STREAM FLOW FORECASTING OF SAVITRI RIVER USING ARIMA MODEL

A Thesis submitted to the

DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH
DAPOLI - 415 712
Maharashtra State (India)

In the partial fulfillment of the requirements for the degree

of

MASTER OF TECHNOLOGY
(AGRICULTURAL ENGINEERING)

in

SOIL AND WATER CONSERVATION ENGINEERING

by

Dhiraj Dhanaji Ahire.

DEPARTMENT OF SOIL AND WATER CONSERVATION ENGINEERING
COLLEGE OF AGRICULTURAL ENGINEERING AND TECHNOLOGY
DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH
DAPOLI- 415 712, DIST.RATNAGIRI, M.S. (INDIA)
2014

STREAM FLOW FORECASTING OF SAVITRI RIVER USING ARIMA MODEL

A Thesis submitted to the

DR. BALASAHEB SAWANT KONKAN KRISHI VIDYAPEETH
DAPOLI - 415 712
Maharashtra State (India)

In the partial fulfillment of the requirements for the degree

of

MASTER OF TECHNOLOGY
(AGRICULTURAL ENGINEERING)

in

SOIL AND WATER CONSERVATION ENGINEERING

Approved by the Advisory Committee

Chairman and Research Guide

(S.B. Nandgude)
Associate Professor
Department Of Soil and Water Conservation Engineering
College of Agricultural Engineering and Technology, Dapoli

Members

(dilip MAHALE)
Professor and Head,
Department Of Soil and Water
Conservation Engineering
College of Agricultural Engineering
and Technology, Dapoli

(B.L.Ayare)
Agriculture Engineer
(Associate Professor) Water
Management Scheme, Wakawali,
Dapoli

(Er. H.N. Bhange)
Assistant Professor,
Department of soil and water conservation
Engineering, College of Agricultural
engineering and technology, Dapoli

CANDIDATE'S DECLARATION

I hereby declare that this thesis or part of thereof has not been submitted

by me or any other person to any other

University or Institute

for a degree or

Diploma

Place: Dapoli

Date: / / **(Dhiraj Dhanaji Ahire)**

Dr. S.B. Nandgude.
M.Tech, Ph. D. (SWCE), IIT Kharagpur.
Chairman and Research Guide,
Department of Soil and Water Conservation Engineering,
College of Agricultural Engineering and Technology,
Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth,
Dapoli- 415 712, Dist. Ratnagiri,
Maharashtra, India.

CERTIFICATE

This is certify the thesis entitled **"STREAM FLOW FORECASTING OF SAVITRI RIVER USING ARIMA MODEL"** submitted to the faculty of Agricultural Engineering, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli, Dist. Ratnagiri (Maharashtra State) in the partial fulfillment of the requirement for the award of the degree of **Master of Technology (Agricultural Engineering) in Soil and Water Conservation Engineering,** embodies the record of a piece of bonafied research work carried out by **Mr. Dhiraj Dhanaji Ahire** under my guidance and supervision. No part of this thesis has been submitted for any other degree, diploma or publication in any other form.

The assistance and help received during the course of this investigation and source of the literature have been duly acknowledged.

Place: Dapoli

Date: / / **(S.B. Nandgude)**

Prof. dilip MAHALE
B. Tech. (Agril. Engg.), M. Tech. (SWCE).
Associate Dean,
Faculty of Agricultural Engineering,
Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth,
Dapoli 415 712, Dist. Ratnagiri,
Maharashtra, India.

CERTIFICATE

This is to certify that the thesis entitled "**STREAM FLOW FORECASTING OF SAVITRI RIVER USING ARIMA MODEL**" submitted to the Faculty of Agricultural Engineering, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli, Dist. Ratnagiri (Maharashtra State) in the partial fulfillment of the requirements for the award of degree of Master of Technology (Agricultural Engineering) in Soil and Water Conservation Engineering is a record of bonafied research work carried out by **Mr. Dhiraj Dhanaji Ahire** under the guidance and supervision of Dr. S.B.Nandgude, Associate Professor Department Of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli. No part of the thesis has been submitted for any other degree, diploma or publication in any other form.

The assistance and help received during the course of this investigation and source of the literature have been duly acknowledged.

Place: Dapoli.

Date: / /

(**dilip MAHALE)**

ACKNOWLEDGEMENT

*It is my proud privilege to place on record, my sincere thanks to research guide **Dr. S. B. Nandgude**, Associate Professor, Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, Dapoli for his precious guidance and valuable help throughout the research work. The pleasant experience of working with him and counsel he gave will be remembered and appreciated for whole life.*

*I mention my sincere gratitude to **Prof. dilip MAHALE**, Associate Dean. College of Agricultural Engineering and Technology, Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli who gave me an opportunity for undergoing this research work and providing necessary facilities for whenever needed.*

*I am thankful to **Dr. B.L.Ayare**, Associate Professor, Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, Dapoli. **Er. H.N. Bhange**, Assistant Professor, Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, Dapoli, **Er. K. D. Garade**, Assistant Professor, Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, Dapoli for encouragement and motivation in the research work.*

*I am especially indebted to **Dr. S.D.Gorantiwar**, Professor and Head, Department of Irrigation and Drainage Engineering, Dr.Annasaheb Shinde College of Agricultural Engineering, Rahuri, for his valuable suggestions and guidance from very inception of the research work.*

*I am thankful to **Mrs. S.S. Nagarkar**, Senior Research Assistant, Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, Dapoli, **Shri. S.S. Idate**, Lab Assistant, Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, Dapoli, for their good co-operation and always rendered me helping hands during the entire cource.*

*Grateful thanks are extended to **Chief Engineer**, Hydrology Project, Nashik, Superintendent Engineer, Data analysis circle, Nashik, providing all necessary help for completion of this research work.*

*I am also grateful to **Shri. Pramod Phophale**, Ph.D. Scholar, Department of Irrigation and Drainage Dr.Annasaheb Shinde College of Agricultural Engineering, Rahuri, for kindly help to completion of this research work.*

I am very glad to express my thanks to my friends that it was impossible to complete the research work without their assistance for the constant encouragement and help during my project work.

*I am extremely obliged to acknowledge the love and affection of my beloved parents **Dhanaji** and **Kusum** and my wife **Sayali**, and my sweet daughter **Barbie** and other family members. I express my sincere thanks to all whom directly and indirectly extended help during the research work.*

Place: Dapoli

Date: / / **(Dhiraj Dhanaji Ahire)**

LIST OF FIGURES

Sr.No.	Title	Page No.
3.1	Schematic representation of the Box-Jenkins methodology for time series modelling	16
4.1	ACF of original monthly data without differencing	27
4.2	PACF of original monthly data without differencing	28
4.3	Streamflow time series ACF and PACF of streamflow d=0;D=0	29
4.4	Streamflow time series ACF and PACF of streamflow d=1;D=0	29
4.5	Streamflow time series ACF and PACF of streamflow d=0;D=1	30
4.6	Streamflow time series ACF and PACF of streamflow d=1;D=1	30
4.7	Streamflow time series ACF and PACF of ARIMA model $(0,1,1)_x(0,1,1)_{12}$	32
4.8	Streamflow time series ACF and PACF of ARIMA model $(0,1,1)_x(1,1,0)_{12}$	32
4.9	Streamflow time series ACF and PACF of ARIMA model $(0,1,1)_x(1,1,1)_{12}$	33
4.10	Streamflow time series ACF and PACF of ARIMA model $(1,1,0)_x(0,1,1)_{12}$	33
4.11	Streamflow time series ACF and PACF of ARIMA model $(1,1,0)_x(1,1,0)_{12}$	34
4.12	Streamflow time series ACF and PACF of ARIMA model $(1,1,0)_x(1,1,1)_{12}$	34
4.13	Streamflow time series ACF and PACF of ARIMA model $(1,1,1)_x(0,1,1)_{12}$	35
4.14	Streamflow time series ACF and PACF of ARIMA model $(1,1,1)_x(1,1,0)_{12}$	35
4.15	Streamflow time series ACF and PACF of ARIMA model $(1,1,1)_x(1,1,1)_{12}$	36

4.16	Streamflow time series ACF and PACF of ARIMA model $(0,0,1)_x(0,1,1)_{12}$	36
4.17	Streamflow time series ACF and PACF of ARIMA model $(0,0,1)_x(1,1,1)_{12}$	37
4.18	Streamflow time series ACF and PACF of ARIMA model $(1,0,0)_x(0,1,1)_{12}$	37
4.19	Streamflow time series ACF and PACF of ARIMA model $(1,0,0)_x(1,1,0)_{12}$	38
4.20	Streamflow time series ACF and PACF of ARIMA model $(1,0,0)_x(1,1,0)_{12}$	38
4.21	Streamflow time series ACF and PACF of ARIMA model $(1,0,0)_x(1,1,1)_{12}$	39
4.22	Streamflow time series ACF and PACF of ARIMA model $(1,0,1)_x(0,1,1)_{12}$	39
4.23	Streamflow time series ACF and PACF of ARIMA model $(1,0,1)_x(1,1,0)_{12}$	40
4.24	Streamflow time series ACF and PACF of ARIMA model $(1,0,1)_x(1,1,1)_{12}$	40
4.25	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,1,1)\times(0,1,1)_{12}$	43
4.26	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,1,1)\times(1,1,0)_{12}$	43
4.27	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,1,1)\times(1,1,1)_{12}$	44
4.28	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,0)\times(0,1,1)_{12}$	44
4.29	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,0)\times(1,1,0)_{12}$	45
4.30	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,0)\times(1,1,1)_{12}$	45
4.31	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,1)\times(0,1,1)_{12}$	46

4.32	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,1)\times(1,1,0)_{12}$	46
4.33	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,1)\times(1,1,1)_{12}$	47
4.34	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,0,1)\times(1,1,0)_{12}$	47
4.35	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,0,1)\times(0,1,1)_{12}$	48
4.36	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,0,1)\times(1,1,1)_{12}$	48
4.37	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,0)\times(1,0,1)_{12}$	49
4.38	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,0)\times(1,1,0)_{12}$	49
4.39	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,0)\times(1,1,1)_{12}$	50
4.40	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,1)\times(0,1,1)_{12}$	50
4.41	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,1)\times(1,1,0)_{12}$	51
4.42	Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,1)\times(1,1,1)_{12}$	51
4.43	Observed and forecasted mean monthly stream flow during January -08 to Decmber-08.	54

TABLE OF CONTENTS

Sr.No.	Title	Page No.
	CANDIDATE'S DECLARATION	iii
	CERTIFICATES	iv-v
	ACKNOWLEDGEMENT	vi-vii
	TABLE OF CONTENTS	viii- ix
	LIST OF TABLES	x
	LIST OF FIGURES	xi-xiii
	LIST OF ABBREVIATIONS AND SYMBOLS	xiv-xv
	ABSTRACT	xvi-xvii
1	INTRODUCTION	1-3
2	REVIEW OF LITERATURE	4-11
	2.1 Forecasting	4
	2.2 Time Series Forecasting	4
	2.3 Streamflow Forecasting by Different Methods	7
	2.4 Streamflow Forecasting by ARIMA	8
3	MATERIALS AND METHODS	11-26
	3.1 Study area	11
	3.1.1 Data collection and Pre-processing	11
	3.1.2 Soil characteristics	11
	3.2 Description of ARIMA	12
	3.2.1 Theoretical Basis of Time series Analysis	13
	3.2.2 Box-Jenkins Methodology for Time series Modelling	15
	3.2.2.1 Phase I (Identification)	16
	3.2.2.2 Phase II (Estimation and Testing)	18
	3.2.2.3 Phase III (Application)	19
	3.3 General Model	19
	3.4 Identification and Specification of the ARIMA part of Model	21
	3.5 Model Estimation and Inference	23
	3.6 Diagnostic Checking Including Detection	24
	3.7 Forecasting	25

	3.8 ARIMA Model for Streamflow Forecasting of Savitri river area	26
4	RESULTS AND DISCUSSIONS	27-54
	4.1 Mean Monthly Streamflow Forecasting of Savitri river	27
	4.1.1 Identification	27
	4.1.1.2 Identification of Model	28
	4.1.2 Estimation and Testing	31
	4.1.3 Mean Streamflow Forecasting with ARIMA	53
5	SUMMARY AND CONCLUSIONS	55-57
6	BIBLIOGRAPHY	58-61

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations	Meanings
%	Percent
<	Less than
>	Greater than
ϕ_i	Non seasonal autoregressive (AR) parameter
Θ_i	Theta, i^{th} nonseasonal moving average parameter
ε	Epsilon, independent variable
Φ	Seasonal AR parameter
σ	Sigma, population standard deviation
β	Beta
Ψ	Psi
α	Alpha, regression coefficient
∇	Backward difference operator
ºC	Degree centigrade
ACF	Autocorrelation Function
AIC	Akaike Information Criterion
AISLUS	All India Soil and Land Use Survey
Agril.	Agriculture
AR	Autoregressive
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARS	Agricultural Research Service
B	Backward linear operator defined by $BZ_t = Z_{t-1}$
BIC	Bayesian Information Criterion
CAET	College of Agricultural Engineering and Technology
d	Order of nonseasonal differencing operator
D	Order of the seasonal differencing operator
Dept.	Department

Equ.	Equation
e.g.	For example
et.al.	And others
etc.	And so forth
Fig	Figure
i.e	That is
k	Lag of k time
Km	Kilo meter
LUP	Land use planning
N	North
No.	Number
PACF	Partial Autocorrelation Function
R^2	Coefficient of Determination
RD	Residual Deviance
RMSE	Root Mean Square Error
SE	Standard Error
Sr.No.	Serial Number
Sq.	Square
Viz.	Namely
Y_t	Standard inflow series, t= 1,2,…N
Z	Standard measure
Z_t	Standardized normal inflow series, t= 1….N

STREAM FLOW FORECASTING OF SAVITRI RIVER USING ARIMA MODEL

BY
Dhiraj Dhanaji Ahire
College of Agriculture Engineering and Technology,
Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth, Dapoli
Dist.Ratnagiri, Maharashtra-2013

Research Guide : Dr.S.B.Nandgude

Department : Soil and Water Conservation Engineering

In India water resources are abundant but the proper utilization of water is not so good. The water and its application in agriculture and other soil practices is more useful. Irrigation projects, which receive water from reservoir, can be challenging to manage since annual fluctuations in runoff from the reservoir's catchment can have considerable impact on the irrigation management strategy. Hence, it is essential to forecast reservoir inflow for proper planning and management of reservoir based irrigation projects. Konkan region of Maharashtra is bestowed with abundance of natural resources. However undulating topography of Sahyadri ranges makes this region vulnerable. This about 30,000 km2 of region is having 22 west flowing rivers which drain to Arebian Sea. Although the rainfall is very high (average annual rainfall 2800mm) water reaches to sea very fast due to steep slope. This also leads to flash floods in lower reaches of river causing heavy damage to life, property and natural resources every year. Early warning about these flash floods can save the life and resources upto large extent. Forecasting models based on historical data can aid the administrator in their efforts.

Irrigation projects, which receive water from reservoir's catchment, can be challenging to manage since annual fluctuation in runoff strategy. Hence, it is essential to forecast reservoir inflow for proper planning and management of reservoir based irrigation projects.

This study focused on the application of ARIMA models for monthly stream flow forecasting of Savitri River Streamflow using Gen-Stat package. First according to the Bayesian Information Criteria (BIC), Akaike Information Criterion (AIC) and many other statistical

parameters, suitable ARIMA models were selected for stream flow forecasting. Here eighteen Models were suitable for monthly streamflow forecasting. It is concluded that the selected models can be used for forecasting mean stream flow to Savitri River with reasonable accuracy But for the forecasting of the streamflow ARIMA (1,1,1) (1,1,1) model is 'good fit'. The AIC Value of this model is 779.764 and the BIC value is 789.762. The R-squared value is 0.9459, and the RMSE value is 28.60, the standard error of this model is 0.00738. The all parameter values are shows that this model is 'good fit 'for forecasting the streamflow of Savitri River.

I. INTRODUCTION

Agriculture in India has a significant history. The pace of advancement of economic super structure of a nation primarily depends on the strength of its agricultural base. India continues to be an agriculture- intensive country with over 65% of population living in rural areas. Agricultural productivity and prospects are inextricably linked to the performance of monsoon rains. Rain fed agriculture continues to constitute nearly 60% of the net sown area. Steering the overall growth of the economy, agriculture sector contributes 26% of the Gross Domestic Product (GDP) (AIC, 2006). The gradual expansion in irrigation development has played a significant role in strengthening the Indian economy. In India, irrigation remains the single largest user of the water resources and accounts for about 84% of all withdrawals (Planning Commission, 2002). For estimating runoff (reservoir inflow), there are basically two types of model, i.e., process-driven and data-driven. Process-driven models are based on physical facts of the problem and are constituted with combination of some experimental equations. The data- driven models are based on the analysis of all the data, characterizing the system under study. A model can then be defined on the basis of connections between the system state variables (input, internal and output variables) with only a limited number of assumptions about the "physical" behavior of the system.

Forecasting is the process of making statements about events whose actual outcomes typically have not yet been observed. The forecasts are either based on the physical understanding of the process or on the statistical analysis of the process popularly termed as stochastic approach. A commonplace example might be estimation of some variable of interest at some specified future date. Prediction is a similar, but more general term. Both might refer to formal statistical methods employing time series, cross-sectional or longitudinal data, or alternatively to less formal judgmental methods. In hydrology, the terms "forecast" and "forecasting" are sometimes reserved for estimates of values at certain specific future times, while the term "prediction" is used for more general estimates, such as the number of times floods will occur over a long period.Forecasting is the process of estimation in unknown situations. Forecasting is used in the practice of Customer Demand Planning in everyday business forecasting for manufacturing companies. Forecasting is commonly used in discussion of time-series data. Recently, a combination of physical and stochastic approaches is gaining more popularity in the field of hydrology forecasting.

The modeling of the river streamflow processing has two approaches; the deterministic or physical simulation of the hydrologic system and the statistical or stochastic simulation of the system. In the deterministic approach, the hydrologic system is described and represented by theoretical and empirical relationships. In the stochastic approach, however, a type of model is assumed, aimed to represent the most relevant statistical characteristics of the historic series. Within this approach, the most widely used models have been the autoregressive models. Regression models are frequently used to forecast stream flows as these are simple and easy to use. Graphical techniques developed by Linsley *et al.* (1975) can be regarded as among the first regression models. Later on, multiple regression model developed by Zuzel *et al.* (1975) and non-parametric regression model developed by Smith (1975) are notable studies carried out on this subject. Other important data-driven models are time serial models, which are different forms of autoregressive integrated moving average (ARIMA) model. Most frequently used ones of this kind are autoregressive moving average (ARMA), autoregressive (AR), autoregressive integrated moving average (ARIMA), partial autoregressive moving average (PARMA) and seasonal autoregressive integrated moving average (SARIMA) models. Among these models, ARIMA model is most suitable for forecasting the inflow.

ARIMA model is an extrapolation method for forecasting and, like any other such method, it requires only the historical time series data on the variable under forecasting. Among the extrapolation methods, this is one of the most sophisticated methods.

Konkan region of Maharashtra in India is bestowed with abundance of natural resources. This region of Maharashtra having an area of 30,394 sq.Km.and has coast length of 720 km. However undulating topography of Sahyadri ranges makes this region vulnerable. This region is having 22 west flowing rivers which drain to Arebian Sea. Although the rainfall is very high (average annual rainfall 2800mm) water reaches to sea very fast due to steep slope. This also leads to flash floods in lower reaches of river causing heavy damage to life, property and natural resources every year. Early warning about these flash floods can save the life and resources up to large extent. Forecasting models based on historical data can aid the administrator in their efforts.

Keeping in view the above facts, the present study was under taken with the following specific objectives.

1. Time series analysis of stream flow and rainfall data

2. Forecasting of mean stream flow at different time steps using ARIMA model.

II. REVIEW OF LITERATURE

This chapter deals with a brief review of significant contributions made by various researchers in time series modelling using ARIMA for the stream flow forecasting.

2.1 Forecasting

Forecasting is the process of estimation in unknown situations. Forecasting is used in the practice of Customer Demand Planning in everyday business forecasting for manufacturing companies. The discipline of demand planning, also sometimes referred to as supply chain forecasting, embraces both statistical forecasting and a consensus process. Forecasting is commonly used in discussion of time-series data.

2.2 Time Series Forecasting

Achela *et al*. (1994) worked on generation and forecasting of monsoon rainfall data. In the study, various stages of decomposing and synthesizing a time series were described and applied to a monthly rainfall data series. The results show that generated data preserved the basic statistical properties of the original series.

Amaha and Sharma (2011) worked on modelling and forecasting of rainfall data of Mekele for Tigray region (Ethiopia) using Univariate Box-Ljung methodology.

Liang and Kershaw (1995) worked on climate change in the Mackenzie Mountains, N.W.T., and Canada. They found that the mean annual air temperature increased by 3.6°C (1968 to 1992) at Ross River, 1.6°C (1974 to 1982) at Tsichu River, 1.8°C (1966 to 1990) at Tungsten and 0.9°C (1943 to 1992) at Norman Wells. Results from an Autoregressive Integrated Moving Average (ARIMA) model suggested shortening of the winter season.

Yurekli and Kurunc (1997) studied the performances of stochastic approaches in generating low stream flow data for drought analysis. They analyzed the monthly-minimum daily discharge data of each month from three gauge stations for forecasting. The two approaches of stochastic modelling, ARIMA and Thomas-Fiering models were used to simulate the monthly minimum daily discharge data of each month. They concluded that ARIMA model performed slightly better than the Thomas-Fiering model. Both approaches were appropriate for simulating the monthly-minimum daily discharge data.

Toth (2000) compared short-term rainfall prediction models for real-time flood forecasting. He compared the accuracy of the short-term rainfall forecasts obtained with time-series analysis techniques, using past rainfall depths as the only input information. The techniques used were linear stochastic auto-regressive moving average (ARMA) models, artificial neural networks (ANN) and the non-parametric nearest-neighbours method. The results indicated that ANN provided a significant improvement in the flood forecasting accuracy in comparison to the use of simple rainfall prediction approaches.

Lin and Chen (2005) worked on time series forecasting by combining the radial basis function network and the self-organizing map. The proposed model was examined using simulated time series data. The results demonstrated that the proposed RBFN was more competent in modelling and forecasting time series than an autoregressive integrated moving average (ARIMA) model. Finally, the proposed model was applied to actual groundwater head data.

Weesakul (2005) worked on rainfall forecast for agricultural water allocation planning. Here ARMA and ARIMA models were used to fit the time series of annual rainfall of 41 years data of 15 stations. It was found that ARIMA model was more suitable for describing the inter-annual variation of annual rainfall.

Koutroumanidis *et al*. (2006) worked on Time-series modelling of fishery landings using ARIMA models and Fuzzy Expected Intervals software. They assessed the model accuracy by comparing model results to the actual monthly fish catches of the year 2000 by using the three different types of models like ARIMA, Fuzzy and DSS model. Optimal forecasting produced by combined modelling scored better than application of the simple ARIMA model. Overall, ARIMA results showed that the Fuzzy Expected Intervals methodology could be used as a reliable tool for short-term predictions of fishery landings.

Jovanovski and Delipetrov (2007) carried out analysis of hydro-meteorological rainfall process by ARIMA models. They found that ARIMA models are good technique for estimation & prediction of hydro-meteorological variations.

Svetlikova *et al.* (2007) analyzed discharge and rainfall time series in the region of the Klastorske Luky Wetland of Slovakia with the objective of forecasting the discharge and rainfall time series. The models tested were linear ARMA, the nonlinear TAR, and TAR with exogenous component and TAR combined with Long Memory models. They concluded that the results obtained could help ecologists in making decisions on wetland management, improving the ecological conditions in the analyzed wetland, and planning future eco-technical measures.

Hung *et al.* (2008) developed an artificial neural network model for rainfall forecasting in Bangkok, Thailand. The developed ANN model was applied to derive rainfall forecast. The developed ANN model was applied to derive rainfall forecast from 1 to 6 h ahead at 75 rain gauge stations in the study area as forecast point from the data of 3 consecutive years (1997–1999). Results were highly satisfactory for 1 to 3 h ahead rainfall forecast.

Robles *et al*. (2008) worked on a hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas. Box–Jenkins Time Series (ARIMA) and multilinear regression (MLR) models were applied to air quality forecasting in urban areas. The hybrid model was able to capture 100% and 80% of alert and pre-emergency episodes, respectively.

Fernandez *et al.* (2009) worked on use of neurofuzzy networks to improve wastewater flow rate forecasting. A neurofuzzy wastewater flow-rate forecasting model (NFWFFM) was developed and tested with actual data measured at the input of two wastewater treatment facilities. Good agreements between forecasted and actual flow-rates were obtained. The artificial intelligence algorithm used only two input variables (day of the week and average daily flow-rate of day before) and one output variable (predicted average daily flow-rate). Results showed that the forecast made by the NFWFFM is more accurate than the one made by the commonly used statistical method.

Liu (2009) worked on an integrated fuzzy time series forecasting system. The objective of this study was to develop an integrated fuzzy time series forecasting system in which the forecasted value will be a trapezoidal fuzzy number instead of a single-point value. Furthermore, this system could effectively deal with stationary, trend, and seasonal time series and increases the forecasting accuracy. Two numerical data sets were selected to illustrate the proposed method and compare the forecasting accuracy with four fuzzy time series methods. The results of the comparison showed that the proposed system could result in more precise forecast than the other four methods.

Amaha and Sharma (2011) worked on modelling and forecasting of rainfall data of Mekele for Tigray region (Ethiopia) using Univariate Box-Ljung methodology.

2.3 Stream Flow Forecasting by Different Methods

Tong (1990) analyzed suitability of nonlinear time series analysis methodology. He used a daily river flow example to illustrate that data with sudden jumps, time irreversibility, asymmetric joint distributions, persistence, lots of high level crossings, and state dependent correlation between lagged flows do not support the assumptions inherent in classical linear ARMA modelling.

Smith *et al.* (1992) presented some interesting applications of Yakowitz's ideas that expose the flexibility of nonparametric methods for seeking relationships between arbitrary functions of possibly linked data sets. This work showed that the nonparametric framework allows one to work directly with the statistics relevant for reservoir operation, rather than worrying about successfully estimating them from a linear model designed to reproduce a serial correlation structure.

Kember *et al.* (1993) worked on Forecasting river flow using nonlinear dynamics. A Nearest Neighbour Method (NNM) was used to forecast daily river flows that were measured at a single location over a time period spanning about seventy years. A parsimonious three parameter NNM was developed in the context of nonlinear dynamics and the dependence between forecast error and length of history used to construct forecasts was investigated. Comparison was made with the Auto-Regressive Integrated Moving Average (ARIMA) models. The NNM was found to provide improved forecasts.

Rasmussen *et al.*(1996) worked on the estimation and validation of contemporaneous PARMA models for stream flow forecasting. Here the periodicity exhibited by the seasonal stream flow in the auto regressive structure was represented by PARMA model. They examined the statistical properties of low order models like PARMA (2, 2). The main problem regarding such model was to determine the covariance matrices of innovation.

Schreider *et al.* (2001) applied Kalman filtering technique for stream flow forecasting. They used IHACRES model and a self-adaptive filtering approach, based on the autoregressive integrated moving average (ARIMA) representation of the model residuals. A Kalman filter forecasting algorithm, incorporating the sub-daily IHACRES model, provide a more flexible approach and yielded even better result than the ARIMA linear filtering approach.

Brath *et al*. (2002) studied neural networks (ANN) and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. Along with traditional linear stochastic models, they used both stationary (ARMA) and nonstationary (ARIMA) models. The application of non-linear time-series models such as Artificial Neural Networks (ANNs) and the 'nearest neighbours' method was proposed. The performance of the models was then compared and the improvement in the efficiency of the discharge forecasts achieved.

Wang *et al*. (2005) worked on constructing prediction intervals for monthly stream flow forecasts. In this study, the empirical approach and bootstrap approach based on the residuals from AR models were applied to construct prediction interval (PI) for monthly stream flow forecasts. The results showed that both empirical approach and bootstrap method work reasonably well, and empirical approach gave results comparable to or even better than bootstrap method.

2.4 Stream Flow Forecasting by ARIMA

Kember *et al*. (1993) connected the neural network predictor to state space reconstruction method, and reconstructed a nonlinear dynamics from the time series. They developed a weighted neighbourhood of exponentially decreasing weights with distance. Finally the lead time forecast was regressed with the vector of past flows at different time lags. Performance of such method was found to be better than the multiplicative, seasonal, ARIMA models for daily stream flow forecasting.

Montanari (2000) worked on a seasonal fractional ARIMA model and applied it to the Nile River monthly flows. The estimation of the parameters was carried out by applying the Whittle's approximation to the Gaussian maximum likelihood function, which yielded asymptotically consistent estimates. This method was particularly useful for hydrological time series. It was applied to the Nile river monthly flows at Aswan to detect whether long memory is present. The results were compared with those obtained by applying heuristic procedures, some of which were developed earlier to see how these performed on seasonal data.

Dimitris and Pantazis (2001) worked on seasonal ARIMA inflow models for reservoir sizing. They used synthetic monthly inflows as an alternative to historical inflow records, and these synthetic series were generated from stochastic SARIMA models. The model forecasted inflow helped in evaluating the real time reservoir operation policies and provided a probabilistic framework for reservoir design.

Karamouz (2004) worked on seasonal stream flow forecasting using autoregressive integrated moving average models (ARIMA). Most important application of this methodology was to forecast stream flow for six to nine months ahead in the summer season of each year.

Krstanovic and Singh (2005) worked on univariate model for long-term stream flow forecasting. They developed a univariate model for long-term stream flow forecasting on five rivers from different regions of the world. The results of the model were compared with the corresponding results of ARIMA and state-space model. The Lagrange multipliers of the univariate model were found similar to autocorrelation coefficients of the ARIMA model. Forecasts by ARIMA and univariate models were comparable for periodic stream flow, but for highly variable stream flow data, the univariate model was found superior.

Mohammadi *et al*. (2005) compared regression, ARIMA and ANN Models for reservoir inflow forecasting. Twenty five years of observed data were used to train or calibrate the models and five years for testing. The performances of models were compared and the ANN model was found to model the flows better.

Yurekli *et al*. (2005) modelled monthly flow using linear stochastic models known as Box-Jenkins or ARIMA model and performed diagnostic checks. The basic statistical properties were compared among the observed and predicted data and found that the generated data preserved the basic statistical properties of the original series.

Modarres and Eslamian (2006) Streamflow time series modeling of Zayandehrud River. Multiplicative seasonal autoregressive integrated moving average models are appropriate for the monthly stream flow of the Zayandehrud River in western Isfahan province, Iran, through the Box and Jenkins time series modeling approach. Among the selected models interpreted from ACF and PACF, seasonal multiplicative ARIMA $(1,1,0) \times (0,1,1)_{12}$ satisfied all tests and showed the best performance.

Yurekl *et al*. (2006) studied performances of stochastic approaches in generating low stream flow data for drought analysis. They used ARIMA and Thomas-Fiering models to simulate monthly-minimum daily discharge data of each month. The error estimates (RMSE and MAE) of forecasts from both approaches were compared to identify the most suitable approach for reliable forecast. The ARIMA model appeared to be better than Thomas-Fiering.

Parviz and Kholghi (2007) applied temporal and spatial disaggregation method for stream flow forecasting. In this study the annual series of stream flow data were disaggregated into semi-annual and monthly series using basic and extended models. They used two models, i.e., Autoregressive Integrated Moving Average (ARIMA) model and disaggregating models for the stream flow forecasting. The results showed that the disaggregating models have good agreement with normal stream flow series.

Solis *et al.* (2008) compared short-term stream flow forecasting using ARIMA and neural network approach. Surprising results showed that for monthly basis, ARIMA had lower prediction errors than the multilayer perception neural network model.

III. MATERIALS AND METHODS

This chapter deals with the description of the study area, its location, topography, soil characteristics etc. and procedure adopted for ARIMA modelling for stream flow forecasting and to estimate the parameter from analysis using Gen-Stat package.

3.1 Study Area

Savitri River in Konkan region of Maharashtra, India has been chosen as the study area. Konkan region of Maharashtra State is bestowed with abundance of natural resource. Savitri River comes under the Middle Konkan region. Konkan region of area comes under the heavy rainfall. There are about 20 western flowing rivers of 1690 km length. Most of the rivers of this region begin in Sahyadri hill ranges at an altitude of 500 to 700 m and snakes their way westwards towards the Arabian Sea. Water reaches to sea very fast due to steep slope. This also leads to flash floods in lower reaches of river causing heavy damage to life, property of natural resources every year. Savitri river originate near Mahabaleshwar of Satara district. Length of this river is 70 km. It passes through Poladpur, Mahad, Mangaon and Shrivardhan talukas in Raigad district. Kal, Gandhari, Nageshwari are some tributaries of Savitri river.

3.1.1 Data Collection

Daily inflows to the reservoir from the catchment were collected for the period of 17 years (1993-2008) from the Office of the SDSC Manager & Superintending Engineer, Data analysis Circle, Water Resources Department, Nasik, Govt. Of Maharashtra. Daily rainfall data of 17 years (1993-2008) measured at Bhave rain-guage station near Khalapur Dist. Raigad (M H).Information on Savitri reservoir catchment area was collected from the Hydrology Data User Group Nasik, Maharashtra.

3.1.2 Soil

The soils of the study area were lateritic sandy clay with moderately fine texture and well drained. The soil are having acidic reaction (pH=4.75 to 6.50). The soils are medium to low in available nitrogen, low in phosphorus and medium to high available potassium content (Anonymous 1990). The lateritic soil is dominant in the region having field capacity of 28% and wilting point of 17.4%.

3.2 Description of ARIMA

Autoregressive Integrated Moving Average (ARIMA) model has been popularized by George Box and Gwilym Jenkins in the early 1970s, and their names have frequently been used synonymously with general ARIMA models applied to time series analysis and forecasting. ARIMA models are the most general class of models for time series forecasting which can be stationeries by transformations such as differencing and logging Box and Jenkins (1970) effectively put together, in a comprehensive manner, the relevant information required to understand and use univariate time series ARIMA model. The theoretical underpinnings described by Box and Jenkins (1970) and later by Box et al. (1994) are quite sophisticated. The IDENTIFY, ESTIMATE, and FORECAST statements perform these three stages.

Detailed description of time series analysis using ARIMA model is described below. In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity. The model is generally referred to as an ARIMA (p, d, q) model where p, d, and q are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. ARIMA models form an important part of the Box-Jenkins approach to time-series modelling. The ARIMA procedure provides a comprehensive set of tools for univariate time series model identification, parameter estimation, and forecasting, and it offers great flexibility in the kinds of ARIMA or ARIMAX models that can be analyzed. The ARIMA procedure supports seasonal, subset, and factored ARIMA models; intervention or interrupted time series models; multiple regression analysis with ARMA errors; and rational transfer function models of any complexity. When one of the three terms is zero, it's usual to drop "AR", "I" or "MA". For example, ARIMA(0,1,0) is I(1), and ARIMA(0,0,1) is MA(1).

3.2.1 Theoretical Basis of Time-series Analysis

A time series is a set of values of a continuous variable Y (Y_1, Y_2,., Y_n), observed over time period t (1, 2... n). The term time-series comes from econometric studies in which the index variable refers to intervals of time measured in a suitable scale. However, it must be clearly stated that this direct reference to time is not required: actually, any different meaning can be attributed to the index variable, provided that it is able to order the Y values. In general, in a given time series the following can be recognized and separated.

(1) A regular long-term component of variability, termed as ‘trend’ that represents the whole evolution pattern of the series.

(2) A regular short-term component whose shape occurs periodically at intervals of s lags of the index variable, known as ‘seasonality’.

(3) An autoregressive component of p order, AR (p) which relates each value $Z_t = Y_t$ – (trend and seasonality) to the p previous Z values, according to the following linear relationship

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \ldots + \emptyset_p Z_{t-p} + \varepsilon_t \qquad \ldots (3.1)$$

where $\phi_i (i = 1, \ldots p)$ are parameters to be estimated and ε_t is a residual term.

(4) A moving average component of q order, MA(q) which relates each Z_t value to the q residuals of the q previous Z estimates

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots . \theta_p \varepsilon_{t-p} \qquad \ldots (3.2)$$

The theory of time-series analysis has developed a specific language and a set of linear operators. According to Box and Jenkins a highly useful operator in time-series theory is the lag or backward linear operator (B) defined by

$$BZ_t = Z_{t-1} \qquad \ldots (3.3)$$

Consider the result of applying the lag operator twice to a series:

$$B(BZ_t) = B(Z_{t-1}) = Z_{t-2} \qquad \ldots (3.4)$$

Such a double application is indicated by B^2, and, in general, for any integer k, it can be written

$$B^k \ Z_t = Z_{t-k} \qquad \ldots (3.5)$$

By using the backward operator, Eqn. (3.1) can be rewritten as

$$Z_t - \emptyset_1 Z_{t-1} - \emptyset_2 Z_{t-2} - \ldots . - \emptyset_p Z_{t-p} = \varepsilon_t = \emptyset(B) Z_t \qquad \ldots (3.6)$$

where Ø (B) is the autoregressive operator of p order defined by

$$\emptyset\ (B) = 1 - \emptyset_1\ (B) - \emptyset_2\ B^2 - \ldots - \emptyset_p\ B^p \qquad \ldots (3.7)$$

Similarly, Eqn. (3.2) can be written as

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ....\theta_q \varepsilon_{t-q} = \theta(B)\varepsilon_t \qquad ... (3.8)$$

where $\theta(B)$ indicates the moving average operator of q order defined by

$$\theta(B) = 1 - \theta_1(B) - \theta_2 B^2 - ... - \theta_q B^q \qquad ... (3.9)$$

The autoregressive and moving average components can be combined in an autoregressive moving average (ARMA) (p, q) model as:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ....\theta_q \varepsilon_{t-q} \qquad ... (3.10)$$

Or in lag operator form

$$\left(1 - \phi_1(B) - \phi_2 B^2 - ... - \phi_p B^p\right) Z_t = \left(1 - \theta_1(B) - \theta_2 B^2 - ... - \theta_q B^q\right) \varepsilon_t \qquad ... (3.11)$$

Finally,

$$\phi(B) Z_t = \theta(B) \varepsilon_t \qquad ... (3.12)$$

In a preliminary analysis of a time series, it is useful to independently evaluate the long- and short-term periodic components, which are essential to define the regular structure of the series. The trend component can be evaluated by fitting a regular function, a polynomial, or a more complicated general function. The seasonal component can be estimated by a seasonal decomposition procedure, which calculates a seasonal index based on the ratio of the observed values to the moving average. In the final stage of series modelling, however, both the trend and the seasonal component will be integrated in the ARMA (p, q) process. For the trend, such integration is obtained by using the difference linear operator ∇, defined by

$$\nabla Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B) Y_t \qquad ... (3.13)$$

A single application of the ∇ operator corrects the data for a linear increasing trend, whereas its repeated use for d times corrects for a trend that can be fitted by a d-order polynomial. The stationary series Z_t obtained as the dth difference $\left(\nabla^d\right)$ of Y_t is given as:

$$Z_t = \left(\nabla^d\right) Y_t = (1 - B)^d \, Y_t \qquad ... (3.14)$$

The stationary series can be then modelled by an ARMA (p, q) process. The combined use of the ∇ operator and the ARMA (p, q) process, results in an ARIMA (p, d, q) model. Further, ARIMA can account for the seasonal component of s lag period, by using both correlations between Z_t and Z_{t-s} values and those between the corresponding residuals ε_t and ε_{t-s}. In mathematical terms, therefore, a seasonal ARIMA model is an ARIMA (p,d,q) model whose residuals ε_t can be further modelled by an ARIMA(P,D,Q)s structure with linear operators (P,D,Q) being functions of the B^s operator.

The operators of a seasonal ARIMA model, defined as ARIMA (p, d, q)×(P, D, Q)$_{12}$, can be expressed by:

AR (p) nonseasonal operator of p order, $\phi(B)=1-\phi_1(B)-\phi_2 B^2-\ldots-\phi_p B^p$;

AR (P) seasonal operator of P order, $\phi(B)=1-\phi_1(B)^s-\ldots-\phi_p B^{sp}$;

MA (q) nonseasonal operator of q order, $\theta(B)=1-\theta_1(B)-\theta_2 B^2-\ldots-\theta_q B^q$;

MA (Q) seasonal operator of Q order, $\theta(B)=1-\theta_1(B)^s-\theta_2(B)^{2s}-\ldots-\theta_Q B^{Qs}$; and

$\nabla^d=(1-B)^d$

3.2.2 Box-Jenkins methodology for Time-Series modelling

The Box-Jenkins methodology for analyzing and modelling time series is characterized by three steps as described in Fig. 3.1. Fitting nonseasonal ARIMA model to the time series of the new data sequences consists of three phases: model identification, parameter estimation and diagnostic testing and forecasting. The first two phases of model building process are typically repeated several times until a satisfactory model is finally selected. Then final selected model is used for prediction purposes.

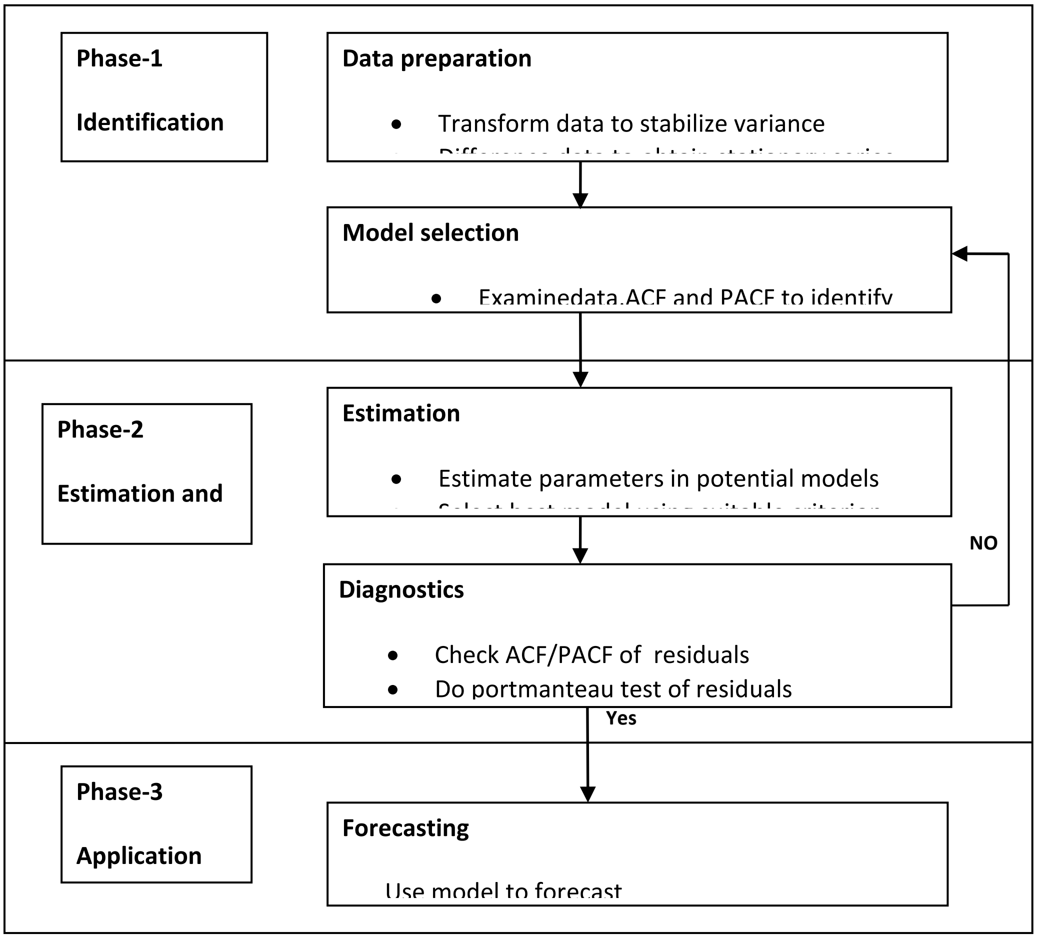

Fig.3.1. Schematic representation of the Box-Jenkins methodology for time series modelling.

3.2.2.1 Phase I (Identification)

The purpose of identification phase is to determine the differencing required for producing stationarity and also the order of nonseasonal AR and MA operators for a given series. Stationarity is a necessary condition in building an ARIMA model that is useful for forecasting. A stationary time series has the property that its statistical characteristics such as the mean and the autocorrelation structure are constant over time. When the observed time series presents trend, differencing and transformation are often applied to the data to remove the trend and stabilize variance before an ARIMA model can be fitted.

The autocorrelation function *(ACFρ(K))* at lag k of the Z_t series is the linear correlation coefficient between Z_t and Z_{t-k}, calculated for k =0, 1, 2... as

$$\rho_k = \frac{\operatorname{cov}\left(Z_t - Z_{t-k}\right)}{\sqrt{\operatorname{var}\left(Z_t\right)\operatorname{var}\left(Z_{t-k}\right)}} \qquad \text{... (3.15)}$$

The PACF is defined as the linear correlation between Z_t and Z_{t-k}, controlling for possible effects of linear relationships among values at intermediate lags. Theoretically, both an AR (p) process and an MA (q) process should be associated with well-defined patterns of ACF and PACF, usually decreasing exponential or alternate in sign or decreasing sinusoidal patterns. A precise correspondence between ARMA (p, q) processes and defined ACF and PACF patterns is more difficult to recognize. When the order of at least one of the two components (AR or MA) is clearly detectable, however, the other can be identified by attempts in the following step of parameter estimation. Finally, the existence of a seasonal component of length s is underlined by the presence of a periodic pattern of period s in the ACF. General steps for identification phase are described below.

(i) Plot the time series and identify any unusual observations. Decide if a transformation is necessary to stabilize the variance. If necessary, transform the data to achieve stationarity in variance.

(ii) Consider if the possible transformed data appear stationary from the time plot and the ACF and PACF. If the time plots are such that the data are scattered horizontally around a constant mean, or equivalently, the ACF and PACF drop to or near zero quickly, and it indicates that the data are stationary. If the time plot is not horizontal or ACF and PACF do not drop to zero, non-stationarity is implied.

(iii) When the data appear non-stationary, then it can be made stationary by differencing for non-seasonal data, taking first differences of the data. For seasonal data, seasonal differences of the data are taken and checked that these appear stationary. If these are still non-stationary, the first differences of the differenced data are taken. For most practical purpose a maximum of two differences will transfer the data into a stationary series.

(iv) When stationarity has been achieved, examine the autocorrelation to see if any pattern remains. There are three possibilities to consider

a) Seasonality may suggest that autocorrelation and partial autocorrelation at the seasonal lags are large and significantly different from zero.

b) AR or MA models may be revealed by the pattern of autocorrelation and partial autocorrelation. If there are no significant autocorrelation after lag q, a MA (q) model may be appropriate. If there are no significant partial autocorrelation after lag p, a AR (p) model may be appropriate.

c) If there is no clear MA or AR model suggested, a mixture model is necessary.

3.2.2.2 Phase II (Estimation and testing)

Once a suitable ARIMA $(p, d, q)\times(P,D,Q)_{12}$ structure is identified, subsequent steps of parameter estimation and testing/diagnostics must be performed. Estimation stage consists of using the data to estimate and make inferences about parameters of tentatively identified model. The parameters are estimated such that an overall measure of residuals is minimized. The testing or diagnostic checking of model adequacy is the last stage of model building. This stage determines whether residuals are independent, homoscedastic and normally distributed. Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit. After identifying tentative model, the process is again followed by the stage of parameter estimation and model verification. Diagnostic information may help to suggest alternative model(s).

Once a suitable ARIMA $(p, d, q)\times(P, D, Q)_{12}$ structure is identified, the parameters are usually obtained by maximum likelihood, which is asymptotically correct for time series. Estimators are usually sufficient, efficient, and consistent for Gaussian distribution and are asymptotically normal and efficient for several non-Gaussian distribution families.

Validation of the goodness of fit of an ARIMA model can be developed according to the following steps:

1) Evaluation of statistical significance of parameters by the usual comparison between the parameter value and the standard deviation of its estimate. For a test statistic that is valid only asymptotically, a parameter whose value exceeds twice its standard error can be considered significant.

2) Analysis of the ACF of residuals. In this step, residuals ε_t are considered as a new time series, and ACF and PACF are estimated to be sure that values at lag k > 0 are not statistically different from zero.

3) Calculation of BIC (Bayesian Information Criterion): In this step, BIC is calculated for all models and according to the lowest BIC value, the appropriate model is selected.

$$BIC\left(\hat{\sigma}^2\right) = -2ln \times L\left(\overrightarrow{\hat{\beta}}, S\left(\overrightarrow{\hat{\beta}}/T\right)\right) + (p+q+1)\ln T \qquad \text{... (3.16)}$$

For prediction purposes, ARIMA models are different from the analytical functions of time: $Z_t = f(t)$, because ARIMA forecasting uses previous values of the series and errors in the previous estimates. Actually, this peculiarity of ARIMA forecasting is valid in the short term because parameters of the model cannot account, in the long term, for changes in the dynamics of the series.

3.2.2.3 Phase III (Application)

After suitable form of ARIMA (p, d, q)×(P,D,Q)$_{12}$ is selected and its parameter are estimated, the model is ready for forecasting future events at different lead times. Now, Eqn. (3.17) can be used for forecasting.

$$(1-B)\times\left(1-B^{12}\right)Y_t = (1-\theta_1 B)\left(1-\theta_1 B^{12}\right)e_t \qquad \text{... (3.17)}$$

3.3.3 General model

ARIMA models, as discussed by Box and Jenkins (1976), are frequently used for seasonal time series. A general multiplicative seasonal ARIMA model for a time series z_t can be written as

$$\phi(B)\Phi\left(B^s\right)\left(1-B^s\right)^{D_{z_t}} = \theta(B)\Psi\left(B^s\right)\alpha_t \qquad \text{... (3.18)}$$

where B is the backshift operator ($Bz_t = z_{t-1}$), s is the seasonal period, $\phi(B) = \left(1-\phi_1 B - .... - \phi_p B^p\right)$ is the nonseasonal autoregressive (AR) operator, $\Phi\left(B^s\right) = \left(1-\Phi_1 B^s - .... - \Phi_p B^s\right)$ is the seasonal AR operator $\theta(B) = \left(1-\theta_1 B - .... - \theta_q B^q\right)$ is the nonseasonal moving average (MA) operator,

$\Psi(B^s) = (1 - \Psi_1 B^s - .... - \Psi_Q B^{Q_s})$ is the seasonal MA operator. The $(1 - B)^d(1 - B^s)^D$ implies nonseasonal differencing of order d and seasonal differencing of order D. If d = D = 0 (no differencing), it is common to replace z_t in Eqn. (3.20) by deviations from its mean, that is, by $z_t - \mu$ where $\mu = E[z_t]$. A useful extension of ARIMA models results from the use of a time-varying mean function modelled via linear regression effects. More explicitly, suppose we write a linear regression equation for a time series y_t as

$$y_t = \sum_i \beta_i x_{it} + z_t \qquad ... (3.19)$$

where y_t is the (dependent) time series, the x_{it} are regression variables observed concurrently with y_t, the β_i are regression parameters, and $z_t = y_t - \sum_i \beta_i x_{it}$, the time series of regression errors, is assumed to follow the ARIMA model in Eqn. (318). Modelling z_t as ARIMA addresses the fundamental problem with applying standard regression methodology to time series data, which is that standard regression assumes that the regression errors (z_t in Eqn. (3.18)) are uncorrelated over time. In fact, for time series data, the errors in Eqn. (3.19) will usually be autocorrelated, and, moreover, will often require differencing. Assuming z_t as uncorrelated in such cases will typically lead to grossly invalid results.

Eqns. (3.18) and (3.19) taken together define the general regARIMA model. Combining Eqns. (3.18) and (3.19), the model can be written in a single equation as

$$\phi(B)\Phi(B^s)(1-B)^d (1-B^s)^D \left(y_t - \sum_i \beta_i x_{it} \right) = \theta(B)\Psi(B^s)\alpha_t \qquad ... (3.20)$$

The ARIMA model in Eqn. (3.20) can be thought of either as generalizing the pure ARIMA model in Eqn. (3.18) to allow for a regression mean function $\sum \beta_i x_{it}$, or as generalizing the regression model in Eqn. (3.19) to allow the errors z_t to follow the ARIMA model in Eqn. (3.18). In any case, notice that the regARIMA model implies that first the regression effects are subtracted from y_t to get the zero mean series z_t, then the error series z_t is differenced to get a stationary series, say ω_t, and ω_t is then assumed to follow the stationary ARMA model, $\phi(B)\Phi(B^s)\omega_t = \theta(B)\Psi(B^s)\alpha_t$. Another way to write the ARIMA model in Eqn. (3.20) is

$$(1-B)^d\left(1-B^s\right)^D y_t=\sum_i \beta_i(1-B)^d\left(1-B^s\right)^D x_{it}+\omega_t \qquad \text{... (3.21)}$$

Where ω_t follows the stationary ARMA model just given. Eqn. (3.21) emphasizes that the regression variables x_{it} in the regARIMA model, as well as the series y_t, are differenced by the ARIMA model differencing operator $(1 - B)^d(1 - Bs)^D$. Notice that the regARIMA model as written in Eqn. (3.20) assumes that the regression variables x_{it} affect the dependent series y_t only at concurrent time points, i.e., model in Eqn. (3.20) does not explicitly provide for lagged regression effects such as $\beta x_{i,t-1}$. Flexibility in the specification of the ARIMA part of a reg-ARIMA model by permitting

(i) more than two multiplicative ARIMA factors,

(ii) missing lags within the AR and MA polynomials,

(iii) the fixing of individual AR and MA parameters at user-specified values when the model is estimated, and

(iv) inclusion of a trend constant, which is a nonzero overall mean for the differenced series $((1 - B)^d(1 - Bs)^D y_t)$.

Detailed discussions of ARIMA modelling are given in the classic book by Box and Jenkins (1976), and also in several other time series texts, such as Abraham and Ledolter (1983) and Vandaele (1983).

3.4 Identification and Specification of the ARIMA Part of the Model

The ARIMA part of a regARIMA model is determined by its orders and structure, e.g., (p d q), (P D Q), and s for model Eqn. (3.20). If no regression variables are included in the model, then determination of the orders for the resulting pure ARIMA model (called ARIMA model identification) can be carried out by following well-established procedures that rely on examination of the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) of the time series y_t and its differences. For regARIMA models, a modified approach is needed, since the presence of regression effects can distort the appearance of the ACF and PACF. Typically, the differencing orders can be identified by examining ACFs of the time series y_t and its differences. Then, one should obtain the residuals from a regression of the differenced data on the differenced regression variables. The ACF and PACF of these residuals can then be examined in the usual way to identify the AR and MA orders of the regression error term in the regARIMA model. This approach to regARIMA model identification is discussed and illustrated in Bell and Hillmer (1983). The key spec used to implement this approach to ARIMA model identification is the identify spec. For illustration, consider a monthly seasonal time series. The usual ACFs examined to determine the differencing needed are those of y_t, $(1-B)y_t$, $(1-B^{12})y_t$, and $(1-B)\,(1-B^{12})y_t$. The identify spec can produce all these ACFs in a single run. Once the differencing has been determined, another run of ARIMA can be made using the identify and regression specs together to (i) regress the differenced y_t series on the differenced regression variables, and (ii) produce the ACF and PACF of the regression residuals for use in identifying the AR and MA orders of the model. For example, if one nonseasonal and one seasonal difference are specified (d = 1 and D = 1), the identify and regression specs will fit the model

$$(1-B)(1-B^{12})y_t = \sum_i \beta_i (1-B)(1-B^{12})x_{it} + \omega_t \qquad \text{... (3.22)}$$

by ordinary least squares (OLS), and will produce the ACF and PACF of the regression residuals ω_t in Eqn. (3.22).

An alternative approach that does not require two runs of the ARIMA program can be used if the maximum differencing orders (nonseasonal and seasonal) that may be required are assumed known. For example, suppose that these maximum differencing orders are d = 1 and D = 1. Then the identify and regression specs can be used to (i) perform OLS regression on (3.21) to produce parameter estimates β_i (ii) compute an estimated (undifferenced) regression error series $\bar{z}_t = y_t - \sum_i \beta_i \bar{x}_{it,}$ and (iii) produce ACFs and PACFs of $\bar{z}_t, (1-B)\bar{z}_t, (1-B^{12})\bar{z}_t$ and $(1-B),(1-B^{12})\bar{z}_t$. These ACFs and PACFs can be examined to determine the orders of differencing required, as well as the orders of the AR and MA operators for the model.

3.5 Model Estimation and Inference

The regression and ARIMA specs specify a regARIMA model. The estimate spec then estimates the model parameters by exact maximum likelihood, or by a variant known as conditional maximum likelihood (Box and Jenkins, 1970), which is sometimes called conditional least squares. Differences between exact and conditional likelihood for MA parameter estimation are more fundamental, with exact likelihood being the recommended approach. The default option is exact maximum likelihood estimation for both the AR and MA parameters.

Whichever choice of estimation method is made, the resulting log-likelihood for a pure ARIMA model is reduced to a sum of squares function that is then minimized by a nonlinear least squares routine MINPACK, discussed by More et al. (1980). To maximize the likelihood for a full regARIMA model, an iterative generalized least squares (IGLS) algorithm, (Otto *et al*., 1987) is used. This algorithm involves two general steps: (i) for given values of the AR and MA parameters, the regression parameters that maximize the likelihood are obtained by a generalized least squares (GLS) regression (using the covariance structure of the regression errors, which is determined by their ARIMA model); and (ii) for given values, β_i, of the regression parameters, the ARIMA model is fitted by maximum likelihood to the time series of regression errors, $z_t = y_t - \sum \beta_i x_{it}$. IGLS iterates between these two general steps until convergence is achieved. (Output options in the estimate spec allow for display of intermediate results from the estimation iterations, if desired). The likelihood function (exact or conditional) is evaluated using an approach derived from those suggested by (Box and Jenkins, 1970; Ljung and Box, 1979; Hillmer and Tiao, 1979; Wilson 1983). Statistical inferences about regARIMA model parameters may be made using asymptotic results for maximum likelihood estimation of ARIMA models by Box and Jenkins (1976); Brockwell and Davis (1991) and ARIMA models by Pierce (1971). These results state that, under suitable assumptions, the parameter estimates are approximately normally distributed with means equal to the true parameter values and with a certain covariance matrix that can be estimated. (The “suitable assumptions” include that the true model form is used, that the model’s AR operators are all stationary and its MA operators are all invertible, and that the series is sufficiently long for the asymptotic results to apply).

3.6 Diagnostic Checking Including Outlier Detection

Diagnostic checking of a ARIMA model is performed through various analyses of the residuals from model estimation, the objective being to check if the true residuals (α_t in (3.21)) appear to be white noise. The check spec is used to produce various diagnostic statistics using the residuals from the fitted model. To check for autocorrelation, ARIMA can produce ACFs and PACFs of the residuals (with standard errors), along with Ljung and Box (1978) summary Q-statistics. ARIMA can also produce basic descriptive statistics of the residuals and a histogram of the standardized residuals.

When a model contains two or more level shifts, including those obtained from outlier detection as well as any level shifts specified in the regression spec, ARIMA will optionally produce t-statistics for testing null hypotheses that each run of two, three, etc. successive level shifts actually cancels to form a temporary level shift. Two successive level shifts cancel to form a temporary level shift if the effect of one offsets the effect of the other, which implies that the sum of the two corresponding regression parameters is zero. Similarly, three successive level shifts cancel to a temporary level shift if the sum of their three regression parameters is zero, etc. (There is a user-specified limit on the number of successive level shifts in the runs tested.) The t-statistics produced are the sums of the estimated parameters for each run of successive level shifts divided by the appropriate standard error. An insignificant temporary level shift t-statistic (say, one less than 2 in magnitude) fails to reject the null hypothesis that the corresponding level shifts cancel to form a temporary level shift. These tests are provided primarily as diagnostics to help users assess the impacts of level shifts in a model. Of course, if one or more of these t-statistics are significant, the user may wish to re-specify the model with the relevant level shift regression variables replaced by appropriate temporary level shift variables. These can be specified as user-defined regression variables, or can be obtained by fixing the coefficient of the level shift regressors so that they sum to zero. The choice between using level shifts (which correspond to permanent changes in the level of a series) versus temporary level shifts could be important for forecasting a series with level shifts near the end of the data.

Once a model has been selected and parameter calculated, the adequacy of the model has to be checked. This process is called diagnostic checking. There are number of diagnostic checking methods to test the suitability of the estimated model.

Examination of Standard Error: A high standard error in comparison with the parameter value point out a higher uncertainty in parameter estimation which question the stability of the model.

ACF and PACF of residual: If the model is adequate at describing behaviour of streamflow time series, the residuals of model should not be correlated i.e. all ACF and PACF should lie within the limits. Akaike Information Criteria (AIC: AIC (Akaike, 1974) is computed by equation (3.23).

The lower value of AIC is desirable.

$$AIC = 2k + [In\ (2\pi V_r /T) + 1]\ T \qquad \text{.... (3.23)}$$

Where,

AIC = Akaike Information Criteria

K = No. Of Model parameters

V_r = Residual Variance

T = Total number of observation

3.7 Forecasting

For a given ARIMA model with parameters estimated by the ARIMA program, the forecast spec will use the model to compute point forecasts, and associated forecast standard errors and prediction intervals. The point forecasts are minimum mean squared error (MMSE) linear predictions of future y_ts based on the present and past y_ts assuming that the true model is used—which means we assume the ARIMA model form is correct, that the correct regression variables have been included, that no additive outliers or level shifts will occur in the forecast period, that the specified ARIMA orders are correct, and that the parameter values used (typically estimated parameters) are equal to the true values. These are standard assumptions, though obviously unrealistic in practical applications. What is more realistically hoped is that the ARIMA model will be a close enough approximation to the true, unknown model for the results to be approximately valid. Two sets of forecast standard errors are produced. One assumes that all parameters are known. The other allows for additional forecast error that comes from estimating the regression parameters, while still assuming that the AR and MA parameters are known. For a reasonably long time series (Box and Jenkins 1970), the contribution to forecast error of the error in estimating the AR and MA parameters is generally small, thus providing a justification for ignoring this source of error when computing the forecast standard errors.

If the series has been transformed, then forecasting results are first obtained in the transformed scale, and then transformed back to the original scale. For example, if one specifies a model of form (3.21) for $y_t = \log(Y_t)$, where Y_t is the original time series, then y_t is forecasted first, and the resulting point forecasts and prediction interval limits are exponentiated to produce point and interval forecasts in the original (Y_t) scale. The resulting point forecasts are MMSE for $y_t = \log(Y_t)$, but not for Y_t under the "standard" assumptions mentioned above. Analogous procedures are followed for other transformations allowed by ARIMA. If any prior adjustments are made, these will also be inverted in the process of transforming the point forecasts and prediction interval limits back to the original scale.

3.8 ARIMA Model for Stream Flow Forecasting of Savitri River Area.

To fit an ARIMA model, a sufficiently large historical data set is required. In this study, inflow data for the period 1993 to 2008 were used. Gen-Stat, Statistical package which has become a leader in predictive analytics technologies, is used here to predictive analysis. The input data is obtained from the Hydrological Department Nasik.

IV. RESULTS AND DISCUSSION

This chapter deals with the appropriate model selection and forecasting of the mean stream flow of Savitri River during June 1993 to June 2008 for monthly.

4.1 Mean Monthly Stream Flow Forecasting of Savitri River

4.1.1 Identification

Stationarity is a necessary condition in building an ARIMA model that is useful for forecasting. Most common method to check stationarity in data series is examining the graph or time plot of the data and ACF and PACF. To test stationarity of monthly data, ACF and PACF were determined against different lags. Figs. 4.1 and 4.2 present ACF and PACF plots respectively, with upper and lower confidence limits. For a stationary series all the ACF and PACF should lie within the confidence limit. It is evident from the figure that both ACF and PACF were having higher values at 12 lag. This indicates seasonality in the time series. Further, PACF also has high value indicating the requirement of differencing the data sequences.

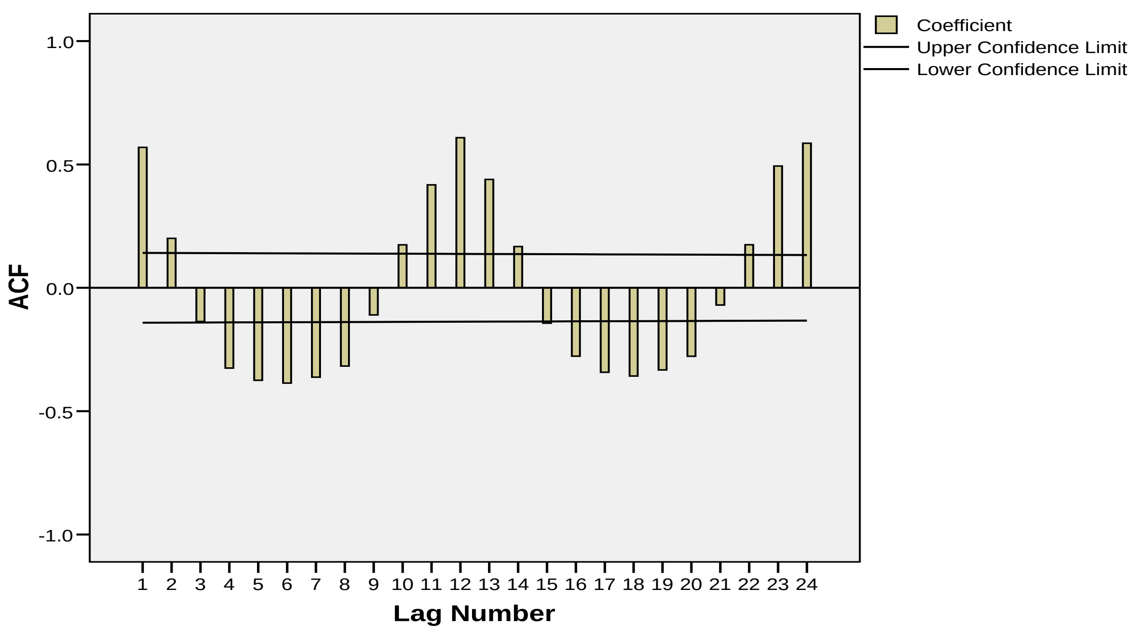

Fig. 4.1. ACF of the monthly data without differencing.

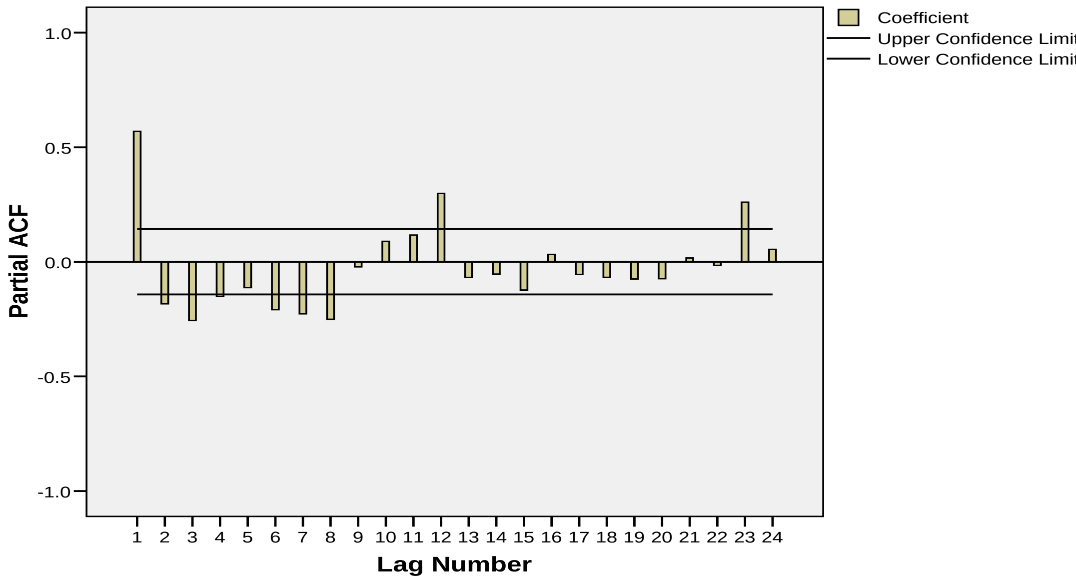

Fig. 4.2.

PACF of the monthly data without differencing.

4.1.1.2 Identification of Model

The first and most important step in the modelling of a time series is the identification of the tentative model type to be fitted to the data set. It is one of the basic conditions for applying ARIMA class of model for particular time series is its stationarity. It is the stage at which graphical methods are particularly useful and judgment must be exercised.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) were examine to know the stationarity of time series. The time series was transformed using following differencing schemes.

1) d = 0 ; D = 0   2) d = 0 ; D = 1,   3) d = 1 ; D = 0 ,   4) d = 1 ; D = 1

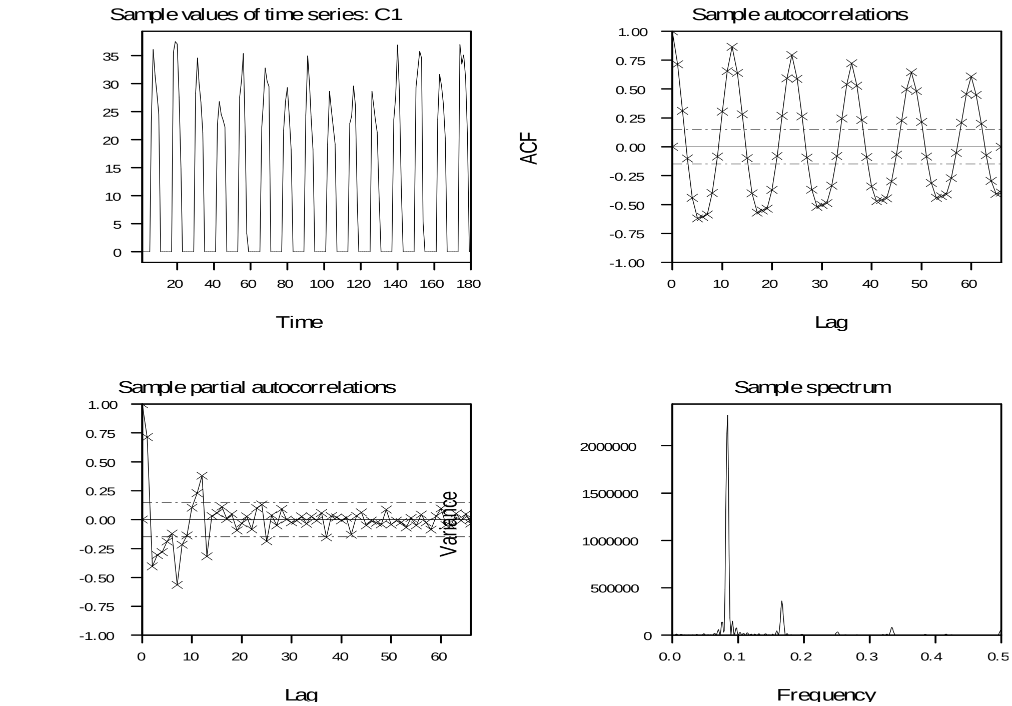

Fig. 4.3 Streamflow time series ACF and PACF of streamflow d=0;D=0

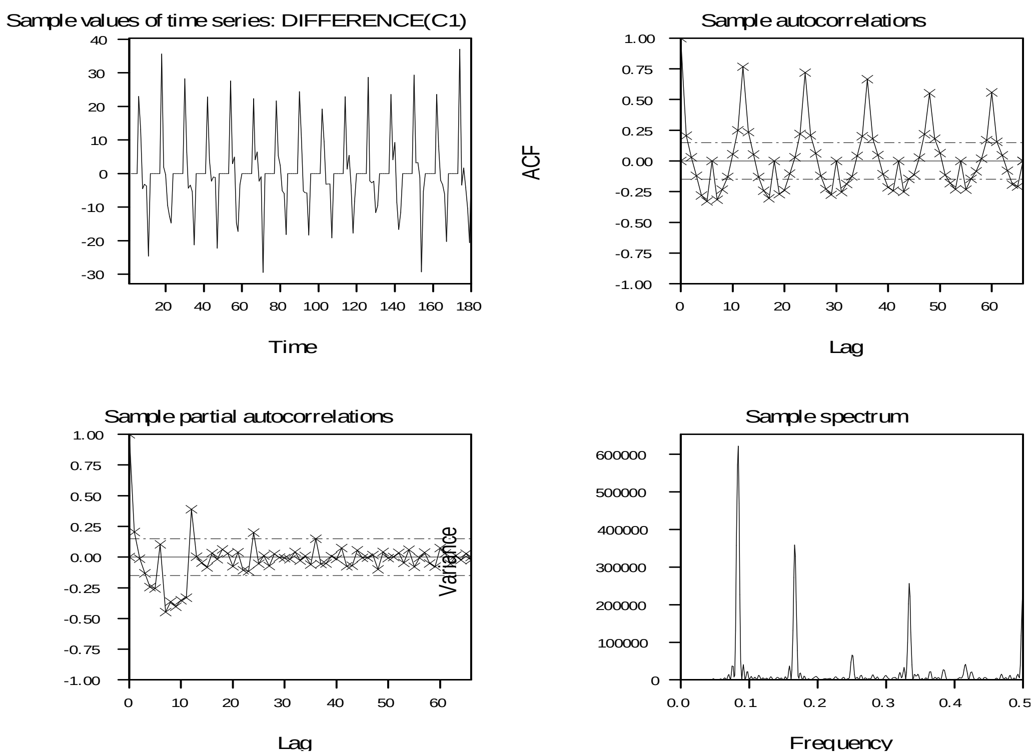

Fig. 4.4 Streamflow time series ACF and PACF of streamflow d=1;D=0

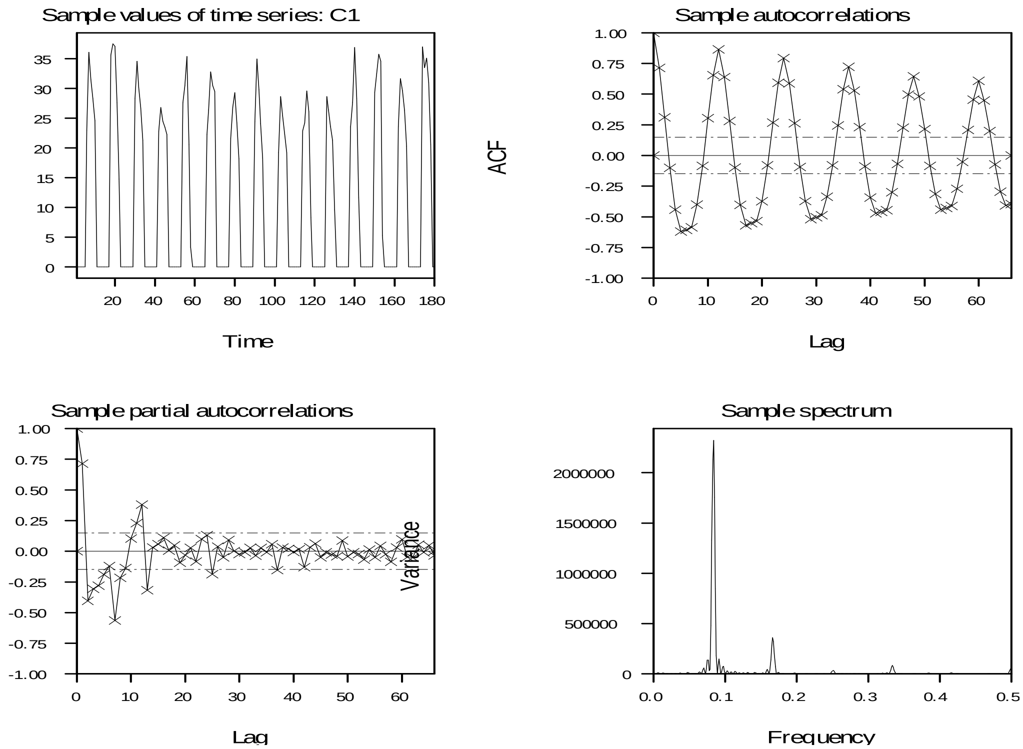

Fig. 4.5 Streamflow time series ACF and PACF of streamflow d=0;D=1

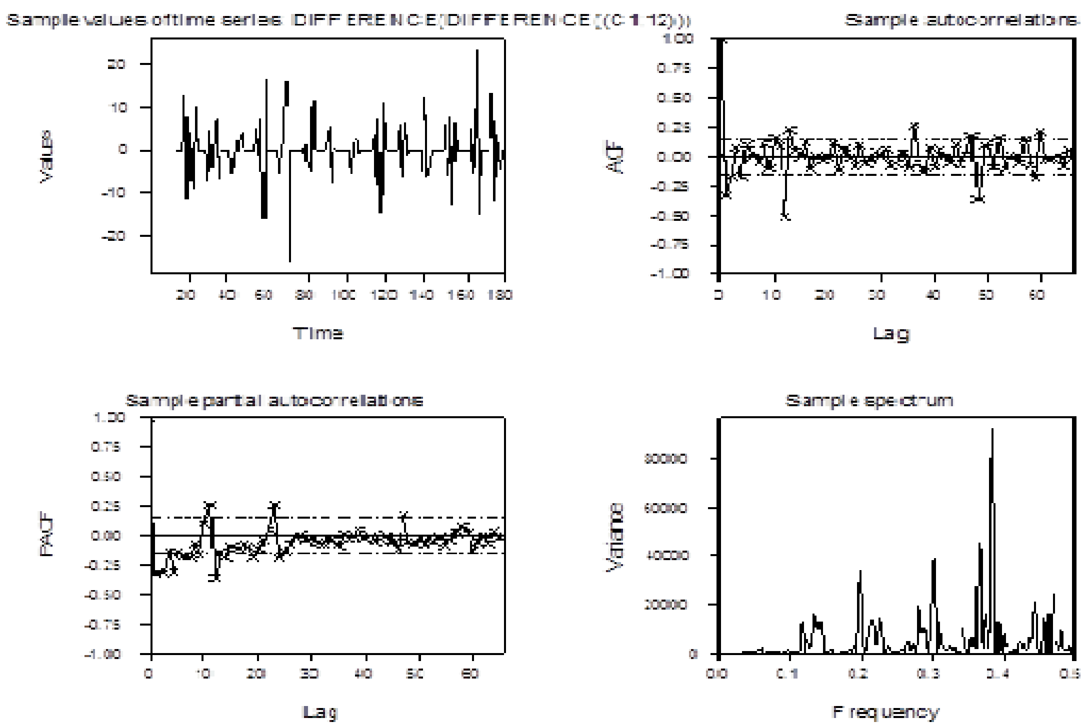

Fig. 4.6 Streamflow time series ACF and PACF of streamflow d=1;D=1

4.1.2 Estimation and testing.

From the fig.4.3-4.6 for estimation and testing d=0; D=1 and d=1; D=1 are stationary. In this case difference of order 1 was sufficient to achieve stationarity in mean. However, at periodic lag of 12, there were peaks, which suggest seasonality in the data. Thus, a more complicated mixture ARIMA model was required. For estimation and testing d=0; D=1 and d=1; D=1.

Alternative ARIMA models were estimated by considering the ACF and PACF graphs for the monthly data series. Here, altogether twenty six ARIMA were analyzed. According to the minimum AIC and BIC criteria, Eighteen models were selected. AIC and BIC value of the selected Eighteen models are presented in Table 4.1. From the table it is seen that ARIMA $(0,0,1)\times(0,1,1)_{12}$; ARIMA $(0,0,1)\times(1,1,1)_{12}$; ARIMA$(1,0,0)\times(0,1,1)_{12}$ ARIMA$(1,0,0)\times(1,1,1)_{12}$; ARIMA$(1,0,1)\times(0,1,1,)_{12}$; ARIMA $(1,0,1)\times(1,1,1)_{12}$;ARIMA$(0,1,1)\times(1,1,1)_{12;}$ARIMA$(1,1,1)\times(0,1,1)_{12}$;ARIMA$(1,1,1)\times(1,1,1)_{12}$ models have comparatively lower BIC and AIC values. Hence, these models were selected and further diagnostic checks were performed to determine the most suitable model from amongst these nine selected models (Table 4.2). Referring to Table 4.2, ARIMA $(1,1,1)\times(1,1,1)_{12}$ was found to be the most suitable model.

To further establish, this conclusion, model verification was performed by examining the autocorrelations and partial autocorrelations of the residuals of various orders. Similarly Figs. 4.7 - 4.24 present the ACF and PACF of residuals of ARIMA $(0,1,1)\times(0,1,1)_{12}$, ARIMA $(0,1,1)\times(1,1,0)_{12}$, ARIMA$(0,1,1)\times(1,1,1)_{12}$, ARIMA $(1,1,0)\times(0,1,1)_{12}$, ARIMA $(1,1,0)\times(1,1,0)_{12}$, ARIMA $(1,1,0)\times(1,1,1)_{12}$, ARIMA $(1,1,1)\times(0,1,1)_{12}$, ARIMA $(1,1,1)\times(1,1,0)_{12}$, ARIMA $(0,0,1)\times(0,1,1)_{12}$, ARIMA $(0,0,1)\times(1,1,0)_{12}$, ARIMA $(0,0,1)\times(1,1,1)_{12}$, ARIMA $(1,0,0,)\times(0,1,1)_{12}$, ARIMA $(1,0,0)\times(1,1,0)_{12}$, ARIMA $(1,0,0)\times(1,1,1)_{12}$, ARIMA $(1,0,1)\times(0,1,1)_{12}$, ARIMA $(1,0,1)\times(1,1,0)_{12}$, ARIMA $(1,0,1)\times(1,1,1)_{12}$ models respectively.

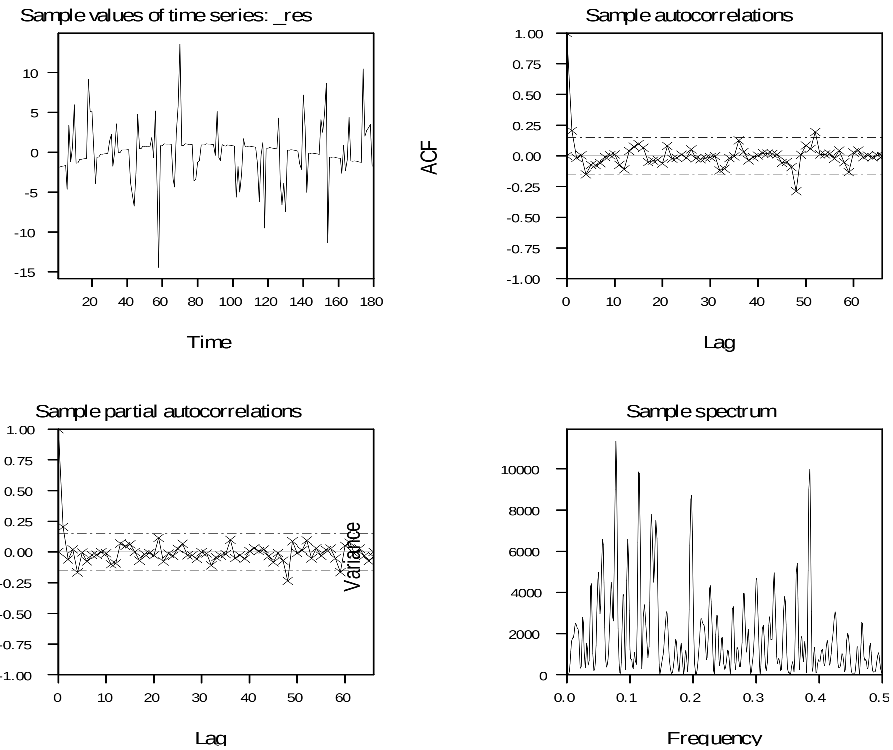

Fig. 4.7 Streamflow time series ACF and PACF of Model (0,1,1) × $(0,1,1)_{12}$

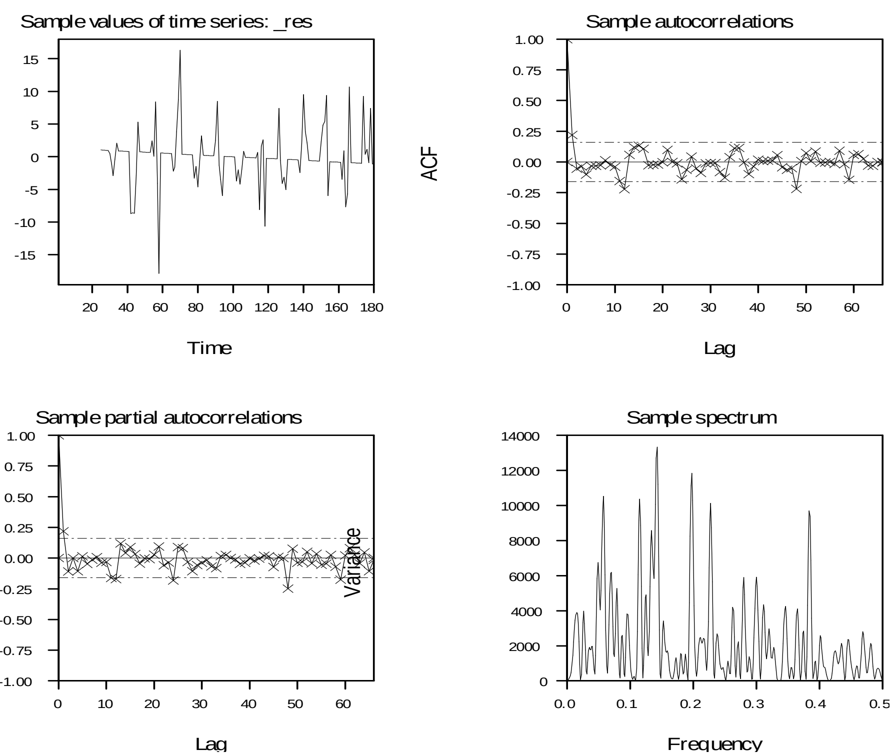

Fig. 4.8 Streamflow time series ACF and PACF of Model (0,1,1) × $(1,1,0)_{12}$

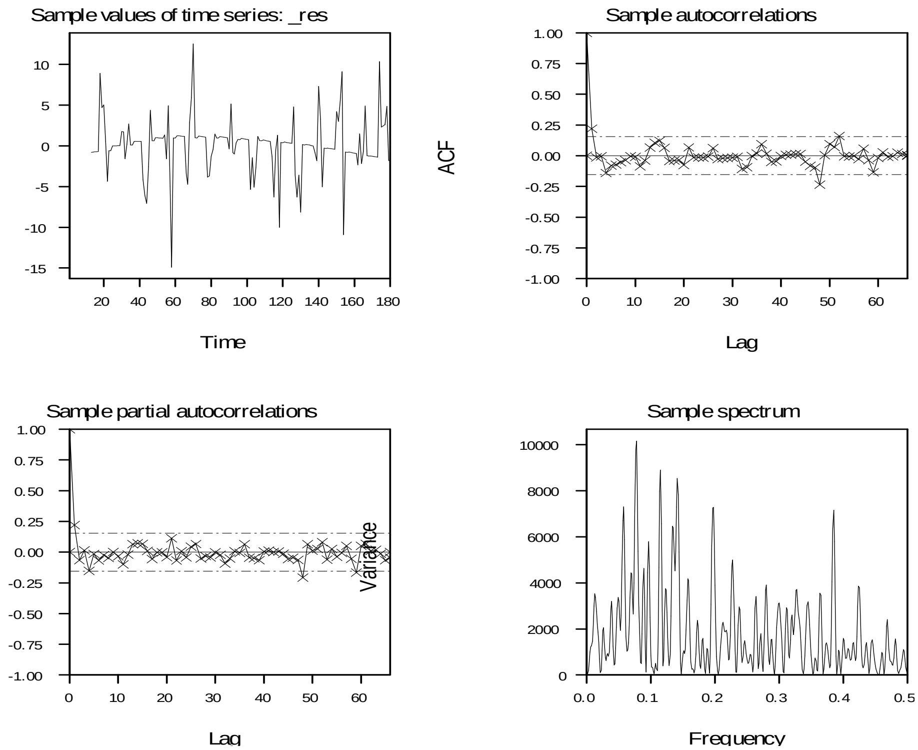

Fig. 4.9 Streamflow time series ACF and PACF of Model (0,1,1) × $(1,1,1)_{12}$

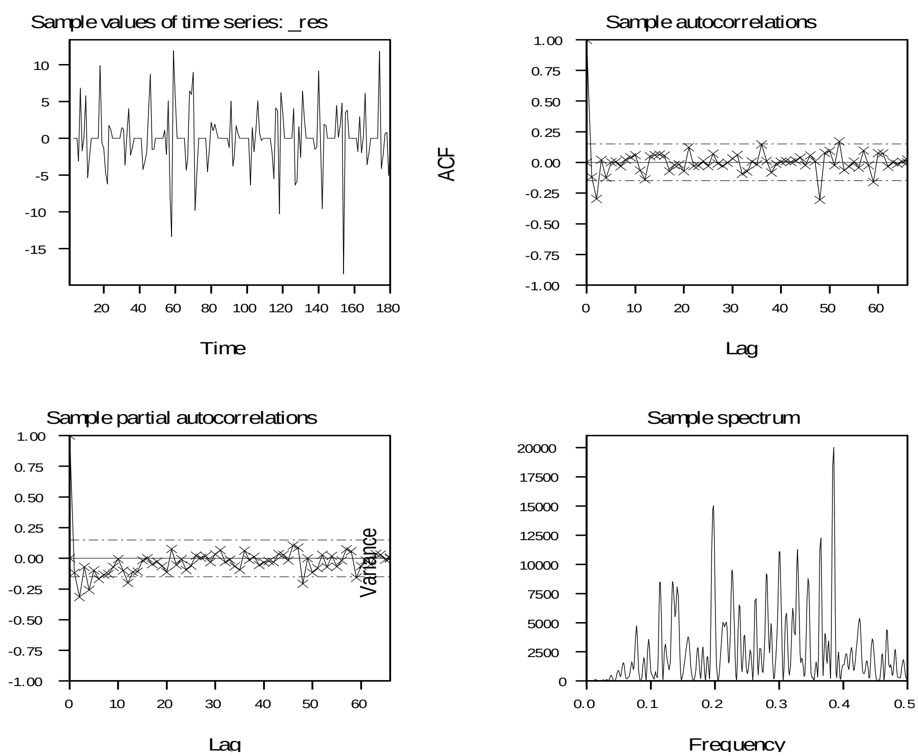

Fig. 4.10 Streamflow time series ACF and PACF of Model (1,1,0) × $(0,1,1)_{12}$

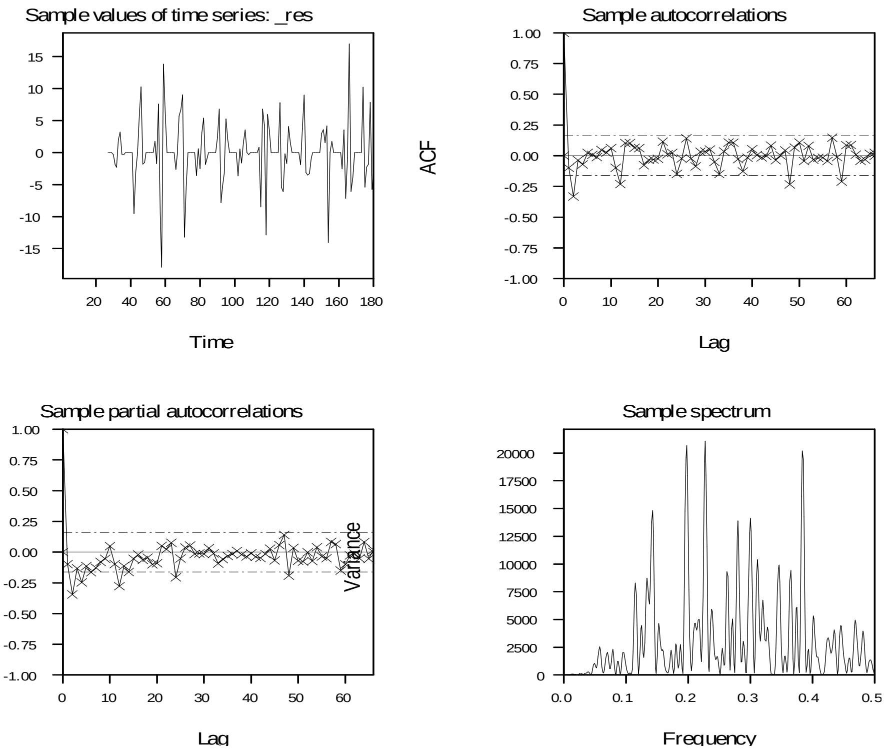

Fig. 4.11 Streamflow time series ACF and PACF of Model (1,1,0) × $(1,1,0)_{12}$

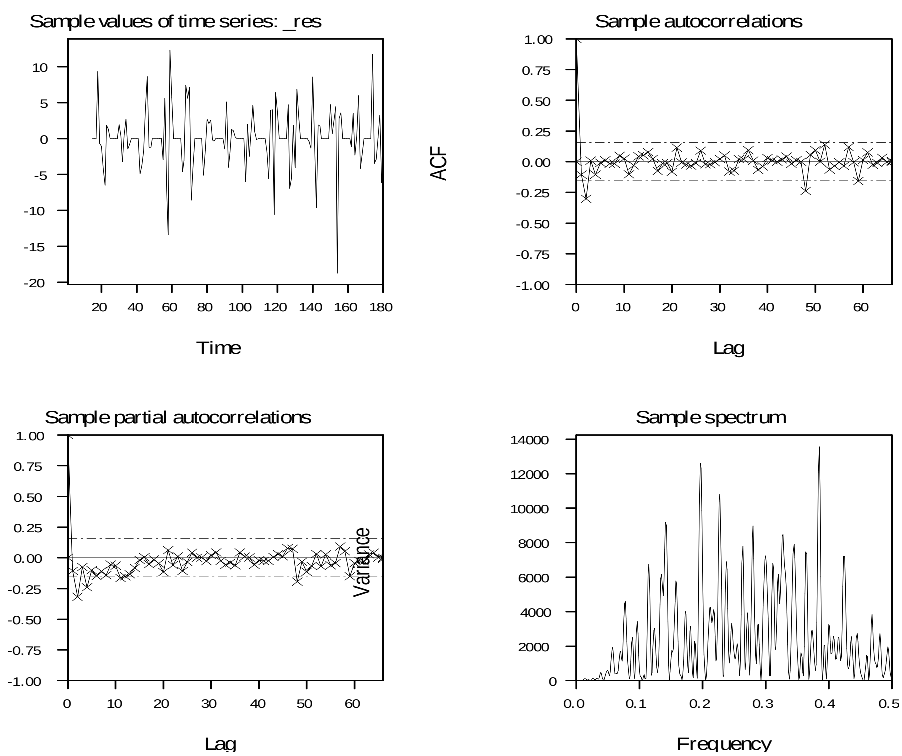

Fig. 4.12 Streamflow time series ACF and PACF of Model (1,1,0) × $(1,1,1)_{12}$

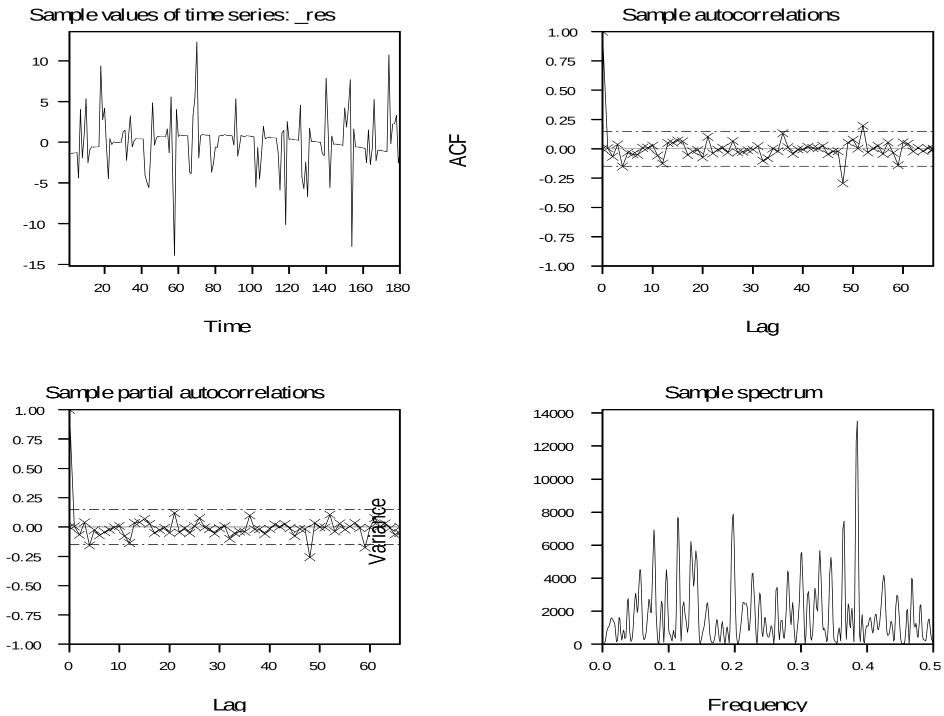

Fig. 4.13 Streamflow time series ACF and PACF of Model (1,1,1) × $(0,1,1)_{12}$

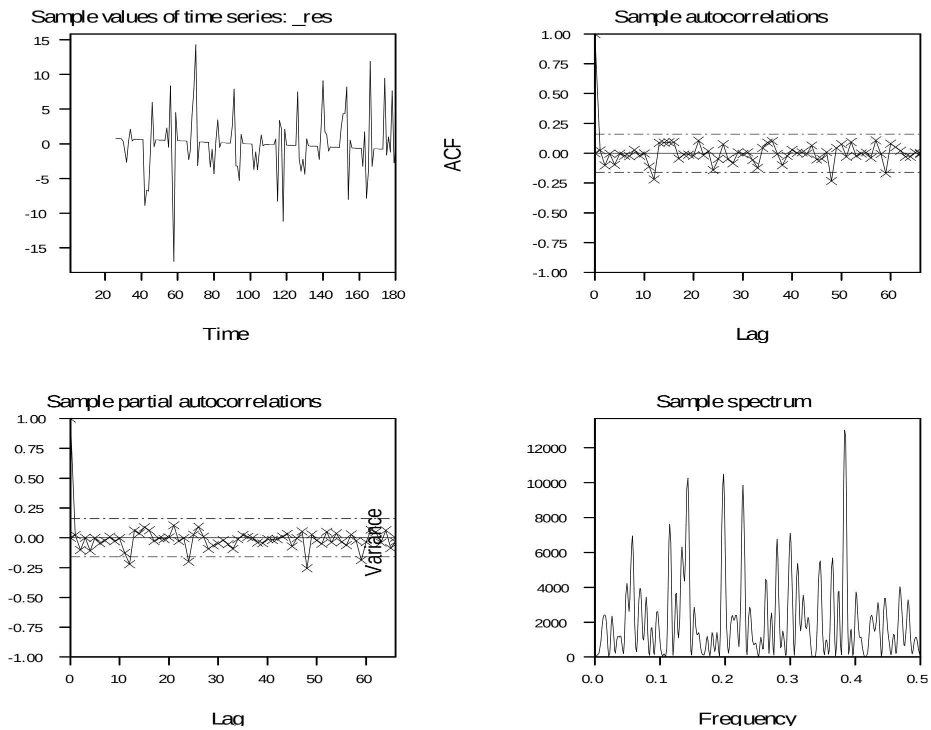

Fig. 4.14 Streamflow time series ACF and PACF of Model (1,1,1) × $(1,1,0)_{12}$

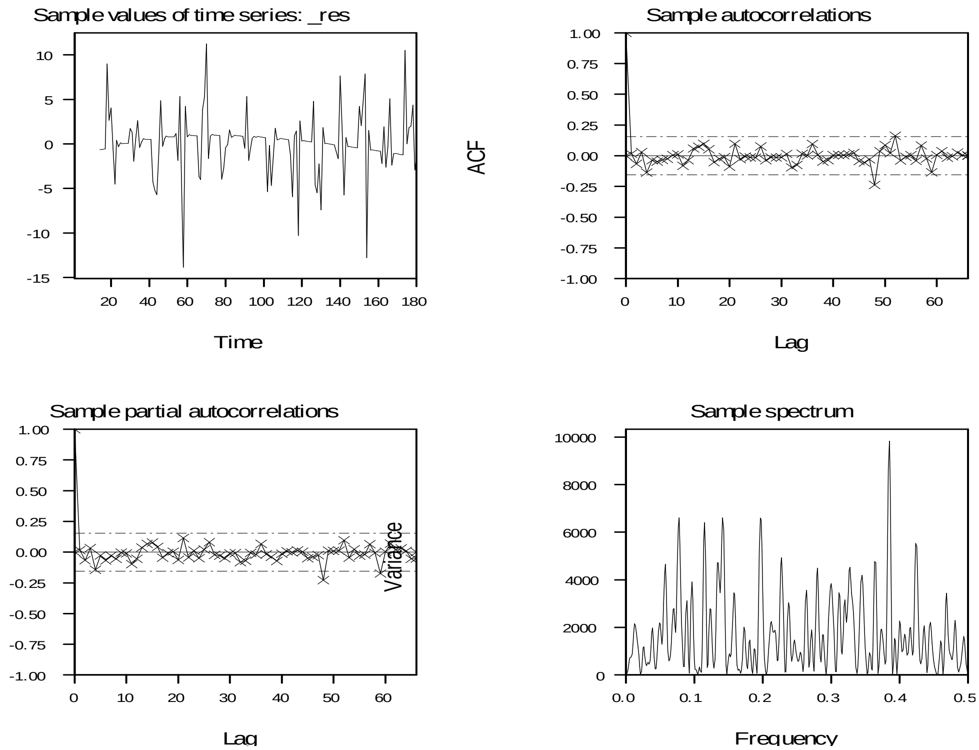

Fig. 4.15 Streamflow time series ACF and PACF of Model (1,1,1) × $(1,1,1)_{12}$

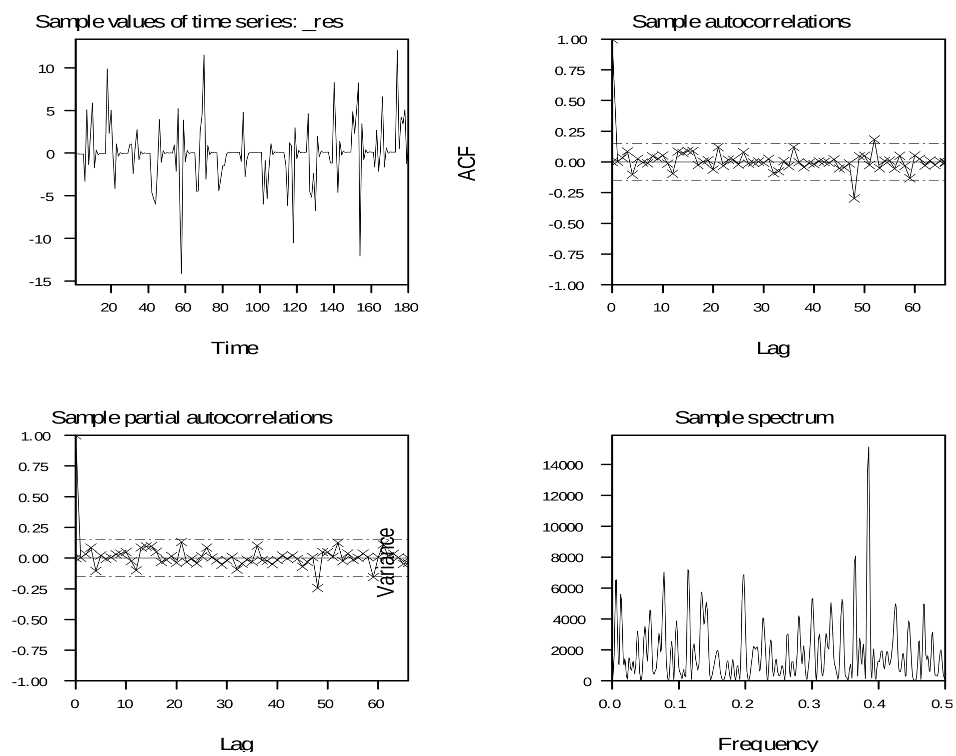

Fig. 4.16 Streamflow time series ACF and PACF of Model (0,0,1) × $(0,1,1)_{12}$

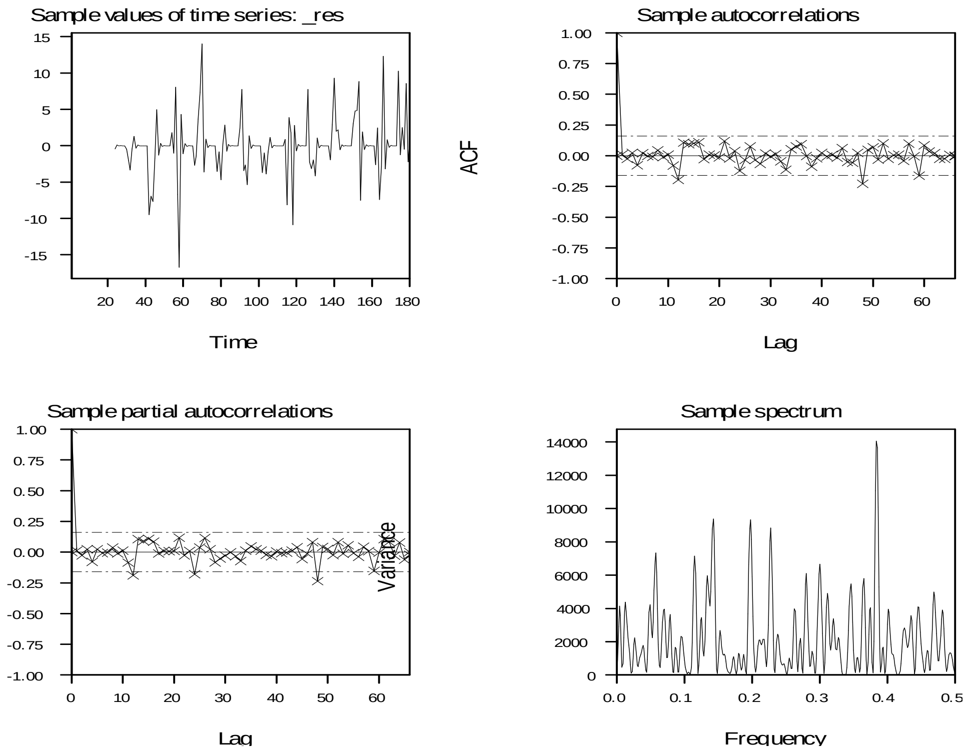

Fig. 4.17 Streamflow time series ACF and PACF of Model (0,0,1) × $(1,1,1)_{12}$

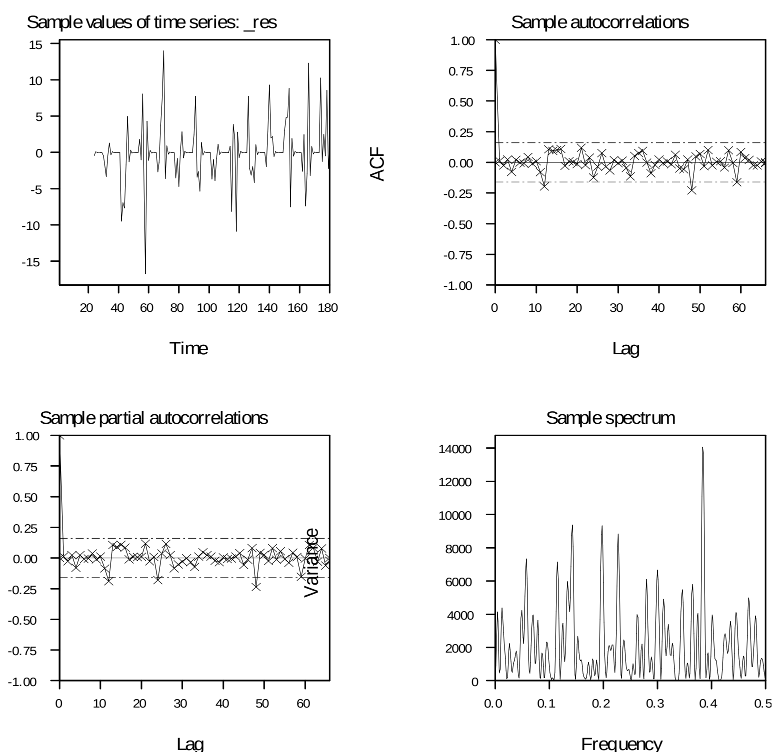

Fig. 4.18 Streamflow time series ACF and PACF of Model (0,0,1) × $(1,1,1)_{12}$

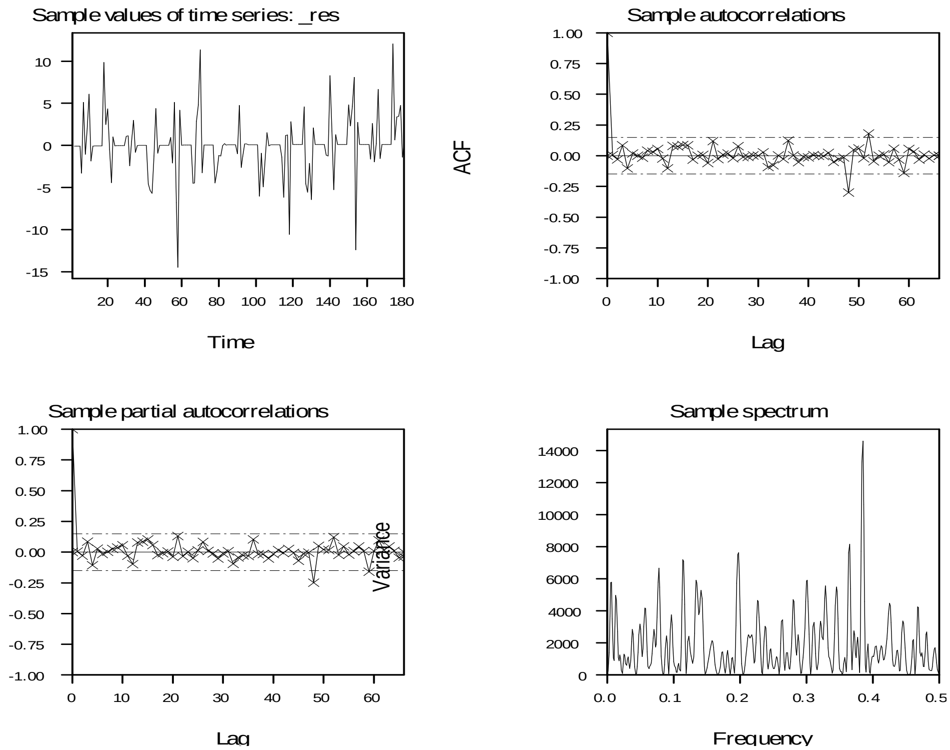

Fig. 4.19 Streamflow time series ACF and PACF of Model (1,0,0,) × $(0,1,1)_{12}$

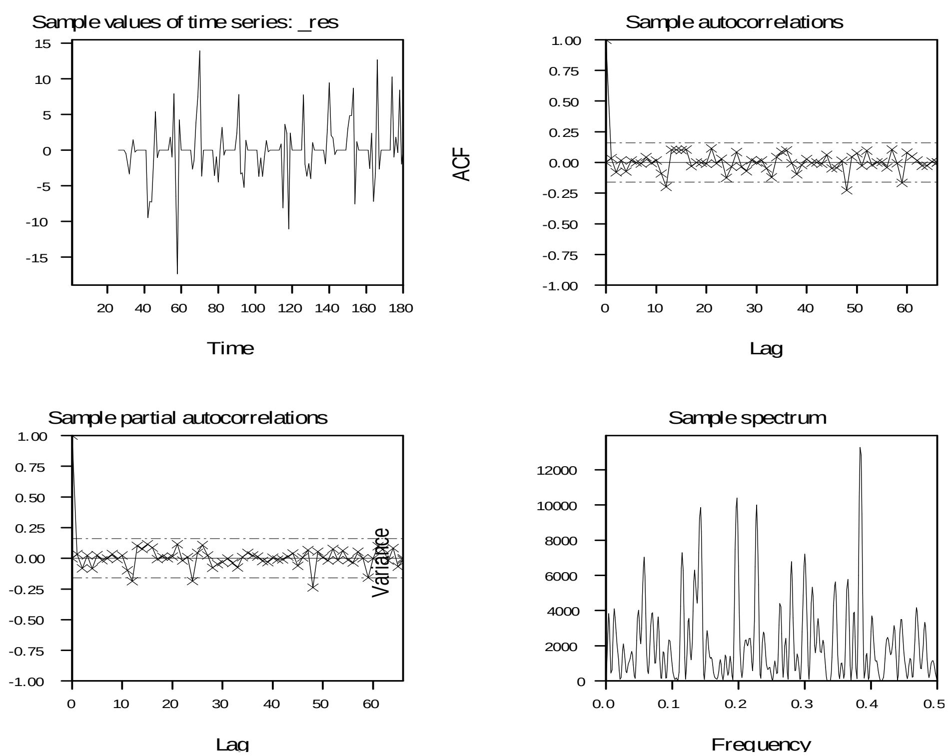

Fig. 4.20 Streamflow time series ACF and PACF of Model (1,0,0,) × $(1,1,0)_{12}$

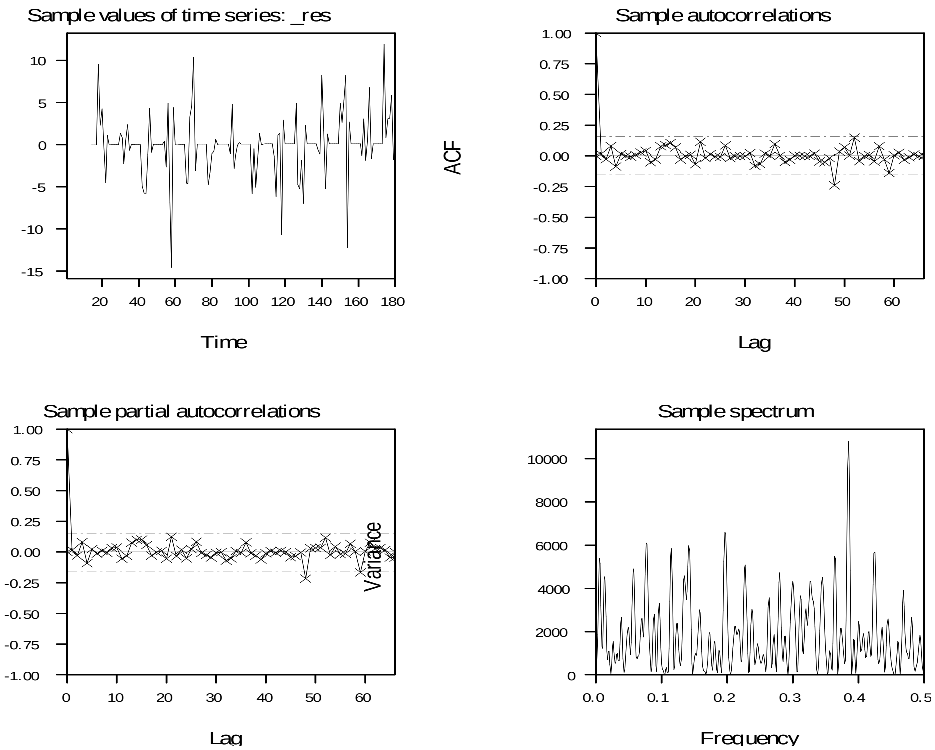

Fig. 4.21 Streamflow time series ACF and PACF of Model (1,0,0,) × $(1,1,1)_{12}$

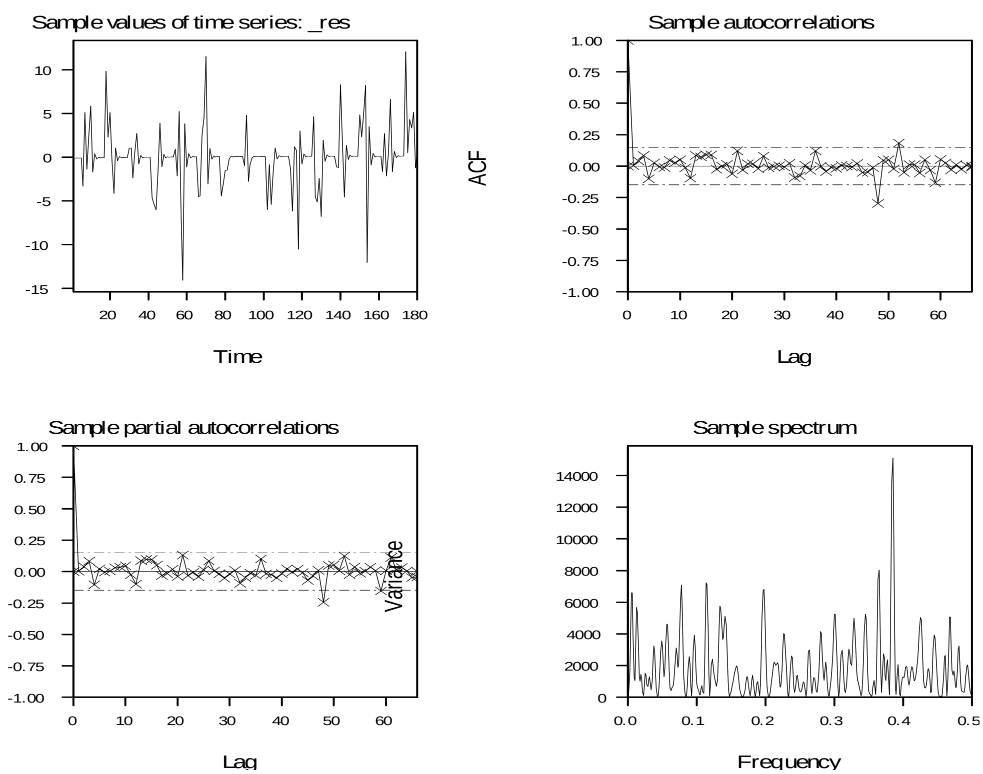

Fig. 4.22 Streamflow time series ACF and PACF of Model (1,0,1) × $(0,1,1)_{12}$

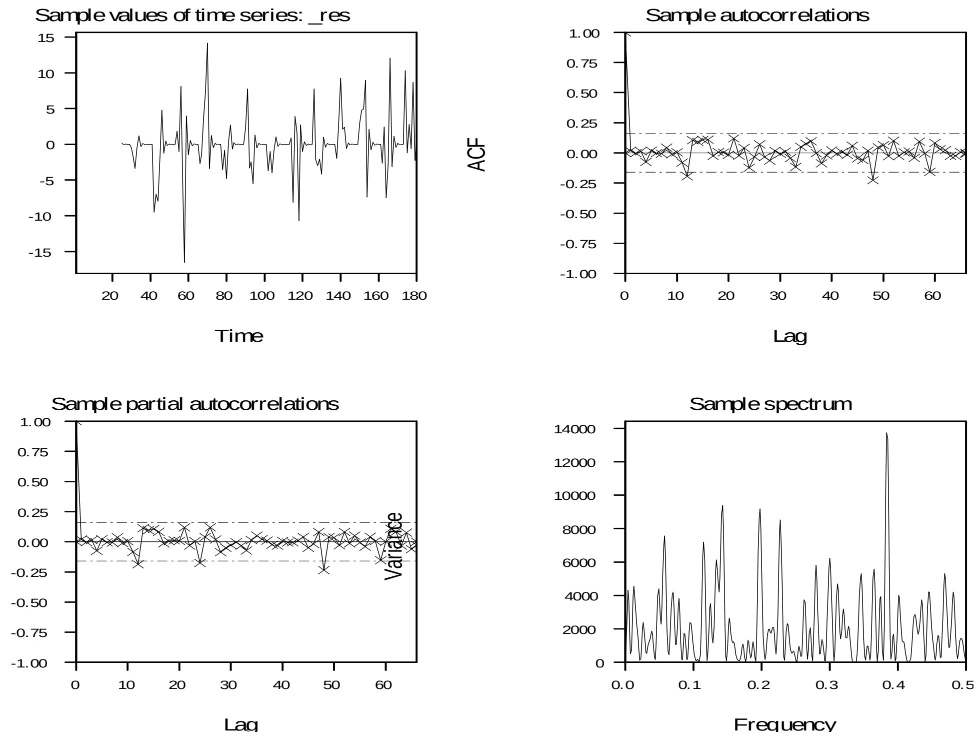

Fig. 4.23 Streamflow time series ACF and PACF of Model (1,0,1) × (1,1,0)$_{12}$

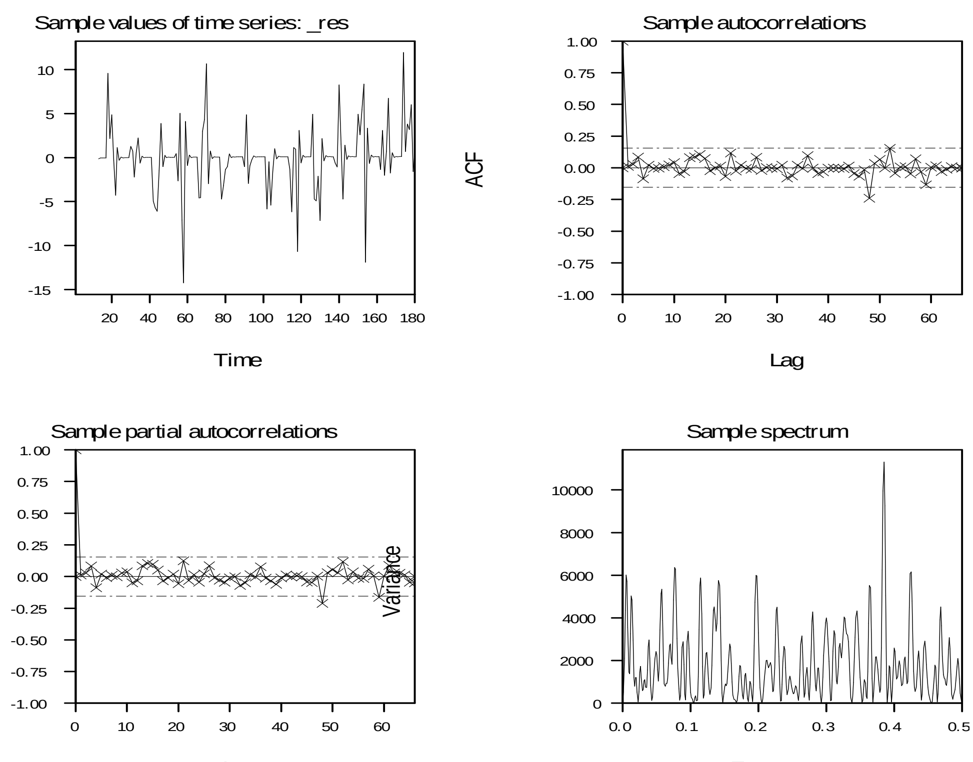

Fig.4.24

Streamflow time series ACF and PACF of Model (1,0,1) × (1,1,1)$_{12}$

Table 4.1. ARIMA models with their corresponding AIC and BIC values

Model Structure $(p,d,q)\times(P,D,Q)_{12}$	AIC	BIC
$(0,0,1)\times(0,1,1)_{12}$	781.125	797.762
$(0,0,1)\times(1,1,0)_{12}$	808.560	824.565
$(0,0,1)\times(1,1,1)_{12}$	780.700	798.702
$(1,0,0)\times(0,1,1)_{12}$	781.125	800.125
$(1,0,0)\times(1,1,1)_{12}$	809.658	827.658
$(1,0,0)\times(0,1,1)_{12}$	780.615	799.645
$(1,0,1)\times(0,1,1)_{12}$	781.125	797.125
$(1,0,1)\times(1,1,0)_{12}$	808.342	828.342
$(1,0,1)\times(1,1,1)_{12}$	780.615	805.615
$(0,1,1)\times(0,1,1)_{12}$	791.941	807.942
$(0,1,1)\times(1,1,0)_{12}$	821.941	825.521
$(0,1,1)\times(1,1,1)_{12}$	788.913	806.912
$(1,1,0)\times(0,1,1)_{12}$	838.314	875.314
$(1,1,0)\times(1,1,0)_{12}$	873.979	897.986
$(1,1,0)\times(1,1,1)_{12}$	836.244	858.261
$(1,1,1)\times(0,1,1)_{12}$	786.244	804.244
$(1,1,1)\times(1,1,0)_{12}$	812.515	831.512

$(1,1,1))\times(1,1,1)_{12}$	779.764	789.762

The table 4.1 shows the AIC and BIC value of eighteen models. All AIC value and BIC value mention in above table 4.1. On the basis of low value of AIC ARIMA $(1,1,1)\times(1,1,1)_{12}$ model was selected.

The comparison of observed streamflow and the predicted streamflow was drawn. The coefficient of determination (R^2) found out from the graph of observed streamflow and predicted streamflow. All graphs of eighteen ARIMA model showed the coefficient of determination of respective models. Fig.4.25 to Fig.4.42 shows the predicted monthly streamflow and observed streamflow data of year 2008 for ARIMA models.

The parameters of ARIMA models showed the differencing in the each ARIMA model. The ARIMA model having high coefficient of determination value is 'good fit'.

The Fig.4.25-Fig4.42 showed the forecasted monthly streamflow was higher than the observed monthly streamflow. The curve lines on the graphs showed the observed monthly streamflow and forecasted streamflow peak value of monthly streamflow in rainy months. The polynomial trend lines showed the fluctuation and the trend of the monthly streamflow data. Fig. 4.33 showed that the ARIMA model $(1,1,1)\times(1,1,1)_{12}$ having highest coefficient of determination value.

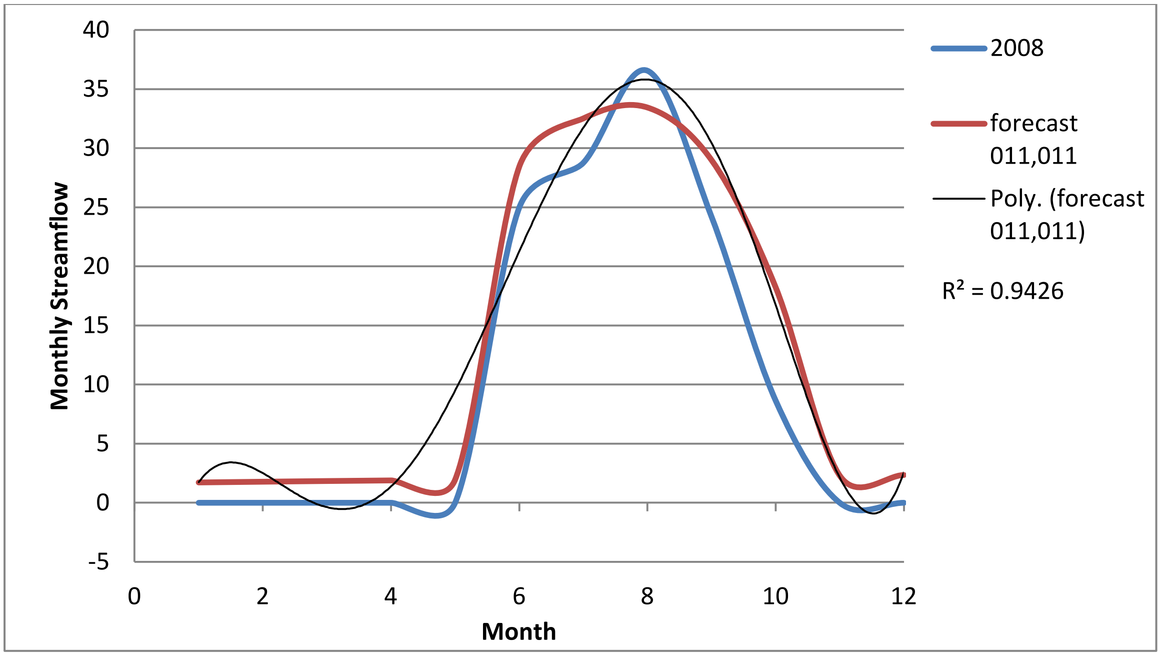

Fig. 4.25 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,1,1)\times(0,1,1)_{12}$

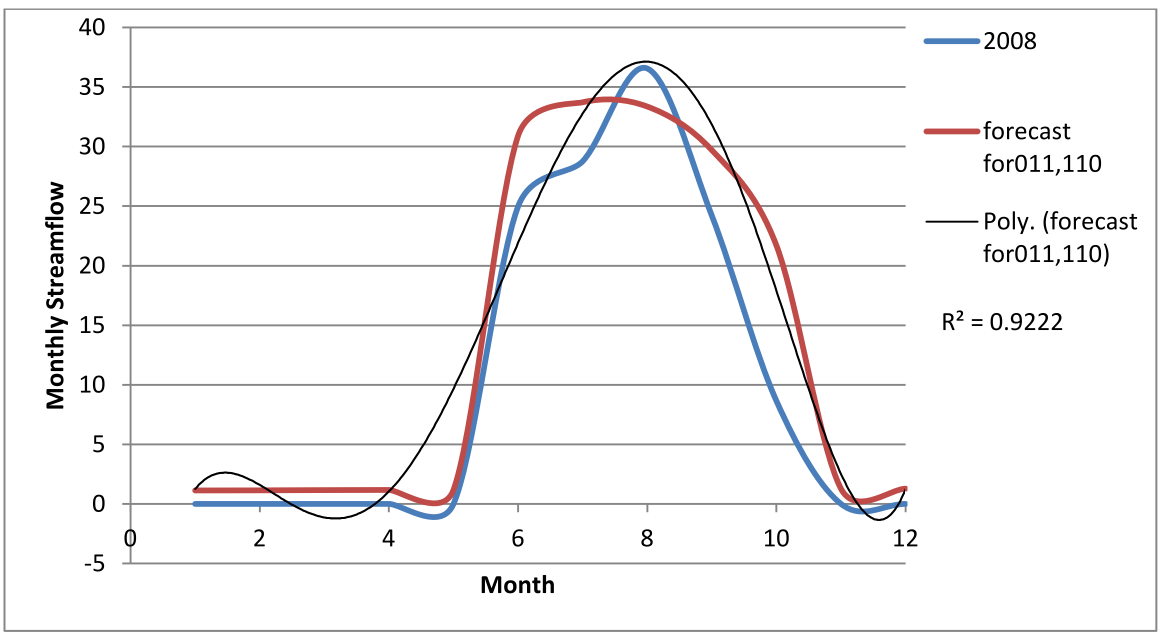

Fig. 4.26 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(0,1,1)\times(1,1,0)_{12}$

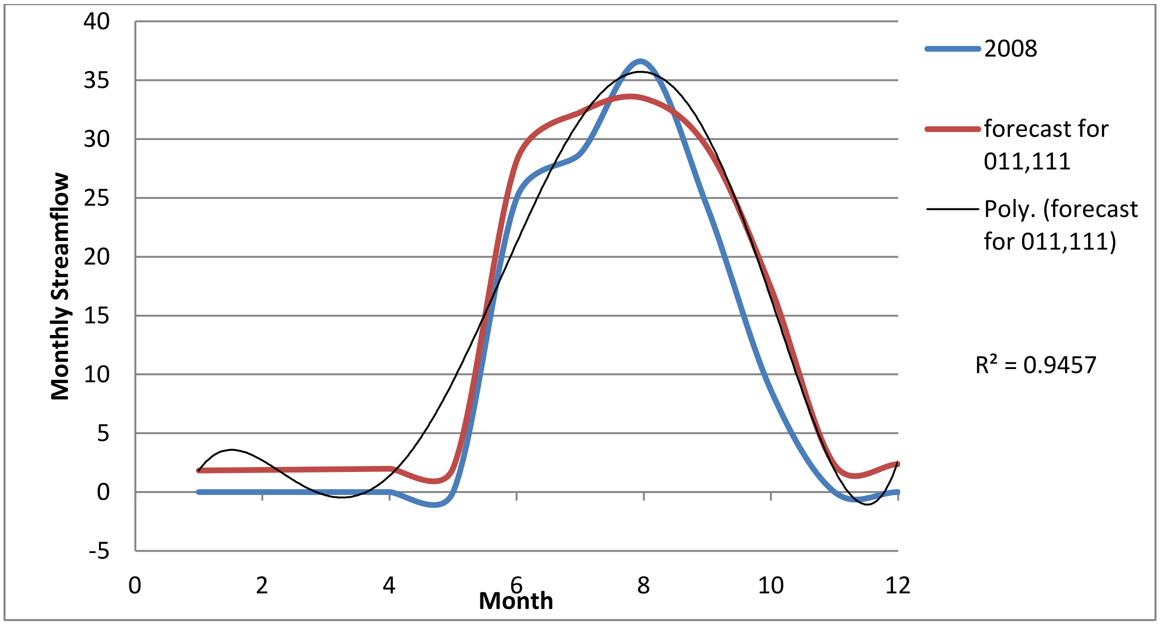

Fig. 4.27 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (0,1,1)×(1,1,1)$_{12}$

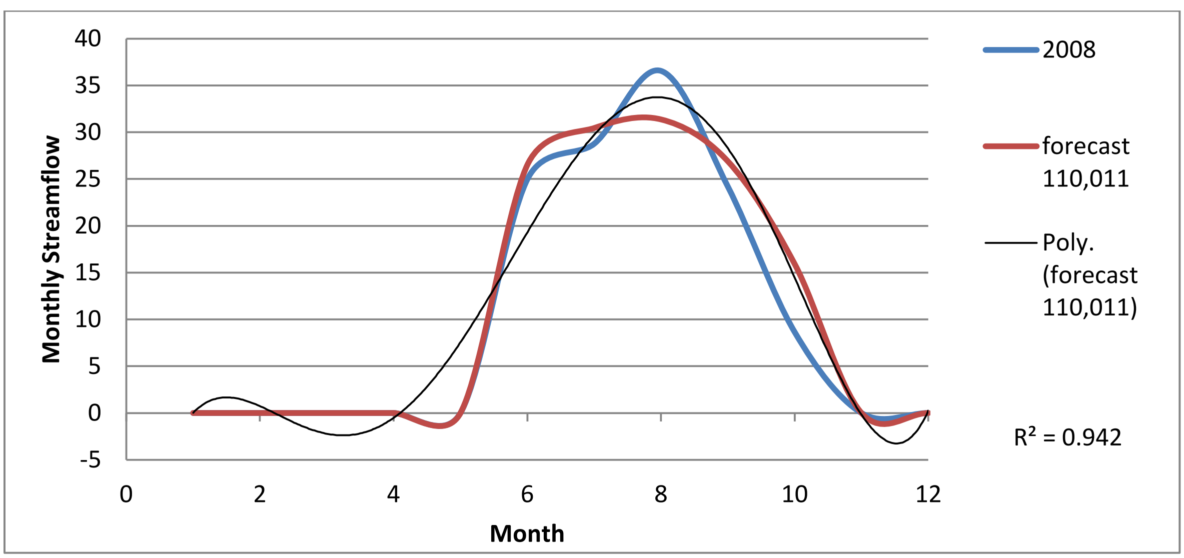

Fig. 4.28 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (1,1,0)×(0,1,1)$_{12}$

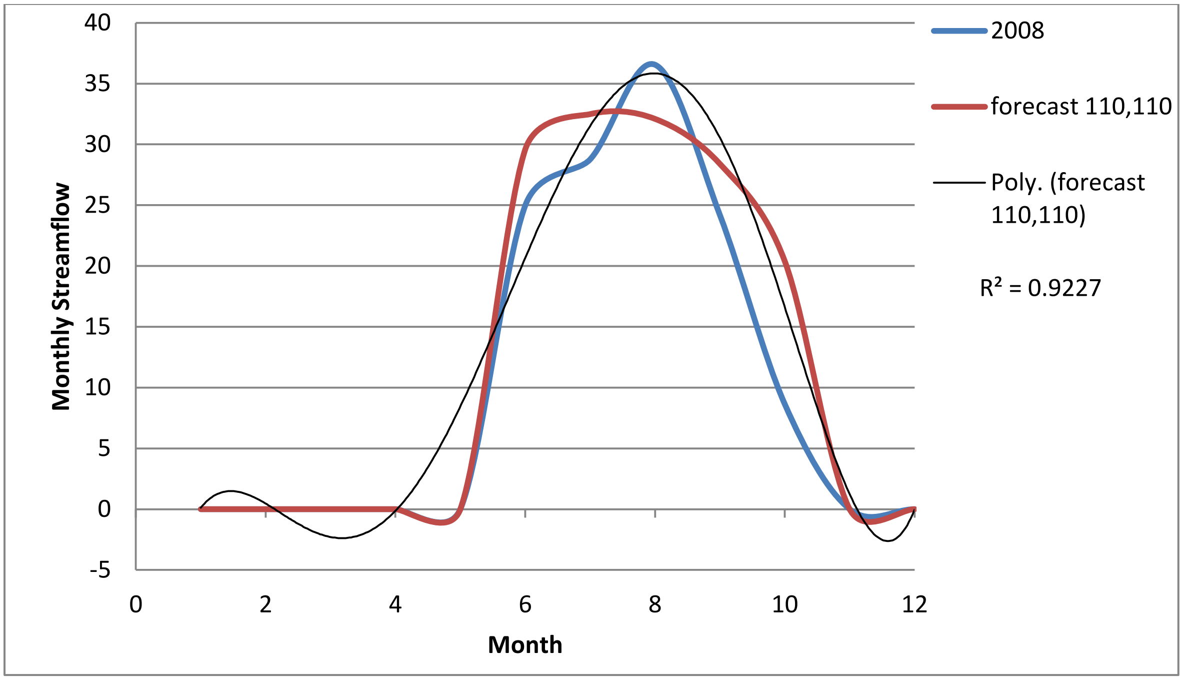

Fig. 4.29 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,0)\times(1,1,0)_{12}$

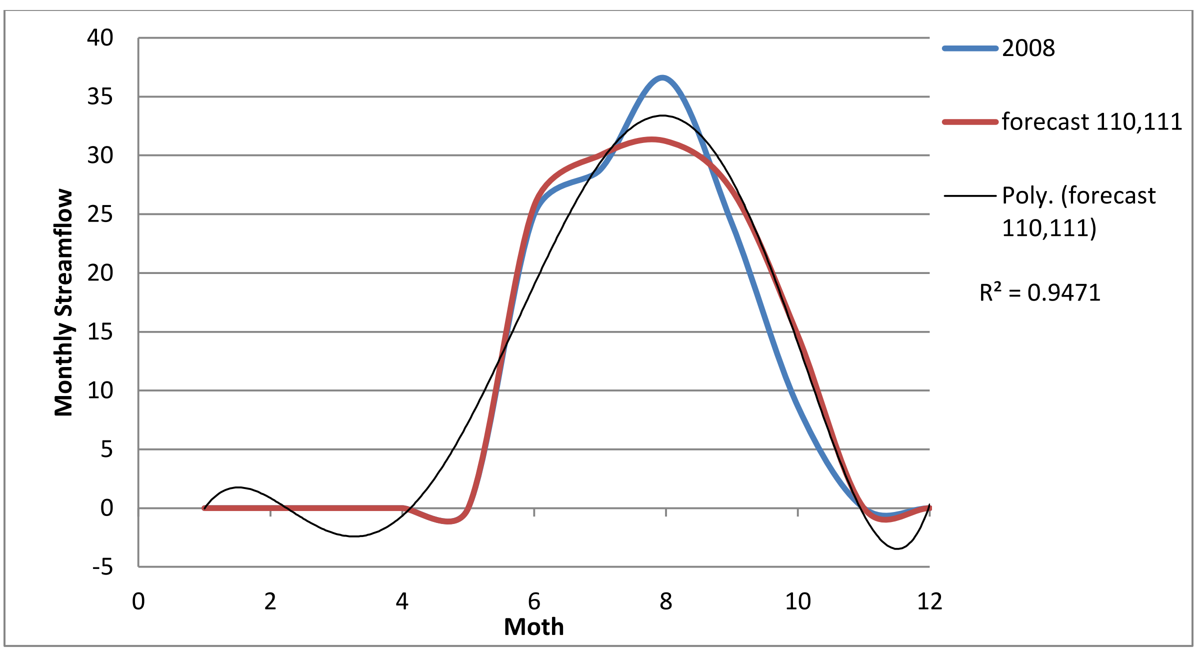

Fig. 4.30 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,1,0)\times(1,1,1)_{12}$

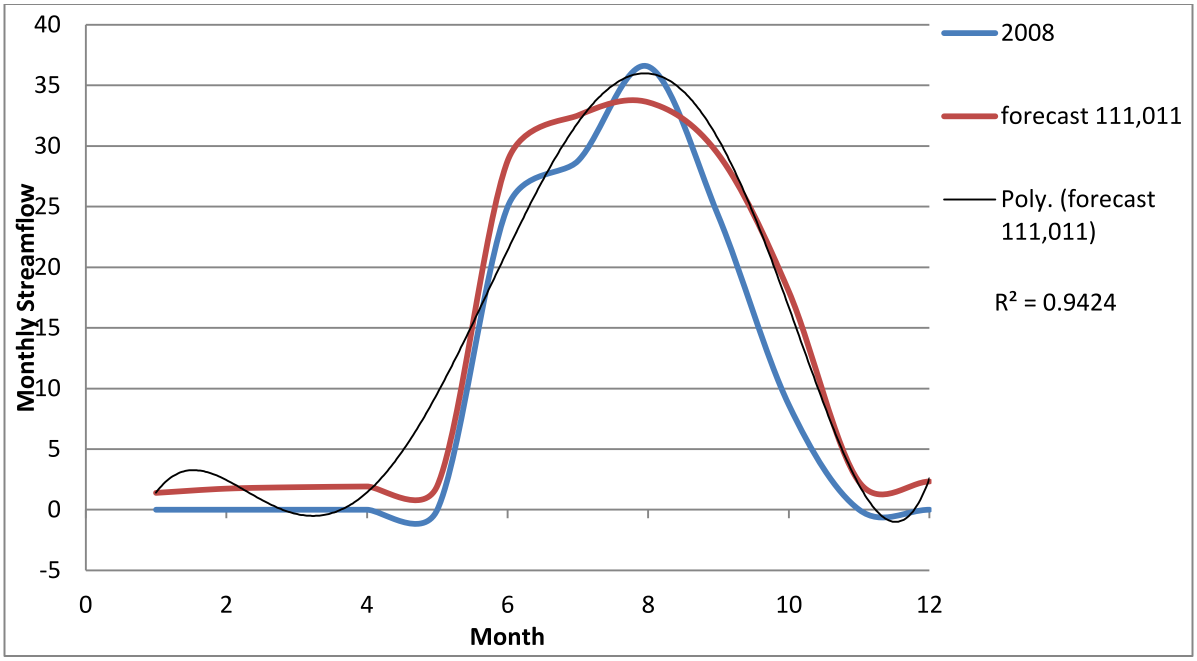

Fig. 4.31 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (1,1,1)×(0,1,1)$_{12}$

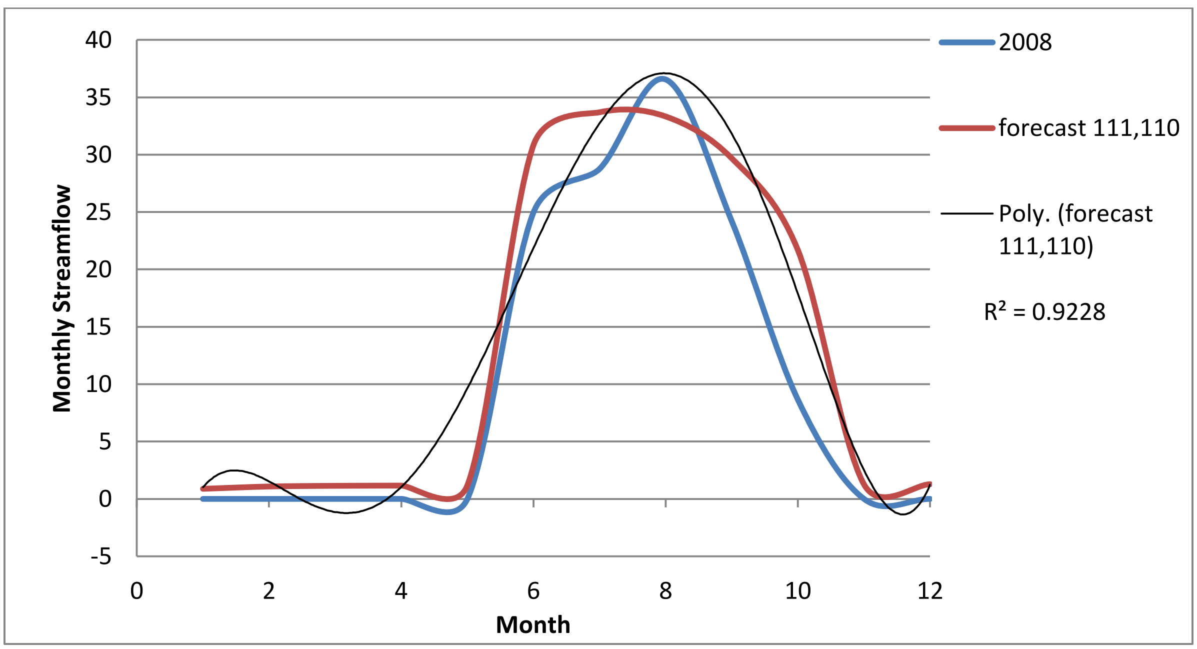

Fig. 4.32 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (1,1,1)×(1,1,0)$_{12}$

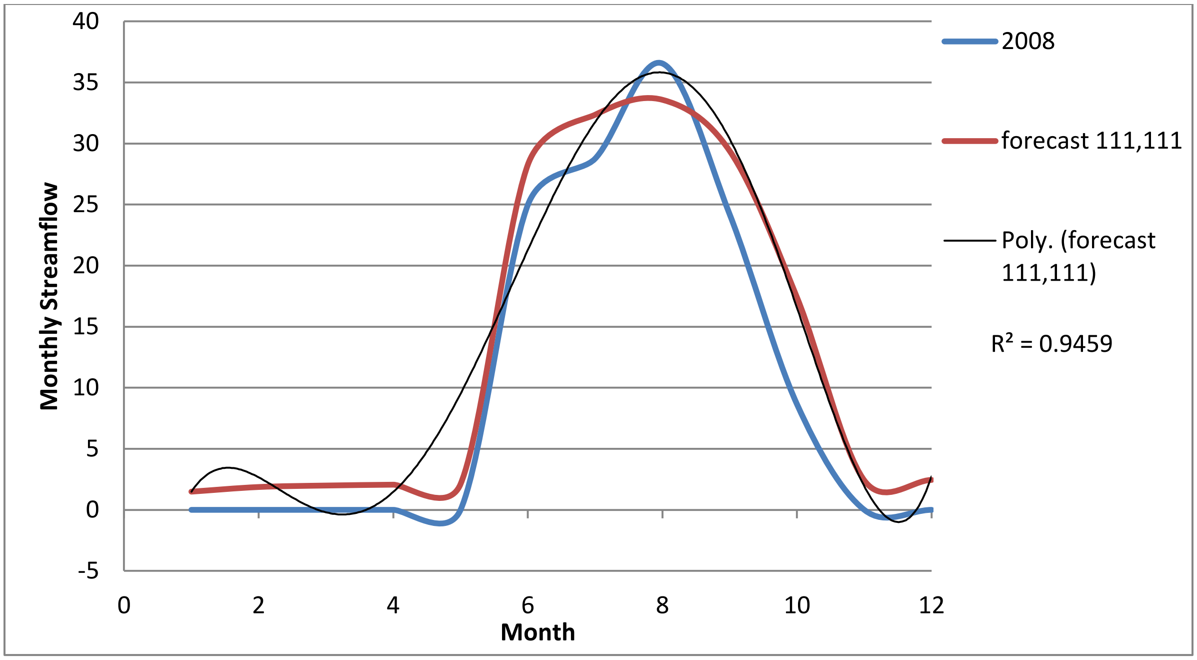

Fig. 4.33 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (1,1,1)×(1,1,1)$_{12}$

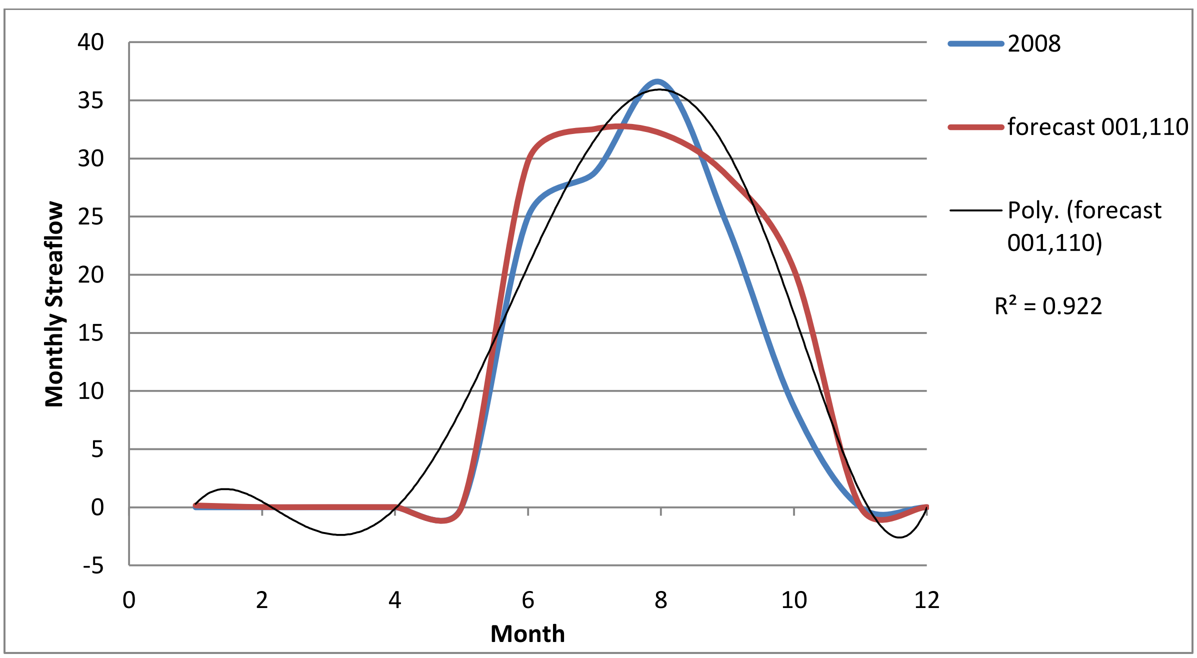

Fig. 4.34 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (0,0,1)×(1,1,0)$_{12}$

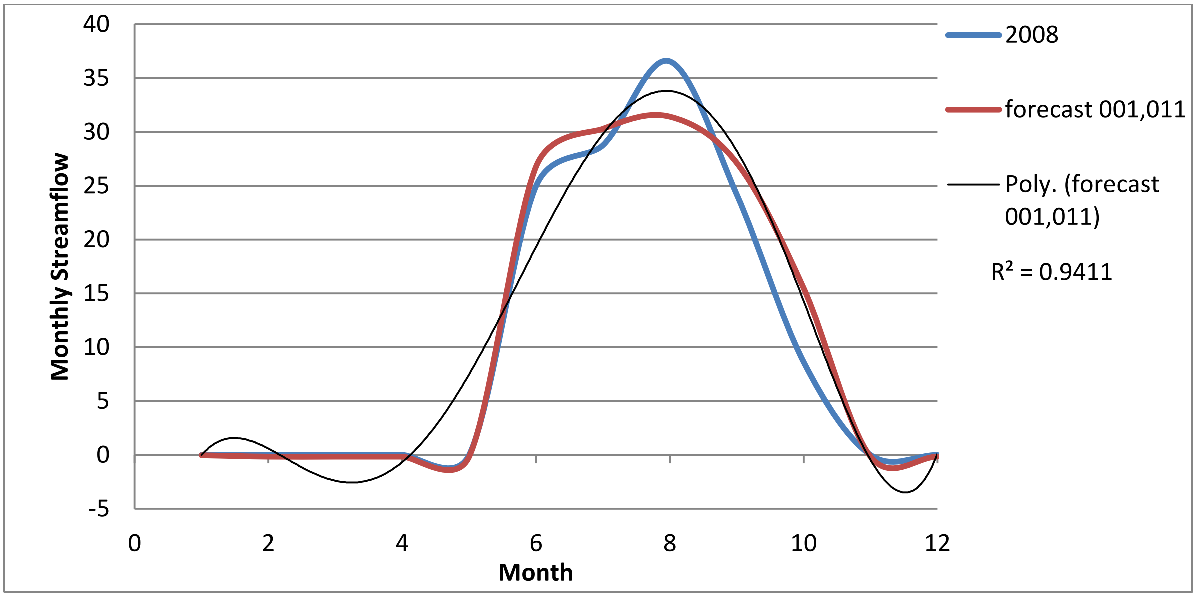

Fig. 4.35 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (0,0,1)×(0,1,1)$_{12}$

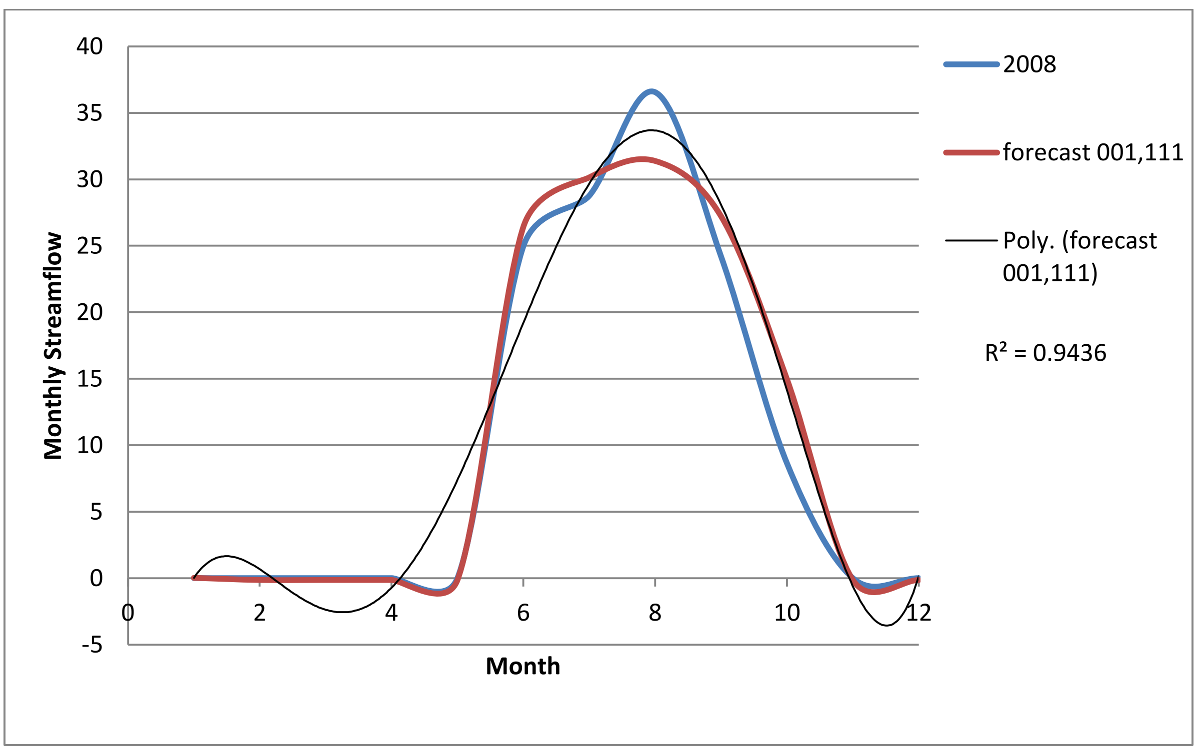

Fig. 4.36 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (0,0,1)×(1,1,1)$_{12}$

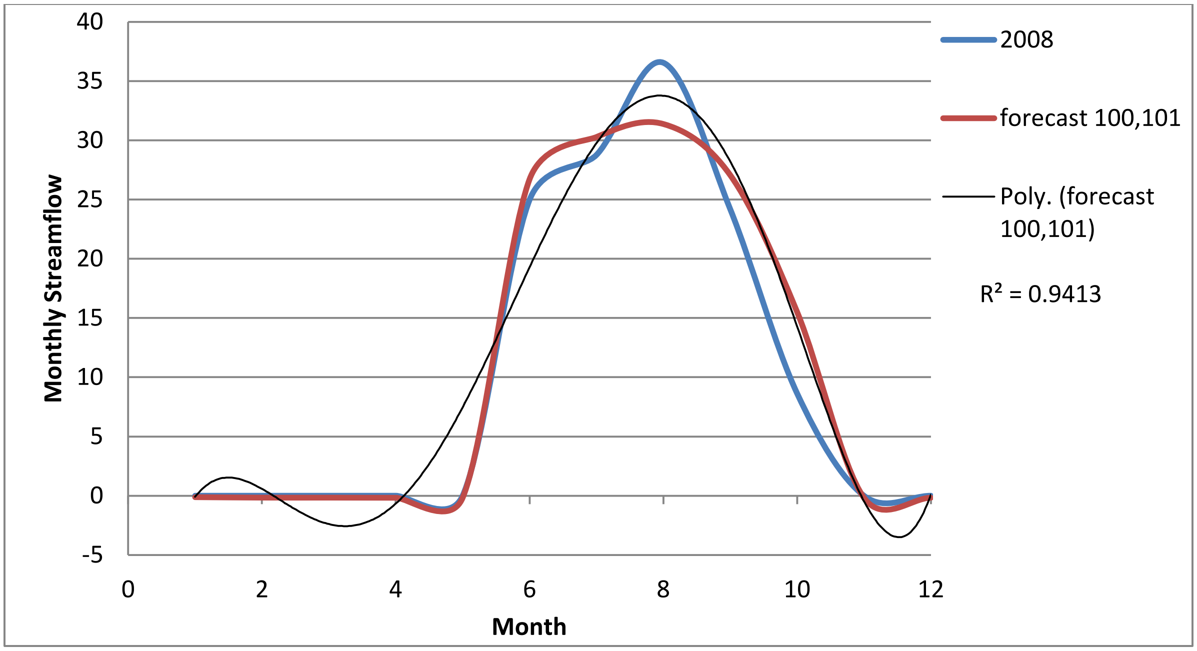

Fig. 4.37 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,0)\times(1,0,1)_{12}$

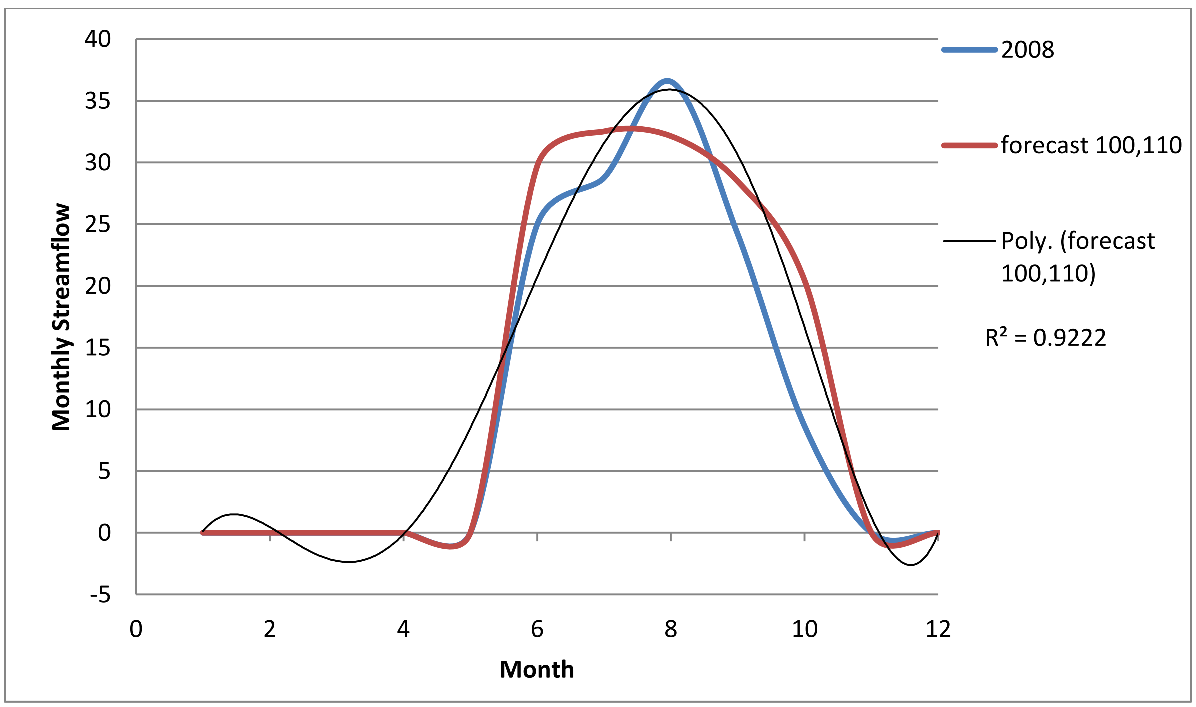

Fig. 4.38 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,0)\times(1,1,0)_{12}$

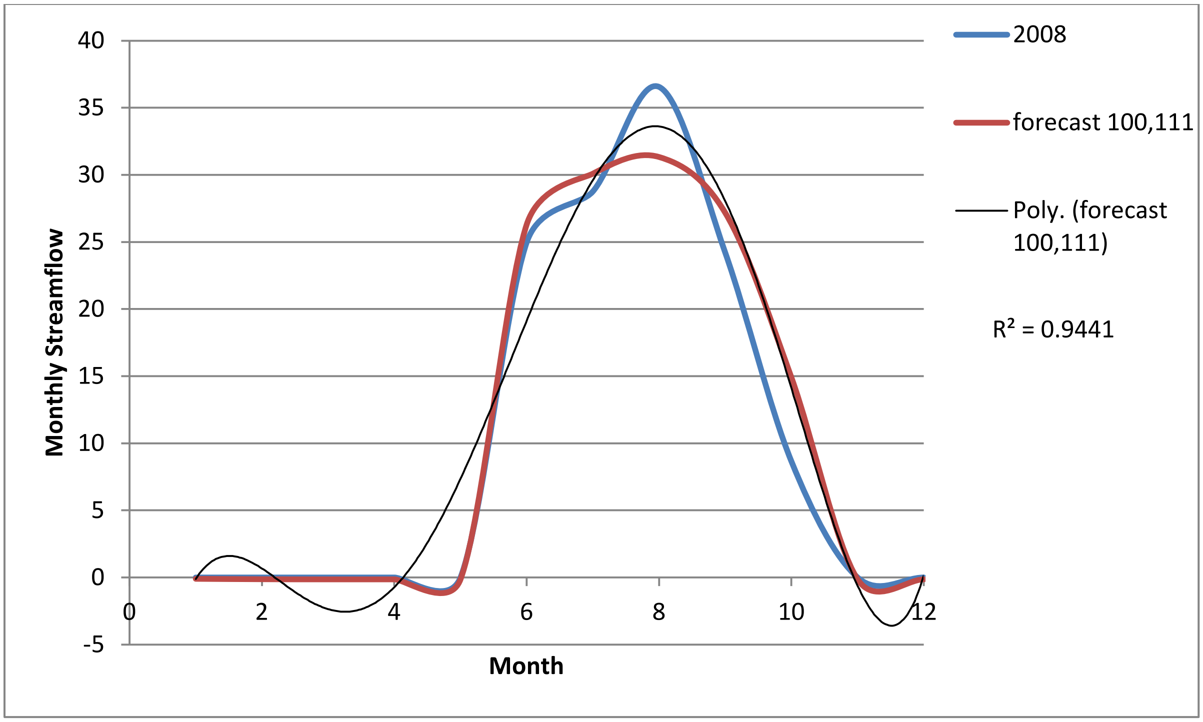

Fig. 4.39 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,0)\times(1,1,1)_{12}$

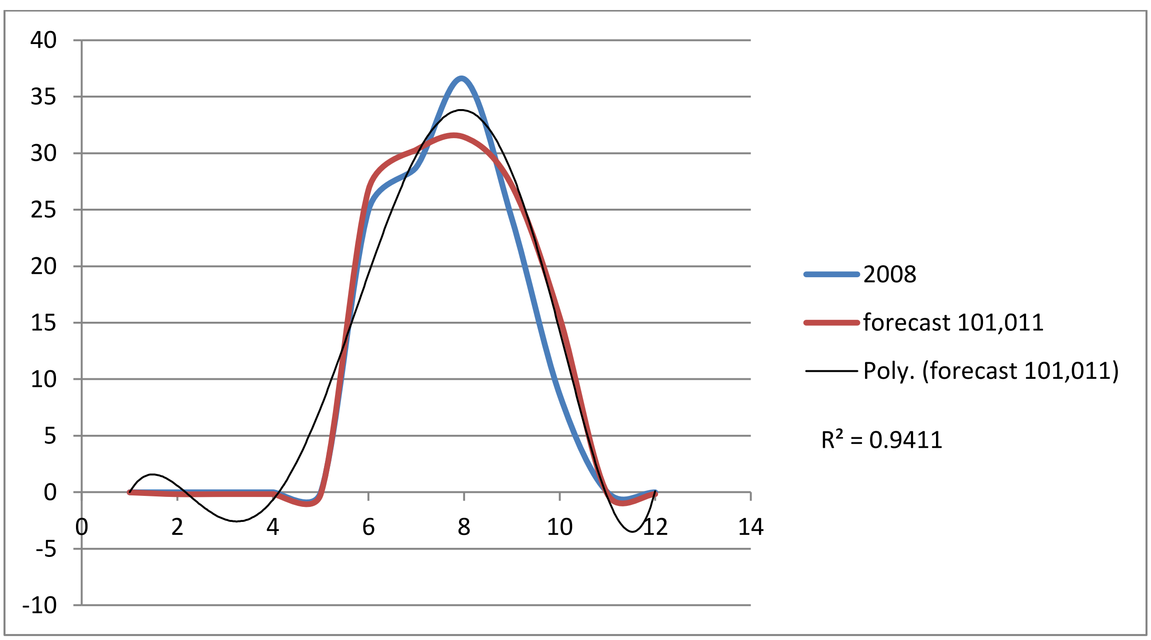

Fig. 4.40 Comparison of observed monthly streamflow and predicted streamflow by ARIMA $(1,0,1)\times(0,1,1)_{12}$

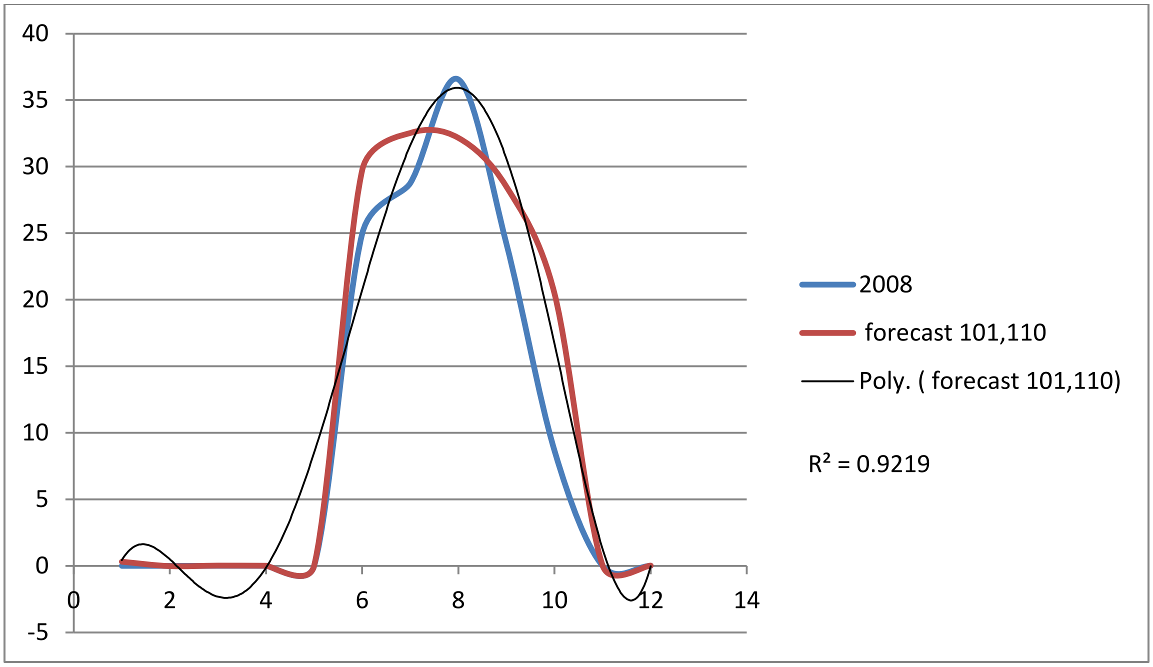

Fig. 4.41 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (1,0,1)×(1,1,0)$_{12}$

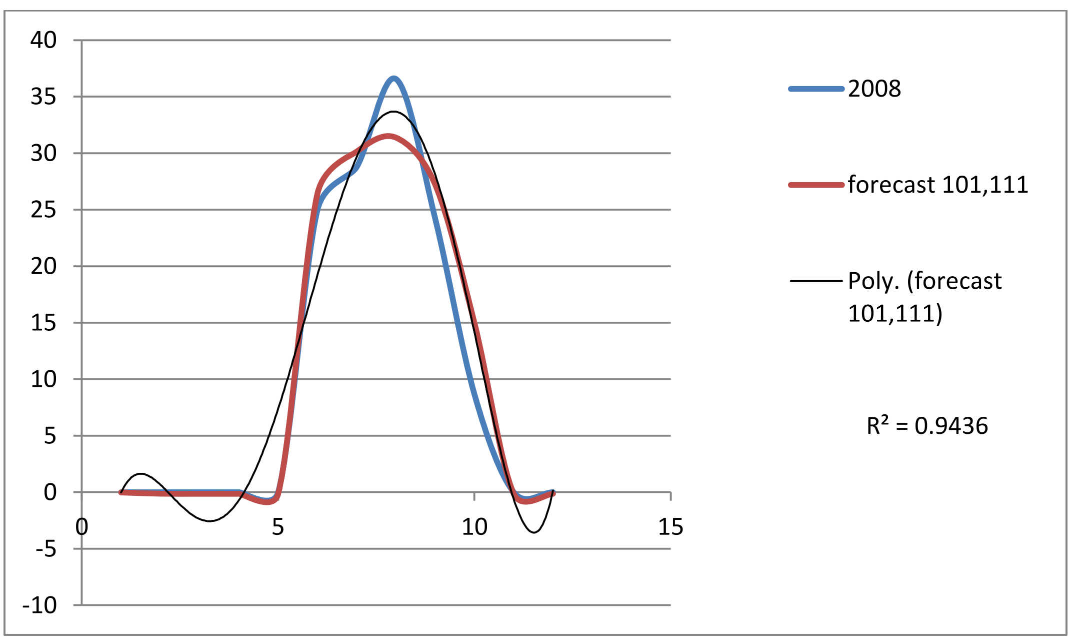

Fig. 4.42 Comparison of observed monthly streamflow and predicted streamflow by ARIMA (1,0,1)×(1,1,1)$_{12}$

Table 4.2. Diagnostic test results for the selected ARIMA

Model statistics	R-squared	RMSE	SE	Box-Ljung Statistic	AIC	BIC
$(0,0,1)\times(0,1,1)_{12}$	0.9411	28.82	0.0622	6.815	781.125	797.125
$(0,0,1)\times(1,1,1)_{12}$	0.9436	28.82	0.0788	6.853	780.700	798.702
$(1,0,0)\times(0,1,1)_{12}$	0.9413	28.79	0.0619	6.853	781.125	800.125
$(1,0,0)\times(1,1,1)_{12}$	0.9441	28.62	0.0781	6.812	780.616	799.645
$(1,0,1)\times(0,1,1)_{12}$	0.9411	28.70	0.0625	6.827	781.125	797.125
$(1,0,1)\times(1,1,1)_{12}$	0.9436	28.57	0.0792	6.581	780.616	805.615
$(0,1,1)\times(1,1,1)_{12}$	0.9457	28.51	0.0727	7.942	788.914	806.912
$(1,1,1)\times(0,1,1)_{12}$	0.9424	28.59	0.0611	6.717	786.244	804.244
$(1,1,1)\times(1,1,1)_{12}$	0.9459	28.60	0.0738	6.937	779.764	789.762

It is evident from Figs. 4.7 - 4.24 that the values of the ARIMA models showed the forecasted streamflow of the respective ARIMA models. The analysis done for the all eighteen ARIMA models and the parameters of nine ARIMA models were selected from that. With the parameters R-square, RMSE value, Standard error, Box-ljunk statistics were find out. In these ARIMA models the AIC value of ARIMA model $(1, 1, 1)\times(1, 1, 1)_{12}$ was lowest. The ARIMA model $(1,1,1)\times(1,1,1)_{12}$ residuals lie within the upper and lower confidence limits (Figs. 4.15). In case of other models, all the values however, do not lie within upper and lower confidence limits. Thus, the ACF and PACF of the residuals also indicated 'best fit' of ARIMA $(1, 1, 1)\times(1, 1, 1)_{12}$ model.

4.1.3 Mean stream flow forecasting with ARIMA

The selected ARIMA $(1, 1, 1)\times(1, 1, 1)_{12}$ is used to forecast the mean monthly stream flow for the period January-08 to December-08 by using the observed data of the period June-1993 to June-07. The forecasted data were compared with the observed data (Table 4.3).

Table 4.3. Comparison of observed and forecasted stream flow during January -08 to Decmber-08.

Month	Observed	Forecasted ARIMA
Jan-08	0.00	2.15
Feb-08	0.00	3.94
Mar-08	0.00	5.71
Apr-08	0.00	3.94
May-08	0.00	6.36
Jun-08	24.96	28.26
Jul-08	28.76	32.36
Aug-08	36.55	33.58
Sep-08	24.19	29.39
Oct-08	0.00	2.41
Nov-08	0.00	2.45
Dec-08	3.33	21.11

The forecasted monthly streamflow has followed the same tend. However the forecasted monthly streamflow data of ARIMA model $(1,1,1)\times(1,1,1)_{12}$ was more than the observed monthly data. (Table 4.3). The ARIMA model $(1,1,1)\times(1,1,1)_{12}$ tends to overestimates the streamflow magnitude of magnitude of Savitri river.

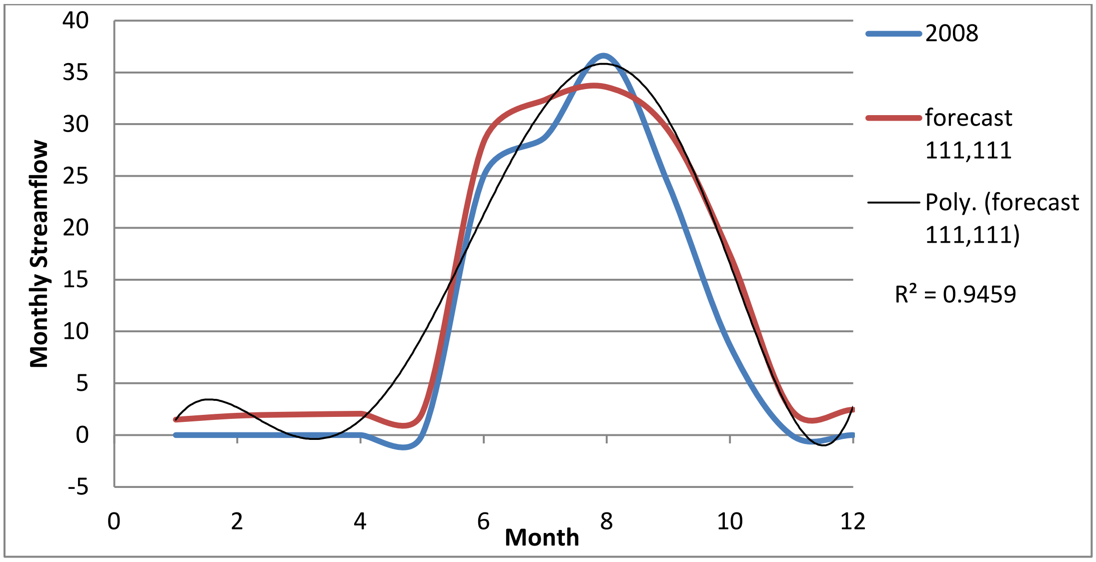

Fig 4.43. Observed and forecasted mean monthly stream flow during January -08 to Decmber-08.

The ARIMA model $(1,1,1)\times(1,1,1)_{12}$ was having Coefficient of determination 0.9459, RMSE value 28.60, Standard error value 0.0738, AIC value 779.764 and BIC value was 789.762. The ARIMA model of $(1,1,1)\times(1,1,1)_{12}$ showed the forecasting streamflow values for January 2008 to December 2008. The ACF and PACF of the residuals also indicated ‘best fit’ of ARIMA model $(1,1,1)\times(1,1,1)_{12}$ of Savitri River.

V. SUMMARY AND CONCLUSIONS

In developing countries, where agriculture is a major economic sector, irrigation accounts for about 85% of water withdrawals. Irrigation has acquired increasing importance in agriculture over the world. India has made considerable progress as far as creation of irrigation potential is concerned. The total irrigation potential has increased from 22.6 Mha in 1951 to 93.95 Mha in 2001, out of which 80.06 Mha has been utilized at the end of the ninth five year plan (Jeyaseelan, 2004). The gap between irrigation potential created and utilized is, however, a matter of concern. This gap has shifted the attention from development of new irrigation projects to the improvement of the performance of the existing irrigation projects. The success of irrigation system operation and planning depends on the quantification of supply and demand and equitable distribution of supply to meet the demand if possible, or, to minimize the gap between the supply and demand. Irrigation projects, which receive water from reservoir, can be challenging to manage since annual fluctuations in runoff from the reservoir's catchment can have considerable impact on the irrigation management strategy. Hence, forecasting reservoir inflow for such projects is essential for proper planning and management.

The planning and designing of water resources projects need information on different hydrologic event including stream flows. While historical streamflow data would indicate the characteristics of the streamflow, it does not give the long term sequence of flow to which the system would be subjected. As the streamflow are the important input to the water resources system, the development of the good stochastic models is essential in order to determine an optimal system operation. The appropriate models is that generate flow sequence are those, which manage to characterize the streamflow ideally representing the statistical and the correlation structure of the corresponding observed streamflow.

In India, irrigation remains the single largest user of the water resources and accounts for about 84% of all withdrawals (Planning Commission, 2002). For estimating runoff (reservoir inflow), there are basically two types of model, i.e., process-driven and data-driven. Process-driven models are based on physical facts of the problem and are constituted with combination of some experimental equations. The data- driven models are based on the analysis of all the data, characterizing the system under study. A model can then be defined on the basis of connections between the system state variables (input, internal and output variables) with only a limited number of assumptions about the "physical" behavior of the system.

Forecasting of the streamflow and generation of the synthetic flows has been one of the important problems for irrigation and other agriculture practices. Because of the inherently non linear relationship between input variables that influence streamflow such as rainfall, land characteristics, human interference etc. This complicates the attempts to forecast streamflow events

ARIMA is most sophisticated extrapolation method for forecasting. ARIMA model has been popularized by George Box and Gwilym Jenkins in the early 1970’s, and there names have frequently been used synonymously with general ARIMA models applied to time series analysis and forecasting. Box and Jenkins (1970) effectively put together, in a comprehensive manner, the relevant information required to understand and use univariate time series ARIMA model. The initial values can also be forecast backward in time. These forecasts are obtained using ARIMA time series model or regression model with ARIMA errors.

Considering the above facts in view, this study focused on the Application of ARIMA models for mean monthly stream flow forecasting of Savitrii River with the following specific objectives.

1. Time series analysis of stream flow and rainfall data.

2. Forecasting of mean stream flow at different time steps.

Monthly streamflow data of Savitri River for the year 1993 to 2008 were collected. The monthly streamflow data used for statistical analysis and find out the stochastic models.

ARIMA model of first order were tried for modelling of monthly streamflow of Savitri River. The Autocorrelation function (ACF) and Partial autocorrelation function (PACF) were obtained. The parameters of the selected models were obtained with the help of maximum likelihood method. Standard error of model, AIC, BIC, RMSE were calculated.

Applying the low AIC value Criteria, the best fit model were choose (1,1,1) (1,1,1). With that model the comparison test were done.

The following specific conclusions are drawn from this study.

1. ACF and PACF achieved stationarity after first difference.
2. ARIMA $(1, 1, 1)\times(1, 1, 1)_{12}$ is found to be the most suitable models for mean monthly stream flow forecasting of Savitri river.
3. ARIMA $(1, 1, 1)\times(1,1,1)_{12}$ tends to overestimate the stream flow magnitude of Savitri river.

VI. BIBLIOGRAPHY

Achela, D., Fernando, A. and Jayawardena, A.W. (1994). Generation and forecasting of monsoon rainfall data. 20th WEDC Conference, Colombo, Sri Lanka.

Amaha and Sharma (2011). Worked on modelling and forecasting of rainfall data of Mekele for Tigray region (Ethiopia) using Univariate Box-Ljung methodology.

AIC (2006). From the Chairman's desk. Agriculture Insurance Company of India Ltd. http://aicofindia.nic.in/cmd2.html (accessed on July 21, 2006).

Box, G.E.P., Reinsel, G.C. and Jenkins, G.M. (1994). Time Series Analysis: Forecasting and Control, Prentice Hall, Englewood Cliffs, NJ, 3rd edition, 1994.

Brath *et.al.* (2002). Studied neural networks (ANN) and non-parametric methods for improving real time flood forecasting through connectional hydrology model. Along with traditional linear stochastic models, they used both stationary (ARMA) and nonstationary (ARIMA) models.

Dimitris, M. and Pantazis, E. (2001). Seasonal ARIMA inflow models for reservoir sizing. Journal of the American Water Resources Association, 37(4): 885-877.

Fernandez, F.J., Seco, A., Ferrer, J. and Rodrigo, M.A. (2009). Use of neurofuzzy networks to improve wastewater flow-rate forecasting. Journal of Environmental Modelling & Software, 24: 686–693

Jovanovski, V. and Delipetrov, T. (2007). Auto-regressive integrated moving average (ARIMA) modeling of rainfall process, estimation and forecast. Geophysical Research Abstracts, 9: 21-34.

Karamouz, M. (2004). Seasonal streamflow forecasting using snow budget and El Nino-southern oscillation climate signals: application to the salt river basin in arizona. Journal of Hydrologic Engineering, 9(6): 523-533.

Kember, G., Flower, A.C. and Holubeshen, J. (1993). Forecasting river flow using nonlinear dynamics. Journal of Stochastic Hydrology & Hydraulics, 7: 205-212

Krstanovic, P.F. and Singh, V.P. (2005). A univariate model for long-term streamflow forecasting. Journal of Stochastic Environmental Research and Risk Assessment, 5(3): 189-205.

Liang, L. and Kershaw, G.P. (1995). Climate change in the Mackenzie mountains. Journal of Climate Research, 5(1):1-13.

Lin, G.F. and Chen, L.H. (2005). Time series forecasting by combining the radial basis function network and the self-organizing map. Hydrological Processes, 19: 1925-1937.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. Journal of Biometrika, 65: 297–304.

Ljung, G.M. and Box, G.E.P. (1979). The likelihood function of stationary autoregressive-moving average models. Journal of Biometrika, 66: 265–270.

Modarres, E. and Eslamian (2006) Streamflow time series modelling of Zayandehrud River. ARIMA model are appropriate for monthly streamflow of the Zayandehrud river in western Isfahan province, Iran.

Mohammadi, K., Eslami, H.R. and Dardashti, D. (2005). Comparison of regression, ARIMA and ANN models for reservoir inflow forecasting using snowmelt equivalent (a case study of Karaj). Journal of Agricultural Science Technology, 7: 17-30.

Montanari, A. (2000). A seasonal fractional ARIMA model applied to the Nile river monthly flows at Aswan. Water Resource Research, 36(5): 1249-1259.

Papamichail, D.M. and Georgiou, P.E. (2001). Seasonal ARIMA inflow models for reservoir sizing. Journal of American Water Resources Association, 37(4): 877-885.

Planning Commission. (2002). Tenth Five Year Plan, Vol. II, Sectoral Policies And Programmes, Section VIII, Infrastructure, Chapter 8.1, Irrigation, Flood Control And Command Area Development. Govt. of India, New Delhi.

Rasmussen, P.F., Salas, J. D., Fagherazzi, L., Rassam, J. C. and Bobee, B. (1996). Estimation and validation of contemporaneous PARMA models for stream flow simulation. Water Resource Management, 32(10): 3151-3160.

Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chowc, J.C., Watson, J.G. and Moncada, J.A. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. Journal of Atmospheric Environment, 42: 8331–8340.

Smith, J.A., Day, G.N. and Kane, M.D. (1992). Nonparametric framework for long-range stream flow forecasting. Journal of Water Resource planning and Management, 118(1): 82-92.

Soils, J.F., Pita, E. and Lagunas, J. (2008). Short-term stream flow forecasting: ARIMA vs neural networks. Mathematics and Computers In Science and Engineering Proceedings of the American Conference on Applied Mathematics, Cambridge, Massachusetts, pp. 402-407, ISBN ~ ISSN:1790-5117 , 978-960-6766-47-3.

Svetlikova, D., Komornikova, M., Kohnova, S., Szolgay, J. and Hlavcova, K. (2007). Analysis of discharge and rainfall time series in the region of the Klastorske Luky wetland in Slovakia. XXIVth Conference of the Danubian Countries on the Hydrological Forecasting and Hydrological Bases of Water Management Bled, Slovenia 2 – 4 June, 2008.

Tong, H. (1990). Non-Linear Time Series: A Dynamical System Approach. Oxford University Press, UK.

Toth, E. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology, 239: 132-147.

Wang, W., Pieter, H.A.J.M., Gelder, V. and Vrijling, J.K. (2005). Constructing prediction intervals for monthly streamflow forecasts. ISSH - Stochastic Hydraulics 2005, 23-24 May, 2005, Nijmegen, the Netherlands.

Weesakul, U. (2005). Rainfall forecast for agricultural water allocation planning in Thailand. Journal of Science & Technology, 10(3):18-27.

Yurekli, K. and Kurunc, A. (1997). Performances of stochastic approaches in generating low streamflow data for drought analysis. Journal of Spatial Hydrology, 5(1): 245-267.

Yurekli, K., Kurunc, A. and Ozturk, F. (2005). Application of linear stochastic models to monthly flow data of Kelkit stream. Journal of Ecological Modelling, 183: 67-75.

Yurekli, K., Kurunc, A. and Ozturk, F. (2005). Testing the residuals of an ARIMA model on the creeker stream watershed in Turkey. Turkish Journal of Engineering and Environmental Sciences, 29: 61-74.

Yurekli, K. and Kurunc, A. (2006). Performances of stochastic approaches in generating low streamflow data for drought analysis. Journal of Spatial Hydrology, 5:21.32.

Zuzel, J.F., Robertson, D.L. and Rawls, W.J. (1975). Optimizing long term streamflow forecast. Journal of Soil and Water Conservation, 30(2):76-78.