RASHI 3 (2) : 65 - 73 (2018) ISSN : 2581 9755

Forecasting monthly rainfall using artificial neural network

Soumen Pal and [1]Debasis Mazumdar

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

[1]Bidhan Chandra Krishi Viswavidyalaya, West Bengal

Received : 10-03-2019 ; Revised : 25-04-2019 ; Accepted : 30-04-2019

ABSTRACT

Rainfall is one of the most difficult climatic variables to be forecasted. Due to complex relationship with other weather parameters, precipitation often comes out to be erratic and irregular in many cases. Being primary source for survival and agricultural production in major parts of the world, rainfall draws attention of the research workers to model and forecast the amount of precipitation precisely. In Indian context, prediction of rainfall is mainly important for the monsoon months (June to September) as more than 75% of the rainfall, in the country, is received during this season. It is always beneficial to study the rainfall in regional scale, such as block or district, for higher precision level that results effective crop and hydrological planning of the region. Artificial Neural Network (ANN) model, in many situations, is capable in explaining the behaviour of complex system, such as rainfall, to a substantial extent. In the present study, modelling of monthly rainfall (during monsoon) time series for Birbhum and North 24 Parganas districts of West Bengal state have been done employing promising Neural Network Autoregression (NNAR) technique. Lagged values of the time series were used as inputs to a multilayer feed-forward neural network with one hidden layer. The weights assigned to each node were learned and a nonlinear (sigmoid) activation function was applied in the hidden layer. NNAR was compared with traditional Autoregressive Integrated Moving Average (ARIMA) method for both in-sample and out-of sample forecast accuracy using different diagnostic measures. Predictive power of NNAR over ARIMA model, for the data under consideration, was established and residuals, obtained from the former, were found to be non-autocorrelated. Finally, forecasting of monthly rainfall during monsoon season was done from 2019 till 2020 for the two districts.

Keywords **:** ANN, ARIMA, Forecasting, Monsoon rainfall, NNAR.

1. INTRODUCTION

Agricultural performance of a country, generally, depends to a large extent on the quantum and distribution of rainfall. In India, the nation's agricultural planning is primarily dependent on the reasonably accurate prediction of the total amount of rainfall from the beginning of June to the end of September. This kind of prediction comes under the category of long range forecast (LRF). The period for LRF varies from 12-30 days up to two years (Das *et al.*, 2010). On the basis of LRF, various precautionary measures may be planned and adopted. If an LRF indicates below-normal rainfall, then necessary agricultural commodities can be purchased from the international market well in advance. Also, adequate arrangements could be made for the transport, storage, and distribution of such products. The government authorities can work out various plans and schemes to counter the adverse situation in well advance, and the strategies can be used at various levels, such as states, districts, and villages. India Meteorological Department (IMD) has been issuing LRF of the southwest monsoon rainfall since 1886. However, many statistical and dynamic models of IMD often fail to predict monsoon accurately (Rajeevan *et al.*, 2007). Prediction of Indian Summer Monsoon Rainfall is of vital importance for Indian economy, and it has remained a great challenge for hydro-meteorologists due to inherent complexities in the climatic systems (Kashid and Maity, 2012).

Gangetic West Bengal (GWB) meteorological sub-division is an agro-based and economically important area of India. It has tremendous climatic complexity due to ecological differences from one place to another and therefore, always experiences variation in spatial distribution of precipitation. Hence, a proper study of rainfall is critical to crop and hydrological planning for this region. In the present study, two sample districts viz. Birbhum and North 24 Parganas of GWB region have been selected randomly for modelling and forecasting of monsoon rainfall.

For the LRF of summer monsoon rainfall, a number of approaches are used. One of the popular methods, among these, is the empirical method based on a time series analysis. This method uses only the time series of past rainfall data (Goswami and Srividya, 1996; Kishtawal *et al.* 2003; Iyengar and Kanth, 2005) and do not use any predictors. This approach can be explored in modelling and accurate forecasting of monsoon rainfall of a region for planning and policy making.

Stochastic modelling and forecasting of monthly rainfall in different regions around the world have been tried since long back by many researchers. Goswami and Srividya (1996) explored a generalized structure of a neural

E-mail : soumen.4345@gmail.com

network to provide consistent prediction of all-India annual mean rainfall with good accuracy. The performance of this methodology was evaluated and the generalized neural network was found to be more consistent than the conventional neural network models. Guhathakurta (2008) employed Artificial Neural Network (ANN) technique to forecast the monsoon rainfall of 36 meteorological sub-divisions of India. In this study, the idea of up-scaling was introduced in monsoon rainfall prediction using the ANN. It was shown that up scaling helped to capture the variability of the all India rainfall better. Chattopadhyay and Chattopadhyay (2010) attempted ARIMA and Autoregressive Neural Network (ARNN) model to develop a univariate model to forecast the summer monsoon rainfall over India. They obtained a three-three-one architecture of the ARNN model and the supremacy of ARNN has been established over ARIMA. Yusof and Kane (2012) found both seasonal ARIMA and state space models based on exponential smoothing adequate for modelling of rainfall time series of two selected weather stations viz. Malacca and Kuantan in Malaysia. Ramana *et al*. (2013) attempted to find an alternative method for monthly rainfall prediction of Darjeeling rain gauge station of India by combining the wavelet technique with ANN. The results of monthly rainfall series modelling indicated that the performances of wavelet neural network models were more effective than the ANN models. Singh and Borah (2013) used feed-forward back-propagation neural network algorithm for Indian summer monsoon rainfall forecasting. They have proposed five neural network architectures which exhibited superiority over the existing models. Farajzadeh *et al*. (2014) applied feed-forward neural network and ARIMA model to forecast the monthly rainfall in Urmia lake basin located in northwestern Iran. The results showed that the estimated values of monthly rainfall through feed-forward neural network were close to ARIMA model.

In the present study, both the ARIMA and ANN methodologies have been adopted for estimating rainfall trend and the superiority of ANN over ARIMA is established. The rest of this paper is organized as follows: the next section describes the data source and methodology used to carry out the research work followed by discussion on results and finally conclusion.

2. MATERIALS AND METHODS

2.1 Data Source

District-wise monthly rainfall data (MRD) for the districts from 1901 to 2003 have been collected from IMD, Pune (NCC Research Report, 2011) for conducting the present study. From 2004 to 2018, the MRD, which is maintained by IMD, has been collected online from *http://www.indiawaterportal.org* (2004-2008) and *http://www.imd.gov.in* (2009-2018) for those districts. Around 2% of the total observations, which are missing in the time series, have been replaced by 100 year's (1901-2000) average monthly rainfall values. All the observations are measured in millimeter (mm).

Both Seasonal ARIMA and ANN methodologies have used in the present study to model and forecast of monthly rainfall volume during monsoon.

2.2 Seasonal ARIMA (SARIMA) Model

When seasonality is added to ARIMA model, then the ARIMA notation can be extended to handle seasonal aspects, and the general shorthand notation is ARIMA (*p, d, q*)(*P, D, Q*)$_s$ where $s =$ number of periods per season (Makridakis *et al.* 1998). The equation for the ARIMA model can be written as follows :

$$\left(1-\phi_1 B-\cdots-\phi_p B^p\right)\left(1-\Phi_1 B^s-\cdots-\Phi_P B^{Ps}\right)\left(1-B^s\right)^D\left(1-B\right)^d Y_t$$
$$=\left(1-\theta_1 B-\cdots-\theta_q B^q\right)\left(1-\Theta_1 B^s-\cdots-\Theta_Q B^{Qs}\right)e_t \tag{1}$$

The estimation procedures for SARIMA model are available only for stationary series. Unit root tests are used to determine the stationarity in the time series. In the present study, *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) test has been used to check the stationarity in the rainfall time series. This test is used for testing a null hypothesis that an observable time series is stationary around a deterministic trend. Kwiatkowski *et al.* (1992) provided straight forward test of the null hypothesis of trend stationarity against the alternative of a unit root. For this, they considered 3-component representation of the observed time series y_1, y_2, ..., y_n as the sum of a deterministic time trend, a random walk and a stationary residual :

$$y_t = \xi t + r_t + \varepsilon_t . \tag{2}$$

Here r_t is a random walk :

$$r_t = r_{t-1} + \mu_t , \tag{3}$$

where the u_t are iid (0, σ_u^2). The initial value r_0 is treated as fixed and serves the role of an intercept. The null and the alternative hypothesis are formulated as follows :

$H_0 : y_t$ is trend stationary or $\sigma_u^2 = 0$,

$H_1 : y_t$ is a unit root process.

To determine the number of seasonal differences required for time series to be made stationary, the Osborn-Chui-Smith-Birchenhall (OCSB) (1988) test has been used (with null hypothesis that a seasonal unit root exits) in the present study. The initial values for the orders of seasonal and non-seasonal parameters in the SARIMA model have been identified using autocorrelation function (ACF) and partial autocorrelation function (PACF). The appropriate SARIMA model is chosen on the basis of minimum Akaike Information Criterion (AIC), due to Akaike (1974), and is given by

$$AIC \approx n\left(1 + \log(2\pi)\right) + n \log \sigma^2 + 2m, \tag{4}$$

where $m = p + q + P + Q$ i.e. the number of parameters estimated in the model, σ^2 is the variance of the residuals, n is the number of observations in the series and L is the likelihood function.

For the specific case of a linear model with homogeneous errors, Hurvich and Tsai (1989) derived a corrected AIC (AIC_c) which includes a correction for small sample sizes as

$$AIC_c = AIC + \frac{2m(m+1)}{n-m-1}. \tag{5}$$

As an alternative to AIC, Bayesian Information Criterion (BIC), due to Schwarz (1978), can also be used and is given by

$$BIC = -2 \log L + m \log(n). \tag{6}$$

Here, L is the likelihood function.

The parameters of the SARIMA model have been estimated using Maximum Likelihood Estimation (MLE).

2.3 Neural Network Autoregression (NNAR)

There is increasing interest in using neural networks to model and forecast time series (Zhang, 2012). In the present study, feed-forward neural network with one hidden layer has been considered to model and forecast monthly rainfall during monsoon. The notation NNAR (*p, k*) is used to indicate there are *p* lagged inputs and *k* nodes in the hidden layer. A NNAR (*p,* 0) model is equivalent to an ARIMA (*p,* 0, 0) model but without the restrictions on the parameters to ensure stationarity. With seasonal data, it is useful to also add the last observed values from the same season as inputs. In general, an NNAR $(p, P, k)_m$ model has inputs $\left(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, y_{t-m}, y_{t-2m}, \ldots, y_{t-Pm}\right)$ and *k* neurons in the hidden layer. An NNAR $(p, P, 0)_m$ model is equivalent to an ARIMA $(p,0,0)(P,0,0)_m$ model but without the restrictions on the parameters to ensure stationarity. Here, the relationship between the output (y_t) and the inputs $\left(y_{t-1}, \ldots, y_{t-p}\right)$ can be represented by the following model (Khashei and Bijari, 2010) :

$$y_t = w_0 + \sum_{j=1}^{q} w_j g\left(w_{0,j} + \sum_{i=1}^{p} w_{i,j} y_{t-i}\right) + \varepsilon_t, \tag{7}$$

where, $w_{i,j}\left(i = 0,1,2,\ldots,p; j = 1,2,\ldots,q\right)$ and $w_j\left(j = 0,1,2,\ldots,q\right)$ are model parameters or connection weights; *p* is the number of input nodes; q is the number of hidden nodes; *g*(.) is the hidden layer sigmoid transfer function; and ε_t is the random error at time *t*. Hence, the ANN model here, in fact, performs a nonlinear functional mapping from past observations to the future value y_t, i.e.,

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

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 nonlinear auto-regressive model. The weights are selected in the neural network framework using a *learning algorithm* that minimizes Mean Squared Error (MSE) (Hyndman and Athanasopoulos, 2014).

2.4 Forecast accuracy measures

A comparative study on performance of two methods viz. SARIMA and NNAR is carried out from the viewpoint of multi-step-ahead forecasts on the basis of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE). If y_t denotes the observation at t-th time point and $\hat{y}_t$, denotes forecast of y_t, then the forecast error is $e_t = y_t - \hat{y}_t$.

Accordingly,

$$\text{MAE} = \text{mean}\ (|e_t|), \tag{9}$$

$$\text{RMSE} = \sqrt{\text{mean}\left(e_t^2\right)}, \tag{10}$$

$$\text{MAPE} = \text{mean}\ (|p_t|), \tag{11}$$

where, $p_t = 100e_t / y_t$. The MASE was proposed by Hyndman and Koehler (2006) as an alternative to using percentage errors when comparing forecast accuracy across series on different scales as below:

$$\text{MASE} = \text{mean}\ (|q_t|), \tag{12}$$

where, $q_t = \dfrac{e_t}{\dfrac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}$.

2.5 Diebold-Mariano (DM) test

DM test is used for in order to test the predictive power of one model against the other in use. If there are two models estimated at the same data sample, generating forecasts $y_1^F(t)$ and $y_2^F(t)$, then forecast errors are defined as

$$\begin{aligned} e_1 t &= y(t) - y_1^F(t), \\ e_2 t &= y(t) - y_2^F(t). \end{aligned} \tag{13}$$

The mean loss is defined as

$$\bar{d} = \frac{1}{H}\sum_{t=1}^{H}\left[g(e_1(t)) - g(e_2(t))\right], \tag{14}$$

where H is the number of forecasts generated by each model. If the loss time series $d(t) = g(e_1(t) - e_2(t))$ is auto correlated of order q, then provided that both forecasts are equally adequate, the DM statistic (Diebold and Mariano, 1995) is defined as

$$S = \frac{\bar{d}}{\sqrt{\dfrac{\gamma_0 + 2\gamma_1 + \ldots + 2\gamma_q}{H-1}}} \sim t_{H-1}. \tag{15}$$

Here, γ_q is the autocovariance at lag q.

As far as the functional form of $g(.)$ is concerned, a common approach is to use the square (Bryzgalova and Gelman, 2008). The null of equal predictive accuracy is rejected at the 5% level if $|S| > 1.96$.

3. RESULTS AND DISCUSSION

Out of total 13 districts of GWB, 2 sample districts have been selected in random for fitting of the monthly monsoon rainfall data to both SARIMA and NNAR models. The whole dataset (1901-2018) is divided into training (1901-2013) and test/holdout sample (2014-2018).

3.1 Time Series Modelling using SARIMA

Before using SARIMA, the stationarity of the series is tested by using unit root tests and differencing is done, if necessary, based on the result of the KPSS and OCSB test. The appropriate SARIMA model, for each of the time series, is selected on the basis of minimum AIC or AIC_c or BIC value. Using maximum likelihood estimation, the estimates of the parameters, along with their standard errors have been calculated (Table 1). The associated values are calculated using a two-sided test from the normal probability table.

Table 1. SARIMA model with parameter estimates for 2 districts

District	SARIMA Model	Parameter	Estimate	Std. Error	*p*
Birbhum	(0,1,1) (2,0,0)	MA1	-0.9895	0.0096	<0.0001
		SAR1	0.1085	0.0475	0.0223
		SAR2	0.1445	0.0478	0.0025
North 24 Parganas	(2,2,0) (0,0,0)	AR1	-0.5219	0.0557	<0.0001
		AR2	-0.8890	0.0322	<0.0001
		MA1	0.5689	0.0370	<0.0001
		MA2	0.9750	0.0151	<0.0001

(SAR: Seasonal Autoregression)

The p value in the above table gives a method of testing the significance of each parameter separately. In this case, all parameters are highly significant since all p values are very small (<0.05) showing each of the terms in the models are required.

3.2 Time Series Modelling using NNAR

For each of the districts, a feed-forward neural network model is fitted with lagged values of the time series data as inputs and a single hidden layer with a number of nodes. If p and P are the number of non-seasonal and seasonal lags respectively, then the inputs are for lags 1 to p , and lags m to mP where m is the frequency of the time series i.e. 4 in this case. For non-seasonal time series, the default is the optimal number of lags (according to the AIC) for a linear AR (p) model (Hyndman and Khandakar, 2008). For seasonal time series, the same method is used but applied to seasonally adjusted data. A total of 20 networks are fitted, each with random starting weights. These are then averaged when computing forecasts. The network is trained for one-step forecasting. Multi-step forecasts are computed recursively. In the present investigation, statistical software *R* (R Core Team, 2019) has been used for analysis of the rainfall data. The *forecast* package in *R* contains *nnetar*() function which fits a NNAR $(p, P, k)_m$ model. For seasonal time series, the default value is $P = 1$ and p is chosen from the optimal linear model fitted to the seasonally adjusted data. If k is not specified, it is set to $k = (p + P + 1)/2$ (rounded to the nearest integer). For each of the districts, respective NNAR model which fits best is shown in table 2.

Table 2. NNAR model for 2 districts

NNAR Model for Districts and GWB	*p*	*P*	*k*
Birbhum NNAR(25,1,13)	25	1	13
North 24 Parganas NNAR(25,1,13)	25	1	13

3.3 Comparing Forecast Accuracy between SARIMA and NNAR

Forecast accuracy is compared between SARIMA and NNAR models for all the time series under study. Both in-sample (training set) and out-of-sample (test set) forecast accuracy associated with the two methods is reported for the sample districts in table 3.

Table 3. In-sample and out-of-sample forecast accuracy for SARIMA and NNAR

District/GWB	Dataset	Method	RMSE	MAE	MAPE	MASE
Birbhum	Training	ARIMA	114.6597	88.2395	40.6518	0.7601
		NNAR	7.6572	5.0658	2.2208	0.0436
	Test	ARIMA	135.2876	99.3877	42.2787	0.8561
		NNAR	126.9125	94.6770	40.6894	0.8156
North 24 Parganas	Training	ARIMA	127.4967	95.8100	38.7816	0.6962
		NNAR	8.9207	6.1335	2.5158	0.0446
	Test	ARIMA	126.8006	97.3919	33.5216	0.7077
		NNAR	119.1246	86.5784	29.4841	0.6291

It can be observed from table 3 that in both cases of in-sample and out-of-sample forecasting, for each of the districts, the values of all the diagnostic measures viz. RMSE, MAE, MAPE and MASE are less in case of NNAR than those of SARIMA. Hence, SARIMA model is less suited for the present dataset.

The DM test has been applied on residuals from SARIMA and NNAR to compare the in-sample and out-of-sample forecast accuracy of two forecast methods and the results are shown in table 4. The alternative hypothesis has been set here is that NNAR is more accurate than SARIMA method.

Table 4. DM test based on in-sample and out-of-sample forecast errors from SARIMA and NNAR

District	in-sample		out-of-sample	
	DM Statistic	*p* -value	DM Statistic	*p* -value
Birbhum	8.6690	<0.0001	1.0059	0.3271
North 24 Parganas	9.7208	<0.0001	0.5668	0.5775

Hence, it is clear from table 4 that for the two districts, NNAR method has more accuracy than SARIMA in case of fitting with in-sample data, however, there is no significant difference between these two methods in case of out-of-sample forecast accuracy. The lack of fitting of SARIMA model with in-sample data can be more prominent if the observations are plotted against the fitted (using SARIMA) values as illustrated in figure 1 and figure 2.

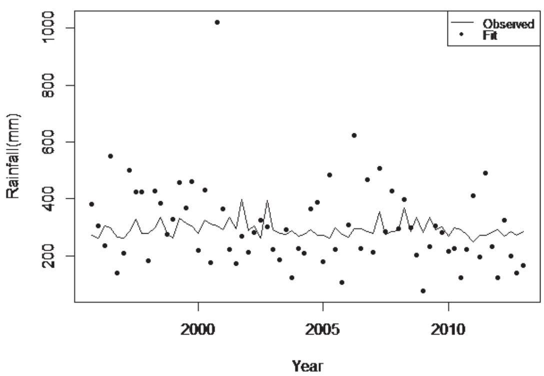

Fig. 1: Observed vs. fitted (SARIMA) monthly monsoon rainfall for Birbhum district

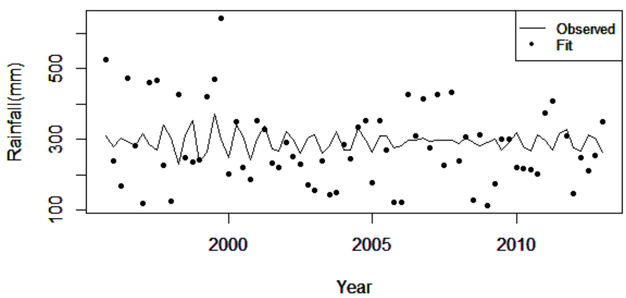

Fig. 2 : Observed vs. fitted (SARIMA) monthly monsoon rainfall for North 24 Parganas district

For both the districts, it is evident from fig.1 and fig.2 that SARIMA model has not been fitted well for the present dataset.

3.4 Diagnostics of Residuals from NNAR Model

The residuals from each of the fitted NNAR model are checked for autocorrelation at different lags. The plots for ACF and PACF do not show any autocorrelations among the residuals as all autocorrelations and partial autocorrelations lie between 95% control limits (Figure 3 and 4).

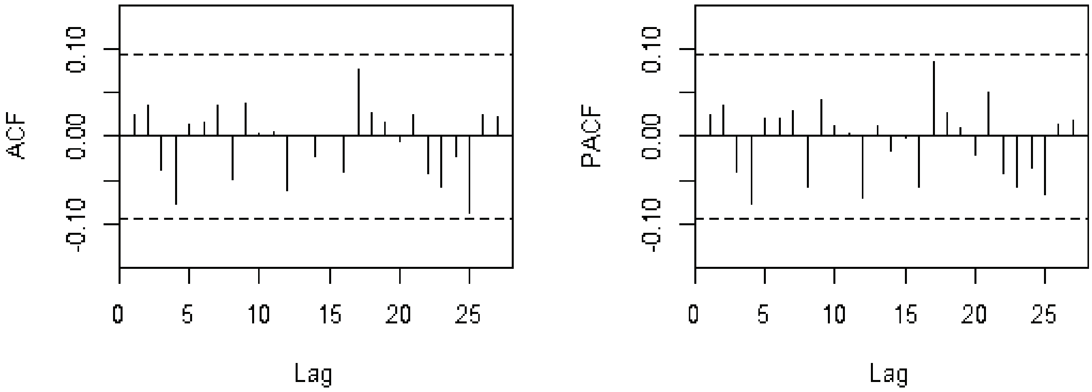

Fig. 3 ACF and PACF of residuals from NNAR model for Birbhum district

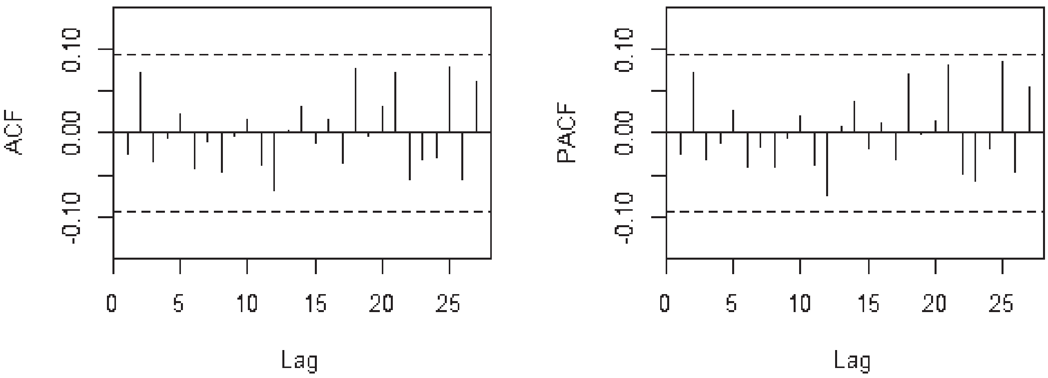

Fig. 4 ACF and PACF of residuals from NNAR model for North 24 Parganas district

3.5 Forecasting using NNAR Model

As the predictive accuracy of NNAR model, in all the cases, is better than the SARIMA model, the NNAR model has been used for forecasting of rainfall volume in the present study. At first, the full time series (1901-2018) for each of the sample districts are fitted with the NNAR model and the models, thus obtained, are similar to those of the models suitable for training dataset.

Finally, the models are then employed to obtain the forecasted values of monthly monsoon rainfall for the year 2019 and 2020 (Table 5).

Table 5 Forecasting of monthly monsoon rainfall for 2 districts

	Forecasting of rainfall (mm)				
Districts	**Year**	**June**	**July**	**August**	**September**
Birbhum	2019	168.39	269.98	263.62	139.19
	2020	260.44	243.50	416.66	263.33
North 24 Parganas	2019	254.01	376.81	316.86	168.34
	2020	227.82	436.74	330.96	247.62

The observed, fitted and forecast values, for each of the districts are illustrated in fig. 5 and 6.

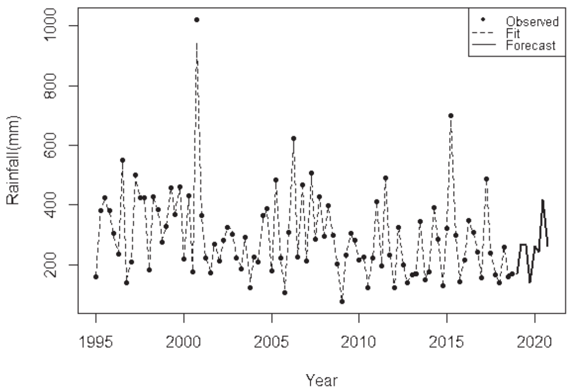

Fig. 5: Observed, fit and forecast values of monthly monsoon rainfall amount of Birbhum district using NNAR model

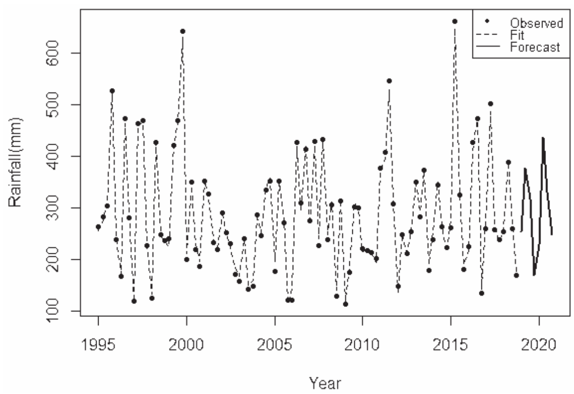

Fig. 6: Observed, fit and forecast values of monthly monsoon rainfall amount of North 24 Parganas district using NNAR model

4. CONCLUSION

The present study critically investigated modelling rainfall volume for the 2 districts in Gangetic West Bengal meteorological sub-division of India. It has been observed that the neural network model over-performed the traditional SARIMA model for the districts. There is future scope to introduce more advanced ANN technique for rainfall prediction to possibly obtain more accurate forecasting result. It has been revealed from the present analysis that there might be shortage of overall monsoon rainfall in the two districts in 2019. Precautionary measures can be taken well ahead to combat the situation accordingly. Finally, similar investigation may also be conducted for the other meteorological sub-divisions as well as agro-climatic zone or state using the NNAR methodology.

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE T. Automat. Contr.*, **19(6)**: 716-723.

Bryzgalova, S. and Gelman, S. 2008. Spurious Long-range Dependence in Emerging Financial Markets. *Unpublished manuscript*.

Chattopadhyay, S. and Chattopadhyay, G. 2010. Univariate modelling of summer monsoon rainfall time series: Comparison between ARIMA and ARNN. *C. R. Geosci.*, **342(2)**: 100-107.

Das, H. P., Doblas-Reyes, F. J., Garcia, A., Hansen, J., Mariani, L., Nain, A., Ramesh, K., Rathore, L. S. and Venkataraman, R. 2010. Weather and climate forecasts for agriculture. *World Meteorological Organization, Guide to Agricultural Meteorological Practices, E-Publishing, Switzerland*, 5-1.

Diebold, F. X. and Mariano, R. S. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, **13(3)**: 253-263.

Farajzadeh, J., Fard, A. F. and Lotfi, S. 2014. Modelling of monthly rainfall and runoff of Urmia lake basin using feed-forward neural network and time series analysis model. *Water Resour. Ind.*, **7**: 38-48.

Goswami, P and Srividya 1996. A novel neural network design for long range prediction of rainfall pattern. *Curr. Sci.*, **70(6)**: 447-457.

Guhathakurta, P. and Rajeevan, M. 2008. Trends in rainfall pattern over India. *Int. J. Climatol.*, **28(11)**: 1453-1469.

Hurvich, C. M. and Tsai, C. L. 1989. Regression and time series model selection in small samples. *Biometrika*, **76(2)**: 297-307.

Hyndman, R. J. and Athanasopoulos, G. 2014. *Forecasting: Principles and Practice*, OTexts.

Hyndman, R. J. and Khandakar, Y. 2008. Automatic time series forecasting: the forecast package for R. *J. Stat. Softw.*, **26(3)**: 1-22.

Hyndman, R. J. and Koehler, A. B. 2006. Another look at measures of forecast accuracy. *Int. J. Forecasting*, **22(4)**: 679-688.

Iyengar, R. N. and Kanth, S. R. 2005. Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall. *Meteorol. Atmos. Phys.*, **90(1-2)**: 17-36.

Kashid, S. S. and Maity, R. 2012. Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using Genetic Programming. *J. Hydrol.*, **454**: 26-41.

Khashei, M. and Bijari, M. 2010. An artificial neural network (*p, d, q*) model for time series forecasting. *Expert Syst. Appl.*, **37**: 479-489.

Kishtawal, C. M., Basu, S., Patadia, F. and Thapliyal, P. K. 2003. Forecasting summer rainfall over India using genetic algorithm. *Geophys. Res. Lett.*, **30(23)**.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. 1992. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *J. Econometrics*, **54(1)**: 159-178.

Makridakis, S., Wheelwright, S. C. and Hyndman, R. J. 1998. *Forecasting Methods and Applications*, 3rd Edition. John Wiley & Sons.

New rainfall series for the districts, meteorological sub-divisions and country as whole of India. 2011. *National Climate Centre Research Report No: 2/2011, Sept*.

R Core Team 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rajeevan, M., Pai, D. S., Kumar, R. A. and Lal, B. 2007. New statistical models for long-range forecasting of southwest monsoon rainfall over India. *Clim. Dynam.*, **28(7-8)**: 813-828.

Ramana, R. V., Krishna, B., Kumar, S. R. and Pandey, N. G. 2013. Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. *Water Resour. Manag.*, **27(10)**: 3697–3711.

Osborn, D. R., Chui, A. P. L., Smith, J. and Birchenhall, C. R. 1988. Seasonality and the order of integration for consumption. *Oxford B. Econ. Stat.*, **50(4)**: 361-377.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.*, **6(2)**: 461-464.

Singh, P. and Borah, B. 2013. Indian summer monsoon rainfall prediction using artificial neural network. *Stoch. Env. Res. Risk A.*, **27(7)**: 1585-1599.

Yusof, F. and Kane, I. L. 2012. Modelling Monthly Rainfall Time Series using ETS State Space and SARIMA Models. *International Journal of Current Research*, **4(9)**: 195-200.

Zhang, G. P. 2012. Neural Networks for Time-Series Forecasting. In *Handbook of Natural Computing*, Rozenberg, G., Back. T. and Kok, J. N. (Ed.) Springer Berlin Heidelberg, New York, pp. 461-477.