FORECASTING OF STREAMFLOW USING TIME SERIES MODELLING

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY

IN

HYDRAULICS AND WATER RESOURCES ENGINEERING

Submitted by

GAURAV KUMAR

(2K18/HFE/21)

Under the supervision of

Prof. Vijay K. Minocha

DEPARTMENT OF CIVIL ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042
JULY, 2020

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, **GAURAV KUMAR**, Roll No. **2k18/HFE/21** of **M.Tech (HWRE),** hereby declare that the project Dissertation titled "**Forecasting of Streamflow using Time Series Modelling**" which is submitted by me to the department Hydraulics and Water Resource Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technological, is original and not copied from any source without proper citation. This work has not previously formed the basis for any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

Date: **(Gaurav Kumar)**

DEPARTMENT OF CIVIL ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the project Dissertation titled "**Forecasting of Streamflow using Time Series Modelling**" which is submitted by **GAURAV KUMAR**, Roll number **2K18/HFE/21** of M.Tech **(HWRE),** Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is record of the project carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or fully for any Degree or Diploma To this University or elsewhere.

Place: Delhi **(Prof. VIJAY K. MINOCHA)**

Date: **(SUPERVISOR)**

ACKNOWLEDGEMENT

I take this opportunity to express my sincere regards and deep gratitude to **Prof. Vijay K. Minocha** (professor, civil engineering department, DTU) for his consistent guidance, monitoring and constant encouragement throughout the course of this project work. Also, I express my gratitude to **Department of Civil Engineering, DTU** for giving me such an opportunity to commence this project in the first instance.

Professors and faculties of the civil engineering, DTU have been very supportive and cooperative. They have been always present for their kind opinions and suggestions regarding this project work and therefore I am deeply obliged to them.

Last but not the least, I would like to thank my family and my colleagues from the department who encouraged me to bring work to a successful close.

Place: Delhi

Date: (GAURAV KUMAR)

ABSTRACT

For the proper management of any hydrological or water resources projects, the primary key is the early availability of the data associated with the project. One of the crucial tools regarding the same is an approach based on time series analysis. Time series analysis for the forecast of the monthly streamflow has vital importance in water resources engineering and act as a fundamental part in planning, designing and management of water resources systems. In this study, autoregressive integrated moving average (ARIMA) model has been used for forecasting the monthly discharge of the Sarda River at Banbassa, Uttarkhand, India. ARIMA model improves the performance of advance information for making planning and maintenance of the available water resources. The behaviour of the streamflow under different level of demand has been analyzed based on autoregressive integrated moving average (ARIMA) model, and it was found that the used model has great efficiency for the fitting and prediction.

In order to implement the model application, a 32 years span of streamflow data from 1976 to 2007 has been used. The First 30 year's data have been used for developing and trending a statistics related ARIMA model and the last two years streamflow data have been used for the validation of the generated model. The working procedure of the ARIMA model is based on the combine operation with various AR and MA orders. The developed model has been selected based on the t-value and the residual of the autocorrelation function (ACF) and partial autocorrelation function (PACF). In this study, the statistical analysis for a developed model has been made with the help of IBM SPSS version 21. The prediction accuracy of various developed models has been examined by comparing their mean absolute percentage error (MAPE), and the coefficient of determination (R^2) values and hence selected the best iterative model based on the above comparison. Furthermore, the selected iterative model has been used to forecast the stream flow up to 3 steps ahead in terms of MAPE. According to, above analysis, the generated model has been found the best solutions for the proper predictions and forecasting for the future usage of the streamflow resources and management.

Table of Contents

List of figures

List of Tables

1.1 General

Streamflow forecasting is a crucial aspect for planning, designing and analysis of future events. It helps to give timely flood warnings, and advance information for making planning, maintenance of the available water resources.

There are two types of forecasting models physical model and stochastic model. Most of the physical model based on the theoretical or empirical equations and it shows a unique coincidence between input and output variables, while stochastic models are based on time series modelling and mostly used for analyzing the river runoff variations. In this study a time series based stochastic model has been used for the forecasting.

A Time series is a sequential set of data points equally spaced and measured typically over successive times. Time series analysis is an essential part of statistics which analyzes data set to study the characteristics of the data and helps in predicting future values of the series based on the observed series. It is mathematically defined as a set of vectors X(t) = 0, 1, 2, 3…….where t represents the time elapsed. The variable X(t) is treated as a stochastic variable. At first the arrangement of the measurement obtained during an event is carried out in the chronological order. Time series modelling is a dynamic research area which has attracted the attention of the researcher's community over the last few decades. The time series approach is used for the carefully collected data and studying the past observations rigorously for the purpose of developing an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past. An auspicious time series forecasting depends on the development of an appropriate robust model. One of the most popular and frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model, based on Box-Jenkins methodology and for seasonal time series forecasting, an enhanced variation of ARIMA model, viz. the Seasonal ARIMA (SARIMA) is used (McLeod and Hipel, 1994). The popularity of the ARIMA model is mainly due to its flexibility to represent several varieties of time series with simplicity as well as the associated Box-Jenkins methodology for the optimal model building process but, the limitation of these models is the pre-assumed linear form of the associated time series, which becomes inadequate in some situations.

1.2 Importance of forecasting

Forecasting plays a vital role in various fields of concern

- It helps to provide advanced information about planning, designing and analysis of future events.
- It is very useful in mountain areas because most of the downstream populations have highly depended upon their livelihood and commercial activities.
- Forecasting plays a crucial role to estimate the future water events.
- Forecasting helps to guide researchers to take advantage of future opportunities.
- It also useful in preparing the contingency and emergency plans.

1.3 Objective of the Thesis

The main aim of this study is to develop a stochastic forecasting model of streamflow based on time series. The specific objectives of this study are:

1. To develop a time series simulation model for streamflow forecasting.
2. To validate the developed simulation model.

1.4 Organization of Thesis

The thesis consist of six chapters detailing about the project and execution of work as well as format of the work, these six topics are further divided into different sub-topics,

Chapter 1 provides the introductory section that gives a brief description of the time series modelling by Box-Jenkins methodology and forecasting methods. This provides the background and terminology necessary for presenting the subsequent of time series modelling. The general and objective of this thesis are also described in this chapter.

Chapter 2 review of relevant literature dealing with the physical and stochastic model is presented. The review chapter represent the forecasting results carried out by many researchers and shows the improvement and standardized terminology. From the literature it is clear that various forecasting models has been used such as ARIMA, SARIMA, Artificial neural network and many more and it classify time series as well as methods of predicting the future values.

Chapter 3 details of the data collection and selection of study area are presented. The Sarda River in Banbassa was selected as the project to implement in this research. It is important for water supply for many Uttarkhand states and the streamflow forecasting of the Sharda River is important for both fulfilling needs and agriculture developments as well as for the purpose of generation of hydroelectric power.

Chapter 4 evaluates the methodology of time series analysis. This chapter includes the use of the SPSS software used for the forecasting purpose. The formation of the ARIMA model using the time series analysis has also been carried out in this chapter.

Chapter 5 describes the model selection based on error estimation. Different ARIMA models were constructed for different regimes in which time series operate and the results were obtained when ARIMA models is applied on streamflow data and then used to make forecasting. The forecast value are compared to the observed value and the different forecast error are reported and this study Determine the best model for forecasting.

Chapter 6 contains summary and conclusions of the thesis. In this study, the stochastic nature of streamflow is analyzed with autoregressive integrated moving average (ARIMA) Stochastic models. The best ARIMA model were estimated using the SPSS software and it makes time series stationary, in both stages trending and validation stage and ARIMA model improved the performance of advance information.

2.1 General

Time series analysis plays an important role in analysing and obtaining better predicted result in water resources engineering. The use of models depending on such approach is not a new concept and has been used by many researchers for the purpose of proper predicted result. In the present chapter the use of the time series model in prediction of various parameters carried out by various scholars has been shown. From the literature it is clear that various forecasting models has been used such as ARIMA, SARIMA, Artificial neural network and many more.

2.2 Time-series Modelling in Past

Various researchers and engineers learning from the past work done in the field of forecasting using the time series approach have conducted various studies and experiment in which they made more regressive effort to make the forecast much near to the accurate. Some of the work carried out by the researchers has been listed below:

Sun and Koch (2001) studied the cross-correlation of autoregressive integrated moving average ARIMA and dynamic regression transfer model to analyse the time series data for forecasting of salinity variations for Apalachicola Bay. They showed that the rational distributed lag transfer functions between the hourly variation of the tidal water levels and salinity can be used to forecast the short-term fluctuation in the salinity. They also concluded that certain important control variables can be highlighted by performing a multivariate similarity evaluation of daily salinity with river discharge. They also concluded that the fluctuation in the tidal water levels results in only short-term periodic variation in salinity. The cross-correlation done by they showed that despite the Apalachicola River being a major fresh water source it strongly affects the current and salinity in the bay for a long term.

Karamouz and Zahraie (2004) conducted a study to propose a method to improve long term statistical streamflow forecast. They conducted a case study for Salt River Basin. They describe the method in three steps. In their first step they use the relationship between the average snow water equivalent to define the hydrologic seasons to study the combined effect of climatic and hydrological characteristics. They then using the autoregressive integrated moving average (ARIMA) developed a seasonal streamflow time series in the next step. Following this, certain

Fuzzy rules are developed so as to modify the statistical forecast, utilizing average snow budget over a watershed and time series of forecasted streamflows.

Mingrong and Hengxin (2008) used a seasonal autoregressive integrated moving average model seasonal (ARIMA) to forecast the seasonal highway traffic volume. They later compared the forecasted result obtained by the seasonal ARIMA model with three seasonal forecasting model that is regression model, variable seasonal index forecasting model and seasonal regression model and it gives the better performance and accurate result.

Landeras *et. al* (2009) compared the result of weekly forecasted evapotranspiration obtained from ARIMA model and Artificial neural network (ANN) model with those obtained from the weekly averages. They described the ARIMA and the ANN model and generated a weekly evapotranspiration time series for a period of 1975-2003, which the further used for the implementation purpose for these models. The comparison result showed that the ARIMA and ANN model so developed resulted in less root mean square difference (RSMD) by 6-8% and also the standard deviation was also reduced by 9-16% in comparison to the result obtained by average model. The also concluded that the performance of the prediction model was depending on the pattern of the weekly evapotranspiration.

Mohan and Arumugam (2009) used the Autoregressive integrated moving average (ARIMA) model approach and winter's exponential smoothing model approach to predict the Evapotranspiration and then compared the result obtained. They collected daily meteorological data for the years 1977-1992 in which they included data of maximum and minimum temperature, wind speed, maximum and minimum relative humidity and Vapour pressure. They referred the reference crop evapotranspiration calculated by the Penman and Pruitt method as the Evapotranspiration ET. They used the first 14 years' time series data for the development of the model and the following two years data for investigating the accuracy of the developed model. Both the model developed to forecast was found suitable with less errors.

Abudu *et. al* (2011) using the mixture of stochastic TFN model and ANN technique forecasting the monthly streamflow Runoff season in Rio Grande Headwaters basin. They for the one month ahead forecast result, compared it with TFN and ANN models that were adjust especially for every month of the runoff season. They concluded that in comparison to a single technique used i.e. either TFN or ANN, the TFN+ANN technique provide much accurate result with high coefficient of determination.

Alnaa and Ahiakpor (2011) used the autoregressive integrated moving average approach to predict the inflation in Ghana. The study was conducted on the monthly time series data from 2000 to 2010. From the data available last eight values were remained fixed for the validation purpose. They used consumer price index as the variable. They then compared the eight forecast values form the models developed to the eight actual observation available. And so they concluded that the ARIMA model can be applied to predict the inflation.

Mondal *et.al* (2014) studied the effectiveness of the time series autoregressive integrated moving average model (ARIMA) in prediction of the stock prices. They applied the model on time series data set of stocks of fifty six companies. They used the Akaike Information Criteria (AIC) as a measure for the statistical model. They validated the predicted value with the data obtained from the same source. The overall predicted results obtained by the model developed showed an accuracy of about 85% which indicated that the model developed by them can be applied suitably for the stock prediction.

Manoj and Madhu applied the time series autoregressive integrated moving average model (ARIMA) to forecast the production of sugarcane in India. To develop the model they used the time series of the data available of sugarcane production of a span of 62 years from 1950 to 2012. They conclude that the successive errors in the model were normally distributed, with mean zero, and so their selected model can be applied to forecast the production.

Shathir and Saleh (2016) conducted a study to forecast the inflow of discharge to Hit station on Euphrates river, Iraq. They developed seven different model by time series autoregressive integrated moving average approach and then tested them for forecasting the discharge. They carried out statistical analysis of the model so developed with the use of IBM SPSS statistics 21 software. They obtained the actual data of the inflow from October 1932 to September 1972 and used the time series approach on same. For detecting any change in mean of any two sample they used the T-test approach and for determining difference in variance they used the F-test approach. And by comparison of their result they concluded that the ARIMA model can be used for forecasting the inflow to the station.

Oliveira and Boccelli (2017) used various approach to forecast the water demand and compared the forecasted results with each other to obtain an optimum model. Initially they used the k-nearest neighbour approach which was applied to utilize the data in hourly demand time series for a five week set. They also evaluated the performance of both the model developed by k-nearest neighbour (KNN) and autoregressive integrated moving average (ARIMA) by creating an experimental design. They concluded that the forecast result by the seasonal autoregressive time series model (SAR) despite with the linearity resulted in much better agreement with the actual one. However the forecast resulted by the KNN model was not so much satisfactory and showed the large prediction error comparatively with that of SAR result.

Ghimire (2017) applied the autoregressive integrated moving average model (ARIMA) to forecast and study the result over Schuylkill River. They used the developed ARIMA model on the time series of know six year data and validated the result with the last one year available data. With their study they concluded that the model developed by them showed a much accurate and agreed result.

Deen Dayal *et.al* (2019) conducted a study to develop an Autoregressive Integrated Moving Average (ARIMA) model which is further used to predict the monthly rainfall over Betwa River Basin. The model was developed using the past available data of rainfall of 52 year span from 1960 to 2012. For framing the most suitable model, the precipitation time series was done for the training purpose with data available for a span from 1960-2000 and the data for rest of the period i.e from 2000-2012 was used for validation purpose of the predicted model so developed. To check the model efficiency using the parameters coefficient of determination (R^2). The ARIMA model developed by them possessed a good efficiency and the forecasted rainfall was highly accurate to the known value of the rainfall.

2.3 Conclusion

From the literature as provided in the chapter we find that the use of time series analysis in the model formation for the purpose of forecasting is not a new approach. Various research whether hydrological or non-hydrological for the prediction of data has been carried out. Also from the literature it is clear that a lot of stochastic time series model have been developed for the forecasting in water resources engineering. We found that the frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model, based on Box-Jenkins methodology. Box-Jenkins methodology have many benefit over other methodology for analysis of time series variables. The results of this study are applicable for streamflow forecasting and water resources managers to schedule optimum operation of river to control floods and for both fulfilling needs and agriculture developments as well as for the purpose of generation of hydroelectric power.

3.1 General

This chapter deals with the description of the study area from which the data required for the time series has been discussed. The study which is Sarda River at Banbassa in Uttarakhand, India has been selected and the required data has been discussed in the following section. In this chapter data collection with topography information and discharge of river are described and the format has been discussed in the following section.

3.2 Watershed Description

The work presented in this thesis constitutes a contribution to modeling and forecast the demand in agriculture and generation of hydroelectric power**.** This work demonstrates how the historical data could be utilized to forecast future demand and these affect the supply**.** The Sarda River in Banbassa was selected as the project to implement in this research. The Sarda River originates at kalapani in the Himalayas at an elevation of 3600m and the basin area of 14871 km^2. It flow along Nepal western border also and joins Ghanghra River, a tributary of the Ganges. It is important for water supply for many Uttarkhand states and the streamflow forecasting of the Sharda River is important for both fulfilling needs and agriculture developments. The river flow through major district in Nepal before joining the Indian administration. Kumaon region and Pithoragarh region are the major basin of the river for the purpose of study. The river is not joined by any tributaries in its course. The water from the river is diverted form the barrage into Right Sharda Canal for the purpose of irrigation as well as for the purpose of generation of hydroelectric power. At certain point downstream of the barrage and another barrage namely lower sharda barrage is located.

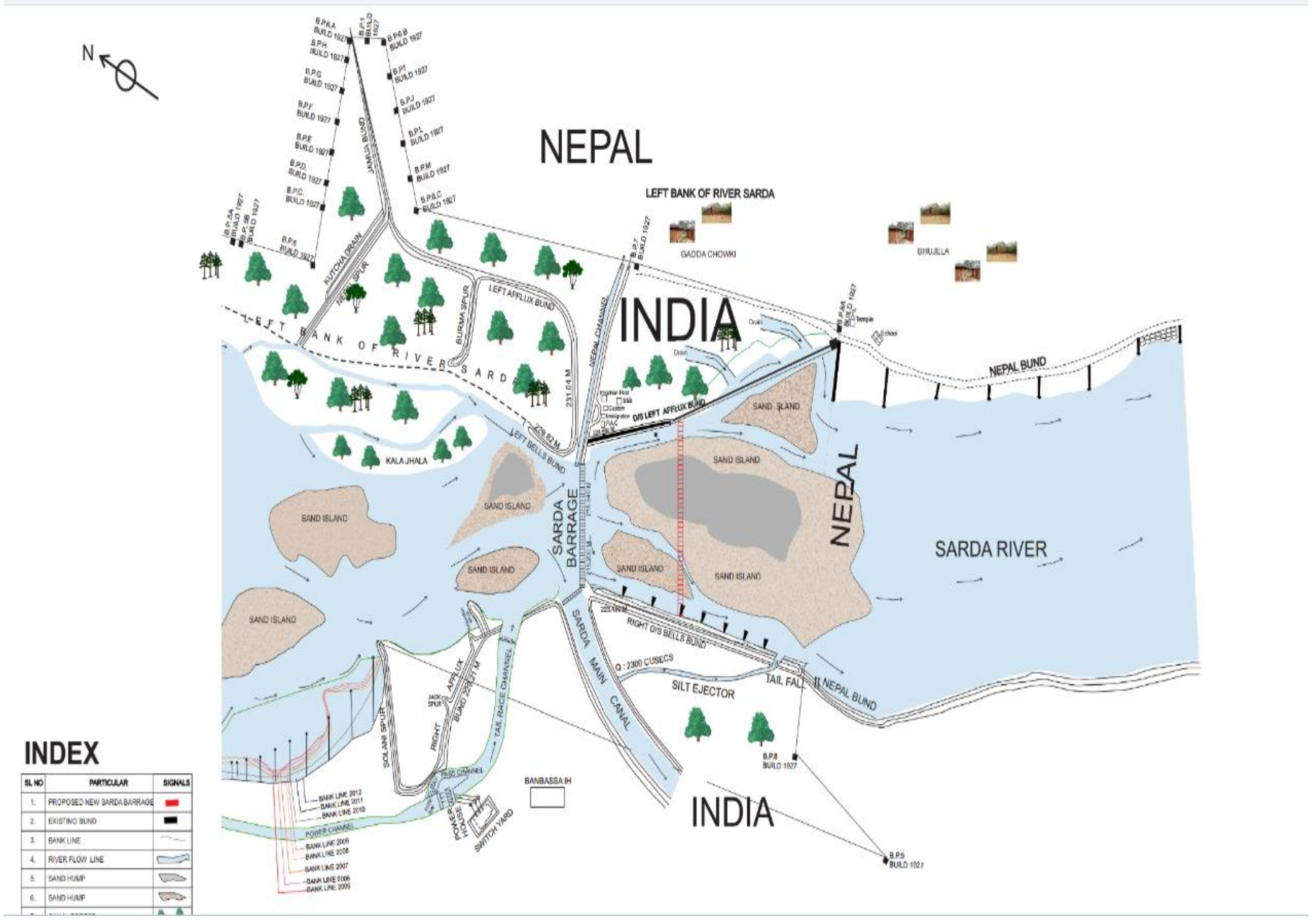

Figure-1 Sarda River

3.3 Data Collection

For the purpose of the study the monthly time series data of discharge (in cusec) is collected for a span of 32 years from 1976 to 2007. The data for the Sarda River at Banbassa were obtained from the Irrigation Department of Uttar Pradesh. The data to be used in preparing the model is taken from the period of 1976-2005(30 years) and the remaining data for the period of 2 years from 2006 and 2007 has been used to validate the model and check the accuracy of the model so prepared.

4.1 General

A Time series is a chronological set of data points corresponding spaced and estimates over successive times. Time series analysis is a vital part of statistics which examine data to study the feature of the data and helps in predicting future values of the series. The measurements taken during an event in a time series modelling are in a sequential order.

4.2 Software used and Data handling

SPSS standing for statistical product and service solution is one of the most widely used computer application software for the statistical analysis in the field of science. The software in addition to the field of science plays an important role in market researchers, survey companies, government, education and others. The software has been extensively used in analyzing the correlation between ACF and PACF for the past and in addition to this various model parameters has been estimated using the SPSS software.

4.3 Time Series Analysis

A set of values that has to be used in the estimation of a variable using a suitable model approach is after suitably fitting it in the model is done by time series analysis. Also the process of a suitable time series models to a correct models is known as Time Series Analysis. It shows that methods attempt to grasp the nature of a series and it is helpful for future forecasting. In forecasting done by the time series modelling approach previous data available are collected and the analyzed and as a result a mathematical model is developed through which the process of further data generation process is carried out. The future events then predicated using the model. The above approach is beneficial when there is a lack of knowledge regarding the statistical pattern and successive observation. Time series forecasting play a vital importance in applications of different fields. Many valuable strategic decisions and estimates are taken based on the good forecast results. Thus making a good forecast, i.e fitting an adequate model to a time series is important.

4.4 Time Series and Stochastic Process

A time series is a non-deterministic in a nature, i.e. we unable to forecast with certainty what can be obtained in future. A time series x(t),t=0,1,2,3….) is to take follow certain probability model which make the probability structure of the time series analysis is called stochastic process.

A constant assumption is that the time series variables X_t are independent and identically distributed following the normal distribution.

4.5 Time Series Forecasting Using Stochastic Models

The most famous techniques used to forecast the time series is Box-Jenkins Methodology, which is based on examining a wide range of models for forecasting a time series and many works have been done by scholar from many years for the development of effective models to give better performance of forecasting accuracy. The outcomes, varied important time series forecasting models have been shown in literature. The most frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) Model, based on Box and Jenkins Methodology. The basic limitation of this model is that the considered time series is linear. ARIMA model has subclasses of the other models, such as the Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) Models.

For seasonal time series forecasting, increase changes of ARIMA model, that is Seasonal ARIMA (SARIMA) is used (McLeod and Hipel, 1994). It accept the ARIMA Model due to flexibility to represent different varieties of time series with associated Box-Jenkins Methodology for optimal model building process. But the limitations of these models is the pre-assumed linear form of the associated time series which becomes inadequate in some conditions.

4.6 The Autoregressive (AR) Model

Autoregressive models are based on the current values of the series, X_t, can be explain as a linear current combination of p past values, X_{t-1}, X_{t-2}…X_{t-p}, together with a random error in the same series.

An autoregressive model of order p, abbreviated AR(p), is of the form

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots \ldots + \Phi_p X_{t-p} + w_t = \sum_{i=1}^{p} \Phi_1 X_{t-i} + w_t$$

Where X_t is stationary, Φ_1, Φ_2 are model parameters.

4.7 The Autoregressive AR (p) Model

The AR(p) process

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots \ldots + \Phi_p X_{t-p} + w_t$$

The autoregressive operator is defined as:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots \ldots .. - \phi_p B^p = 1 - \sum_{j=1}^{p} \phi_j B^j$$

then the AR(p) can be written as:

$$\phi(B) X_t = w_t$$

4.8 The Moving Average (MA) Models

A moving average model of order q, or MA (q), is defined to be

$$X_t = w_t + \phi_1 w_{t-1} + \theta_2 w_{t-2} + \cdots \ldots + \theta_p w_{t-p} = w_t + \sum_{j=1}^{q} \theta_j w_{t-j}$$

Moving average operator

Equivalent to autoregressive operator define as

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots \ldots \ldots .. \theta_q B^q$$

Therefore the moving average model can be written as:

$$X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots \ldots \ldots . + \theta_q B^q) w_t$$

$$X_t = \theta(B) w_t$$

4.9 The Autoregressive moving Average (ARMA) Models

ARMA (p, q) model is the mix of AR (p) and MA (q) models and are appropriate for univariate time series modeling. In AR (p) model the coming value of a variable is consider to be a linear mixture of p past observations and random error together with constant term. AR (p) model is written as:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t = c + \varphi_1 X_{t-1} \varphi_2 X_{t-2} + \ldots \ldots \ldots \ldots \ldots \ldots \varphi_p X_{t-p} + \varepsilon_t$$

Where, X_t and ϵ_t are respectively the actual value and random error at time period t, φ_i (i=1, 2,3,………..p) are model parameters and c is a constant. The integer constant p is known as order of the model.

Autoregressive (AR) and Moving average (MA) models is a combination to form and helpful class of time series models, known as the ARMA models. Mathematically ARMA (p,q) model is

$$X_t = c + \epsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-1} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

Here the model orders p,q refer to p autoregressive and q moving average terms.

Usually ARMA models are manipulated using the lag operator notation. The backshift is defined as $LX_t = X_{t-1}$.

AR (p) model: $\varepsilon_t = \varphi(L) X_t$.

MA (q) model: $X_t = \theta(L)\varepsilon_t$

ARMA (p, q) model: $\varphi(L)X_t = \theta(L)\varepsilon_t$

Here $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i \; and \; \theta(L) = 1 + \sum_{j=1}^{q} \theta_j L_j$

4.10 The Autoregressive integrated moving Average (ARIMA) Models

The generalized form of the autoregressive moving average approach ARMA is what is called autoregressive integrated moving average ARIMA. For the purpose of forecasting if any of the model used, it has the same base of time series or for the purpose of better understanding. For certain cases where there are chances of occurrence of non-stationarity in the data, there ARIMA model is applied, also for such areas an approach of initial differencing can be applied one to more times to eliminate the non-stationarity.

The AR part of ARIMA shows that variables of interest is regressed on its own lagged values. The MA part shows that regression error is actually a linear combination of the errors terms whose values occurred contemporaneously and various times in the past. The I in the ARIMA shows that data values have been replaced with difference between their values and the previous values.

By Box-Jenkins methodology, the ARIMA model can be estimated by a three stage approach.

Procedures

1. Model identification stage: check stationarity and seasonality, performing differencing if requires.
2. Parameters estimation stage: computing coefficients that best fit the selected ARIMA model using maximum likelihood estimation.
3. Model checking stage: testing whether the obtained model conforms to the specifications of a stationary univariate process.

The general form of ARIMA (p,d,q) model can be written as

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots \ldots + \phi_p X_{t-p} + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}$$

Or the backshift notations,

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots \ldots . -\phi_p B^p\right) X_t = c + (1 - \theta_1 B - \theta_2 B^2 - \cdots . . -\theta_q \epsilon_{t-q}) \varepsilon^t$$

Where c=constant term, ϕ_1=i autoregressive parameters, θ_j=j moving average parameters ϵ_t=the error term at time t.

4.11 The Seasonal Autoregressive integrated moving Average (SARIMA) Models

The ARIMA model is for the non-seasonal non-stationary data. Box-Jenkins methodology shows that this model is deal with seasonality and also called as Seasonal ARIMA (SARIMA) model. In this method seasonal differencing of order is used to remove non-stationarity from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the previous year and is calculated as $Zt=Yt-Y_{t-s}$. for monthly time series s=12.

This model is termed as the SARIMA(p,d,q)x$(P,D,Q)^s$ model.

The mathematical formulation of a SARIMA (p,d,q)x$(P,D,Q)^s$ model in terms of lag polynomials is given as:

$$\Phi_p(L^s)\varphi_p(L)(1-L)^d(1-L^s)^D y_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t$$

$$\text{i.e.} \quad \Phi_p(L^s)\varphi_p(L)Z_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t$$

Here Z_t=is the seasonally differenced series.

4.12 Forecast performance measures

4.12.1 ACF and PACF

Autocorrelation and partial autocorrelation are estimation of association between current values and past values and shows that past values are more helpful in predicting future values.

Autocorrelation function (ACF): at lag K, Correlation between series values that are K intervals apart.

Partial autocorrelation function (PACF): at lag K, Correlation between series values that are K intervals apart, accounting for the values of interval between.

4.12.2 Coefficient of Correlation

A very important part of statistics is describing the relationship between one or two variables. If two variables are correlated, this means it can us information about one variable to predict the values of the other variable is given by

$$r = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}}$$

Where, r = is correlation coefficient, X_i= observed value and Y_i= forecasted value

$\bar{x}$ *and* $\bar{y}$ are corresponding means.

The value range between -1 to 1.

4.12.3 Mean absolute percentage error (MAPE)

The mean absolute percentage error is defined as MAPE=$\frac{\sum \left(\frac{|Actual - Forecast|}{Actual} \right) * 100}{N}$

N= Number of observations.

4.12.4 LJUNG-BOX

The LJung-box is a statistical test of a group of autocorrelations of a time series analysis are different from zero, it checks the overall randomness based on a number of lags, and it is portmanteau test.

$$Q(m) = n(n+2) \sum_{j=1}^{m} \frac{r^2}{n-j}$$

Where n=is the sample size, r=is the autocorrelation at lag j, m=number of lags being tested.

4.12.5 Bayesian information criterion

BIC is a criterion for model selection among a finite set of models. It is based in part on the likelihood function.

BIC=-2*LL+log(N)*K

Where N=is the number of examples in the training dataset, LL=is the log-likelihood of the model on the training dataset, and k=is the number of parameters in the model.

4.12.6 Forecasting equation formation

The general form of equation is

$$\Phi_p(B)\phi_p\nabla_s^D\nabla^d x_t = \theta_q(B)\Theta(B^s)a_t + c$$

Where,

p = Highest order AR parameter in the model.

d = Number of times the series is differenced.

q = Highest order MA parameter in the order.

P = Highest order seasonal AR parameter in the model.

D = Number of times the series is seasonally differenced.

Q = Highest order seasonal MA parameters in the model.

The notation ARIMA $(p,d,q)(P,D,Q)_S$ is used to donate the generalized seasonal.

4.12.7 t-value

t-value calculate the size of the many relative to variation in the sample data and greater the magnitude of t, greater the evidence against the null hypothesis.

It is computed as the ratio of the standard deviation of the sample to the mean of the sample, express in the percentage. Add up the values in dataset and divide the results by the number of values to get sample mean.

t-value greater than+2 or less than -2 is acceptable.

5.1 General

This chapter presents results from ARIMA modelling and forecasting of selected streamflow data sets. On the Streamflow data ARIMA models has be applied from the literature; the results carried out from the chosen models and for forecasting are consider residual of ACF, graphical, quantitative evaluation criteria like R^2, MAPE, BIC, LJUNG-BOX. These more estimate provide a more accuracy in evaluation of the forecasting performance of ARIMA models.

5.2 Methods of Analysis

Analyses was performed using the IBM SPSS Software; specifically the ARIMA procedure. For the data sets, an ARIMA model was developed using the Box-Jenkins methodology for identification, estimation, and diagnostic checks.

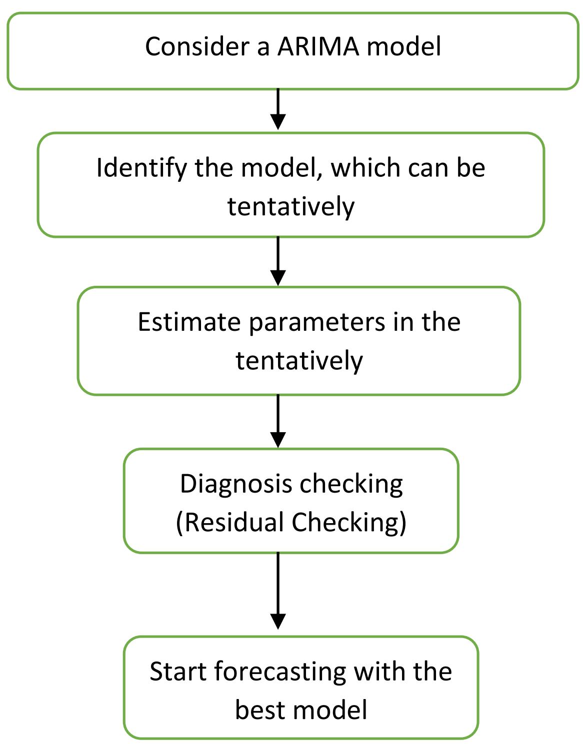

Figure-2: The Box-Jenkins methodology for model selection

In appropriate model selection is the determination of the optimal model parameters. Selection process is that the sample of Residual of ACF, R^2, MAPE, BIC and LJUNG-BOX. The forecasting results were plotted and then evaluated by examining the structure and magnitude of the errors.

5.3 Model identification

In model formation the discharge of 360 months has been used for making training model. On discharge data various ARIMA model has been tried and parameters and performance are discussed below. The tentative models have been chosen based on ACF and t-values of the estimated parameters. Monthly discharge data indicates seasonal variation having order of 12.

S.NO	Data used	Time period	Starting to Ending year	Number of Observation	Remarks
1	Monthly discharge	30 years	1976-2005	360	Trending data
2	Monthly discharge	2 years	2006-2007	24	Validation data

Table-1 Data sets used for the development of ARIMA model

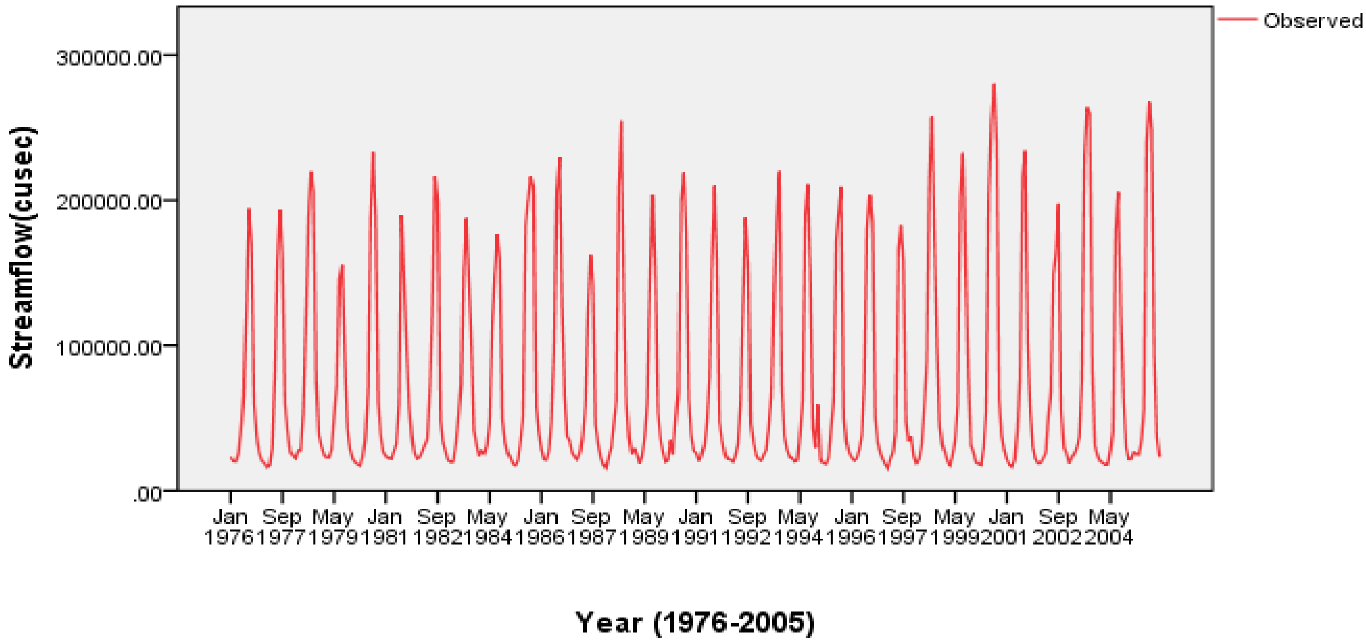

Figure-3. Original streamflow time series

After plotting the Residual ACF of raw data are examined in order to identify the structure of model. According to correlations plot residual ACF suggest that start model is AR(2) model.

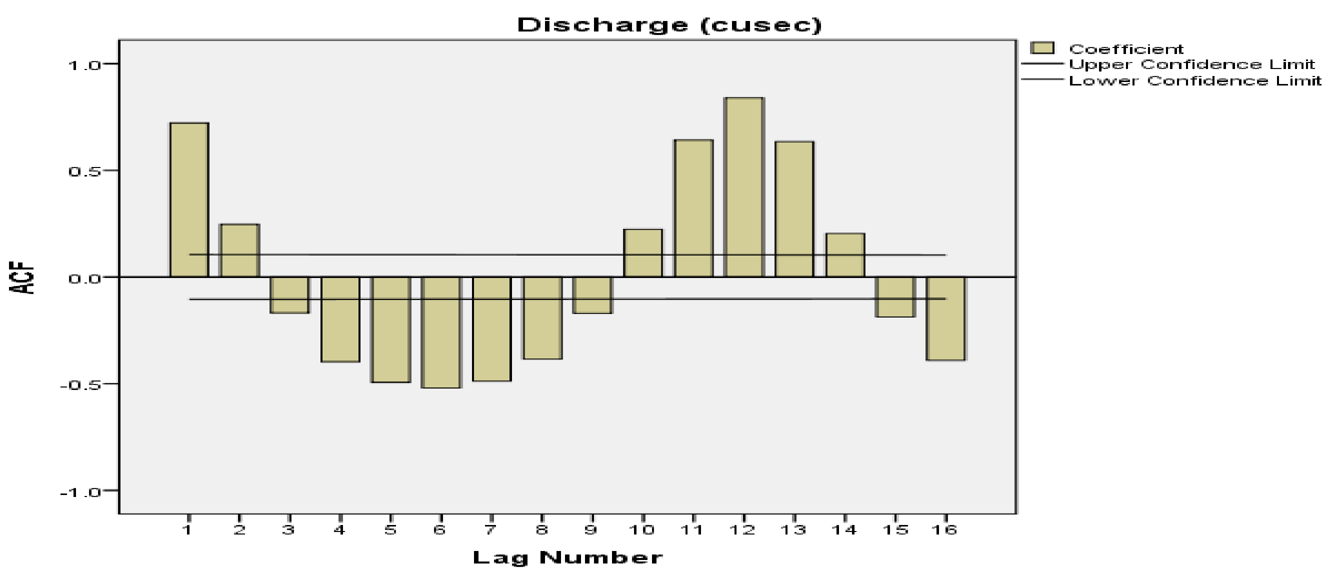

Figure-4: ACF of Raw data.

5.4 AR(2) Model

ARIMA(2,0,0)(0,0,0)$_{12}$ it is a starting model and first we plot the residual of ACF&PACF of this model and from the figure it is clear visible that residual of ACF and PACF of this model correlation values are not within the prescribed acceptable error band and the parameters of AR1 and AR2 both models are significant and this model is rejected due to poor parameters quality and residual variance. The value of R^2=0.588, LJUNG BOX=244.302, BIC=21.472, MAPE=44.669 and the value of AR1 =1.25 it means when value comes more than 1 the modelling is not good.

Next trial model is ARIMA(2,1,0)(0,0,0)$_{12}$.

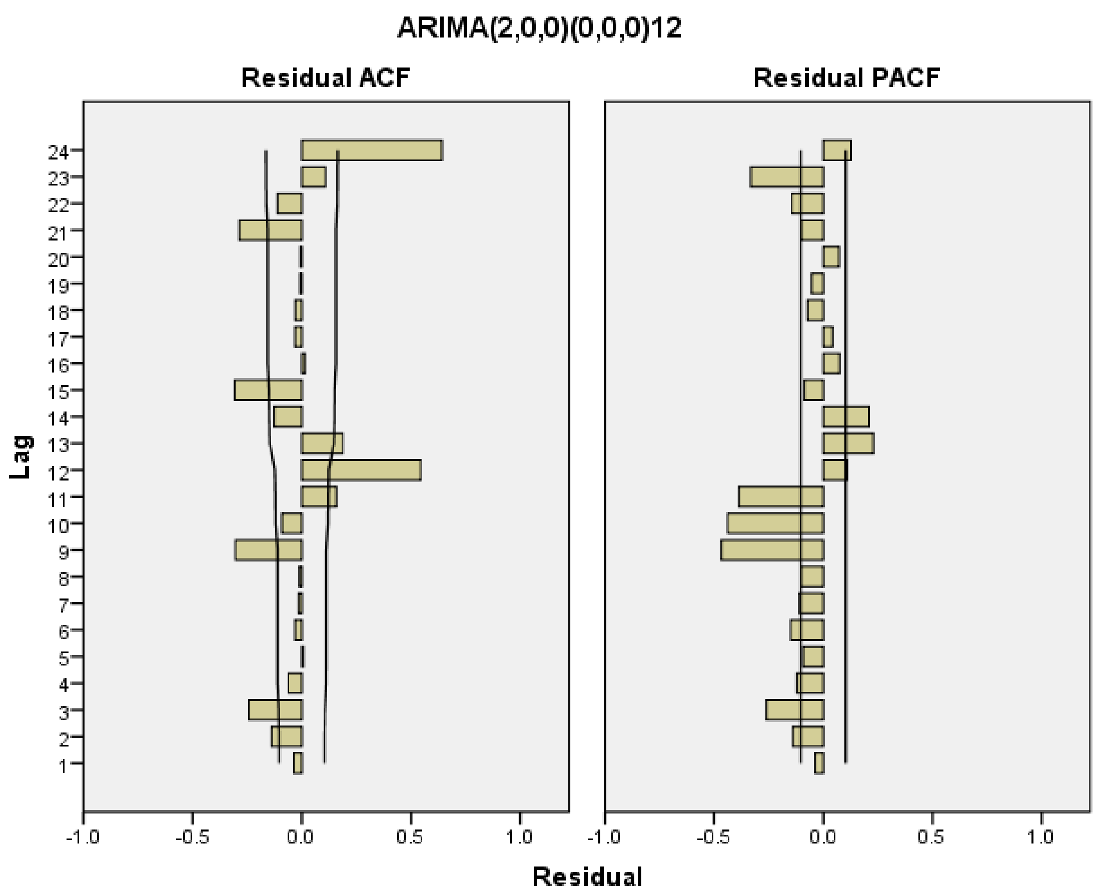

Figure-5: Residual of ACF and PACF of ARIMA(2,0,0)(0,0,0)$_{12}$.

ARIMA(2,1,0)(0,0,0)12 first we plot the Residual of ACF&PACF of this model and from the figure it is clear visible that ACF and PACF of this model are not within the prescribed acceptable error band and the parameters of AR1 and AR2 both models are significant and this model is rejected due to poor parameters quality and residual variance. The value of R^2=0.551, LJUNG BOX=237.630, BIC=21.561 and MAPE=47.447.

Next trial model is ARIMA(2,1,1)(0,0,0)12.

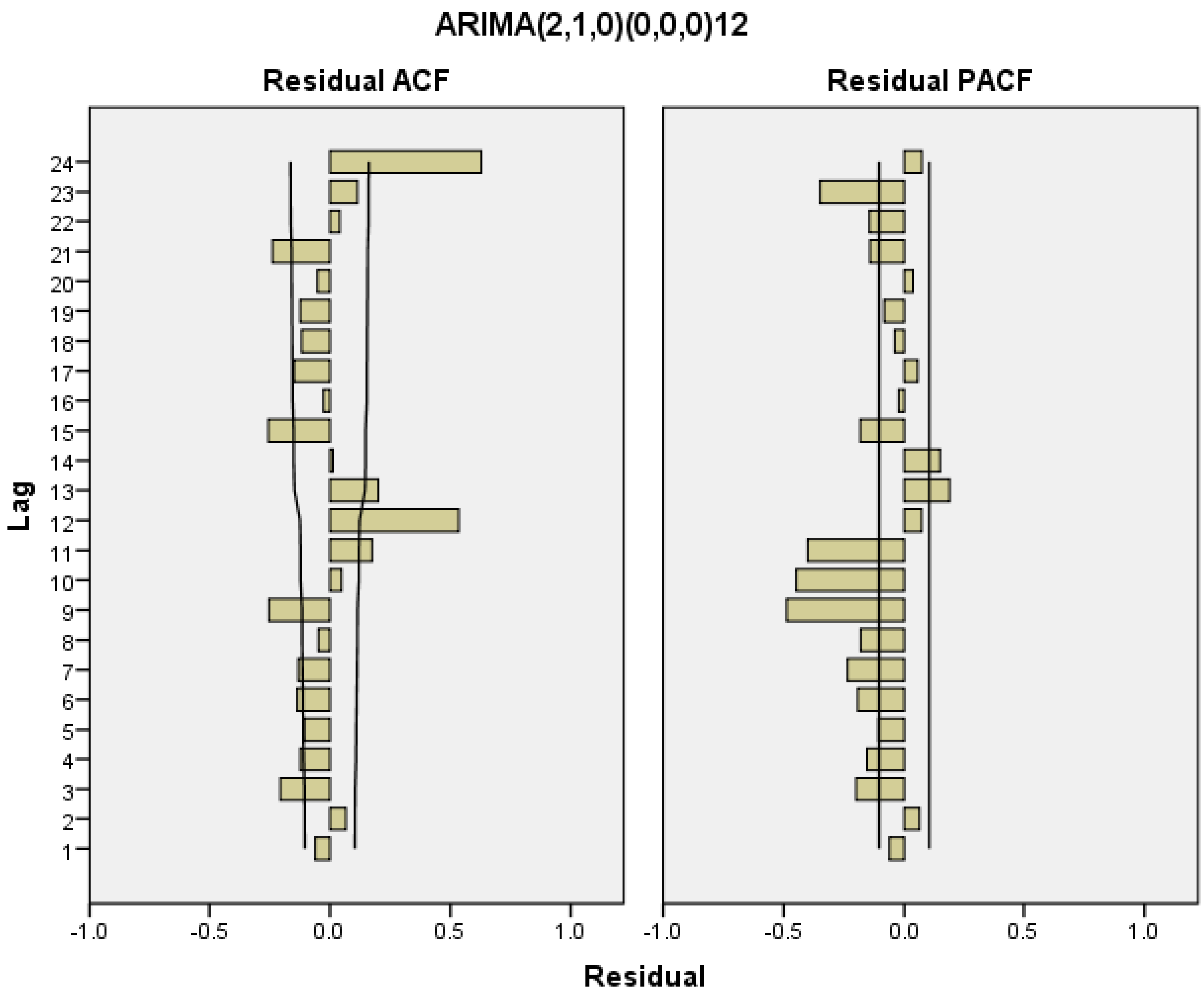

Figure-6: Residual of ACF and PACF of ARIMA(2,1,0)(0,0,0)12

ARIMA(2,1,1)$(0,0,0)_{12}$ and we plot the residual of ACF&PACF of this model and from the figure it is clear visible that ACF and PACF values are not within the prescribed acceptable error band. The value of R^2=0.677, LJUNG BOX=199.346, BIC=21.250, MAPE=57.340 and the value of AR1 =1.141 and MA=1 it means when value comes more than 1 the modelling is not good.

Next trial model is ARIMA(2,1,0)$(1,0,0)_{12}$.

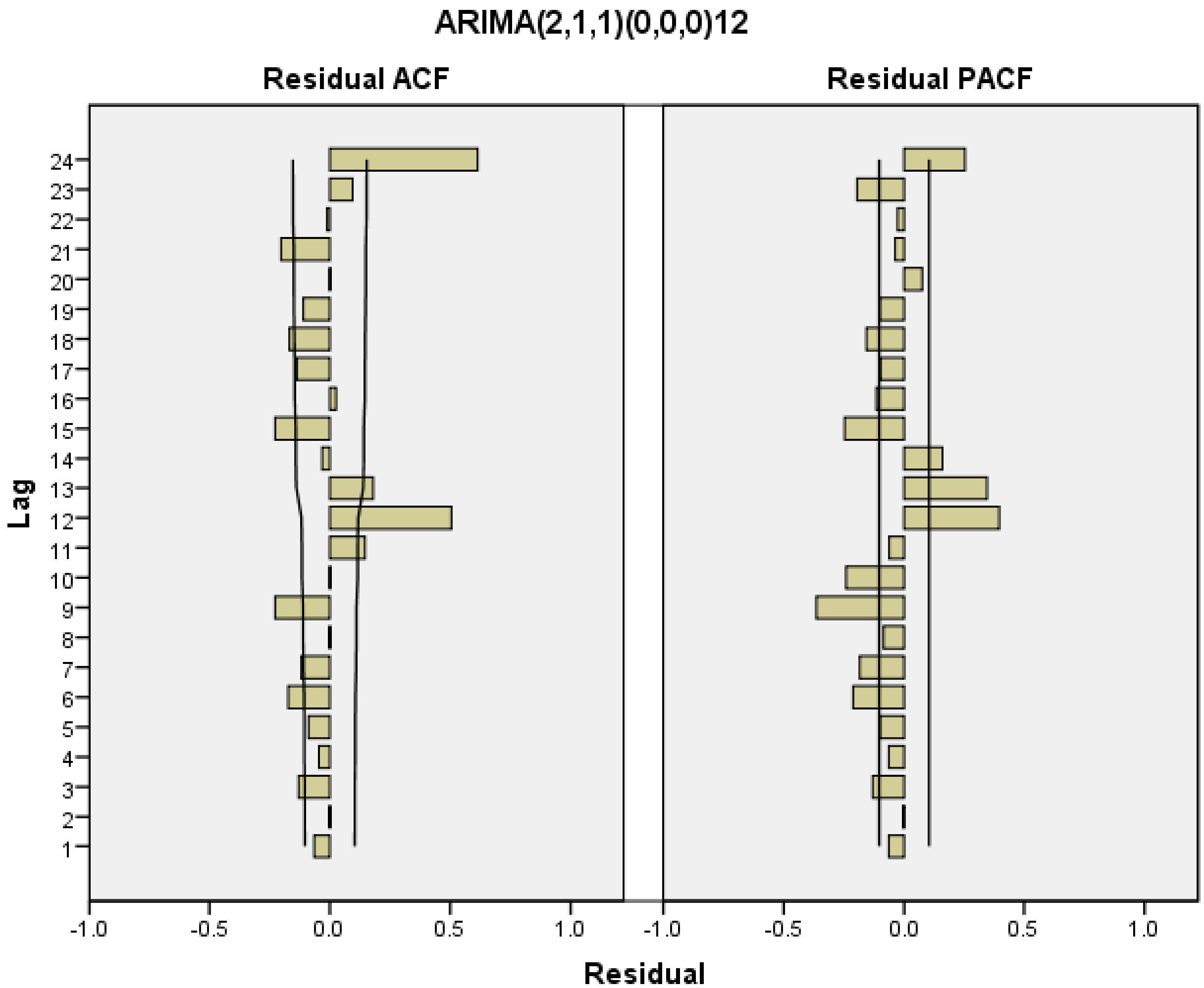

Figure-7: Residual of ACF and PACF of ARIMA(2,1,1)$(0,0,0)_{12}$

ARIMA(2,1,0)(1,0,0)12 and first we plot the Residual of ACF&PACF of this model and from the figure it is clear visible that at lag 4, 12 there is spike and the parameters of AR1 and AR2 both models are significant and this model is rejected due to poor parameters quality and residual variance. The value of R^2=0.815, LJUNG BOX=106.52, BIC=20.714, MAPE=27.074 and the value of MA=1 it means when value comes more than 1 the modelling is not good.

Next trial model is ARIMA(2,1,0)(1,1,0)12.

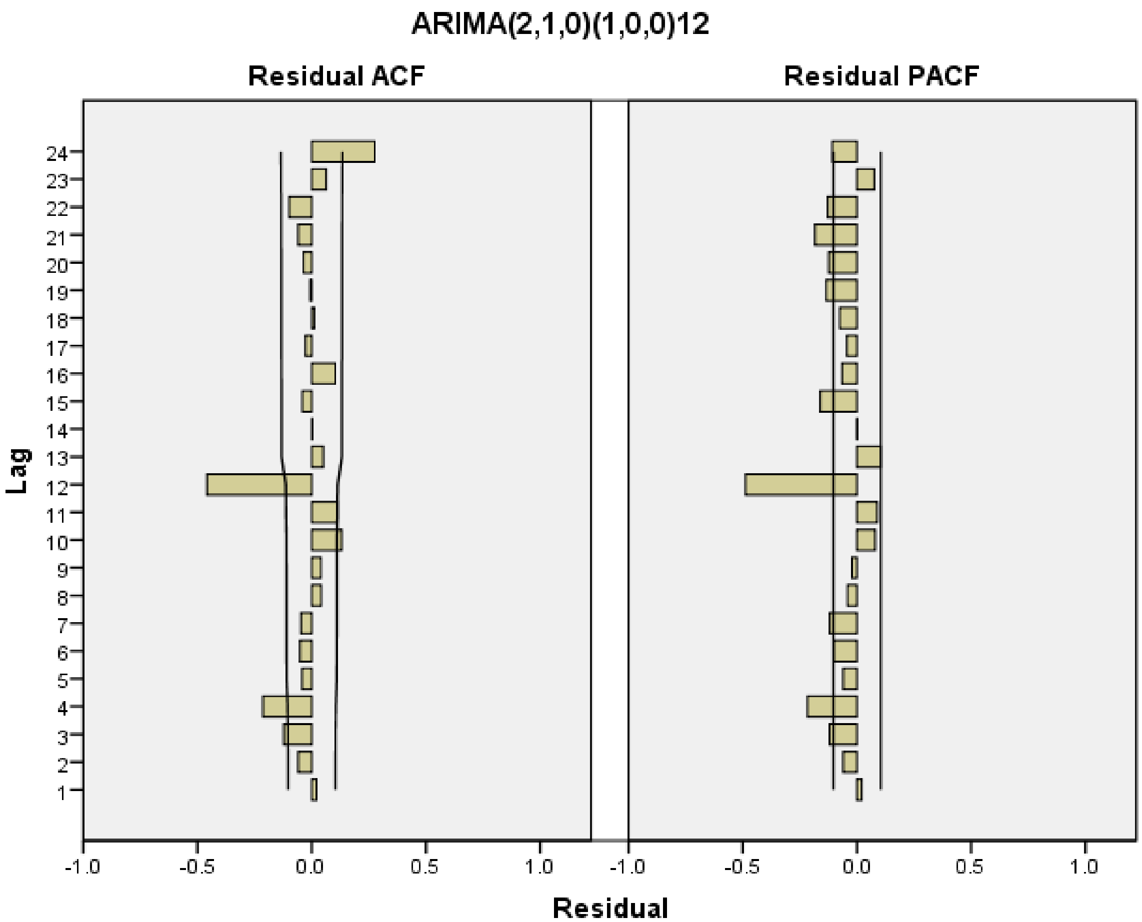

Figure-8: Residual of ACF and PACF of ARIMA(2,1,0)(1,0,0)12

ARIMA(2,1,0)(1,1,0)$_{12}$ and first we plot the residual of ACF&PACF of this model and from the figure it is clear visible that at many lags there and this model is rejected due to poor parameters quality and residual variance. The value of R^2=0.892, LJUNG BOX=39.804, BIC=20.188, MAPE=18.173.

Next trial model is ARIMA(2,1,0)(1,1,1)$_{12}$.

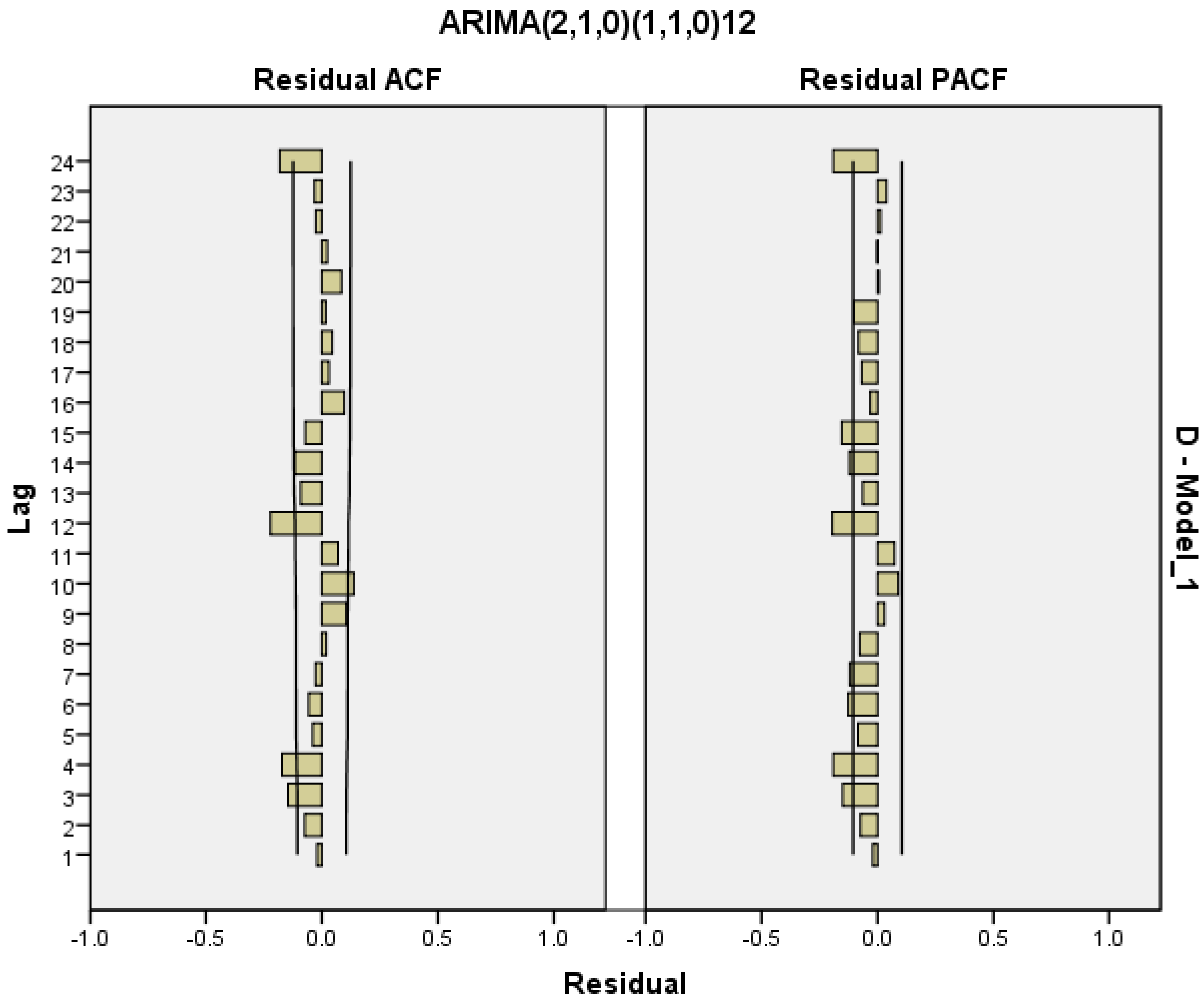

Figure-9: Residual of ACF and PACF of ARIMA(2,1,0)(1,1,0)$_{12}$

ARIMA(2,1,1)$(1,1,1)_{12}$ and first we plot the ACF&PACF of this model and from the figure it is clear visible that ACF of this model is satisfactory as all the series correlation values are within the prescribed acceptable error band and the parameters of AR1 and AR2 both models are significant and AR2 model is rejected due to poor parameters quality and residual variance. The value of R^2=0.912, LJUNG BOX=15.296, BIC=20.001, MAPE=18.467. but t- value is not more 2.

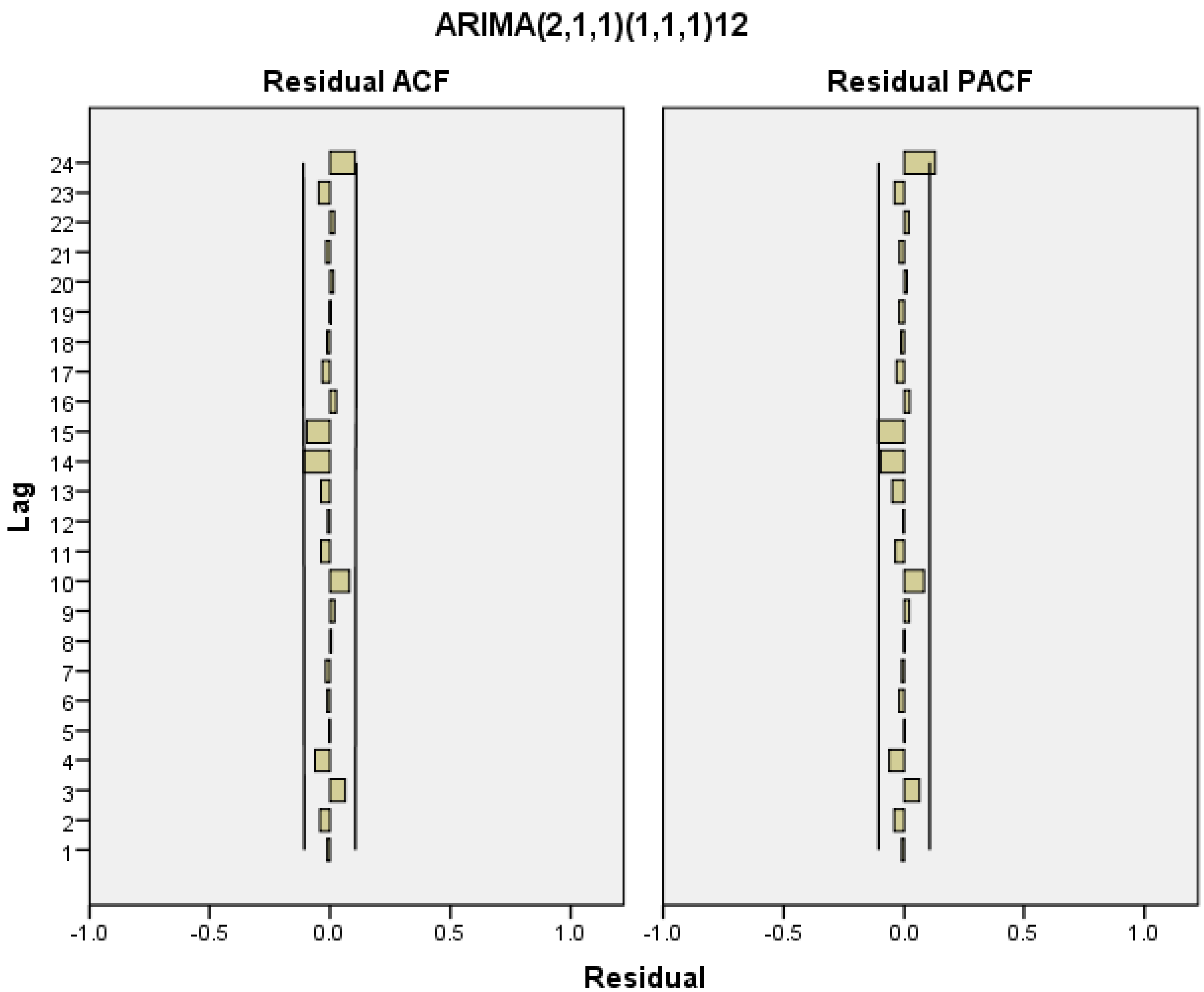

Figure-10: Residual of ACF and PACF of ARIMA(2,1,1)$(1,1,1)_{12}$

Table-2: Summary of Estimation Results of AR(2) Models

S.NO	Parameters	ARIMA(2,0,0)(0,0,0)12		ARIMA(2,1,0)(0,0,0)12		ARIMA(2,1,0)(1,0,0)12		ARIMA(2,1,0)(1,1,0)12		ARIMA(2,1,1)(1,1,1)12	
		Value	T	Value	T	Value	T	Value	T	Value	T
1	AR1	1.259	26.6	0.450	8.81	-0.28	-5.38	-0.390	-7.48	0.423	7.54
2	SE	0.047		0.051		0.053		0.052		0.057	
3	AR2	-0.449	-9.51	-0.263	-5.16	-0.24	-4.61	-0.254	-4.8	-0.22	0.397
4	SE	0.047		0.051		0.052		0.054		0.057	
5	MA1									0.981	44.77
6	SE									-0.022	
7	SAR1					0.854	29.45	-0.652	-15.9	-0.259	18.45
8	SE					0.029		0.042		0.061	
9	SMA1									0.840	5.78
10	SE									0.045	
11	MAPE	44.66		47.44		57.34		18.17		18.48	
12	R^2	0.588		0.551		0.678		0.892		0.912	
13	BIC	21.48		21.57		21.2		20.18		20.003	
14	LJUNG BOX	244.3		237.8		199.3		39.84		15.29	

AR(2) model is rejected due to poor parameters quality and fail in t-test.

5.5 AR(1) model: AR(2) model fail in test. Now we move on AR(1) model.

ARIMA(1,0,0)(0,1,0)$_{12.}$ It is a starting model and first we plot the ACF and PACF of residuals and after seeing the ACF and PACF of this model we see there is a spike on both ACF and PACF at lag 12 and the parameters values are R^2=0.810, MAPE=23.033, LJUNG BOX=174.598, AR1=0.458. t=9.6.

So next trial model we take ARIMA(1,0,0)(1,1,0)$_{12}$.

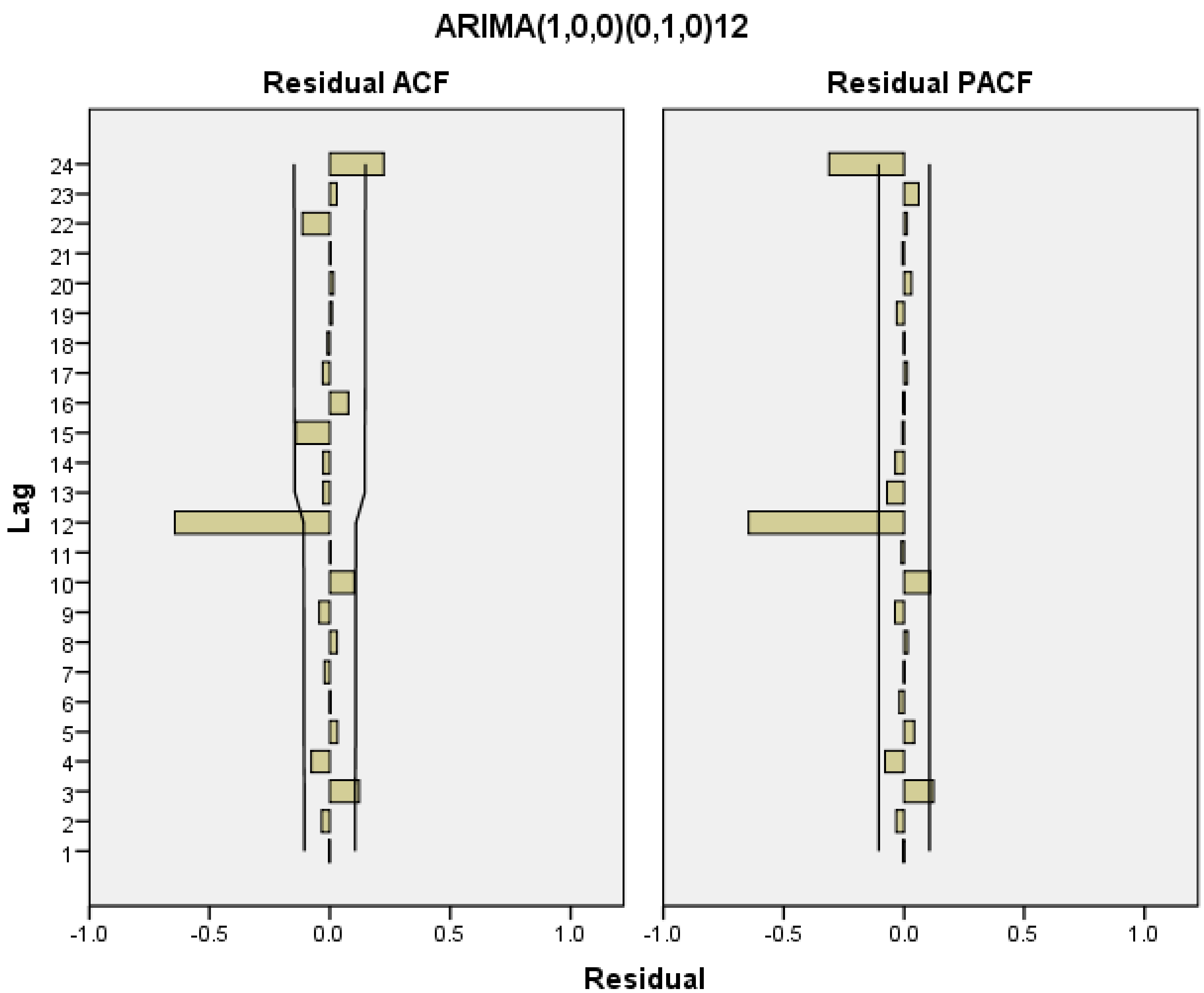

Figure-11: Residual of ACF and PACF of ARIMA(1,0,0)(0,1,0)$_{12}$

ARIMA(1,0,0)(1,1,0)$_{12}$. First we plot the ACF and PACF of residuals and after seeing the ACF and PACF of this model we see there is a spike on both ACF and PACF at lag 12 and the parameters of this model is better than ARIMA(1,0,0)(0,1,0)$_{12}$ values are R^2=0.893, MAPE=17.651, LJUNG BOX=40.678, AR1=0.431, . t=8.884 and SAR1=-0.656, t=-15.922.

So next trial model we take ARIMA(1,1,0)(1,1,0)$_{12}$

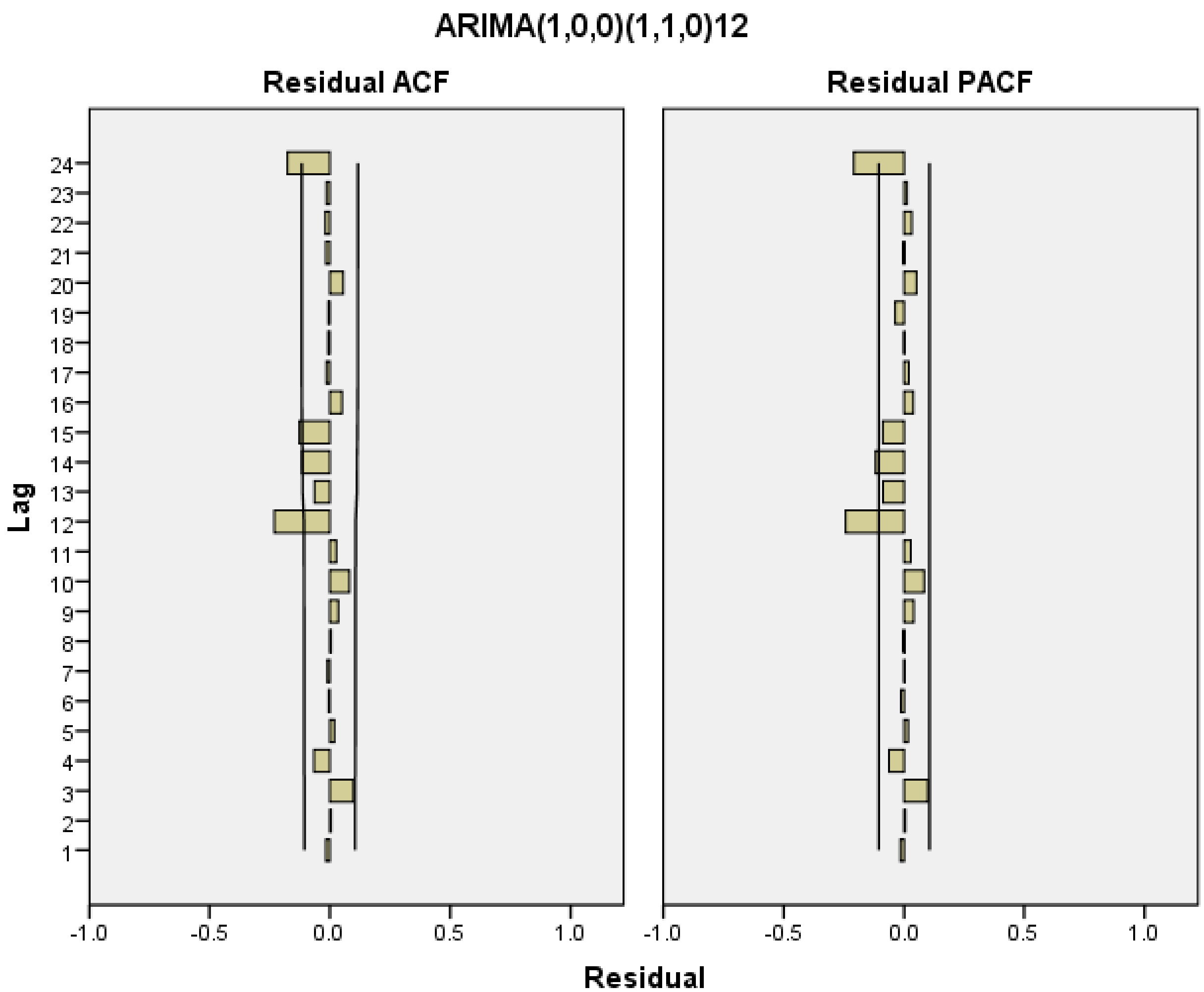

Figure-12: Residual of ACF and PACF of ARIMA(1,0,0)(1,1,0)$_{12}$

ARIMA(1,1,0)(1,1,0)$_{12.}$ First we plot the ACF and PACF of residuals and after seeing the ACF and PACF of this model we see there is a spike on both ACF and PACF at lag 2,12 and it is increasing also in PACF and the parameters of this model is not good than ARIMA(1,0,0)(1,1,0)$_{12}$ because the value of MAPE and LJUNG BOX is increased and the values are R^2=0.481, MAPE=24.796, LJUNG BOX=66.851, AR1=-0.311, . t=-6.075 and SAR1=-0.656, t=-15.988.

So next trial model we take ARIMA(1,0,1)(1,1,0)$_{12}$

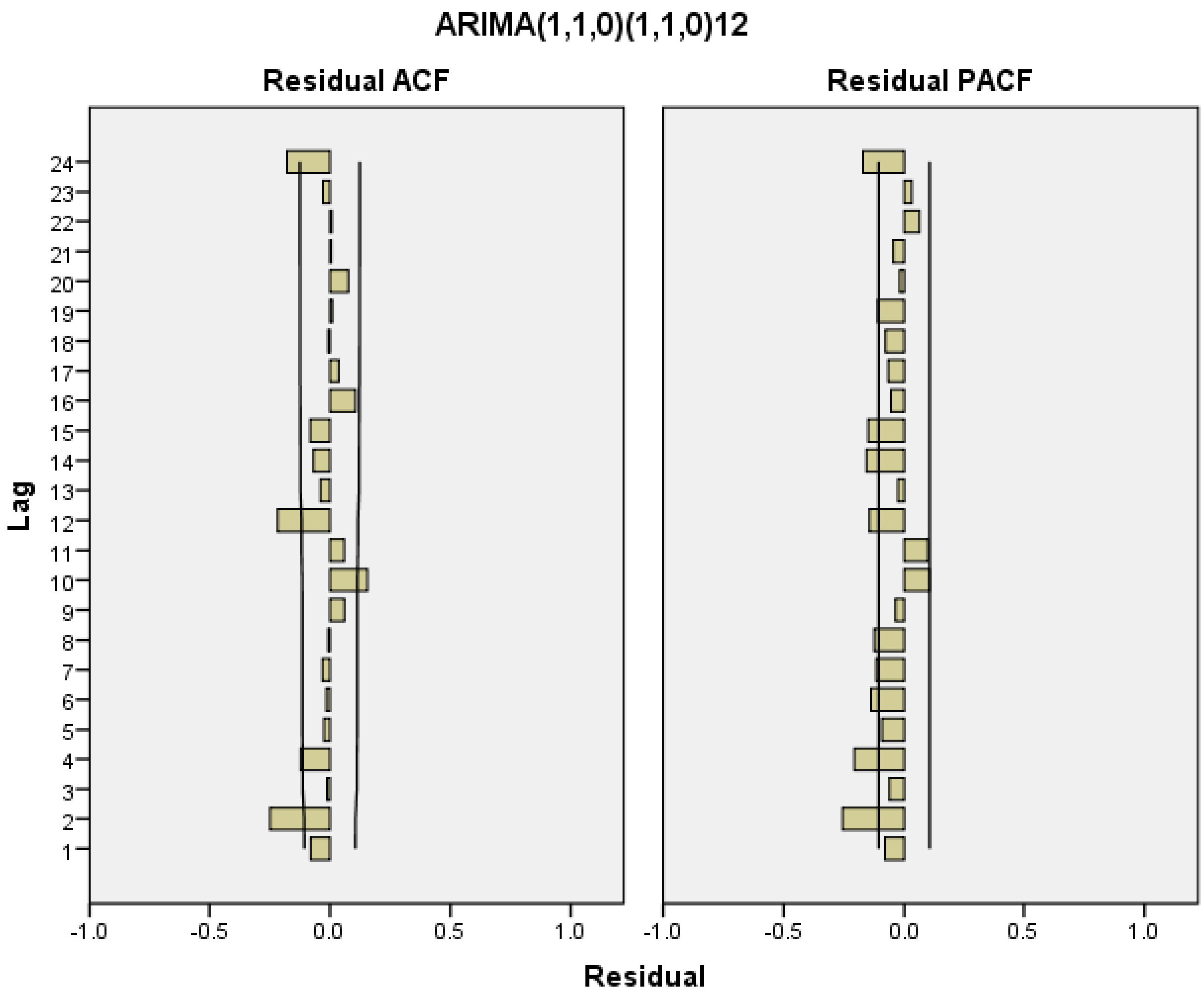

Figure-13: Residual of ACF and PACF of ARIMA(1,1,0)(1,1,0)$_{12}$

ARIMA(1,0,1)(1,1,0)$_{12}$. We plot the ACF and PACF of residuals and after seeing the ACF and PACF of this model we see there is a spike at lag 12, 24 only. And the parameters of this model is not good because this Model t- value is less 2. And values are R^2=0.893, MAPE=17.685, LJUNG BOX=40.507, AR1=0.519, t=4.872, MA1=0.109, t=0.870, SAR1=-0.657, t=-15.867.

So, the next trial model is ARIMA(1,0,1)(1,1,1)$_{12}$.

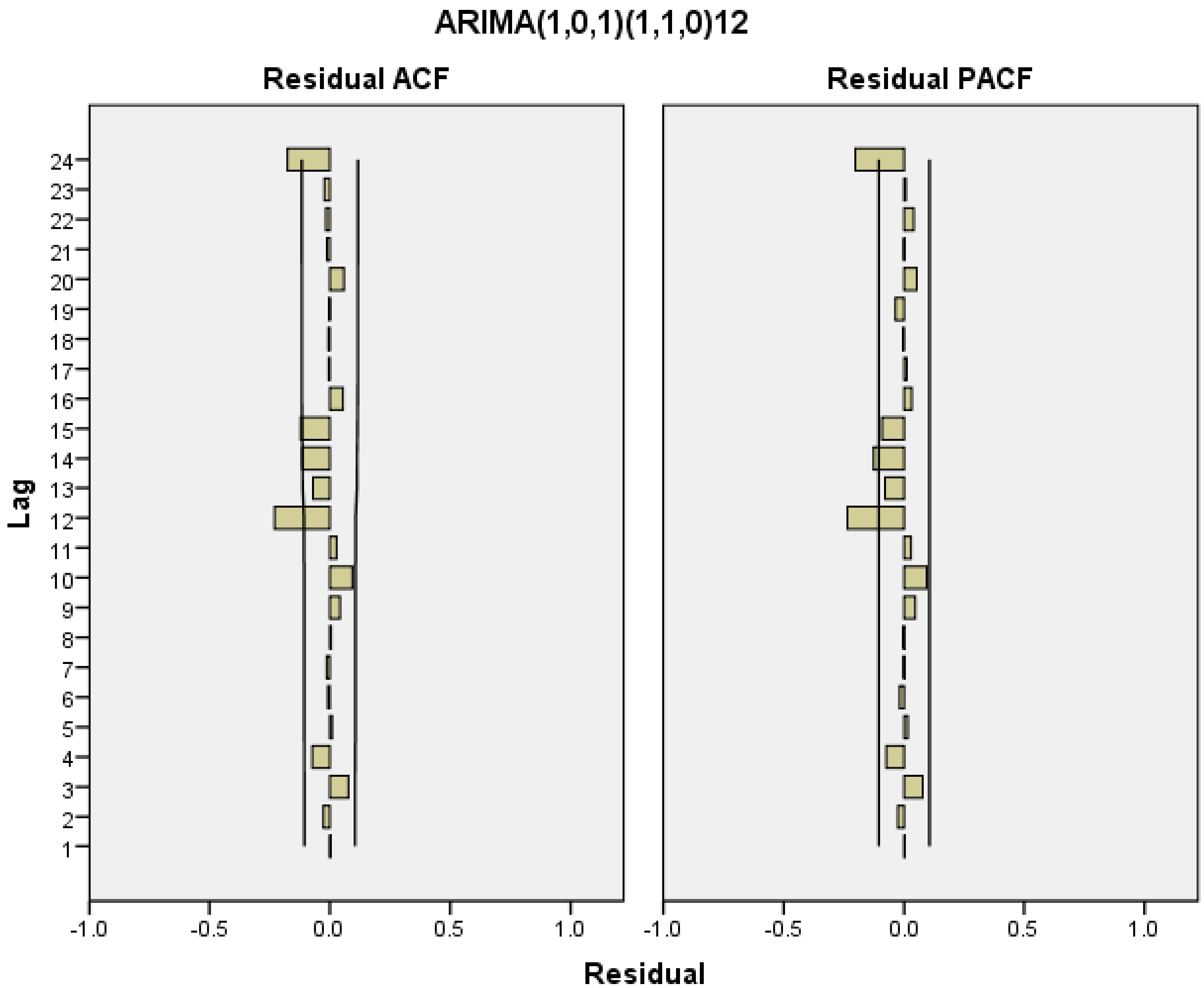

Figure-14: Residual of ACF and PACF of ARIMA(1,0,1)(1,1,0)$_{12}$

ARIMA(1,0,1)(1,1,1)$_{12}$. We plot the ACF and PACF of residuals and it is perfect. The parameters of this model is not good because t-value is less than 2. And values are R^2=0.914, MAPE=16.515, LJUNG BOX=14.926, AR1=0.459. t=4.010, MA1= 0.047, t=0.365, SAR1=-0.258, t=-4.293, SM1=0.822, t=18.782.

So, the next trial model is ARIMA(1,0,0)(1,1,1)$_{12}$.

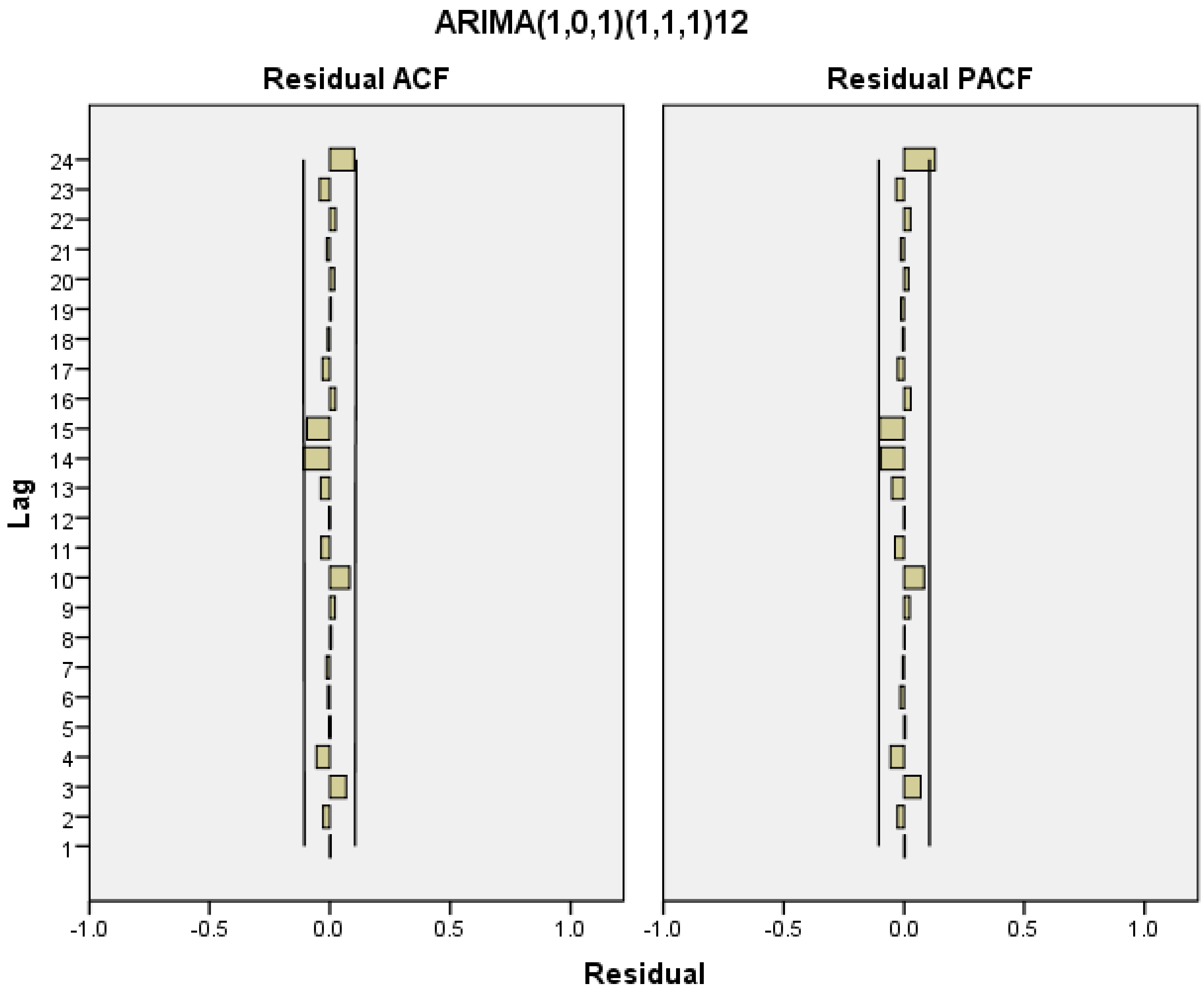

Figure-15: Residual of ACF and PACF of ARIMA(1,0,1)(1,1,1)$_{12}$

ARIMA(1,0,0)(1,1,1)12. We plot the ACF and PACF of residuals and it is ok. The parameters of this model is better than previous model and the values is R^2=0.914, MAPE=16.566, LJUNG BOX=14.985, AR1=0.420, t=8.606, SAR1=-0.258, t=-4.293, SMA1=0.823, t=18.899.

This model is good overall and this model comes from logic. So we try plus and minus of this neighbourhood model.

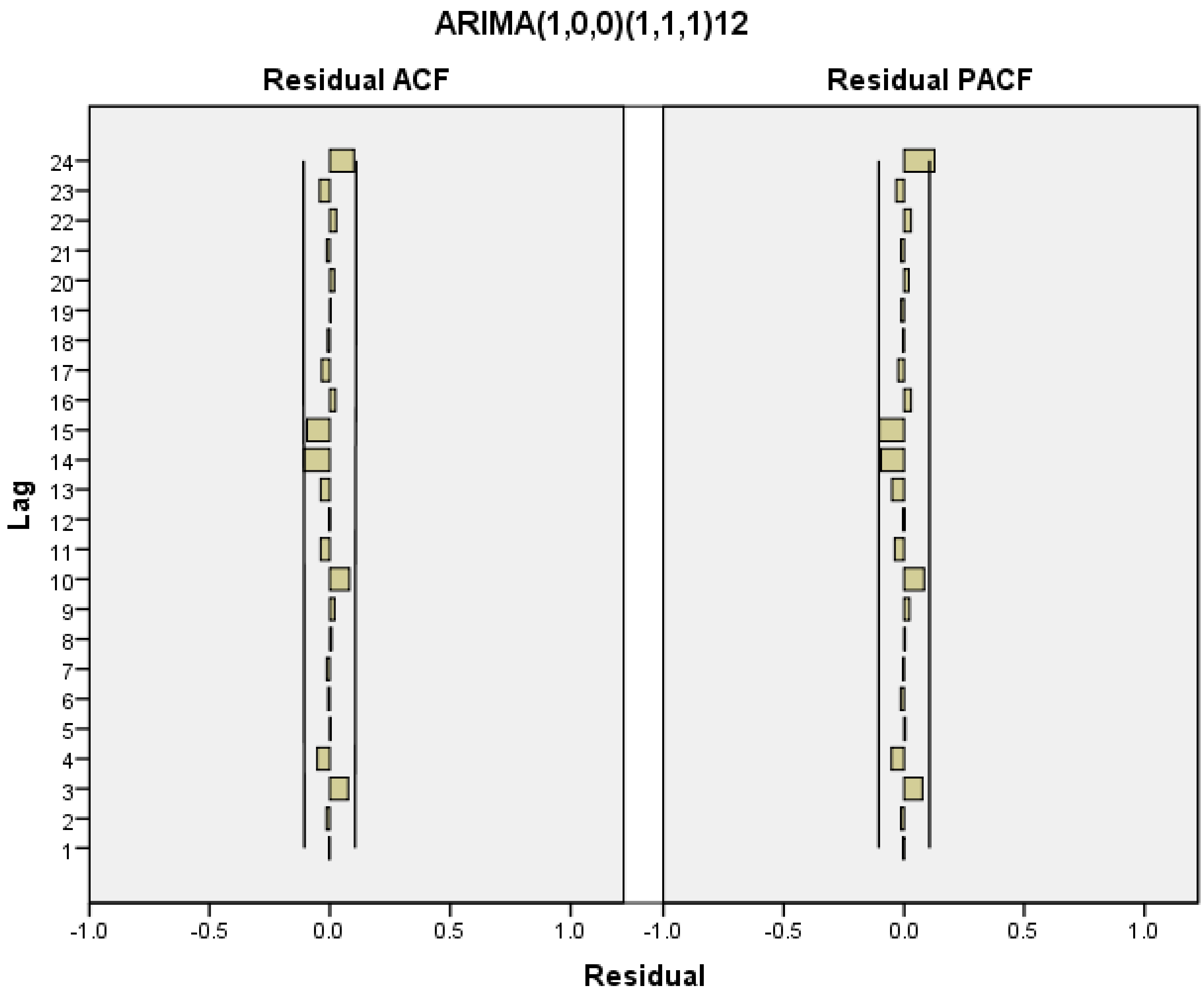

Figure-16: Residual of ACF and PACF of ARIMA(1,0,0)(1,1,1)12

Table-3: Summary of Estimation Results of tentative AR(1) Models

S.NO	Parameters	ARIMA(1,0,0)(0,1,0)$_{12}$		ARIMA(1,0,0)(1,1,0)$_{12}$		ARIMA(1,1,0)(1,1,0)$_{12}$		ARIMA(1,0,1)(1,1,1)12		ARIMA(1,0,0)(1,1,1)12	
		Value	T	Value	T	Value	T	Value	T	Value	T
1	AR1	0.458	9.6	0.431	8.89	-0.31	-6.07	0.459	4.010	0.420	8.6
2	SE	0.048		0.049		0.051		0.114		0.048	
3	MA1							0.048	0.36		
4	SE							0.129			
5	SAR1			-0.657	-15.9	-0.65	-6.09	-0.258	-4.29	-0.258	-4.29
6	SE			0.041		0.041		0.0060		0.060	
7	SMA1							0.83	18.78	0.84	18.89
8	SE							0.044		0.044	
9	MAPE	23.034		17.651		24.79		16.52		16.144	
10	R^2	0.810		0.893		0.865		0.914		0.914	
11	BIC	20.69		20.135		20.36		19.95		19.93	
12	LJUNG BOX	174.59		66.851		66.78		14.93		14.89	

ARIMA(1,0,0)(1,1,1)$_{12}$. This model is good overall and this model comes from logic so we try plus and minus of this neighbourhood model and there are 9 model.

Table-4: Various ARIMA model remarks

ARIMA Model	Remarks
ARIMA(0,0,0)(1,1,1)	ACF and PACF of residuals are not good and parameters are also not good.
ARIMA(1,0,0)(1,1,1)	ACF and PACF of residuals are good and parameters are of good quality.
ARIMA(2,0,0)(1,1,1)	This model fails in t-test.
ARIMA(1,0,1)(1,1,1)	This model fails in t-test
ARIMA(1,0,2)(1,1,1)	This model fails in t-test
ARIMA(1,0,0)(0,1,1)	ACF and PACF of residuals is not good.
ARIMA(1,0,0)(2,1,1)	The ACF and PACF of residuals is good but poor parameters quality
ARIMA(1,0,0)(1,1,0)	ACF and PACF of residuals are not good and parameters are also not good
ARIMA(1,0,0)(1,1,2)	ACF and PACF of residuals are good and parameters are of good quality.

Therefore ARIMA(1,0,0)(1,1,1)$_{12}$. Is selected due to good parameters quality and residual variance. Based on values of residual variance MAPE, LJUNG-BOX, BIC criteria and other significant tests. To check that, we test the null hypothesis about the parameters by calculating t-values. The t- value should be more than 2.0 and this model pass the t-test. Residual of ACF and PACF of this candidate model have been plotted below. The ACF and PACF of this model is good as all the series correlation values are within the prescribed acceptable error band.

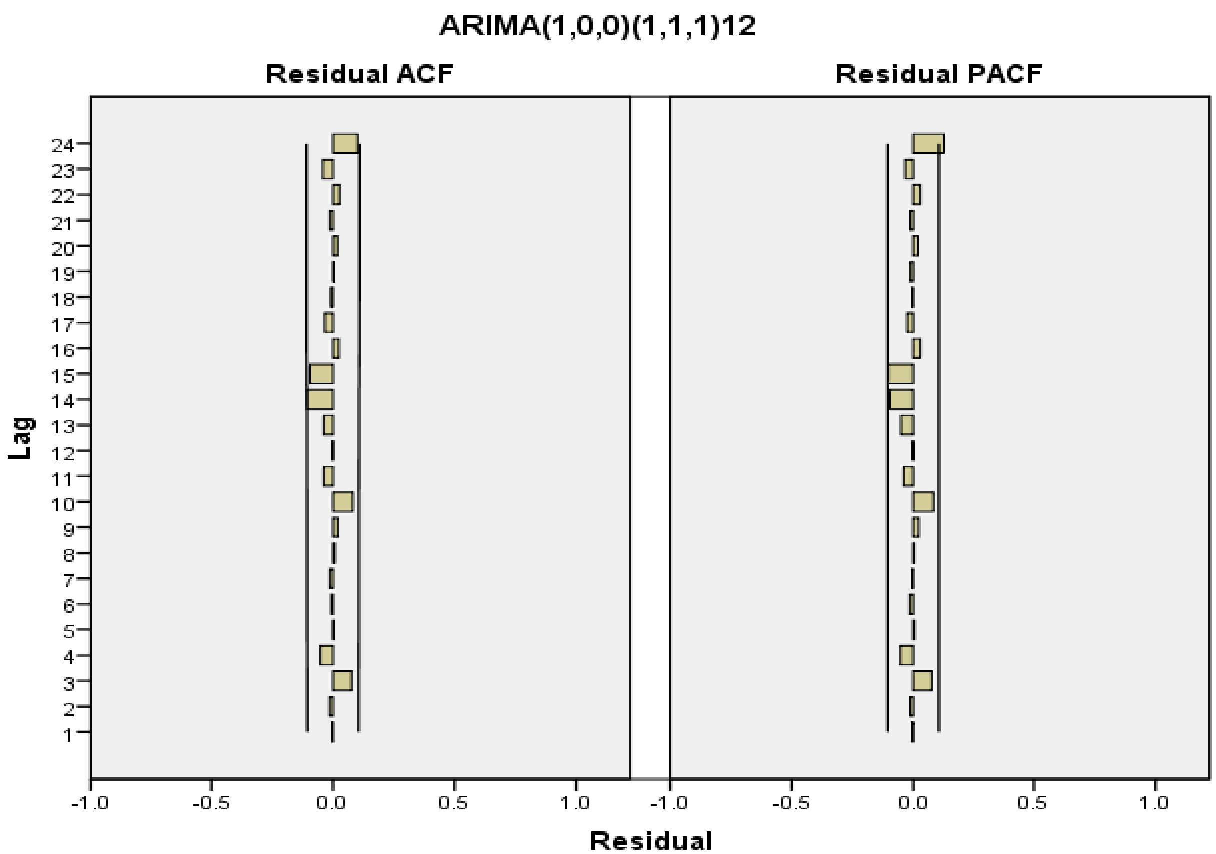

Figure-17: Residual of ACF and PACF of ARIMA(1,0,0)(1,1,1)$_{12}$

Therefore ARIMA(1,0,0)(1,1,1)$_{12}$. Model has been selected for forecasting of monthly streamflow and from this candidate model we forecast 24 months streamflow. This model equation for forecasting can be written as in the form of original series.

$$X_t = X_{t-12} + \Phi_1(X_{t-1} - X_{t-13}) + \phi_1(X_{t-1} - X_{t-13}) + a_t - \Theta_1 a_{t-12}$$

Table- 5: Comparison of forecast values and observed values for selected model ARIMA(1,0,0)$(1,1,1)_{12}$.

S.NO	Month	Observed values	Forecasted values	% Forecast Error
1	M361	19569	19260	1.58
2	M362	18130	18768	-3.52
3	M363	17843	19644	-10.09
4	M364	21684	24810	-14.42
5	M365	41942	40751	2.84
6	M366	71277	77346	-8.51
7	M367	184048	189370	-2.89
8	M368	207521	216911	-4.52
9	M369	169361	170719	-0.80
10	M370	59824	63581	-6.28
11	M371	25218	32462	-28.73
12	M372	19999	26664	-33.33
13	M373	17237	21460	-24.50
14	M374	16831	21361	-26.91
15	M375	25219	21323	15.45
16	M376	29066	24933	14.22
17	M377	32310	39293	-21.61
18	M378	65089	71949	-10.54
19	M379	210314	202453	3.74
20	M380	220553	229992	-4.28
21	M381	203561	190638	6.35
22	M382	88128	71546	18.82
23	M383	32256	33538	-3.97
24	M384	24616	25815	-4.87

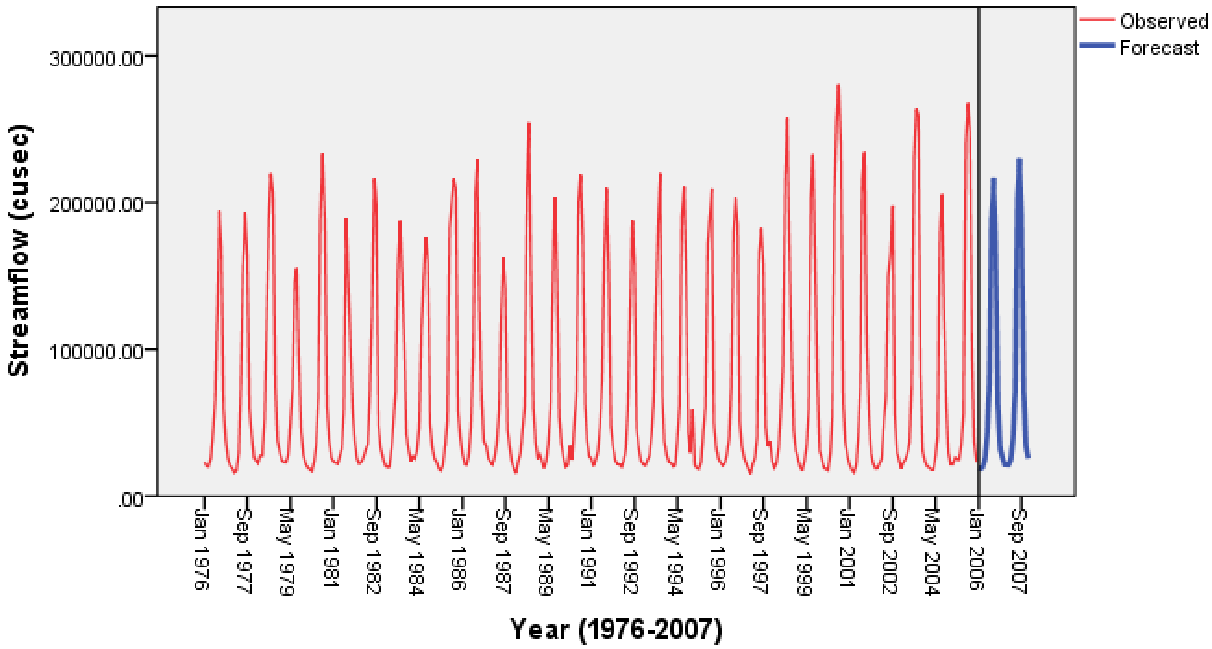

Figure- 18: Observed and forecasted streamflow

Observed streamflow 1976-2005 and forecasted streamflow 2006-2007 was done by the ARIMA(1,0,0)(1,1,1)$_{12}$.

Comparison of Mean absolute percentage error (MAPE) of forecast values (up to 3 steps ahead) and observed values for selected model ARIMA $(1,0,0)(1,1,1)_{12}$.

Table-6: 1- Step ahead forecasting

S NO	Month	Observed Values	Forecast Value	% Error
1	M361	19569	19260	1.58
2	M362	18130	18768	3.52
3	M363	17843	19644	10.09
4	M364	21684	24810	14.42
5	M365	41942	40751	2.84
6	M366	71277	77346	8.51
7	M367	184048	189370	2.89
8	M368	207521	216911	4.52
9	M369	169361	170719	0.80
10	M370	59824	63581	6.28
11	M371	25218	32462	28.73
12	M372	19999	26664	33.33
Mean absolute percentage error				9.79

Table-7: 2- Step ahead forecasting

S NO	Month	Observed Values	Forecast Value	% Error
1	M361	18130	18768	3.52
2	M362	17843	19644	10.09
3	M363	21684	24810	14.42
4	M364	41942	40751	2.84
5	M365	71277	77346	8.51
6	M366	184048	189370	2.89
7	M367	207521	216911	4.52
8	M368	169361	170719	0.80
9	M369	59824	63581	6.28
10	M370	25218	32462	28.73
11	M371	19999	26664	33.33
12	M372	17237	21460	24.50
Mean absolute percentage error				11.70

Table-8: 3-Step ahead forecasting

S NO	Month	Observed Values	Forecast Value	% Error
1	M361	17843	19644	10.09
2	M362	21684	24810	14.42
3	M363	41942	40751	2.84
4	M364	71277	77346	8.51
5	M365	184048	189370	2.89
6	M366	207521	216911	4.52
7	M367	169361	170719	0.80
8	M368	59824	63581	6.28
9	M369	25218	32462	28.73
10	M370	19999	26664	33.33
11	M371	17237	21460	24.50
12	M372	16831	21361	26.91
Mean absolute percentage error				13.65

In this study, the stochastic nature of streamflow is analyzed with autoregressive integrated moving average (ARIMA) Stochastic models. The best ARIMA model were estimated using the SPSS software. The ARIMA model give a better performance because it makes time series stationary, in both stages trending and validation stage and ARIMA model improved the performance of advance information. This ARIMA$(1,0,0)(1,1,1)_{12}$. Model have been chosen based on residual of ACF, PACF and t-values of the estimated parameters and have been selected for making predictions for up to 2 years of streamflow forecasting. Based on this model having least error at trending and validation stage also having least number of parameters is been selected because of its robustness for forecasting. This indicated the superiority of ARIMA model and it can be concluded that ARIMA model could be used for forecasting (up to 3 steps ahead) of streamflow. The prediction accuracy of model is examined by comparing the percentage of forecasting error that is mean absolute percentage error (MAPE).

References

Abudu, S., King, J. P., & Bawazir, A. S. (2011). Forecasting Monthly Streamfow of Spring-Summer Runoff Season in Rio Grande Headwaters Basin Using Stochastic Hybrid Modelling Approach. *Journal of Hydrologic Engineering*, 384-390.

Adnan, R. M., Yuan, X., Kisiu, O., & Curtef, V. (2017). Application of Time Series Models for Streamflow Forecasting. *Civil and Environmental Research*, 56-63.

Alnaa, S. E., & Ahiakpor, F. (2011). ARIMA approach to predicting inflation in Ghana. *Journal of Economics and International Finance*, 328-336.

Dayal, D., Swain, S., Gautam, A. K., Palmate, S. S., Pandey, A., & Mishra, S. (2019). Development of ARIMA model for Monthly Rainfall Forecasting over an Indian River Basin. *World Environment and Water Resources Congress*, 264-271.

Karamouz, M., & Zahraie, B. (2004). Seasonal Streamflow Forecasting Using snow Budget and El nino-1southern Oscillation climate signal. *Journal of Hydrologic engineering*, 523-533.

Kumar, M., & Anand, M. (n.d.). *An Application of Time series ARIMA Forecasting model for Predicting Sugarcane Prouction in India.*

Landeras, G., Ortiz-Barredo, A., & Lopez, J. J. (2009). Forecasting Weekly Evapotranspiration with ARIMA and Artificial Neural network model. *Journal of Irrigation and Drainage Engineering*, 323-334.

Mingrong, T., & Hengxin, X. (2008). Highway Traffic Volume Forecasting Based on Seasonal ARIMA model. *Journal of Highway and Transportation Research and Development*, 109-112.

Mohan, S., & Arumugam, N. (2009). Forecasting Weekly refrence crop Evapotranspiration series. *Hydrologica; sciences journal*, 689-702.

Mondal, P., & Shit labani, G. S. (2014). Study of Effectiveness of Time Series Modelling ARIMA in Forecasting Stocl Prices. *Internation al Journal of Computer Science*, 13-29.

NS, G. B. (2017). Application of ARIMA model for River Discharge Analysis. *Journal of Nepal Physics Society*, 27-32.

Oliveria, P. J., & Boccelli, D. (2017). k-Nearest Neighbour for Short Term Water Demand Forecasting. *World Environmental and Water Resource Congress*, 501-510.

Shathir, A. K., & Saleh, L. M. (2016). Best ARIMA models for Forecasting Inflow of Hit Station. *Basrah Journal for Engineering Sciences*, 62-72.

Sun, H., & Koch, M. (2001). Case Study: Analysis and Forecasting of Salinity in Apalachicola Bay, Florida, Using Box0-Jenkins ARIMA model. *Journal of Hydraulic Engineering* , 718-727.

Annexure-I

DATA

The Sarda River in Banbassa was selected as the project to implement in this research. The Sarda River originates at kalapani in the Himalayas at an elevation of 3600m and the basin area of 14871 km^2. For the purpose of the study the monthly time series data of discharge (in cusec) is collected for a span of 32 years from 1976 to 2007. The data for the Sarda River at Banbassa were obtained from the Irrigation Department of Uttar Pradesh. The data to be used in preparing the model is taken from the period of 1976-2005(30 years) and the remaining data for the period of 2 years from 2006 and 2007 has been used to validate the model.

Trending Data-30 years

YEAR	JAN	FEB	MAR	APR	MAY	JUNE	JULY	AUG	SEP	OCT	NOV	DEC
1976	23344	20432	20145	25436	42323	65849	126137	194269	169763	59870	37306	25586
1977	20665	19142	16154	17664	29141	77857	157141	193385	163301	60329	40855	26196
1978	24327	22345	27617	27665	53784	116081	187572	219690	206834	75616	38032	30560
1979	24607	23080	23030	28322	50951	71791	145726	155395	91667	43907	28953	22558
1980	19635	18428	17187	23774	35267	67294	187088	233247	187833	61325	38152	26463
1981	23810	22753	22025	27420	32729	59572	189780	144232	97719	57007	36610	25697
1982	22155	23263	26959	31705	35347	69081	123568	216531	198169	47959	33046	26723
1983	20973	19666	19988	31457	50777	71509	153030	187718	145569	102234	41712	31744
1984	24059	27378	25592	31320	47143	113762	148733	176746	158828	49443	33615	25583
1985	23204	18573	17443	21049	34074	51022	182902	200652	216283	208947	57752	39663
1986	28239	21705	21257	27419	47513	96447	206057	229622	128154	66298	37226	34316
1987	25797	23464	21507	26242	36494	58993	126090	162583	140811	44966	32612	23423
1988	18018	15686	23396	29420	45954	61145	208119	254453	141071	60365	35581	25499
1989	28555	24347	19075	24483	36874	59914	134542	203974	151693	54189	35862	26370
1990	19899	21124	34578	24808	49630	67556	200667	218939	175993	63462	37408	27248
1991	25795	20570	24977	31054	48566	79608	160380	209789	151900	47740	34118	24885
1992	21865	21716	19963	23998	31962	59060	125244	188114	157192	46318	32976	24370
1993	21797	20869	24633	27777	43231	64961	134656	187782	219878	68127	37017	28155
1994	23222	22723	20213	21096	38812	62302	190920	210904	149438	43568	29627	59407
1995	21049	19015	18365	22961	42898	61281	171692	191775	209268	49704	33083	25619
1996	22784	20789	22608	28677	38625	76032	179809	203532	183940	50656	33181	25901
1997	21769	18014	15378	22065	26430	40897	166600	182541	157078	47269	34056	37269
1998	24045	19313	22160	32082	57781	90159	213706	257403	174917	101323	44726	31497
1999	26053	20070	17427	24297	34682	59376	202819	232567	156942	79107	31052	26299
2000	19647	18546	17723	28984	54996	202902	254878	280201	238018	63293	37146	27674
2001	21637	17873	16260	21049	41306	78308	216047	234096	103448	55510	30069	22180
2002	18997	19193	22712	26082	51415	67557	149585	164699	197609	55597	30008	25142
2003	18931	23389	25208	29675	37166	76514	230505	263832	259100	83718	31049	24113
2004	20293	19470	17924	18154	27706	40738	179218	205718	121984	64805	30081	21725
2005	21733	26287	25092	24837	34901	56310	240131	267684	248051	94502	36639	23370

Validation Data- 2 years

2006	20569	21130	21843	26984	42942	80277	191948	221121	175361	66824	35218	29999
2007	25237	23831	25219	28066	43310	74089	205314	231953	193561	74128	36256	28616