for both ACF and PACF are as given below figure 9 and 10 respectively;



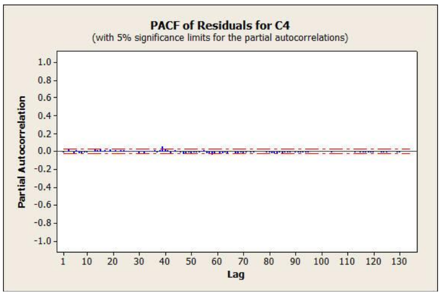


Fig. 9 Autocorrelation functions of residuals for daily flow (ARIMA) Fig. 10 Partial autocorrelation functions of residuals for daily flow (ARIMA)

IV. CONCLUSIONS

The study was able to conclude that the Shiroro river shows some correlated properties, but can still be used all year round with some degree of scheduling. This is because the amount of flow from the dam in each month of the year is still sufficient enough for optimum usage, though irrigation has to be regulated during the dry season, so as not to affect the dam reservoir level. Thus, in conclusion, the Shiroro River can be used for agriculture activities all year round but with some scheduling during the PICK dry season.

With proper regulations, there is a chance that if another dam is built along the down-stream after some contributing tributaries, there will be sufficient water to still produce Hydro power and water supply for the surrounding areas and Niger state as a whole.

REFERENCE

[1] Box, G. E. P., and Jenkins, G. M., (1976): Time Series Analysis Forecasting and Control, Holden-Day, San Francisco, CA.

[2] Cigizoglu, H. K., (2003): Incorporation of ARMA models into flow forecasting by artificial neural networks. *Environmetrics; 14: 417–427*

[3] Gangyan, Z., Goel, N. K., and Bhatt, V. K. (2002): Stochastic modeling of the sediment load of the upper Yangtze River (China). *Hydrological Sciences-Journal, 47: 93 – 105*

[4] Ghanbarpour, M. R., Abbaspour, K. C., Jalalvand, G., and Moghaddam, G. A (2010): Stochastic modeling of surface stream flow at different time scales: Sangsoorakh karst basin, Iran. *Journal of Cave and Karst Studies, v. 72, no. 1, p. 1–10*

[5] Halim, S., Bisono, I.N. Melissa, and Thia, C., (2007): Automatic Seasonal Auto Regressive Moving Average Models and Unit Root Test Detection, *IEEE International Conference on Industrial Engineering and Engineering Management*, Singapore. Pp. 1129-1133.

[6] Higgins, J. J., (2004): Introduction to Modern Nonparametric Statistics (Brooks/Cole, Pacific Grove, CA, 366 pp).

[7] Krishna, B., Satyaji Rao, Y. R., and Nayak, P. C., (2011): Time Series Modeling of River Flow Using Wavelet Neural Networks. *Journal of Water Resource and Protection, 3, 50-59*

[8] Otache Y.M, M.A. Sadeeq, and I.E Ahaneku (2011), ARMA modeling River flow Dynamics: Comparative study of PAR model. *Open Journal of Modern hydrology, vol. 1 no. 1, pp 1-9*

[9] Otache Y.M., and Bakir M., L, Z (2008): Analysis of stochastic characteristics of the Benue River Flow process. *Chin J. oceanol Limnol 26 : 151*

[10] Rajagopalan, B., Salas, J.D., and Lall, U., (2010): Stochastic methods for modeling precipitation and stream flow. In: *Advances in Data-based Approaches for Hydrologic Modeling and Forecasting.* B. Sivakumar and R. Berndtsson, Eds., World Scientific, Singapore, 17-52.

[11] Saremi, A., Pashaki, M. H. K., Sedghi, H., Rouzbahani, A., and Saremi, A., (2011): Simulation of River Flow Using Fourier Series Models. *International Conference on Environmental and Computer Science vol.19: 133 - 138*

[12] Sharif, H. O., Crow, W., Miller, N. L., and Wood, E. F., (2007): Multidecadal High-Resolution Hydrologic Modeling of the Arkansas–Red River Basin. *Journal of Hydrometeorology, Vol 8: 1111 – 1127*

[13] Szilagyi, J., Balint, G., and Csik, A., (2006): Hybrid, Markov Chain-Based Model for Daily Stream flow Generation at Multiple Catchment Sites. *Journal of Hydrologic Engineering. Pp. 245 -256*

[14] Valipour, M. (2012): Parameters Estimate of Autoregressive Moving Average and Autoregressive Integrated Moving Average Models and Compare Their Ability for Inflow Forecasting. *Journal of Mathematics and Statistics 8 (3): 330-338*

[15] Vandaele, W. (1983): *Applied time series and Box-Jenkins models* (Academic Press, New York).