Time Series Analysis and Forecasting of Rainfall for Agricultural Crops in India: An Application of Artificial Neural Network

Debasis Mithiya[1,*], Kumarjit Mandal[2] & Simanti Bandyopadhyay[3]

[1]Department of Business Administration, International School of Hospitality Management, Maulana Abul Kalam Azad University of Technology, West Bengal, India

[2]Department of Economics, University of Calcutta, Kolkata, India

[3]Department of Economics, Victoria Institution (College), University of Calcutta, Kolkata, India

*Corresponding author: Department of Business Administration, International School of Hospitality Management, Maulana Abul Kalam Azad University of Technology, West Bengal, India. E-mail: drdebasis.mithiya@gmail.com

Abstract

Indian agriculture depends heavily on rainfall. It not only influences agricultural production but also affects the prices of all agricultural commodities. Rainfall is an exogenous variable which is beyond farmers' control. The outcome of rainfall fluctuation is quite natural. It has been observed that fluctuation in rainfall brings about fluctuation in output leading to price changes. Considering the importance of rainfall in determining agricultural production and prices, the study has attempted to forecast monthly rainfall in India with the help of time series analysis using monthly rainfall data. Both linear and non-linear models have been used. The value of diagnostic checking parameters (MAE, MSE, RMSE) is lower in a non-linear model compared to a linear one. The non-linear model - Artificial Neural Network (ANN) has been chosen instead of linear models, namely, simple seasonal exponential smoothing and Seasonal Auto-Regressive Integrated Moving Average to forecast rainfall. This will help to identify the proper cropping pattern.

Keywords: artificial neural network, mean absolute error, mean square error, root mean square error, simple seasonal exponential smoothing, seasonal auto-regressive integrated moving average, time series analysis

GEL Classification: C22, Q25

1. Introduction

Rainfall is one of the major factors of agricultural production. In the last few years, the world's climate change has tremendously affected rainfall. The distribution of rainfall has fluctuated throughout the world. In some places it has increased whereas it has decreased elsewhere. Farmers are finding it difficult to cope with this change, primarily because most of the crops are seasonal and rainfall dependent. Presently, about 60 percent of world's staple food production and 90 percent of that of Sub-Saharan Africa depend directly on rainfall. Agricultural productivity can move up with increased rainfall and improvement in agricultural technology. But the regional variation in rainfall across different parts of the world makes it difficult to suggest any general kind of agricultural technology to be adopted throughout the world. The agriculture in the regions of the world with high rainfall does not face any kind of problem arising out of water scarcity. Here, rainwater can be stored through dams, ponds and reservoirs which might help to irrigate the surrounding areas and therefore lead to an increase in crop production. However, regions with low rainfall suffer due to poor availability of water for cultivation (www.sourcetrace.com). Therefore, rainfall influences cropping pattern in different parts of the world in different ways.

The Economic Survey 2017-2018 pointed out that agriculture in India continues to remain vulnerable to the fluctuations of monsoon till date. This is because of the fact that all India percentage of net irrigated area to total cropped area is as low as 34.5 percent. Around 52 percent (73.2 million hectares area out of 141.4 million hectares net sown area) agricultural land is still un-irrigated and dependent on monsoon (Financial Express Bureau, 2018). According to the Agricultural Ministry of India, the kharif (June - September) cultivation has reduced by 7 percent in 2019 compared to 2018 due to rainfall shortage. The kharif cultivated area in 2019 was 56.7 million hectares whereas it was 60.9 million hectares in 2018. The lack of rainfall has affected kharif paddy and pulses cultivation more than any other crop. The shortfall of kharif, paddy and pulses cultivation of 2019 is about 9 percent and 16 percent respectively which are larger than 2018 (Chowdhury, 2019). The monsoon rainfall influences not only kharif cultivation but also impacts groundwater storage and reservoir levels (thus also emerging critical for irrigation of rabi crops).

The average income of a farmer in India is estimated at Rs 77,976 per year (according to the Dalwai committee report 2016) and the Indian government has targeted to double the farmers' income by 2022. After independence, Indian agricultural production has been influenced by the Green Revolution, land reforms, agricultural credit, agricultural extension and other factors in the long-run. However, in the short-run, year to year fluctuations of agricultural production still has been influenced by rainfall. It is not only the volume of rainfall but also its distribution at different stages of the cultivation process that influences the total output. Similarly, the volume and distribution of rain at the time of sowing, flowering, maturing, and harvesting affect crop productivity (Gulati et al., 2013). In other words, rainfall is the major source of water that is essential for plant growth and development. But it is an exogenous variable which is beyond farmers' control. Rainfall varies from time to time. Sometimes it is more than what is needed and sometimes it falls short of the minimum requirement. There can be untimely rainfall as well. Hence, it is not the total amount of rainfall, but its

distribution across regions in a particular season is what matters. Therefore, a fluctuation in the distribution of rainfall over the cropping season adversely impacts rain-fed crops. If severe drought or flood occurs during the reproductive, stages the possibility of crop failure is inevitable. So, there is a strong positive impact of rainfall on crop production. The effect of poor monsoon or monsoon failure is generally understood by looking at the changes in output between the years of monsoon failure and the normal years (Chand and Raju, 2009). Hence, rainfall pattern is one of the most important limiting factors for rain-fed crop production.

So, it can be said that in order to select the proper cropping pattern in India, it is very important to determine the exact relationship between rainfall and crop production. At the same time, it is equally important to know the expected monthly rainfall to select the most suitable crops for cultivation.

The present study intends to analyse and forecast the monthly rainfall based on the basis of time series data on rainfall. Monthly rainfall forecasts will in advance inform the farmers of the volume of rainfall. This information will help them to decide what to cultivate and by how much during a particular season. The specific objective of the study is as follows.

i) To analyze the nature of monthly rainfall in India and forecast the future monthly rainfall.

2. Review of Literature

Gulati Ashok et.al. (2013), attempted to project the likely impact of robust monsoon rains of 2013 on the Agricultural Gross Domestic Product (GDP) growth in India. The model hypothesizes that the performance of agriculture in India depends upon (1) investments in agriculture (private and public); (2) agricultural price incentives; and (3) rainfall. A log-linear model fitted over 1996-97 to 2012-13 period can explain 95 percent of the variations in agri-GDP with all variables being statistically significant. The model also forecasts that the agri-GDP growth rate for the agricultural year (July-June) 2013-14 is likely to be between 5.2% and 5.7%. They also estimates that agri-GDP growth in 2013-14 is likely to be about three times higher than previous year. This growth in agri-GDP is likely to come mainly from oilseeds, pulses, cotton, and coarse cereals belt of central and western parts of the country, which is less irrigated and thereby more dependent on rain. It is very likely that any damage to kharif crops due to excess rainfall (with extended monsoons and cyclones) would be offset by a bumper rabi crop harvest, given that there is excellent soil moisture and ample surplus water in reservoirs.

Olatayo. T. O. and Taiwo. A. I. (2014), have tried to predict rainfall on the basis of annual rainfall data of 31 years (1982 to 2012) of Ibadan South West in Nigeria. According to T. O. Olatayo and A. I. Taiwo, the climate and rainfall are highly non-linear and complicated phenomena, which require classical, modern and detailed models to obtain accurate prediction. In order to attain precise forecast, they used a modern method i.e, fuzzy time series (FST) which belongs to the first order time-variant method. They also used other two fundamentally different approaches which are autoregressive integrated moving average (ARIMA) and the non-parametric method (Theil's regression). The performance of the model

has been evaluated based on lowest value of mean squared forecast error (MAE), root mean square forecast error (RMSE) and Coefficient of determination. The study reveals that FTS model is an appropriate forecasting tool to predict rainfall, since it outperforms the ARIMA and Theil's Regression.

Karuiru Elias Kimani et al. (2016), have analyzed and forecasted precipitation in Keneya. They have used linear time series model of Seasonal Autoregressive Integrated Moving Average (SARIMA) and non-linear model of Time Lagged Feed forward Neural Network (TLFN). In their study, the values of diagnostic checking parameters including Mean Absolute Deviation (MAD), Mean Squared Deviation (MSD) and Mean Absolute Percentage Error (MAPE) have been found lower in TLFN than SARIMA. They conclude that the Time Lagged Feed forward Neural Network model has performed better than Seasonal Autoregressive Integrated Moving Average for forecasting precipitation.

Kamath R.S. and Kamat R.K. (2018), have analysed time series rainfall data during January 2006 to December 2016 of Iddukki district in Kerala to forecast rainfall. They used both linear and non-linear time series models namely the Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN) and Exponential Smoothing State Space (ETS). Their investigation shows that ARIMA model performed better than the other models. The analysis is based on Root Mean Squared Error (RMSE).

Pongdatu. G. A. N. and Putra. Y. H. (2018), in their study have compared SARIMA and Holt-Winter's Exponential Smoothing methods for time series analysis and forecasting of clothing sales in retail stores from 2013 to 2017. The SARIMA model yields more accurate results for short period time series data. However, Holt Winter's Exponential Smoothing is more appropriate in forecasting seasonal time series data, whatever its pattern and trend be. In their study, they have compared the Mean Absolute Deviation (MAD) of the two models and have concluded that the model with smaller MAD, i.e., SARIMA $(1, 1, 0) (0, 1, 0)_{12}$ is a better one.

Kistner Erica et.al. (2018), have found that the temperature and precipitation fluctuations across the Midwest directly impact quantity and quality of specialty crops (generally more sensitive to climatic stressors and require more comprehensive management compared to traditional row crops) and indirectly influence the timing of crucial farm operations. They have observed that increasingly variable weather and climate condition pose a serious threat to specialty crop production in the Midwest. Their results indicate that weather-induced losses vary between Midwestern states with excessive moisture resulting in the highest total number of claims across all states followed by freeze and drought events. They observe that the specialty crop growers are aware of the increased production risk under changing climate condition and have identified the need for crop specific weather, production, and financial risk management tools and increased crop insurance coverage.

Renato Rossetti (2019) has discussed time series approaches to forecast the sales of console games for the Italian market. More specifically he evaluates two univariate techniques namely, Exponential Smoothing and the SARIMA techniques. The aim is to exploit the capabilities of these statistical methods in order to have a comparison of the results and to

choose the most accurate model through an ex-post evaluation. Using monthly time-series data from November 2005 to September 2017, the selection of the most suitable model has been indicated by the smallest value of the measures of accuracy (MAPE, MAPE, RMSE) for the out-of-sample observations during the period October 2017-September 2018. In terms of performance, the smallest values concerning these measures represent the best choice between the models studied, and the preferred option for this time series is SARIMA (2, 1, 0) (1, 1, 0)$_{12}$. Therefore the SARIMA model is the most reliable and relevant model for this analysis.

3. Data and Methodology

The study covers the period between 1992-93 and 2016-17. All India monthly, seasonal and annual rainfall series were constructed based on the area weighted rainfall of all the 306 stations during 1989-2013 and 2014-2016 based on IMD Sub-divisional rainfall (Kothawale and Rajeevan, 2017). The different time series (linear and non-linear) models have been used to analyze the objective of our research.

3.1 Linear Time Series Models

In the analysis of time series, several methods of forecasting are used. Most common methods are the Moving Average method, Linear Regression with Time, Non-seasonal and Seasonal Exponential Smoothing, Autoregressive Integrated Moving average etc.

The monthly rainfall data of India exhibit seasonality (which is explained later). According to the nature of the data, this study concentrates on the Simple seasonal Exponential Smoothing and Seasonal Autoregressive Integrated Moving Average models among the linear models.

3.1.1 Simple Seasonal Exponential Smoothing

When the sequences of data show no trend but seasonality factors exist, the simple seasonal exponential smoothing method is the best option for forecasting. The simple seasonal exponential method comprises the forecast equation and two smoothing equations - one for the level L_t, and other for the seasonal component denoted by S_t, with smoothing parameters α and δ.

$$\text{Level } L_t = \alpha (Y_t - S_{t-s}) + (1-\alpha) L_{t-1} \qquad \text{……… (1)}$$

$$\text{Seasonal } S_t = \delta (Y_t - L_t) + (1-\delta) S_{t-s} \qquad \text{………. (2)}$$

Forecast $F_{t+k} = L_t + S_{t+k-s)}$

Where s (s = 12 for monthly data) is the length of the seasonal cycle, for $0 \leq \alpha \leq 1$ and $0 \leq \delta \leq 1$.

3.1.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

The ARMA is an extrapolation (techniques for forecasting using only the past data) method which requires historical time series data of the underlying variable. The model in specific and general forms may be expressed as follows.

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \delta_1 \varepsilon_{t-1} - \delta_2 \varepsilon_{t-2} - \dots - \delta_q \varepsilon_{t-q} \qquad (3)$$

Box and Jenkins argue that a non-stationary series can be transformed either into a stationary or an almost stationary series if it is differenced (d) an appropriate number of times. After transforming the series into a stationary or to an almost stationary series, the model transforms to ARIMA.

The mathematical equation, of the ARIMA (p, d, q) model is defined in Equation (4) as follows:

$$\phi p\ (B)\ (1-B)^d\ Y_t = \theta\ q\ (B)\ \varepsilon_t \qquad (4)$$

If a time series is seasonal of period **s**, Box and Jenkins made a proposal that such a model may be defined as Equation (5)

$$\phi_p\ (B)\ \varphi_P\ (B^s)(1-B)^d\ (1-B^s)^D\ Y_t = \theta_q\ (B)\ \delta_q\ (B^s)\ \varepsilon_t \qquad (5)$$

Where B is the backshift operator (i.e. $BY_t = Y_{t-1}$, $B^2 Y_t = Y_{t-2}$, $B^s Y_t = Y_{t-s}$ and so on),'s' is the seasonal lag and 'ε' is a sequence of independent normal error variables with mean 0 and variance σ^2. ϕs and φs are the non-seasonal and seasonal autoregressive parameters respectively. θ s and δs are non-seasonal and seasonal moving average parameters respectively. P and q are orders of non-seasonal autoregressive and moving average parameters, whereas P and Q are that of the seasonal auto regressive and moving average parameters. Here, 'd' and 'D' denote non-seasonal and seasonal differences respectively. In its general form, the Seasonal ARIMA (SARIMA) model is represented as SARIMA (p, d, q) (P, D, Q) s. Box and Jenkins (1976) have suggested four basic steps namely (i) Identification of the model (ii) Estimation of parameters of the model (iii) Diagnostic Checking of the model and (iv) Forecasting- to complete the procedure of model building (details of this model has been described in our published article, Mithiya et al., 2019).

Both models (SARIMA and Simple Seasonal Exponential Smoothing) are based on their past values and error terms. One is parametric model and the other one is non parametric. There is a controversy between the researchers over which model is more appropriate. Some literatures have described that simple seasonal exponential smoothing should be considered more useful than SARIMA due to its weight assigning capacity. Others have opined that SARIMA is better alternative than simple seasonal exponential smoothing. However, the best of the two is the one that describes time series the best without over-fitting. But both are suitable in describing linearity of the data.

3.2 Non-linear Time Series Model

Sometimes the time series often contains nonlinear components; under such situations the linear models are not adequate for modeling and forecasting. To overcome this difficulty, a non-linear model has been successfully used. When the linear restriction of the model form is relaxed, the possible number of nonlinear structures (ARCH, GARCH, EGARCH, TAR, NAR, NMA model, etc.) that can be used to describe and forecast a time series are enormous. A good nonlinear model should be "general enough to capture some of the nonlinear phenomena in the data" (De Gooijer and Kumar, 1992). Artificial neural networks (ANNs)

model is one such that is able to approximate various nonlinearities in the data. To deal with uncertainty, linearity or nonlinearity of time series data, artificial neural networks model is the most effective method.

3.3 Artificial Neural Networks (ANNs)

Artificial neural networks are a class of flexible nonlinear models that can discover patterns adaptively from the data. Theoretically, it has been shown that given an appropriate number of nonlinear processing units, neural networks can learn from experience and estimate any complex functional relationship with high accuracy (Zhang and Qi, 2005). The most widely used ANNs in forecasting problems are multi-layer perceptron (MLPs), which use a single hidden layer feed forward network (Zhang et.al., 1998). The model is characterized by a network of three layers, viz. input, hidden and output layers, connected by acyclic links. There may be more than one hidden layers. The nodes in various layers are also known as processing elements. The three-layer feed forward architecture of ANN models can be diagrammatically depicted as follows:
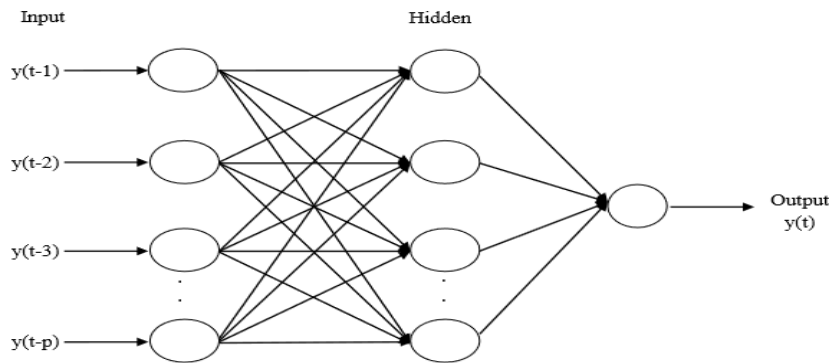

Figure 1. Structure of ANNs

The relationship between the output (y_t) and the inputs (y_{t-1}, y_{t-3}.......y_{t-p}) have the following mathematical representation:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \, g \, (\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} \, y_{t-i}) + \varepsilon_t \tag{6}$$

Here y_{t-i} (i=1, 2, ..., p) are the p inputs and y_t is the output. α_j (j = 1, 2, 3,......, q) and β_{ij} (i= 1 ,2, 3,......p, j = 1,2,3,.........q) are the connection weights and ε_t is the random shock; α_0 and β_{0j} are the bias terms. The integers p, q are the number of input and hidden nodes respectively. Usually, the logistic sigmoid function is often used as the hidden layer transfer function (nonlinear activation function), that is,

$$g(x) = {1}/{1 + e^{-x}} \tag{7}$$

Hence, the ANN model of (6) in fact performs a nonlinear functional mapping from the past observations (y_{t-1}, y_{t-2}, y_{t-3},......,y_{t-p},) to the future value y_t, i.e.,

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, y_{t-p}, w) + \varepsilon_t \tag{8}$$

Where, w is a vector of all parameters and f is a function determined by the network structure and connection weights. Thus, the neural network is equivalent to a ***non-linear autoregressive model***. Note that expression (6) implies one output node in the output layer which is typically used for one-step-ahead forecasting.

The choice of q is data dependent and there is no systematic rule in deciding this parameter. In addition to choosing an appropriate number of hidden nodes, another important task of ANN modeling of a time series is the selection of the number of lagged observations, p (the dimension of the input vector). This is perhaps the most important parameter to be estimated in an ANN model. It plays a major role in determining the nonlinear autocorrelation structure of the time series. There is also no standard rule to guide the selection of p. Hence, experiments are often conducted to select an appropriate p as well as q. Once a network structure (p; q) is specified, the network is ready for training- a process of parameter estimation. The parameters are estimated so that an overall accuracy criterion like the mean squared error is minimized.

In other words, the two essential elements that determine the ANNs are architecture structure and learning algorithm. The architecture is determined by deciding the number of layers and number of neuron nodes in each layer (Aladag et al., 2009) and there is no general rule for determining the best architecture (Zurada, 1992). The links that connect the neurons of a layer to the neurons of another layer are called weights. These weights are determined by a learning algorithm that updates their values. Non-linear least square procedures are used to estimate the connection weights. Back Propagation algorithm is one of the most used learning algorithms which updates the weights based on the difference between the output value of the ANN and the desired real value.

4. Results and Discussion

4.1 Descriptive Statistics

The study considers monthly rainfall statistics of India during the period from January 1992 to December 2016 for time series analysis. The study has taken into account 300 observations. The descriptive statistics of monthly, seasonal and yearly rainfall in India are also presented in Table 1. The table depicts that the minimum monthly rainfall is 0.80 mm whereas the maximum monthly rainfall is 334.1 mm. Therefore the range is 333.30 mm. The mean and standard deviation of monthly rainfall in India are 87.88 and 141.26 respectively.

The seasonal and total rainfall have also been considered in this study. The minimum rainfall is 91.00 mm in rabi (October-March) season, 646.80 mm in kharif (June- September) season and 856.10 mm during the year as a whole while the maximum rainfall is 226.90 mm in rabi, 957.50 mm in kharif and 1229.10 mm in whole year respectively. The mean and standard deviation of rainfall in rabi season are 154.49 mm and 35.06, while in kharif season 829.96 mm

and 77.74. The values of mean and standard deviation are 1066.49 mm and 88.09 for the entire period.

Table 1. Descriptive Statistics of Rainfall (mm)

Category		Minimum	Maximum	Mean	Range	Std. Deviation
Monthly Rainfall	(300)	0.80	334.1	87.88	333.30	92.72
Rabi Rainfall	(25)	91.00	226.90	154.49	135.90	35.06
Kharif Rainfall	(25)	646.80	957.50	829.96	310.70	77.74
Total Rainfall	(25)	856.10	1229.10	1066.49	373.00	88.09

Source: Authors' calculation, mm: Millimeters

4.2 Time Series Models

While analyzing time series data for forecasting purposes, it is very important to know the pattern of the data. The series of data may have trends, seasonality, cyclical and random variability. The trend can be easily identified and confirmed with a computer analysis of the data using statistical techniques such as fitting a trend line or use of the autocorrelation function. The trend may be positive or negative. Seasonality, defined as a structured pattern of changes within a year, is the subject matter i.e., regular wavelike fluctuations of constant length, repeating themselves. Cyclicality is in the realm of economics and is not dealt with here (Davey et al., 1993). The graphical presentation of data in figure 2 indicates that data have no trend but the seasonality is identified. The values of the autocorrelation function (Annexure Table1) have been calculated from our observations and it is clearly indicated that the series has seasonality. So the study needs to consider those types of time series models which have the power to handle the seasonal factor.
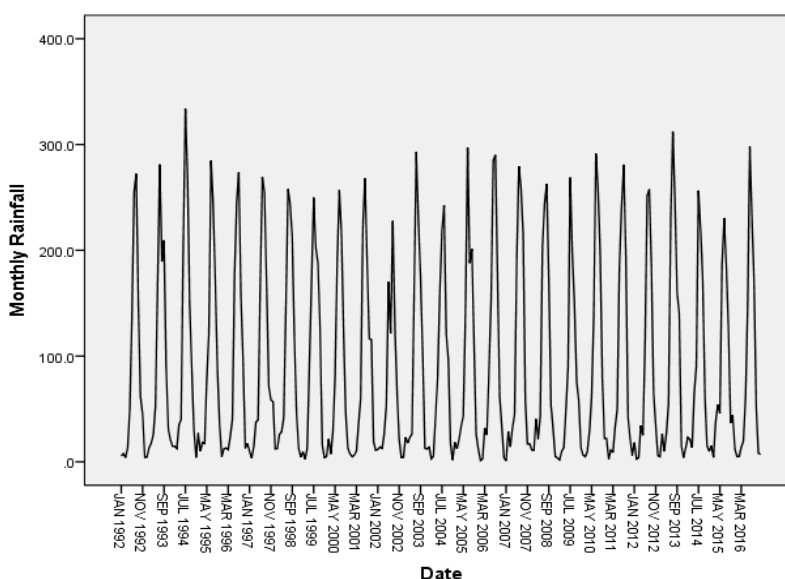

Figure 2. Sequence Chart of Monthly Rainfall of Ind

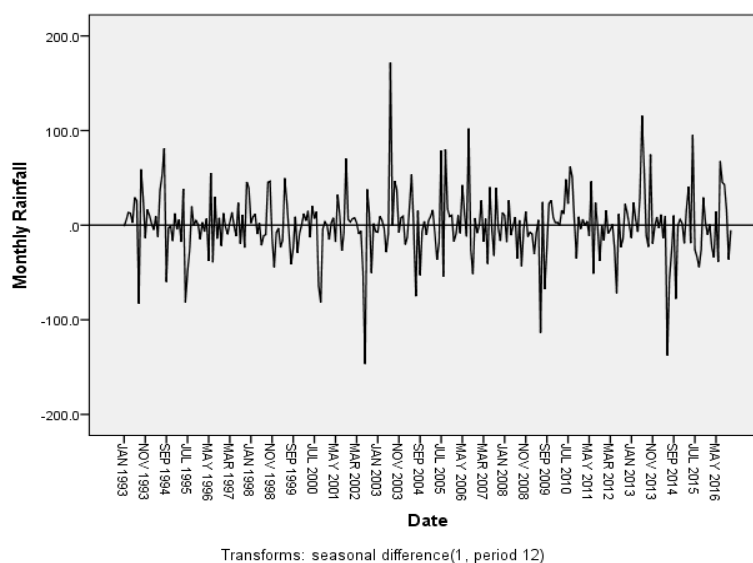Transforms: seasonal difference(1, period 12)

Figure 3. Sequence Chart of Monthly Rainfall of India after Seasonal Differences

4.2.1 Linear Model of Time Series

4.2.1.1 Simple Seasonal Exponential Smoothing

After analyzing the nature of data, the study tries to choose the appropriate linear model to explain the monthly rainfall and forecasting of rainfall in India. By using SPSS expert modeler, the study selects the simple seasonal model as the best model among the all. All parameters and model summary of the simple seasonal model are given in Table 2.

Table 2. Parameters and Model Summary of Simple Seasonal Exponential Smoothing

Division	Variable	Coefficient	SE	t-value	R^2	RMSE	MAE	BIC
Indian Monthly	Alpha (α)	0.10*	0.031	3.241	0.937	23.25	13.23	6.32
Rainfall	Delta (δ)	0.000027	0.029	0.001				

Source: Authors' Calculation

The results of the selected model indicate that the coefficient of level factor (α) is statistically significant but the coefficient of seasonal factor (δ) is insignificant. At the same time, the residual of ACF and PACF are not in the prescribed range. Hence, the study has not accepted the simple seasonal exponential smoothing model for the analysis of rainfall data and forecasting**.**

Therefore, the study has tried to find the appropriate alternative model which can explain the data showing exhibiting seasonality. Seasonal Auto Regressive Moving Average model might be an alternative.

4.2.1.2 Seasonal Autoregressive Integrated Moving Average

In the SARIMA model, the series need to be stationary. In Figure 2 above, it can be clearly observed that the pattern of the graph indicates series is stationary. There is no upward and downward trend but clear seasonal fluctuation is present.

Augmented Dickey-Fuller (ADF) test has been used to test for stationary. The Dickey-Fuller test statistic is -7.60 and the p-value is around 0.000 of the data. Since, the test statistics is less than that of critical value, the study, therefore, fails to accept H_0 (series is non-stationary) at 5 percent level and hence concludes that the alternative hypothesis H_1 (series is stationary) is accepted. Hence, the series is stationary in its mean and variance. Consequently, there is no need for non-seasonal differences. The Table 1 in the Appendix indicates that the autocorrelation coefficients at lags between t and t-12 are significant. The series has seasonal fluctuation. To remove this seasonal fluctuation from the series, seasonal differences have to be taken into account. The sequence of de-seasonal data are presented in figure 3.

4.2.1.2.1 Identification of Model

The first step of SARIMA is model identification. Here the selected SARIMA model is $(0, 0, 0)(0, 1, 0)_{12}$ which has been done using SPSS expert modeler. An identification process has been used to determine the initial values of the orders of non-seasonal and seasonal parameters, p, q, and P, Q which is obtained from the ACF and PACF plots. The ACF and PACF plots (Figure 1 & 2 in the Appendix) of the SARIMA model $(p,0,q)(P, 1, Q)_{12}$ suggest that the model will be $(0,0,0)(1,1,1)_{12}$. The alternative models are $(0,0,0)(4,1,0)_{12}$, $(0,0,0)(5,1,0)_{12}$ which have also been taken into consideration.

4.2.1.2.2 Model Estimations

Table 3. Coefficients of Fitted SARIMA Models of Indian Rainfall

Sl. No	Model	SAR(1)	SAR(2)	SAR(3)	SAR(4)	SAR(5)	SAR(6)	SMA(1)
1	SARIMA $(0,0,0)(1,1,1)_{12}$	-0.084	-	-	-	-	-	0.999
	SE	0.66	-	-	-	-	-	2.998
	t- value	-1.267	-	-	-	-	-	0.333
2	SARIMA $(0,0,0)(4,1,0)_{12}$	-0.866*	-0.592*	-0.431*	-0.193*	-	-	-
	SE	0.059	0.77	0.079	0.065	-	-	-
	t- value	-14.581	-7.665	-5.477	-2.959	-	-	-
3	SARIMA $(0,0,0)(5,1,0)_{12}$	-0.902*	-0.672*	-0.540*	-0.353*	-0.190*	-	-
	SE	0.060	0.081	0.086	0.085	0.067	-	-
	t- value	-15.149	-8.303	-6.245	-4.155	-2.847	-	-

Source: Authors' Calculation

The results of these SARIMA models are summarized in Table 3. The estimated parameters of SAR (1) and SMA (1) are not statistically significant in the model (0, 0, 0) (1, 1, 1)$_{12}$, So SARIMA (0, 0, 0) (1, 1, 1)$_{12}$ is not a appropriate model.

Whereas in the alternative model, all the coefficients are statistically significant. In addition to this, the ACF and PACF of residuals have fallen inside the 95 percent confidence levels in both models. The appropriate model is selected from two alternative SARIMA models [(0,0,0)(4,1,0) and (0,0,0)(5,1,0)] on the basis of diagnostic checking.

4.2.1.2.3 Diagnostic Checking

After estimating the parameters, the study tests the adequacy of the model using Ljung-Box (Q) statistics. In both these models, Q statistics are statistically significant. Comparing these two models, the SARIMA (0, 0, 0) (5, 1, 0)$_{12}$ model is found to be best suited for forecasting Indian rainfall. The values of RMSE, MAE and BIC in SARIMA (0,0,0)(5,1,0)$_{12}$ are lower than that in SARIMA (0,0,0)(4,1,0). The summary of the results are given in Table 4.

Table 4. RMSE, MAPE, BIC Values and Q Statistics of Fitted SARIMA Models

Sl. No	Model	RMSE	MAE	BIC	R^2	Ljung Box Q Statistics	Df
1	SARIMA (0,0,0)(4,1,0)$_{12}$	26.244	18.146	6.633	0.921	12.281	14
2	SARIMA (0,0,0)(5,1,0)$_{12}$	25.885	17.877	6.625	0.923	10.963	13

Source: Authors' Calculation

On the basis of above statistics SARIMA (0, 0, 0) (5, 1, 0)$_{12}$ is the best model. The model is as follows:

$$\Phi_5 (B^{12}) (1-B^{12})^1 Y_t = \varepsilon_t \qquad (6)$$

This model is a special case of the SARIMA model, which is called a *Seasonal Integrated Auto-Regressive Model*.

4.2.1.2.4 Forecasting

Using the identified model SARIMA (0, 0, 0) (5, 1, 0)$_{12}$ the study has forecasted the monthly rainfall in India up to December 2021. The results of the forecasted monthly rainfall have been presented in Table 5.

Table 5. Forecasting of Monthly Rainfall (mm.) in India Using SARIMA

Year	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
2019	8.6	12.9	18.4	29.2	52.4	164.0	268.9	223.2	156.7	71.2	21.3	6.8
2020	9.3	10.3	19.1	27.5	53.0	146.3	263.9	219.7	158.0	58.0	20.7	7.3
2021	8.6	8.21	19.5	31.6	50.3	161.0	263.5	219.0	154.2	57.9	23.1	6.9

Source: Authors' Calculation

The forecasted results show that the rainfall of July-August is maximum and December-January is minimum which coincides with actual series. The annual rainfall will be 1034.318 mm. and 993.715 mm. and 1004.34 mm in 2019, 2020 and 2021 respectively.

4.2.2 Non Linear Model of Time Series

4.2.2.1 Artificial Neural Network

To find the optimal number of input node (time lags) and hidden node, the study attempts to train the network with different values of learning rates and momentum rates until the mean square error and mean absolute error of the training set is very minimum and checked the testing set performance. For whatever values of the above parameters, the mean square errors of training set and testing set is minimum.

The different learning rates and momentum rates have been considered. However, it has been observed that, the different learning rates in the hidden layer and an output layer with a constant momentum rate leads to fast training. The study has been considered fixed momentum rate (value 0.05) with changing learning rates. Learning rates of 0.005 gives better results than learning rates of 0.009. The Zaitun software has been used to train and test the network.

The study developed and trained different time-lagged feed-forward neural networks (TLFN) models for forecasting monthly rainfall. It is observed that 36 -12-1 is the best model among the several TLFN models. The input layer has 36-time lags; and the output layer has one node whereas the hidden layer has 12 hidden nodes. In the model 36-12-1, 74.36 percent observations are used as training data and 25.64 percent are used as testing data. All the results are given in Table 6.

Table 6. Different ANNs Models with RMSE, MAE Statistics of Monthly Rainfall in India

Sl. No	Model	Learning Rate	Momentum	MAE	MSE	RMSE
1	48-12-1	0.005	0.05	14.33	388.15	19.701
2	48-06-1	0.005	0.05	15.49	471.54	21.714
3	36-12-1	0.005	0.05	14.13	359.49	18.960
4	36-06-1	0.005	0.05	14.90	455.89	21.351
5	24-12-1	0.005	0.05	15.61	491.63	22.172
6	24-06-1	0.005	0.05	16.11	537.89	23.192
7	12-06-1	0.005	0.05	17.46	626.84	25.036
8	36-12-1	0.009	0.05	14.59	391.44	19.785

Source: Authors' calculation

To compare the diagnostic checking parameter of the model 36-12-1 at different learning rates, 0.005 gives better results than 0.009.

Hence, 36-12-1 is the best ANN model for Indian monthly rainfall with 0.005 learning rate and 0.05 momentum.

4.2.2.2 Forecasting in ANN

Table 7 has shown that forecasted monthly rainfall in India during 2019, 2020 and 2021 using the ANN model (36-12-1).

Table 7. Forecasting of Monthly Rainfall (mm.) in India Using ANN

Years	Jan	Feb	Mar	April	May	June	July	Aug	Sep	Oct	Nov	Dec
2019	13.81	13.34	15.46	21.90	66.71	186.14	263.72	252.44	179.78	70.05	27.12	9.18
2020	10.65	13.18	20.53	30.44	65.89	171.38	244.36	243.93	160.01	61.93	26.58	11.28
2021	11.43	9.62	17.60	21.50	67.34	188.12	275.83	250.79	181.19	74.24	28.27	8.61

Source: Authors' Calculation

This rainfall is graphically represented in fig. 4. The Y-axis represents the volume of rainfall in millimeters and X-axis indicates time i.e, years. The black and white part of the figure is indicates actual values of rainfall and blue or highlighted portion of the figure is representing predicted values of rainfall.
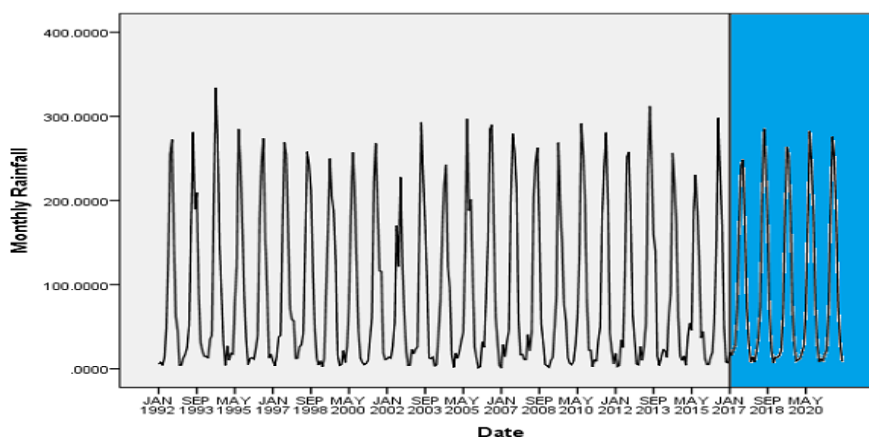

Figure 4. Forecasted Monthly Rainfall Using ANN Model

4.3 Linear Model (3.1) Vs. Non Linear Model (3.2)

Now the question is which model performs better when it comes to analyzing time series data of rainfall and forecasting future rainfall. To select this, the study has considered the values

of model selection parameters like MAE, MSE, and RMSE. The values of all these parameters are comparatively lower in the case of non-linear (ANN) model than the linear (SARIMA) model which are presented in Table 8. Hence, it can be conclude that Non-linear Artificial Neural Network model is the best one to explain monthly rainfall of India and to forecast rainfall accurately.

Table 8. Comparison of SARIMA and ANN in Diagnostics Checking Statistics

Sl. No	Model	MAE	MSE	RMSE
1	ANN (36-12-1)	14.13	359.49	18.960
2	SARIMA $(0,0,0)(5,1,0)_{12}$	17.877	670.033	25.885

Source: Authors' Calculation

4.4 Discussion

The study has estimated the year to year fluctuations in rainfall and agricultural production. The results are presented in Table 1 in the Appendix. The table shows that the years 1995, 2002, 2004, 2009, 2012 and 2014 have received less (deviation is more than 10 percent) rainfall than their respective previous years. While the years 1994, 2003, 2010, 2013 and 2016 have received more (deviation is more than 10 percent) rainfall compared to the preceding years. The agricultural production has changed with change in rainfall. It is obvious that the fluctuation in rainfall has influenced the level of output in most of the cases and agricultural production has increased with an increase in rainfall and vice versa.

From table 1in the appendix, it can be observed that a good rainfall positively influences kharif production and at the same time provides water for irrigation during rabi season. Hence, good and timely rainfall indicates better production and stable prices. Keeping in mind the impact of rainfall on agricultural production, the study has attempted to forecast monthly rainfall for India with the help of time series analysis based on monthly rainfall data.

The diagnostic checking parameters of analyzed linear (SARIMA) and non- linear (ANN) time series models have been presented in Table 8. It is clear that the non-linear model performs better in explaining monthly rainfall data and accurate forecasting. The results match those obtained by Karuiru et al. (2016).

The forecasted values of monthly rainfall in our selected model under Artificial Neural Networks are given in Table 7. The volume of precipitation is higher in the month of July-August and lower in December- January. The nature of forecasted monthly rainfall is similar to that of actual monthly rainfall. The expected yearly rainfall will be 1119.65 mm, 1136.22 mm and 1134.54 mm in 2019, 2020 and 2021 respectively.

5. Conclusion

After 72 years of independence, Indian agriculture still depends on rainfall. The volume of rainfall and its distribution over time and region are uneven. Such vagaries of monsoon affects not only agricultural production but prices of agricultural commodities as well. The study has observed that the fluctuation in rainfall and that of output are highly correlated. The variation in rainfall leads to output fluctuations and thereby raises prices. The study has attempted to forecast monthly rainfall for the country as a whole with the help of time series analysis based on monthly rainfall data. Both linear and non-linear models have been used. The value of diagnostic checking parameters MAE, MSE, RMSE are lower in non-linear model compared to the linear ones. The Non-linear model of Artificial Neural Networks (ANNs) has given better results than linear models namely simple seasonal exponential smoothing and Seasonal Auto-Regressive Integrated Moving Average in explaining rainfall data. It can therefore be concluded that forecasting of rainfall with ANN is more efficient than that using SARIMA model.

References

Aladag, H., Basaran, M. A., Egrioglu, E., Yolcu, U., & Uslu, V. R. (2009). Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations. *Expert Systems with Applications*, *36*(3), 4228-4231.

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.

Chand, R., & Raju, S. S. (2009). Dealing with Effects of Monsoon Failure. *Economics and political weekly, 44*(41-42).

Chowdhury, P. (2019). *Anandabazar Patrika*. Bengali daily Newspaper in West Bengal.

Davey, A. M., & Flore, B. E. (1993). Identification of Seasonality in Time Series: A Note. *Mathematical and Computer Modelling, 18*(6), 73-81. https://doi.org/10.1016/0895-7177(93)90126-J

De Gooijer, J. G., & Kumar, K. (1992). Some recent developments in non-linear time series modelling, testing, and forecasting. *Int. J. Forecasting, 8*, 135-156. https://doi.org/10.1016/0169-2070(92)90115-P

Financial Express Bureau (2018). Economic Survey 2017-18: Agriculture-Climate change likely to lower farmers' income by 25%, New Delhi, January 30, 2018. Retrieved from https://www.financialexpress.com/budget/economic-survey-2017-18-agriculture-climate-change-likely-to-lower-farmers-income-by-25/1035560/

Gulati, A., Saini, S., & Jain, S. (2013). *Monsoon 2013: Estimating the Impact on Agriculture, Working Paper No. 269*. Indian Council for Research on International Economic Relations. Retrieved from http://www.indiaenvironmentportal.org.in/files/file/monsoon%202013.pdf

Kamath, R. S., & Kamat, R. K. (2018). Time-series Analysis and Forecasting of Rainfall at Idukki district, Kerala: Machine Learning Approach. *Disaster Advances, 11*(11), 27-33.

Karuiru, E. K., Orwa, G. O., & Kihoro, J. M. (2016). Sarima Versus Time Lagged Feed forward Neural Networks in Forecasting Precipitation. *American Journal of Theoretical and Applied Statistics, 5*(6), 359-364. https://doi.org/10.11648/j.ajtas.20160506.15

Kistner, E., Kellner, O., Andresen, J., Todey, D., & Morton, L. W. (2018). Vulnerability of specialty crops to short-term climatic variability and adaptation strategies in the Midwestern USA. *Climatic Change, 146*, 145-158. https://doi.org/10.1007/s10584-017-2066-1

Kothawale, D. R., & Rajeevan, M. (2017). Monthly, Seasonal and Annual Rainfall Time Series for All-India, Homogeneous Regions and Meteorological Subdivisions: 1871-2016, Indian Institute of Tropical Meteorology (IITM), Earth System Science Organization (ESSO), Ministry of Earth Sciences, PUNE, INDIA Research Report No.RR-138, ESSO/IITM/STCVP/SR/02(2017)/189. Retrieved from http://www.tropmet.res.in/

Mithiya, D., Mandal, K., & Datta, L. (2019). Forecasting of Potato Prices of Hooghly in West Bengal: Time Series Analysis Using SARIMA Model. *International Journal of Agricultural Economics, 4*(3), 101-108. https://doi.org/10.11648/j.ijae.20190403.13

Olatayo, T. O., Taiwo, A. I., & Iabisi, O. (2014). Statistical Modelling and Prediction of Rainfall Time Series Data. *Global Journal of Comuter Science and Technology: G Interdisciplinary, 14*(1). Retrieved from https://globaljournals.org/item/2989-statistical-modelling-and-prediction-of-rainfall-time-series-data

Pongdatu, G. A. N., & Putra, Y. H. (2018). *Seasonal Time Series Forecasting using SARIMA and Holt Winter's Exponential Smoothing, IOP Conf.* Series: Materials Science and Engineering , 407. https://doi.org/10.1088/1757-899X/407/1/012153

Renato, R. (2019). Forecasting the Sales of Console Games for the Italian Market, Econometrics. Ekonometria. *Advances in Applied Data Analysis*, *23*(3). https://doi.org/10.15611/eada.2019.3.07

Zhang, G., Patuwo, B. E., & Hu, M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*, 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zhang, P. G., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research, Computing, Artificial Intelligence and Information Technology, 160*, 501-514. https://doi.org/10.1016/j.ejor.2003.08.037

Zurada, J. M. (1992). *Introduction of Artificial Neural Systems*. St. Paul: West Publishing.

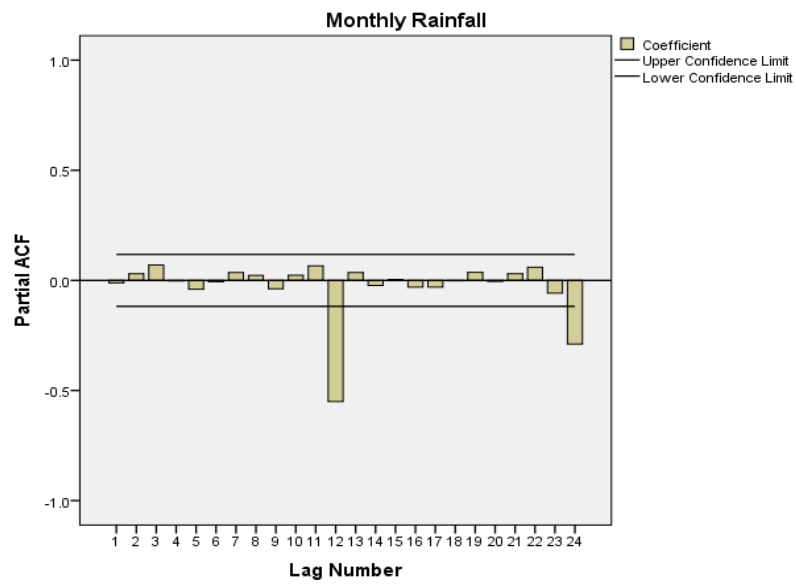**Appendix:**

Table 1. Autocorrelation Monthly Series

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	0.75	0.06	170.88	1.00	0.00
2	0.32	0.06	201.23	2.00	0.00
3	-0.14	0.06	206.88	3.00	0.00
4	-0.46	0.06	270.86	4.00	0.00
5	-0.62	0.06	388.44	5.00	0.00
6	-0.65	0.06	519.38	6.00	0.00
7	-0.60	0.06	629.29	7.00	0.00
8	-0.43	0.06	685.84	8.00	0.00
9	-0.12	0.06	690.11	9.00	0.00
10	0.32	0.06	721.46	10.00	0.00
11	0.73	0.06	887.40	11.00	0.00
12	0.89	0.06	1136.90	12.00	0.00
13	0.72	0.06	1298.99	13.00	0.00
14	0.30	0.06	1326.90	14.00	0.00
15	-0.14	0.06	1332.93	15.00	0.00
16	-0.44	0.06	1395.19	16.00	0.00
17	-0.59	0.06	1508.42	17.00	0.00
18	-0.63	0.06	1635.66	18.00	0.00
19	-0.58	0.06	1742.45	19.00	0.00
20	-0.41	0.06	1798.04	20.00	0.00
21	-0.11	0.06	1801.71	21.00	0.00
22	0.31	0.06	1832.68	22.00	0.00
23	0.70	0.06	1993.04	23.00	0.00
24	0.87	0.06	2241.37	24.00	0.00
25	0.68	0.06	2395.53	25.00	0.00
26	0.28	0.05	2421.56	26.00	0.00
27	-0.14	0.05	2428.02	27.00	0.00
28	-0.43	0.05	2489.15	28.00	0.00
29	-0.57	0.05	2599.00	29.00	0.00
30	-0.60	0.05	2720.56	30.00	0.00
31	-0.55	0.05	2822.88	31.00	0.00
32	-0.39	0.05	2874.25	32.00	0.00
33	-0.10	0.05	2877.91	33.00	0.00
34	0.29	0.05	2906.14	34.00	0.00
35	0.67	0.05	3059.79	35.00	0.00
36	0.82	0.05	3291.44	36.00	0.00

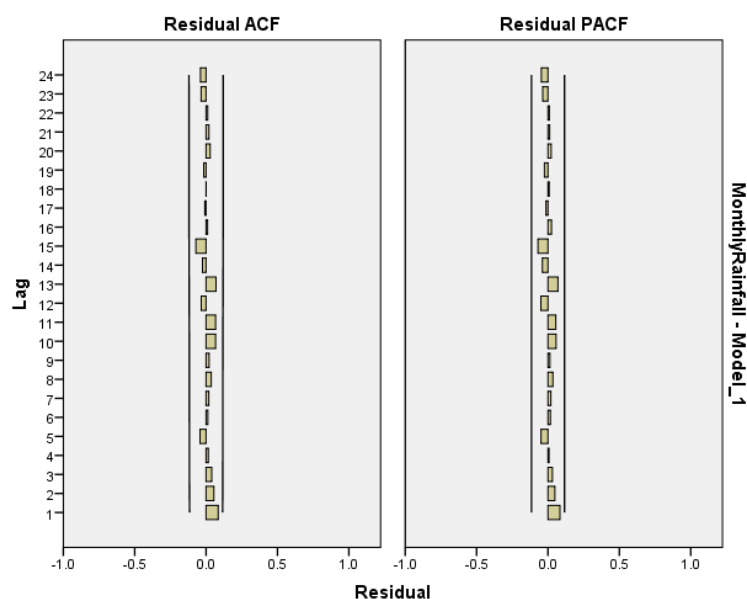a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.
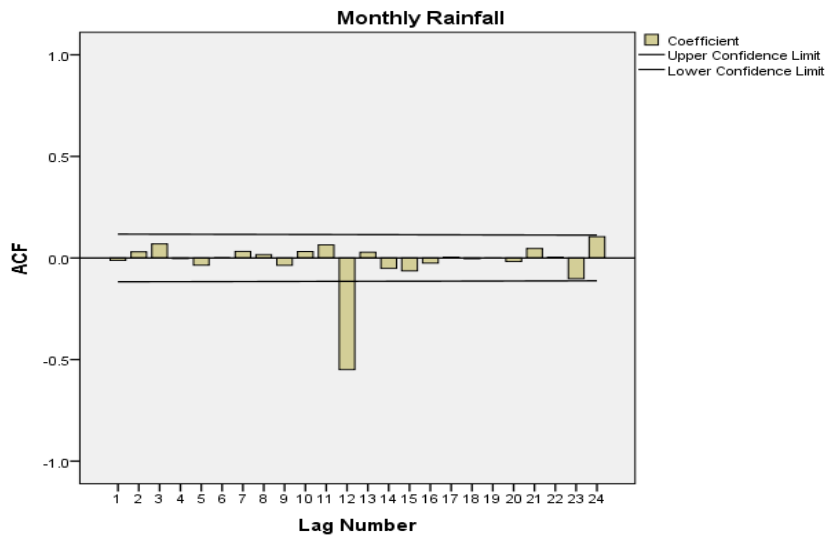
Appendix Fig 1. ACF plots after seasonal differentiation

Appendix Fig 2. PACF plots after seasonal differentiation

Appendix Fig. 3: Residual ACF and PACF

Appendix Table 2. Year to Year Fluctuation in Rainfall and Rain-fed Crops

	Kharif rainfall	Kharif Rice	Maize	Groundnut	Sugarcane	Kharif Cereals	Kharif Pulses	Kharif Oilseeds
1992								
1993	3.83	8.40	-4.64	-14.32	0.71	-0.86	-9.34	2.69
1994	13.18	2.66	-10.38	6.19	19.98	1.47	-11.11	3.22
1995	**-14.40**	**-6.51**	**8.33**	**-0.11**	**2.02**	**-6.16**	**-4.39**	**9.55**
1996	3.83	5.07	10.80	2.48	-1.26	8.78	-19.17	4.35
1997	0.41	1.70	2.79	-4.87	0.71	-1.15	-20.40	3.49
1998	0.34	0.26	1.15	17.11	3.28	0.49	19.85	11.74
1999	-5.78	6.55	1.76	-45.01	3.67	2.99	-6.32	-21.01
2000	-2.72	-6.07	5.29	29.21	-1.12	-3.04	-7.64	-4.33
2001	1.11	10.64	10.07	14.51	0.42	9.83	8.77	10.72
2002	**-18.59**	**-21.66**	**-17.57**	**-44.95**	**-3.31**	**-22.53**	**-14.20**	**-32.10**
2003	32.36	24.63	37.34	121.65	18.62	33.42	48.51	85.74
2004	**-13.36**	**-8.13**	**-9.88**	**-23.29**	**1.38**	**-11.05**	**-23.48**	**-15.13**
2005	11.70	8.36	5.92	19.68	18.59	6.51	3.13	18.50
2006	9.27	2.43	-4.93	-47.69	26.44	0.74	-1.43	-16.43
2007	4.62	3.16	30.72	123.50	-2.06	8.33	33.53	47.83
2008	-6.85	2.72	-6.53	-23.71	-18.14	-0.96	-26.82	-14.03
2009	**-20.43**	**-10.59**	**-12.94**	**-31.42**	**2.55**	**-12.07**	**-10.29**	**-11.68**
2010	26.17	6.24	35.34	72.44	17.13	14.01	69.36	39.38
2011	1.94	14.94	-0.91	-22.82	5.45	10.05	-14.92	-5.62
2012	**-10.83**	**-0.42**	**-81.94**	**-37.81**	**-5.49**	**-2.44**	**-2.35**	**0.28**
2013	17.66	-0.93	20.16	152.75	3.21	0.44	1.35	8.98
2014	**-22.26**	**-0.11**	**5.85**	**-26.40**	**2.89**	**-0.29**	**-4.38**	**-15.14**
2015	-1.13	0.02	-21.25	-9.49	-3.83	-2.27	-3.53	-13.08
2016	16.04	5.35	0.20	12.67	-12.16	7.68	73.31	29.05

ABBREVIATION:

ADF: Augmented Dickey-Fuller

ANNs: Artificial Neural Networks

ARIMA: Auto Regressive Integrated Moving Average

ETS: Exponential Smoothing State Space

IMD: India Meteorological Department

MAE: Mean Absolute Error

MLPs: Multi-Layer Perceptron

Mm: Millimeter

MSE: Mean Square Error.

RMSE: Root Mean Square Error

SARIMA: Seasonal Auto Regressive Integrated Moving Average

TLFN: Time-Lagged Feed Forward Neural Networks