A comparative study on ARIMA and ANN for rainfall pattern of Bangalore rural district

S. Bhavyashree and Banjul Bhattacharryya

Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur-741252, Nadia, West Bengal

Received : 24-01-2019 ; Revised : 26-05-2019 ; Accepted : 31-05-2019

ABSTRACT

Rainfall is the primary source of water for the world. It is very important in our day to day life right from daily consumption to agriculture and crop planning. In this paper, we have used two attractive tools for modeling and predicting the behavioral pattern in rainfall phenomena based on past observations e., Autoregressive Integrated Moving Average (ARIMA) and the a machine learning technique Artificial Neural Network (ANN). For the study, The monthly rainfall data over Bangalore rural district of Karnataka, India for active monsoon period (May31st to September 30th) was collected for the period of 34 years (1980-2013) from AICRP on Agro Meteorology, University of Agricultural Sciences, GKVK, Bengaluru and India Meteorological Department, Pune. The model efficiencies are tested and compared.

***Keywords* :** ARIMA, ANN, monthly monsoon rainfall**.**

1. INTRODUCTION

'Rainfall' is a natural climatic phenomenon, which is a major component of water cycle whose prediction is challenging and demanding. Its forecast is of particular relevance to agriculture sector, and is responsible for depositing most of the fresh water on the Earth. It provides suitable conditions for many types of ecosystem as well as water for hydroelectric power plants and crop irrigation. which contributes significantly to the economy of the nation. Numerous attempts have been made to predict its behavioral pattern using various techniques worldwide. The accuracy of forecasting rainfall is very important due to the current Global climate change. In the present work, a comparative study of rainfall behavior as obtained by autoregressive integrated moving average (ARIMA) and the artificial neural network (ANN) technique is made to compare the efficiency of models in fitting and future prediction. The former is basically a linear statistical technique and has been quite popular for modeling the time series and rainfall forecasting due to ease in its development and implementation. In contrast, the application of the ANN in time series for forecasting is relatively new (Mirko and Christian 2000).

Many researchers attempted to predict using the ARIMA and ANN method in the field of hydrology. Somvanshi V.K *et al.* (2006) conducted a study to evaluate prediction efficiency of the Hyderabad region by using the mean annual rainfall data of 104 years. To evaluate the prediction efficiency they basically used two different approaches for designing a model viz., ARIMA and ANN. The approaches of ARIMA and ANN are used to the data to obtain regression coefficients and weights respectively. The experiment concluded that the ANN model outperformed the ARIMA and found appropriate forecasting tool for rainfall prediction. Abhishek, *et al*. (2012), conducted research using neural networking of rainfall prediction in Udupi district of Karnataka, India. In this paper, we take advantage of the seasonal nature of rainfall to use Seasonal ARIMA model and ANN technique to fit and predict future rainfall of Bangalore rural district of Karnataka, India. Farajzadeh *et al.* (2014) applied feed-forward neural network and ARIMA model to forecast the monthly rainfall in Urmialake basin located in northwestern Iran. Their results showed that the estimated values of monthly rainfall through feed-forward neural network were close to the ARIMA model.

However, literature has shown different view on the relative performance and superiority of ARIMA and ANNs models to time series prediction, especially for different data used; hence the need for further study that can help unified a coherent view on the better methodology. This paper therefore seeks to further clarify conflicting opinions reported in literature on the superiority of ANN model over ARIMA model and vice versa in the effective prediction of rainfall pattern.

E-mail: bhavyashree5591@gmail.com

MATERIALS AND METHOD

Data

The data for monthly rainfall data over Bangalore rural district of Karnataka, India for active monsoon period (May31st to September 30th) was collected for the period of 34 years (1980-2013) from AICRP on Agro Meteorology, University of Agricultural Sciences, GKVK, Bengaluru and India Meteorological Department, Pune.

Both the models were tested using the training data set for the period 1980-2010 and holdout dataset 2011-2013.

ARIMA Model

Box and Jenkins (1970) developed this forecasting technique which is still very popular among hydrologists. The autoregressive integrated moving average ARIMA (p,d,q) model of the time series $\{r_1, r_2 \ldots..\}$ is defined as,

$$\varphi(B)\Delta^d r_t = \theta(B) e_t$$

where,

r_t and e_t respectively represent mean annual rainfall time series and random error terms at time t.

B is the backward shift operator defined by Bry = ry-1 and related to Δ by Δ = 1- B;

d is the order of difference.

The φ(B) and θ(B) of order p and q are defined as,

$$\Phi(B)=1- \varphi_1 B\ ,\ \varphi_2 B^2\ ,\ \ldots\ldots \varphi_p B^P$$
$$\theta(B)=1- \theta_1 B\ ,\ \theta_2 B^2\ ,\ \ldots..\theta_q B^q$$

Where,

φ_1 , φ_2 , ……φ_p are the autoregressive coefficients and

θ_1 , θ_2 , …..θ_q are the moving averages coefficients

In this ARIMA (p,d,q) modeling, the first step is to determine whether the time series is stationary or non stationary. If it is non stationary it is transformed into a stationary time series by applying suitable degree of differencing by selecting proper value of d. The appropriate values of p and q are chosen by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series.

ANN Model

ANN is a data-driven mathematical model that was developed to imitate the structure of a human brain neural network whose speed and efficiency has been always fascinating to researchers for quite a long time. It has been widely applied to solve problems such as prediction and discrimination. Its power comes from the parallel processing of the information from data. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data. Single hidden layer feed forward network is the most widely used model form for time series modeling and forecasting. The back propagation network (BPN) is one of the neural network algorithm which is formalized by Lippmann (1987) and Rummelhart and McClelland (1986). It has been extensively used for inversion, prediction that consist of two passes: a forward pass and a backward pass. In the forward pass the input is applied to input layer and its effect is propagated through network, layer by layer. The net effect is computed as the weighted sum of the output of the neurons of the previous layer. The sum of squared deviation of the output from the target value at the nodes of the output layer defines the error signal that is to be propagated back to previous layers such that the parameters are adjusted to minimize the error in further computations.

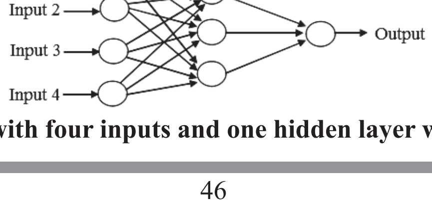

Fig. 1. A neural network with four inputs and one hidden layer with three hidden neurons

As shown in Fig.1, an ANN consists of layers of neurons. The model is characterized by a network of three layers of simple processing units, which are connected to each other. The first layer, which receives input information, is called an input layer. The last layer, which produces output information, is called an output layer. Between output and input layers are hidden layers. There can be one or more hidden layers. Information is transmitted through the connections between nodes in different layers. There can be one or more hidden layers. Information is transmitted through the connections between nodes in different layers. For the analysis R 3.2.2 has been used.

Neural Network Auto regression (NNAR)

The lagged values can be used as inputs to a neural network for time series data. In the present study, feed-forward networks with one hidden layer have been considered. The notation NNAR (p,k) is used to indicate there are p lagged inputs and k nodes in the Material and Methods | 52 hidden layer. NNAR (p, 0) model is identical to an ARIMA (p,0,0) model but here, there will not be any restrictions on the parameters to ensure stationarity.

RESULTS AND DISCUSSION

Table 1 : ARIMA model

ARIMA Model	Parameter	Estimate	Std Error	p	σ^2	AIC	Box Pierce	
							X^2	P Value
$(1,0,2)(1,0,1)_{12}$	AR1	-0.7539	0.8989	0.0000	4934	1945.10	7.14	0.850
	MA1	0.8199	0.8954	0.0000				
	MA2	0.0576	0.0831	0.0000				
	SAR1	0.9800	0.0298	0.0000				
	SMA1	-0.9016	0.0770	0.0000				
	Intercept	113.475	15.661	0.0000				

Table 2: NNAR model

NNAR Model	*p*	*P*	Size (*k*)	Maxit	σ^2
NNAR(13,1)	13	1	1	500	3233

Table 3: Forecast accuracy of training and testing period for ARIMA and NNAR

Dataset	Method	RMSE	MAE	MAPE
Training	ARIMA	75.5501	58.3967	106.1344
	NNAR	56.8575	47.5606	94.5135
Testing	ARIMA	69.5945	55.6432	97.2820
	NNAR	22.7015	18.3413	20.5675

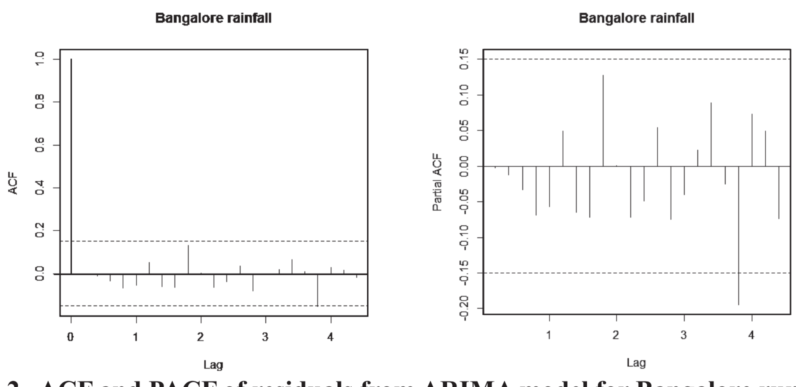

Fig. 2. ACF and PACF of residuals from ARIMA model for Bangalore rural district

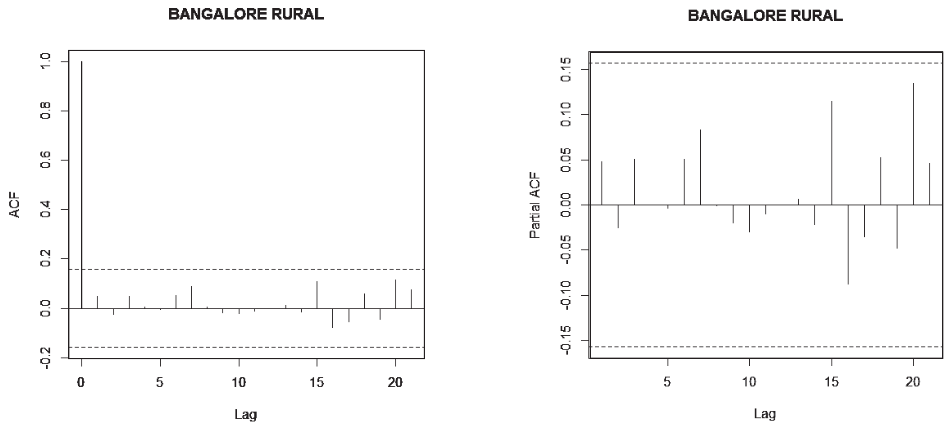

Fig. 3. ACF and PACF of residuals from NNAR model for Bangalore rural district

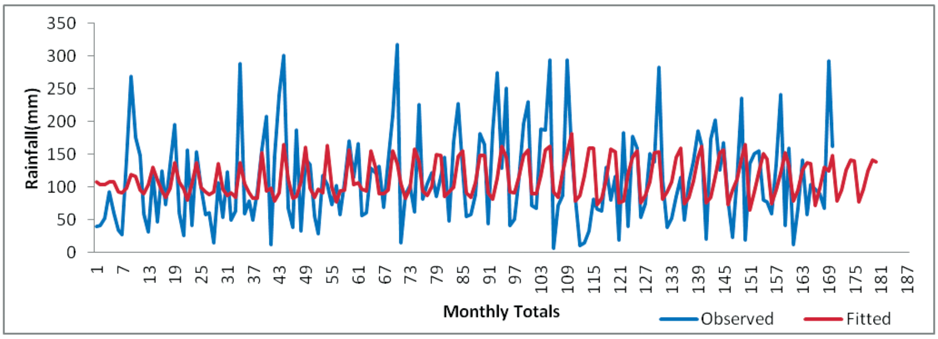

Fig. 4. Observed vs. fitted (ARIMA) monthly monsoon rainfall for Bangalore rural district

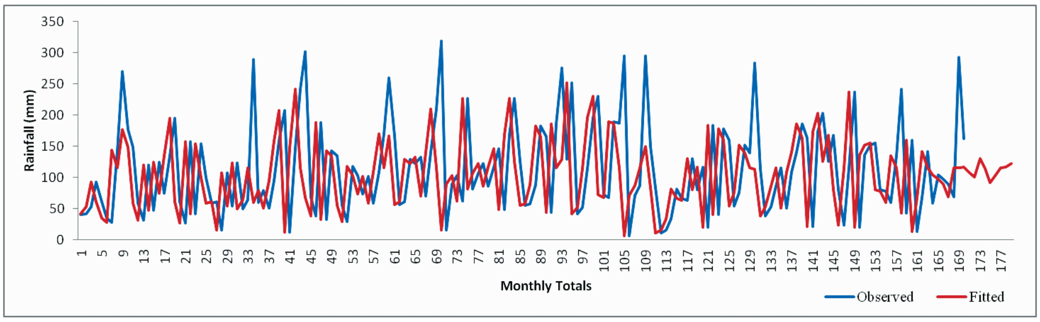

Fig. 5. Observed vs. fitted (ANN) monthly monsoon rainfall for Bangalore rural district

Table 1 and 2 represents ARIMA and ANN models with parameter estimates and from Fig. 2 and 3 it is clearly seen that when compared to ARIMA, the plots for ACF and PACF of ANN do not show considerable autocorrelations among the residuals as all autocorrelations and partial autocorrelations lie between 95% control limits

From Fig. 3 and 4, it shows the plot of observed versus predicted values of the monthly rainfall data for active monsoon period for 34 years (1980-2013) by the ARIMA and ANN models respectively. The ANN model fits extremely well with the actual data values as compared to the ARIMA model.

The performance measures of ARIMA and ANN models in terms of numerical computations are shown in table 3. The table indicates that the ANN model is superior then ARIMA model. The MAE error for training data set and test data set for ARIMA model is 58.3967 and 55.6432 respectively. While the same error measure is considerably lower at 47.5606 and 18.3413 for the ANN model. Similarly, in case of RMSE and MAPE also, the ANN forecast outperforms when compared with ARIMA forecast.

Hence the study clearly depicts that, ANN model can be preferred over ARIMA model an appropriate forecasting tool for modeling and forecasting of monthly monsoon rainfall.

REFERENCES

Kumar, A., Abhay Kumar, Rajeev Ranjan, Sarthak Kumar. (2012) A Rainfall Prediction Model using Artificial Neural Network. *IEEE Control and System Graduate Research Colloquium.*

Somvanshi, V. K., Pandey, O. P., Agrawal, P. K., Kalanker1, N. V., Ravi Prakash, M. and Ramesh, C. (2006). Modelling and prediction of rainfall using artificial neural network and ARIMA techniques. *Journal of Indian Geophysical Union,* **10:** 141-151.

Mirko Van der Baan and Christian Jutten, 2000. Neural network in geophysical Applications, *Geophysics*, **65 (4)**:1032-47.

Jamileh Farajzadeh, Ahmad Fakheri Fard and Saeed Lotfi (2014). Modeling of monthly rainfall and runoff of Urmia lake basin using "feed-forward neural network" and "time series analysis" model. *Water Resources and Industry,* 7**–8**:38-48.

Box, G. and Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco. *Open Journal of Statistics*, 7**(4)**:575.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4–22.

Rumelhart DE and McClelland (1986) Parallel Distributed Processing: Explorations in the Microstructure of Cognition MIT Press, Cambridge, MA.