

Forecasting monthly rainfall using autoregressive integrated moving average model (ARIMA) and artificial neural network (ANN) model: A case study of Junagadh, Gujarat, India

D. K. Dwivedi*

College of Forestry, ACHF, Navsari Agricultural University, Navsari (Gujarat), India

J.H. Kelaiya

College of Agricultural Engineering and Technology, Junagadh Agricultural University, Junagadh (Gujarat), India

G. R. Sharma

Polytechnic in Agricultural Engineering, Junagadh Agricultural University, Targhadia (Gujarat), India

*Corresponding author. E-mail: dhavaldwivedi42@gmail.com

Abstract

The onset, withdrawal and quantity of rainfall greatly influence the agricultural yield, economy, water resources, power generation and ecosystem. Time series modelling has been extensively used in stochastic hydrology for predicting various hydrological processes. The principles of stochastic processes have been increasingly and successfully applied in the past three decades to model many of the hydrological processes which are stochastic in nature. Time lagged models extract maximum possible information from the available record for forecasting. Artificial neural network has been found to be effective in modelling hydrological processes which are stochastic in nature. The ARIMA model was used to simulate and forecast rainfall using its linear approach and the performance of the model was compared with ANN. The computational approach of ANN is inspired from nervous system of living beings and the neurons possess the parallel distribution processing nature. ANN has proven to be a reliable tool for modelling compared to conventional methods like ARIMA and therefore ANN has been used in this study to estimate rainfall. In this study, rainfall estimation of Junagadh has been attempted using monthly rainfall training data of 32 years (1980-2011) and testing data of 5 years (2012-2016). A number of ANN model structures were tested, and the appropriate ANN model was selected based on its performance measures like root mean square error and correlation coefficient. The correlation coefficient Seasonal ARIMA (1,0,0)(3,1,1)₁₂ and ANN back-propagation model (5-12-1) on the testing data was found to be 0.75 and 0.79 respectively. Seasonal ARIMA (1,0,0)(3,1,1)₁₂ and ANN back-propagation model (5-12-1) were used for forecasting rainfall of 5 years (2017-2021).

Keywords: Artificial Neural Network, Autoregressive Integrated Moving Average, Forecasting, Rainfall, Time series modelling

Article Info

DOI:10.31018/jans.v11i1.1951

Received: November 21, 2018

Revised: January 23, 2019

Accepted: January 27, 2019

How to Cite

Dwivedi, D.K. *et al.* (2019). Forecasting monthly rainfall using autoregressive integrated moving average model (ARIMA) and artificial neural network (ANN) model: A case study of Junagadh, Gujarat, India. *Journal of Applied and Natural Science*, 11(1): 35-41

INTRODUCTION

Rainfall forecasting is an important issue for countries whose economies mainly depend on agriculture. Numerous efforts have been devoted to developing and improving the existing time series weather forecasting models by using different techniques. The role of statistical methodology for predicting the weather parameters is important for their precise estimates. Since decades, many efforts have been made by researchers to identify the appropriate and reliable weather forecasting models. The prediction of rainfall is a challenging task and it is highly useful to the sector of agriculture (Nayak *et al.*, 2013).

In this study, ARIMA and ANN have been used to

forecast rainfall. ARIMA is a conventional linear statistical model and due to its simplicity in its implementation, it has remained popular for time series modelling of rainfall. On the other hand, ANN is a self adaptive technique which has the ability to deal with nonlinear functions and approximate them. Basically, ANN is a method similar to the neurological system of humans consisting of neurons which are the basic computing elements and these neurons are interconnected to form a network (Rummelhart & McClelland 1996). Time series modelling is a process of forecasting using the past data. French *et al.* (1992) noted that the complex nature of atmospheric processes made rainfall modelling and forecasting a very difficult task. Time series models may be univariate or

multivariate. In this study, univariate models have been used. In univariate models, there is a single time series whose lags are utilized for modelling. Time series analysis involves procedures for fitting a given time series to an appropriate model by understanding the inherent characteristics of the time series. Such models fitted to the time series are useful for simulating the time series and hence forecasting.

Somvanshi *et al.* (2006) evaluated the prediction efficiency of ARIMA model and ANN model for Hyderabad by utilizing 104 years of annual data from year 1901 to 2003. The ARIMA model used 93 years as training data and the performance of the model was evaluated on testing data of 10 years. ARIMA (4,0,0) was identified as the appropriate ARIMA model for rainfall forecasting and its appropriate coefficients were derived. The root mean square errors and coefficients of determination was found to be 262.57 and 0.9402 respectively for the identified ARIMA model and 145.14 and 0.9695 respectively for ANN model. Kaushik and Singh (2008) formulated a seasonal ARIMA model analysing 12 years (1995-2006) of monthly rainfall data of Mirzapur, Uttar Pradesh and predicted monthly rainfall of next 5 years (2007-2011) using Box Jenkins methodology. The results indicated that the seasonal ARIMA model provided reliable and satisfactory predictions for rainfall on a monthly scale. The seasonal ARIMA (3,0,2) (2,0,1)₁₂ was found to be the suitable model with coefficient of determination of 0.72 and root mean square error of 62.40.

Chattopadhyay and Chattopadhyay (2010) developed a univariate model to forecast the summer monsoon (June–August) rainfall over India based on the data pertaining to the period 1871–1999. After randomness and non-stationarity within the time series were observed, the autoregressive integrated moving average (ARIMA) models were used and the ARIMA (0,1,1) was identified as a suitable representative model producing minimum Akaike Information Criterion (AIC) of 1228.99. Abdul (2013) predicted monthly rainfall in Kirkuk, Iraq using ANN, ARIMA and MLR. The ANN, ARIMA and MLR approaches were applied to the data to derive the weights and the regression coefficients respectively. The performances of the models were evaluated by using remaining (13 years) of data. By comparing R^2 values (0.91, 0.85, and 0.823) of the models, the study revealed that ANN model can be used as an appropriate forecasting tool to predict the monthly rainfall, which is preferable over the ARIMA model and MLR model.

Farajzadeh *et al.* (2014) adopted ARIMA model for monthly rainfall forecast of 6 years (2006-2011) for Urmia Lake basin located in Iran by utilizing 27 years of monthly rainfall training data. ARIMA (2,0,2)(4,1,2)₁₂ was identified as the optimum model with correlation coefficient of 0.654

and root mean square error of 21.4. Bari *et al.* (2015) built a seasonal ARIMA model using Box and Jenkins method to forecast long term rainfall in Sylhet using rainfall data from the year 1980 to 2010. Rainfall data from the year 1980 to 2006 were used to develop the model while data from year 2007 to 2010 were used to verify the prediction precision. Validity of the model was tested using standard graphical explanation of residuals given by Box and Jenkins. As a second step of validation, forecasted values of monthly rainfall were checked using actual data series. After completion of necessary checking and forecast observation, the ARIMA(0, 0, 1) (1,1, 1)₁₂ was found to be the most effective to predict future precipitation with a 95% confidence interval. The objective of the study was to identify optimum ARIMA model and ANN model structure for forecasting rainfall of Junagadh based on the training data of 32 years monthly rainfall.

MATERIALS AND METHODS

The present study was based on a time series rainfall data observed at Junagadh located in Gujarat State of India. Geographically Junagadh is situated at 21°31' N latitude and 70°28' E longitudes with an elevation of 82.8 m above M.S.L. It records annual rainfall of 800 mm. Highest rainfall (2800 mm) in a calendar year was recorded back in 1983. July month has the highest average monthly rainfall of 341.82 mm. September has the lowest average monthly rainfall of 161.35 mm followed by the June and August with the average monthly rainfall of 167.35 mm and 192.63 mm respectively. The highest average rainfall of 31.72 mm in the non-seasonal months was observed in October followed by rainfall of 9.72mm and 6.53 mm in May and November respectively.

The detailed methodology of the models namely ARIMA and ANN is given in this section. The rainfall monthly data were obtained from Agrometeorological department located in Junagadh Agricultural University. In this study, rainfall estimation of Junagadh has been attempted using monthly climatic training data of 32 years (1980-2011) and testing data of 5 years (2012-2016). The selected models were used for forecasting rainfall of 5 years (2017-2021). The distribution of the monthly rainfall data in training, testing and forecasting periods is given in Table 1.

Autoregressive moving average model (ARIMA) model: An ARIMA model predicted values in the monthly rainfall time series as a linear combination of its own past values and past

Table 1. Training and testing period used for models.

Training Period	Testing Period	Forecasting Period
1980-2011 (32 years)	2012-2016 (5 years)	2017-2021 (5-years)

errors. The first component was the autoregressive (AR) term, the second component was the integration (I) term which was responsible for making the data stationary and the third component was the moving average (MA) term of the forecast errors. As seasonality was inherent in the time series of monthly rainfall, the seasonal ARIMA model was utilized in this study in which there were two more terms introduced namely seasonal autoregressive term and seasonal moving average term in addition to the other terms mentioned in the regular ARIMA model as given in equation (1). Box-Jenkins methodology was adopted in this study which involved model identification, parameter estimation and diagnostic check (residual analysis). In the process of model identification, the order of the model was identified which indicated the number of autoregressive and moving average components to be included in the ARIMA model. The order of the model was determined using the using autocorrelation and partial autocorrelation graphs. The Yule walker approach was utilized to evaluate the parameters of the ARIMA model and the best model was selected based on root mean square error criteria. The residuals of the selected model were tested for significant spikes and if they were found to be uncorrelated, the model was finalized. A number of candidates ARIMA models were formed whose performances were checked on testing data and finally the most appropriate model was selected for forecasting. The model type was decided based on the autocorrelation and partial autocorrelation patterns as given in Table 2. The approach of ARIMA modeling given by Box and Jenkins was used (Box and Jenkins, 1976).

$$\Phi_p(B^s)\phi_p(B)\nabla_s^D\nabla^d z_t = \theta_q(B)\theta_q(B_s)a_t \quad \dots\text{Eq.}(1)$$

$\Phi_p(B^s)$ = seasonal autoregressive operator of order P

ϕ_p = regular autoregressive operator of order p

∇_s^D = seasonal differences

∇^d = regular differences

$\theta_q(B_s)$ = seasonal moving average operator of order P

$\theta_q(B)$ = regular moving average operator of order p

a_t = white noise process

The Ljung–Box test is a type of typical statistical test of whether any autocorrelations of a time series are different from zero.

The Ljung–Box test may be defined as:

H_0 : The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).

H_a : The data are not independently distributed; they exhibit serial correlation.

The test statistic is:

$$Q = n(n + 2) \sum_{k=1}^h \frac{\rho_k^2}{n-k} \quad \dots\text{Eq.}(2)$$

n = sample size

ρ_k = sample autocorrelation at lag k

h = number of lags being tested

Artificial neural network (ANN) model: The basic structure of ANN involves a system of layered, interconnected nodes and neurons. The nodes were arranged to form an input layer, with neurons in each hidden layer connected to all neurons in neighbouring layers. The typical ANN model structure using backpropagation algorithm is shown in Figure in which lags are used as inputs. The time series data of inputs and output were normalized between 0 and 1. Appropriate process of normalization and denormalization of data was carried out before and after the program execution. The data was divided by the maximum value for normalization and the result was multiplied by the same amount once the model gave its output.

The errors obtained were propagated backward until best validation performance could be achieved by the model given in Fig. 1. Feed forward backpropagation neural network with Levenberg Marquardt algorithm was used to train the network. The backpropagation neural network architecture for monthly forecast were formed using appropriate input data and number of neurons in the hidden layer. Each model was retrained several times with different combination of weights until the performance improved. A model architecture gave a different correlation coefficient value each time it was retrained and finally the model architecture that gave better correlation coefficient value was selected. Performance of each candidate model was observed, and the residuals were also analysed before selecting the optimum model for forecasting. The predicted values with the developed models were compared with observed data for checking their accuracy of prediction.

The network used a dot-product activation function that, for each neuron Bj (j = 1, 2,,n) in the hidden layer.

$$B_j = \sum_{i=1}^m w_{ij}A_i + w_{0j}A_0 \quad \dots\dots\dots\text{Eq.}(3)$$

Where input nodes A_i (i = 1, 2,, m) and weights W_{ij} between nodes A_i and neurons B_j . Each neuron value was subsequently passed through a transfer function given below;

$$\sigma(u) = [1 + e^{-u}]^{-1} \quad \dots\dots\dots\text{Eq.}(4)$$

Nonlinearities were incorporated into the network via the activation and transfer functions in each neuron. Complexities in the data were captured through the number of neurons in the hidden layer. If increasing a given weight lead to more error, the weights downwards were adjusted and if increasing a weight lead to less error, the weights upwards were adjusted. Adjustment of all the up or down continued throughout this process until the weights and error settled down.

Table 2. Determination of model by using ACF and PACF patterns.

MODEL	ACF	PACF
AR (p)	Dies down	Cut off after lag q
MA (q)	Cut off after lag p	Dies down
ARMA (p, q)	Dies down	Dies down

Table 3. Parameter estimates of the SARIMA model (1,0,0)(3,1,1)₁₂.

Term	Lag	Coef	SE Coef	t ratio	P value
AR	1	-0.0036	0.0569	-0.06	0.950
SAR	12	-1.5763	0.1101	-14.31	0.000
SAR	24	-1.4036	0.096	-14.61	0.000
SAR	36	-0.6836	0.0763	-8.96	0.000
SMA	12	-0.7307	0.1248	-5.85	0.000

RESULTS AND DISCUSSION

ARIMA model: Seasonal ARIMA model (1,0,0)(3,1,1)₁₂ was selected as the appropriate model after assessing its performance and analysing its residuals. The parameters of the selected model are given in Table 3. The scatter plot of actual and predicted monthly rainfall data is given in Fig. 2 while Fig. 3 and Fig. 4 shows the time series plot of actual and predicted rainfall for the training and testing data, from which it was observed that the model was able to capture certain extreme events successfully, however it can be said that it generally underpredicted the rainfall. The Seasonal ARIMA model was then employed to forecast the monthly rainfall for the next five years 2017-2021 as shown in Fig 5.

SARIMA (1,0,0)(3,1,1)₁₂ was checked to see whether the residuals were independently distributed and they did not have significant correlations. Ljung box test was performed on the residuals for checking significant correlations and they were found to be insignificant as given as shown in Fig. 6.

Artificial neural network model: The artificial neural network (5-12-1) was selected as the appropriate model as it simulated rainfall data well

both in training and testing data. The selected architecture of the model is shown in Fig. 7. This model also ensured that there was not overfitting in the training data. The root mean square error of training, validation and testing at various epochs is shown in Fig. 8 which showed that the mean square errors for training, validation and testing were close to each other around epoch 13 at which point the training was halted. The scatter plot of observed and predicted values on testing data is shown in Fig. 9 while the time series comparison of observed and predicted values for training and testing data is shown in Fig. 10 and Fig. 11 respectively. The final weights of the model used for simulating the time series are given in Table 4. The model was then used to forecast rainfall for the next five years from 2017-2021 as shown in Fig.12. The error histogram after execution of the ANN model is shown in Fig. 13. The observed and predicted monthly rainfall (2017) is given in Appendix I.

The performance analysis of the ARIMA and ANN model on the testing data revealed that the correlation coefficient was found to be 0.75 and 0.79 respectively while the root mean square error was 89.70 and 85.06 respectively as given in Table 5. Therefore, the performance of the ANN model was found to be better than ARIMA model. Both ARIMA and ANN models were used for forecasting monthly rainfall of 5 years (2017-2021). Abdul (2014) and Somvanshi *et al.* (2006) also found that the ANN model was better compared to ARIMA model in predicting rainfall. Somvanshi *et al.* (2006) had used training data of 93 years and testing data of 10 years and obtained coefficient of determination of 0.96 for ANN model (correlation coefficient of 0.97) which showed that as the training data increased, the performance of the model also improved. The correlation coefficient of 0.84 (coefficient of determination = 0.72) was obtained by Kaushik and Singh (2008) after modelling monthly rainfall with ARIMA model.

Table 4. Final Weights in the model architecture.

Neurons	IW1	IW2	IW3	IW4	IW5	LW	Bias 1	Bias 2
1	-1.034	3.39	0.284	3.674	3.958	0.204	2.875	0.5247
2	3.651	-5.054	3.486	-1.568	1.919	0.645	-1.988	
3	-1.86	-2.44	-0.12	-2.56	-2.68	0.487	1.57	
4	0.991	1.502	-0.458	1.043	0.585	-0.187	1.39	
5	1.746	-0.076	0.68	0.604	1.285	1.585	1.8284	
6	-1.6	1.614	-2.264	-0.65	-2.915	0.427	0.6514	
7	2.374	-0.908	2.009	1.101	3.11	-1.555	-0.6214	
8	0.779	-2.929	-3.014	-5.943	-8.957	-0.545	2.0965	
9	-0.501	0.484	-2.649	-2.165	-4.814	-0.753	2.5136	
10	0.703	-2.374	-1.957	-4.331	-6.288	0.14	3.6515	
11	3.548	-1.626	1.742	0.789	3.701	1.756	-0.0065	
12	-1.092	-1.677	3.458	-1.627	-0.179	-0.729	-0.6231	

IW- input weight; LW-Layer weight

Table 5. Performance of ARIMA and ANN models in modelling monthly rainfall.

Performance criteria	Period	ARIMA	ANN
Root Mean square error (RMSE)	Training	97.12	93.97
	Testing	89.70	85.06
Correlation coefficient (CC)	Training	0.82	0.84
	Testing	0.75	0.79

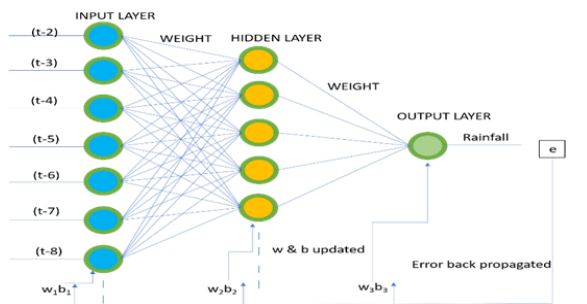


Fig. 1. Typical structure of back propagation ANN with input lags.

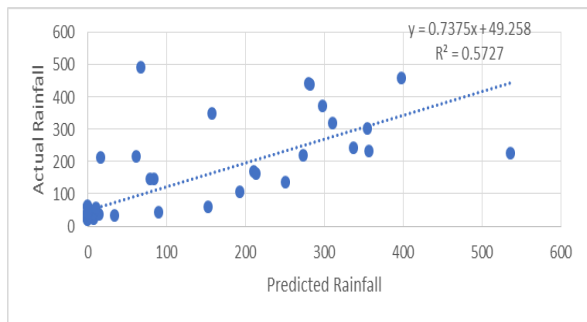


Fig. 2. Scatter plot of actual and predicted rainfall for the testing data.

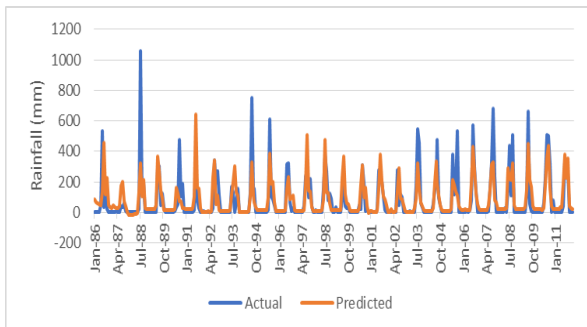


Fig. 3. Comparison of actual and predicted monthly rainfall for the training data.

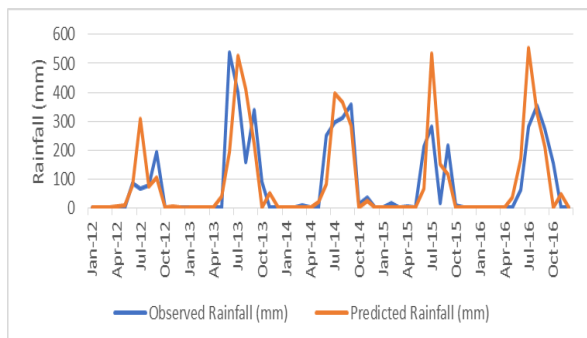


Fig. 4. Comparison of observed and predicted monthly rainfall (A2).

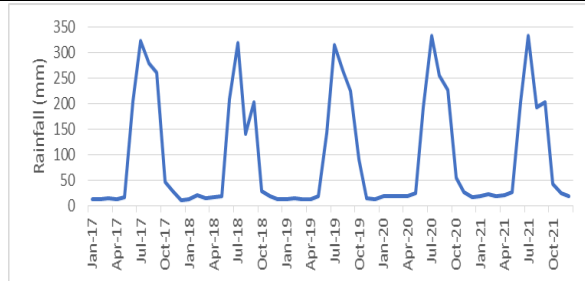


Fig. 5. Rainfall monthly forecast for the year 2017-2021 using SARIMA model.

Lag	AutoCorr	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	Ljung-Box Q	p-Value	Lag	Partial	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8		
0	1.0000										0	1.0000	0	1.0000											
1	-0.0199										0.1523	0.6963	1	-0.0199											
2	0.0020										0.1540	0.9259	2	0.0016											
3	0.0218										0.3386	0.9526	3	0.0219											
4	-0.0296										0.6799	0.9538	4	-0.0288											
5	-0.0121										0.7372	0.9809	5	-0.0134											
6	-0.0062										0.7522	0.9933	6	-0.0070											
7	-0.0222										0.9455	0.9957	7	-0.0212											
8	-0.0169										1.0580	0.9979	8	-0.0181											
9	0.0279										1.3658	0.9980	9	0.0269											
10	0.0061										1.3804	0.9993	10	0.0076											
11	-0.0813										4.0020	0.9699	11	-0.0822											
12	-0.0240										4.2301	0.9789	12	-0.0303											
13	0.0355										4.7333	0.9806	13	0.0360											
14	-0.0193										4.8828	0.9874	14	-0.0141											
15	-0.0139										4.9597	0.9925	15	-0.0193											
16	-0.0103										5.0021	0.9957	16	-0.0149											
17	-0.0054										5.0137	0.9977	17	-0.0035											
18	-0.0092										5.0476	0.9988	18	-0.0137											
19	-0.0234										5.2700	0.9992	19	-0.0290											
20	-0.0392										5.8942	0.9990	20	-0.0366											
21	-0.0070										5.9140	0.9995	21	-0.0059											
22	-0.0029										5.9175	0.9997	22	-0.0128											
23	0.1182										11.6433	0.9756	23	0.1140											
24	0.0022										11.6453	0.9836	24	0.0100											
25	-0.0332										12.0979	0.9858	25	-0.0379											
26	0.0429										12.8577	0.9852	26	0.0300											
27	-0.0449										13.6917	0.9841	27	-0.0406											
28	-0.0077										13.7161	0.9891	28	-0.0062											
29	-0.0093										13.7519	0.9926	29	-0.0109											
30	-0.0123										13.8154	0.9949	30	-0.0090											

Fig. 6. ACF and PACF of the residuals.

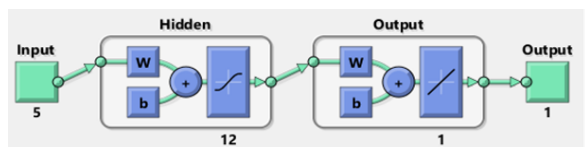


Fig. 7. Architecture of the ANN model 5-12-1.

Appendix I. Observed and predicted values of monthly rainfall of Junagadh (2017).

Month	Observed	Predicted
January	0	0
February	0	0
March	0	0
April	0	0
May	0	22
June	147.8	186
July	330.5	502
August	282.6	206
September	43.5	72
October	0	0
November	0	16
December	2.5	8
Mean	806.9	1011
Standard Deviation	120.1	150.5856
Highest	330.5	502
Lowest	0	0
RMSE	56.62	

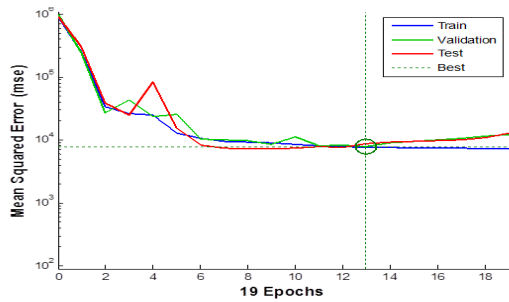


Fig. 8. Mean square error of training, validation and testing data.

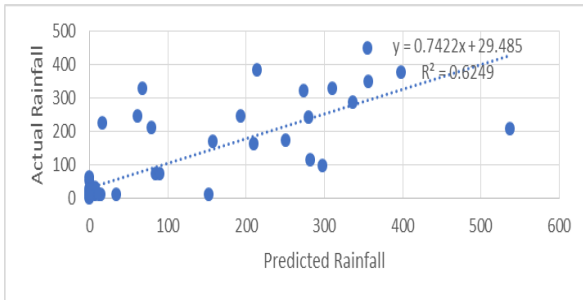


Fig. 9. Scatter plot of actual and predicted rainfall for the testing data.

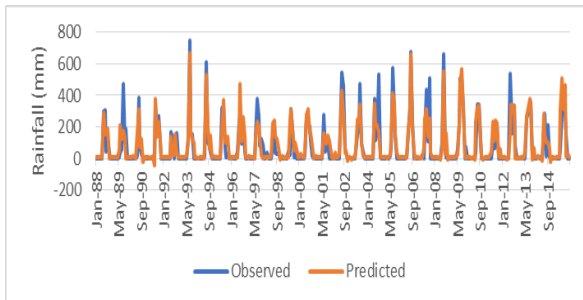


Fig. 10. Comparison of actual and predicted monthly rainfall for the training data.

Conclusion

Seasonal ARIMA (1,0,0)(3,1,1)12 was found to be the appropriate ARIMA model after assessing its performance and analysing its residuals while ANN back-propagation model (5-12-1) was found to be the appropriate ANN model for estimating monthly rainfall of Junagadh city, Gujarat, India. The correlation coefficient obtained after comparing observed and predicted monthly rainfall revealed that ANN model was better than ARIMA model. The selected models were then used to forecast monthly rainfall of 5 years (2017-2021) which could be used for crop planning as well as design and management of water conservation structures.

REFERENCES

1. Abdul-Aziz, A. R., Anokye, M., Kwame, A., Munyakazi, L. and Nsawah-Nuamah, N. N. N. (2013). Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: the case of Ashanti region.

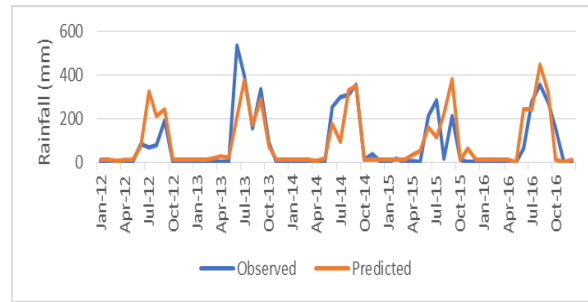


Fig. 11. Comparison of observed and predicted monthly rainfall for the testing data.

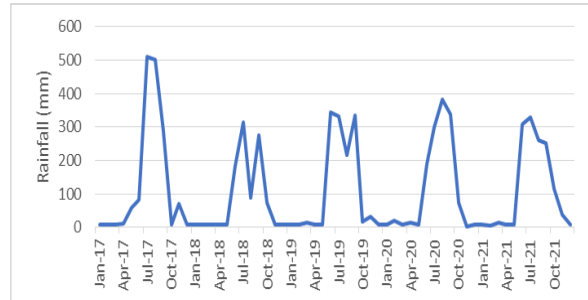


Fig. 12. Rainfall monthly forecast for the year 2017-2021 using ANN model.

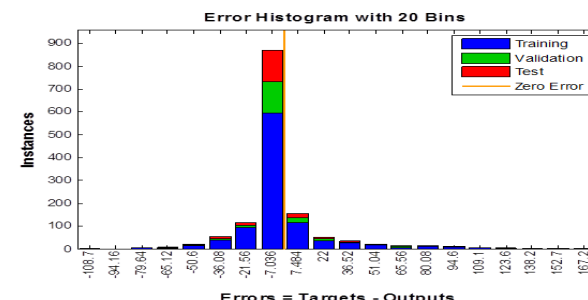


Fig. 13. Error histogram after execution of the ANN model.

International Journal of Humanities and Social Science, 3(3): 224-233

2. Bari, S. H., Rahman, M. T., Hussain, M. M. and Ray, S. (2015). Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*, 7(1): 69-77
3. Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-day Publication, San Francisco.
4. Chattopadhyay, S. and Chattopadhyay, G. (2010). Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. *Comptes Rendus Geoscience*, 342(2):100-107
5. Kaushik, I. and Singh, S. M. (2008). Seasonal ARIMA model for forecasting of monthly rainfall and temperature. *Journal of Environmental Research and Development*, 3(2)
6. Nayak, D. R., Mahapatra, A. and Mishra, P. (2013). A survey on rainfall prediction using artificial neural network. *International Journal of Computer Applications*, 72(16)
7. Farajzadeh, J., Fard, A. F., and Lotfi, S. (2014). Modeling of monthly rainfall and runoff of Urmia lake ba-

- sin using “feed-forward neural network” and “time series analysis” model. *Water Resources and Industry*, 7(1):38-48
8. French MN, Krajewski WF, Cuykendal RR (1992). Rainfall Forecasting in Space and Time Using a Neural Network. *Journal of Hydrology*, 137: 1–37.
 9. Nayak, D. R., Mahapatra, A. and Mishra, P. (2013). A survey on rainfall prediction using artificial neural network. *International Journal of Computer Applications*, 72(16). DOI: 10.5120/12580-9217
 10. Rumelhart, D. E., and McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, London, UK, The MIT Press.
 11. Somvanshi, V. K., Pandey, O. P., Agarwal, P. K., Kalanker, N. V., Prakesh, M. R., and Chand, R. (2006). Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *Journal of Indian Geophysical Union*, 10(2) : 141-151.