ISSN: 2229-6959 (ONLINE)
DOI: 10.21917/ijsc.2017.0196

STOCHASTIC MODELLING BASED MONTHLY RAINFALL PREDICTION USING SEASONAL ARTIFICIAL NEURAL NETWORKS

S.M. Karthik[1] and P. Arumugam[2]
Department of Statistics, Manonmaniam Sundaranar University, India
E-mail: [1]gemofstate@gmail.com, [2]sixfacemsu@gmail.com

Abstract

India is an agrarian society where 13.7% of GDP and 50% of workforce are involved with agriculture. Rainfall plays a vital role in irrigating the land and replenishing the rivers and underground water sources. Therefore the study of rainfall is vital to the economic development and wellbeing of the nation. Accurate prediction of rainfall leads to better agricultural planning, flood prevention and control. The seasonal artificial neural networks can predict monthly rainfall by exploiting the cyclical nature of the weather system. It is dependent on historical time series data and therefore independent of changes in the fundamental models of climate known collectively as manmade climate change. This paper presents the seasonal artificial neural networks applied on the prediction of monthly rainfall. The amounts of rainfall in the twelve months of a year are fed to the neural networks to predict the next twelve months. The gradient descent method is used for training the neural networks. Four performance measures viz. MSE, RMSE, MAD and MAPE are used to assess the system. Experimental results indicate that monthly rainfall patterns can be predicted accurately by seasonal neural networks.

Keywords:

Seasonal Artificial Neural Networks, Annual Rainfall, Rainfall Prediction, Matlab, Stochastic Modelling

1. INTRODUCTION

India is an agrarian society where 13.7% of GDP and 50% of workforce are involved with agriculture. Rainfall is an important element of Indian economy [1]. Agriculture is the main occupation of 50% of India's huge workforce. Most of the agriculture depends directly or indirectly in rainfall. In India, rainfall is highly erratic and varies from region to region and from year to year. The average annual rainfall in 125 cm. Most of the rainfall occurs in the months of June to September and is known as the monsoon. Monsoon winds carry the moisture from the Arabian Sea and Bay of Bengal and causes rainfall across the Indian plateau. The Meteorological department records show that Cherrapunji in Meghalaya is the wettest region in Asia receiving around 1000 cm of rainfall. Around June, the higher temperatures in northern India creates a vacuum which is filled by seasonal winds from the oceans. This moisture laden winds are arrested by the Himalayan range causing precipitation across the plains of India. The south west monsoons in June and north east monsoons in September are two different monsoon systems bringing rainfall in India.

The monsoons are indispensable to Indian economy. Good monsoon leads to better gains from agriculture boosting the rural consumption and job creation. Half of India's fledgling population are dependent of agriculture and allied activities. Good forecast of monsoon coincides with rising stock markets in Mumbai. In addition to irrigating the lands, rainfall replenishes the ground water supply. The ground water in India has come under immense pressure from the growing population. Several deep wells dot the entire rural landscape of India and rainwater harvesting techniques are only beginning to be implemented. This raises the significance of the annual rainfall.

Rainfall prediction is also useful for the purpose of flood prevention and control [2]. The city of Mumbai faced a record rainfall in the monsoon of 2005. The city infrastructure at the time was inadequate to handle flash floods and as a result, the financial capital of India came to standstill. When the floods abated, more than 900 people died and the damages amounting to thousands of crores of rupees. Similar scenes were witnessed in Chennai in 2015. Accurate prediction of impending rainfall can go in a long way in handling flood situations particularly in big cumbersome cities.

2. DATA

The rainfall data set is provided by data.gov.in website of the Ministry of Earth Sciences [3]. It contains the monthly rainfall in mm from 1901 till 2014. The data is described briefly in this section. The Fig.1 shows the total annual rainfall in India from 1901 to 2014. The highest rainfall of 1463.9mm occurred in 1917 and the lowest of 947.1mm in 1972. Overall the amount of rainfall does not show any particular pattern. However it is highly erratic with alternating dry and wet spells and is a result of a chaotic weather system. The Fig.2 shows the box plot of the monthly variation in rainfall. The months from June to October forms the monsoon period marked by both high amount and variability in rainfall. July brings the highest rainfall and is also the most variable. The other months bring less than 60mm in rainfall and are less erratic. The Fig.3 also shows the monthly variability in the rainfall. The Fig.4 demonstrates the study of correlation between the rainfall patterns within few years. The box plot shows three groups: the first group is the distribution of Correlation between the rainfall patterns of first year with the second, the second group between the patterns of one year with the third, the third group between the patterns of one year with the fourth. The correlations among consecutive years show the highest correlation with most of the values above 0.9. The outliers in this case are the years 1905, 1918 and 1911 with the correlations 0.8619, 0.8682 and 0.8728. These years do not have better correlations with any other years. This shows that the best accuracy can be obtained by using one year's rainfall to predict the immediate next year's pattern.

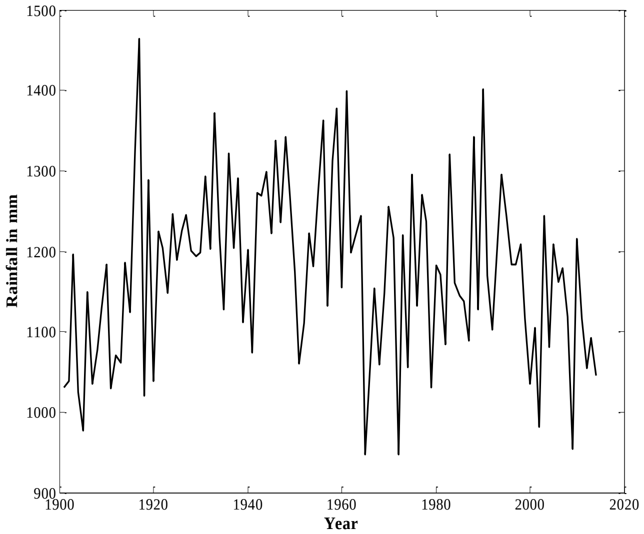

Fig.1. Average Annual Rainfall in mm from 1901 to 2014

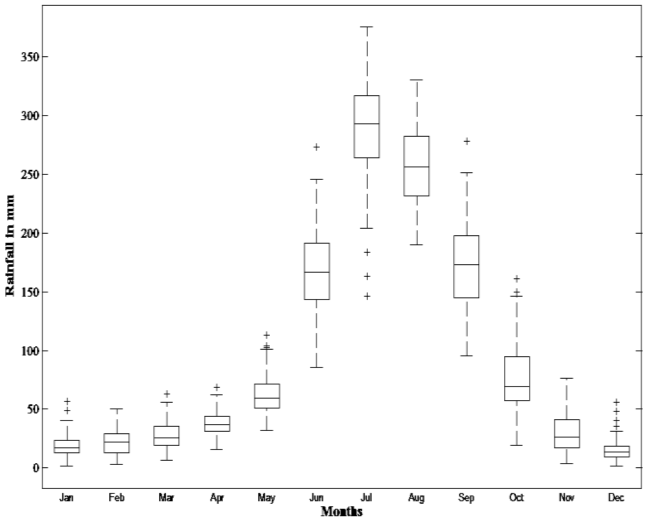

Fig.2. Box Plot of Monthly Variation in Rainfall

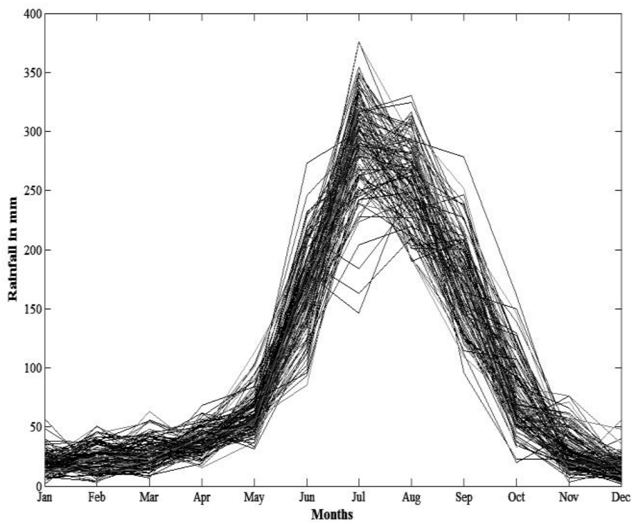

Fig.3. Line Plot of Monthly Variation in Rainfall

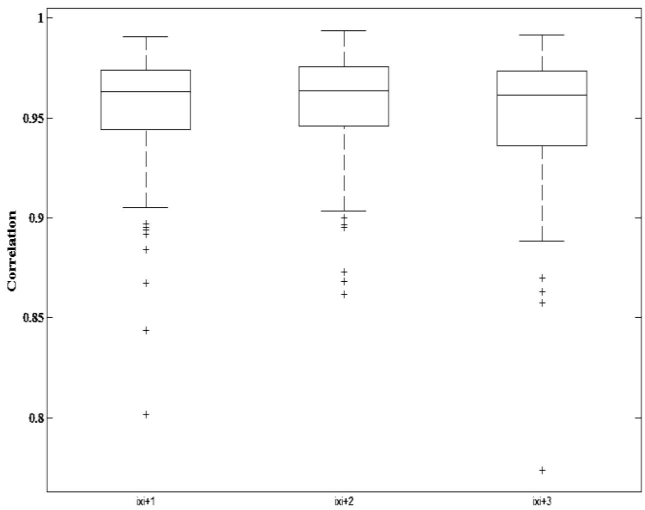

Fig.4. Correlation of Monthly Rainfall with Succeeding Years

3. SEASONAL ARTIFICIAL NEURAL NETWORKS (SANN)

Seasonal Artificial Neural Networks (SANN) was proposed by Hamzacebi in [4]. It is suitable to learning and forecasting non-linear dependence in time series data. Unlike Autoregressive Integrated Moving Average (ARIMA) [5] models, SANN does not eliminate seasonal variation. It learns the seasonal pattern and effectively predicts the future values in the time series. It has been observed empirically to perform very well in applications such as weather, stock market prices etc. All these applications, the measured quantity is subjected to several cyclical patterns with different time periods and phase differences. The Artificial Neural Network (ANN) [6] are supervised learning tools and have the capability to model nonlinear dependencies. The input layer of SANN consists of s nodes, where s is the period of the seasonal variation. In monthly rainfall data, s is taken to be 12 corresponding to the twelve months of the year. The measurements of every month of the year are fed through the input neurons. A hidden layer with fixed number of neurons is taken. The number of neurons in the hidden layer determines the complexity and modeling power of the SANN. However if the number of hidden neurons is too high, the training data might become insufficient. In our experiments, 6 neurons are taken in the hidden layer. The output layer corresponds to the prediction for the monthly rainfall for the next year. Figure 5 illustrates the SANN used in this work.

Let $R_{i,j}$ be the amount of rainfall observed in the j^{th} month of the i^{th} year. The input neurons are fed with $R_{i,j}$, j 1,2,…,12. The output neurons are trained to give $R_{i+1,j}$, j = 1,2,…,12. The weights $U_{j,k}$, j = 1,2,…12; k = 1,2,…,h are present in the connections between the 12 neurons in the input layer and the h neurons in the hidden layer. The weights $V_{k,j}$, j = 1,2,…,12 ; k = 1,2,…,h are present in the connections between the h neurons in the hidden layer and the 12 nodes in the output layer. In addition to these neurons, bias neurons with the weights α_k and β_j are present. The relationship between the output and input neurons are captured by Eq.(1),

$$\hat{R}_{i+1,j} = \beta_j + \sum_{k=1}^{h}\left(V_{k,j}\, f\left(\alpha_k + \sum_{j=1}^{12} U_{jk} R_{i,j}\right)\right). \qquad (1)$$

Here f is the activation function. The choice of a proper architecture is crucial in the accuracy of all types of ANNs. The only parameter here is h which, is the number of neurons in the hidden layer. It is decided on an empirical basis. Common activation functions used include logistic function, softmax function and the gaussian function. In this work, the gaussian function is used because the predicted variable is continuous.

The training of the SANN is done by stochastic gradient descent algorithm [7] given below.

Algorithm 1: Stochastic Gradient Descent for Back propagation training of SANN

initialize V_{kj}, U_{jk}, α_k, β_j for j = 1,2,…,12 and k = 1,2,…,h.

Do for each set of $R_{i,j}$, $R_{i+1,j}$ in training of SANN

initialize V_{kj}, U_{jk}, α_k, β_j for j = 1,2,…,12 and k = 1,2,…,h.

Do for each set of $R_{i,j}$, $R_{i+1,j}$ in training set

Prediction $\hat{R}_{i+1,j} = \beta_j + \sum_{k=1}^{h}\left(V_{k,j}\, f\left(\alpha_k + \sum_{j=1}^{12} U_{jk} R_{i,j}\right)\right).$

Error = $\hat{R}_{i+1,j} - R_{i+1,j}$

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 PERFORMANCE MEASURES

Let $R_{i,j}$, i = 1,2,…n; j = 1,2,…,12 be the actual rainfall recorded in the test set for n years. Let $\hat{R}_{i,j}$ be the prediction for the same period from the SANN. Then the following performance measures are used to assess the accuracy of the prediction.

Mean Squared Error (MSE) is defined as,

$$MSE = \frac{1}{12n}\sum_{i=1}^{n}\sum_{j=1}^{12}\left(R_{i,j} - \hat{R}_{i,j}\right)^2. \qquad (2)$$

Root Mean Squared Error (RMSE) is defined as,

$$RMSE = \sqrt{MSE} \qquad (3)$$

Mean Absolute Deviation (MAD) is defined as,

$$MAD = \frac{1}{12n}\sum_{i=1}^{n}\sum_{j=1}^{12}\left|R_{i,j} - \hat{R}_{i,j}\right|. \qquad (4)$$

It is the average of all absolute deviations of the predicted from the actual values.

Mean Absolute Prediction Error (MAPE) [8] is also known as Mean Absolute Percentage Deviation (MAPD). It is defined as,

$$MAPE = \frac{1}{12n}\sum_{i=1}^{n}\sum_{j=1}^{12}\left|\frac{R_{i,j} - \hat{R}_{i,j}}{R_{i,j}}\right|. \qquad (5)$$

MAPE can be used in our current application since there are no zero values in the predicted variable and does not cause division by zero error. When MAPE is multiplied by 100, it is expressed as a percentage. MSE, RMSE, MAD and MAPE must be low for a good prediction. RMSE, MAD and MAPE are expressed in the same units as the predicted variable i.e., in mm in this case.

4.2 SEASONAL NEURAL NETWORK TRAINING AND RESULTS

The SANN was trained with number of neurons in the hidden layer ranging from 4 to 9. The Fig.6 shows the training parameters with the MATLAB Neural Network Toolbox. The results are summarized in Fig.7 to Fig.12. The prediction values are plotted against the actual values for one representative year out of the 57 years (i.e., 12 months) in the test set.

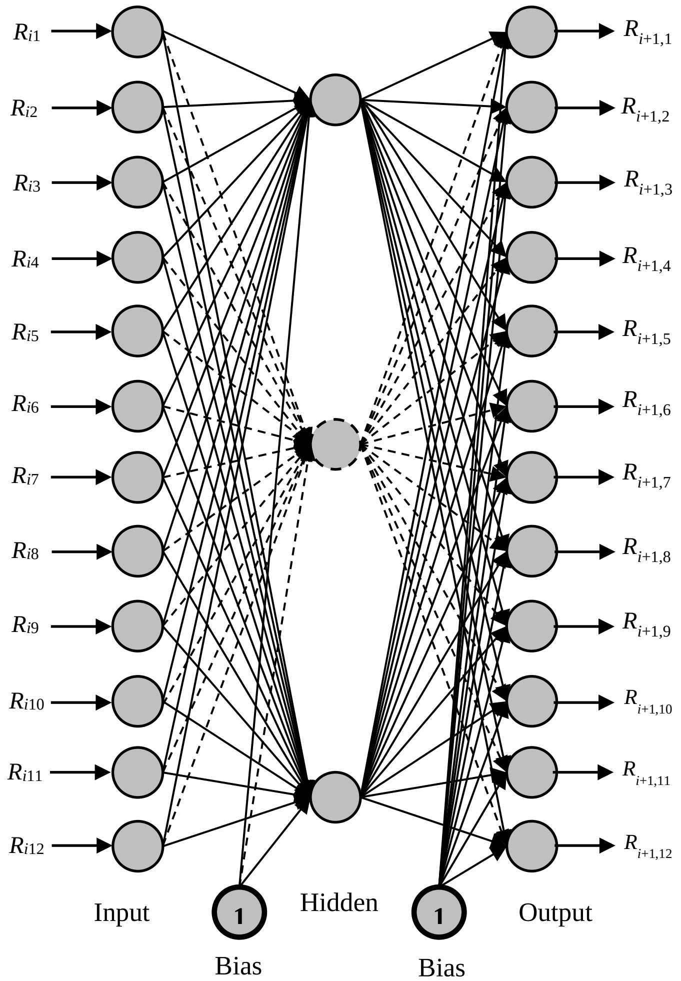

Fig.5. Seasonal Neural Network with s = 12

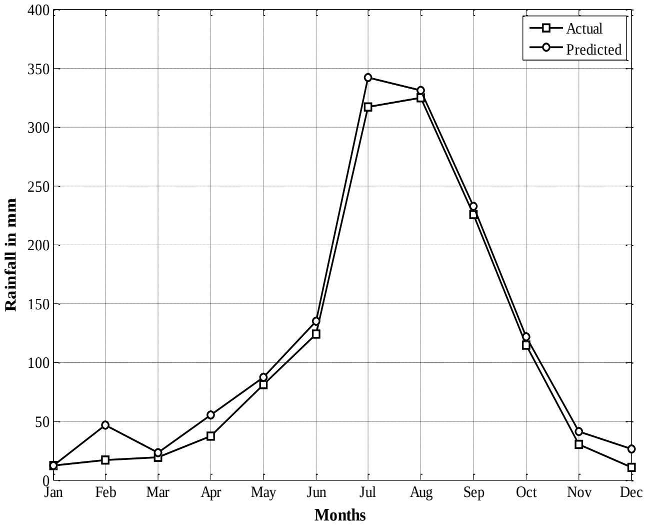

Fig.6. *h* = 4

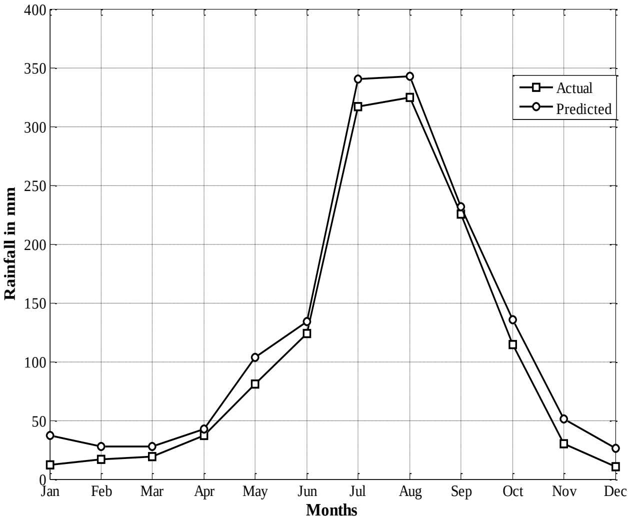

Fig.7. h = 5

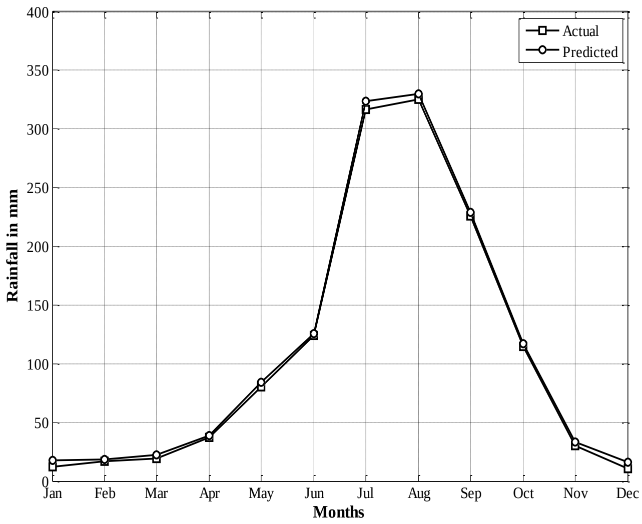

Fig.8. h = 6

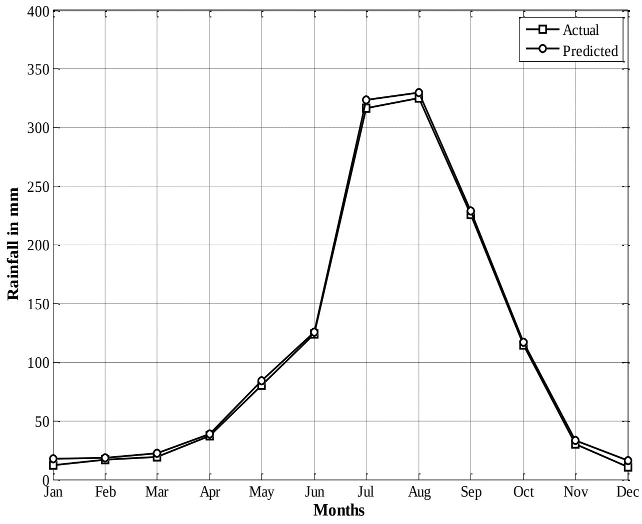

Fig.9. h = 7

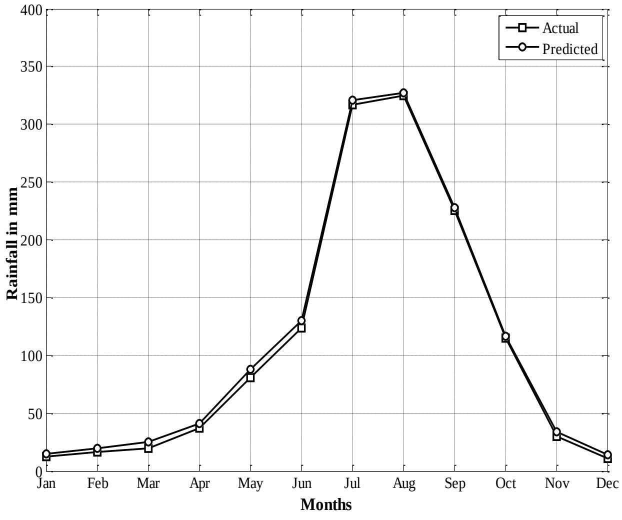

Fig.10. h = 8

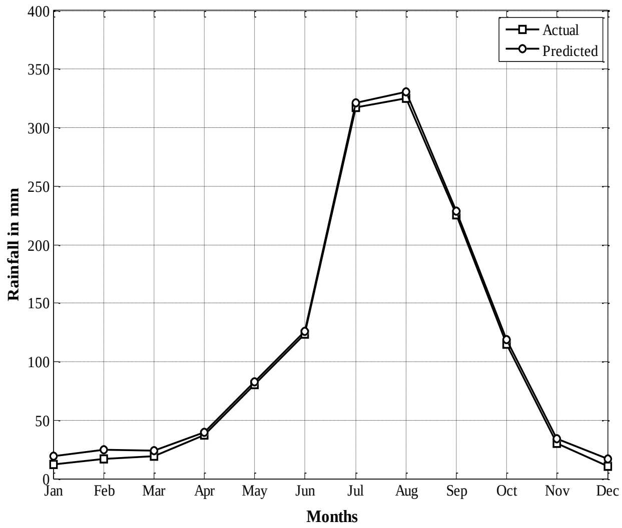

Fig.11. h = 9

4.3 PERFORMANCE ANALYSIS

The performance analysis is summarized in Table.1. It gives the MSE, RMSE, MAD and MAPE values for the parameter *h* ranging from 4 to 12.

Table.1. Performance Measures for SANN

H	MSE	RMSE	MAD	MAPE
4	551.20	23.48	20.60	0.4349
5	303.75	17.43	15.00	0.3213
6	135.48	11.64	10.11	0.2128
7	30.98	5.57	4.80	0.1047
8	32.89	5.73	4.96	0.1027
9	33.32	5.77	5.00	0.1059

As per the performance measures, the accuracy improves with more number of neurons in the hidden layer. More neurons in the hidden layer captures more complex dependence among the time series data. But beyond a certain number, there will be less improvement in accuracy. From Table.1, it can be seen that there is only minimal improvement for *h* more than 7. The optimal performance is given by 8 hidden neurons. The RMSE is 5.77 which means rainfall can be predicted on an average with an error limit of ±6mm. The MAPE is around 10% for *h* = 8. Therefore reliable forecasting of rainfall can be made using SANN. The MAD is below 5.00 for *h* = 8. The maximum deviation observed in the prediction of the test set is 9mm.

A comparative analysis is done and the results are summarized in Table.2. The SANN with *h* = 8 is compared against Autoregressive Moving Average (ARMA) [9], Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA) [10] and Hidden Markov Models (HMM) [11]. The SANN gives the best performance higher than HMMs.

Table.2. Comparative Analysis

Method	MSE	RMSE	MAD	MAPE
SANN (*h*=8)	32.89	5.73	4.96	0.1027
HMM	38.69	6.22	5.34	0.1131
ARMA	39.38	6.28	5.37	0.1134
ARIMA	40.36	6.35	5.55	0.1194
SARIMA	43.42	6.59	5.82	0.1200

4.4 FUTURE RAINFALL PREDICTION

The SANN is used to predict the future rainfall from 2016 till 2020. The results are summarized below. The Fig.13 shows the average annual rainfall predicted for the years. The Fig.14 shows the monthly rainfall prediction for the same period.

The forecast for longer periods are not reliable because the error in the forecast build up. However it is a common property of chaotic systems like weather that predictions become unreliable for longer periods [12-21]. However the entire next year can be predicted and it can be suitable for agricultural planning and flood management.

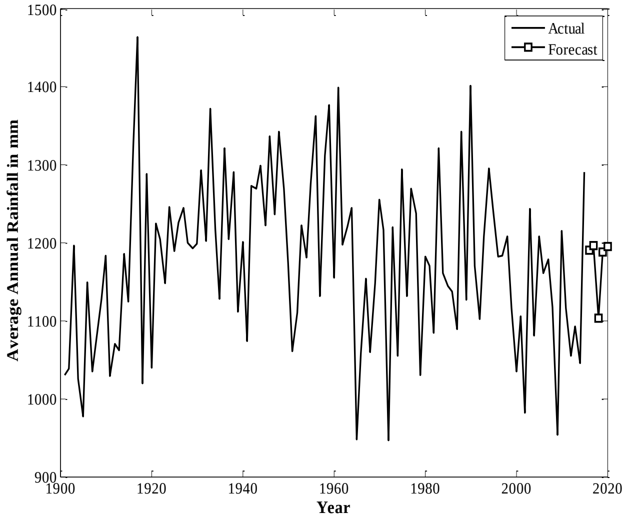

Fig.12. Average Annual Rainfall forecasted till 2020

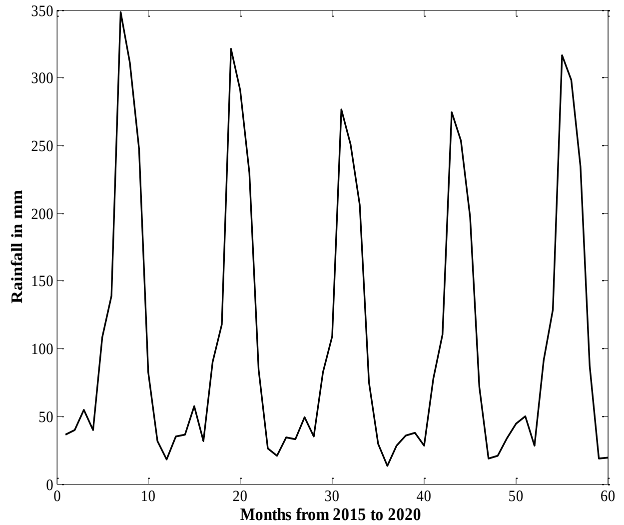

Fig.13. Monthly Rainfall forecast for 2016 to 2020

5. CONCLUSION

Rainfall prediction is an important aspect of weather management. This work presented rainfall forecasting using seasonal artificial neural networks. The seasonal nature of the data is captured in SANN. The concept of SANN was explained briefly and its mathematical models given. The experiments were setup using monthly rainfall data for the period 1901 to 2014. The statistical nature of the data was explained with the help of visualization. Performance measures indicate that the prediction is reliable within an error bound of ±6mm. This is suitable for agricultural planning and flood management. The rainfall forecasts till 2020 is presented. The presented work is compared against common forecasting techniques such as ARMA, ARIMA, SARIMA and HMMs. The higher accuracy of the SANN is demonstrated. In the future, research can be done to combine the higher accuracy of SANN with the stability of HMMs to forecast for longer periods. Also the prediction can be extended to other measured variables such as surface temperature, humidity, cyclones etc.

ACKNOWLEDGEMENT

The Corresponding author thanks the University Grants Commission, New Delhi for awarding fellowship under the scheme of UGC-Basic Science Research fellowship to carry out this work.

REFERENCES

[1] B. Parthasarathy, H.F. Diaz and J.K. Eischeid, "Prediction of All-India Summer Monsoon Rainfall with Regional and Large-scale Parameters", *Journal of Geophysics Research*, Vol. 93, No. D5, pp. 5341-5350, 1988.

[2] Stefan Hastenrath and Arnold Rosen, "Patterns of India Monsoon Rainfall Anomalies", *Tellus*, Vol. 35, No. 4, pp. 324-331, 1983.

[3] Open Government Data Platform India. https://data.gov.in/.

[4] Coşkun Hamzaçebi, "Improving Artificial Neural Networks' Performance in Seasonal Time Series Forecasting", *Information Sciences*, Vol. 178, No. 23, pp. 4550-4559, 2008.

[5] Konstantinos Kalpakis, Dhiral Gada, and Vasundhara Puttagunta, "Distance Measures for Effective Clustering of ARIMA Time-Series", *Proceedings of the IEEE International Conference on Data Mining*, pp. 273-280, 2001.

[6] Martin T. Hagan, Howard B. Demuth, Mark H. Beale and Orlando De Jesus, "*Neural Network Design*", Vol. 20, Boston: PWS Publishing Company, 1996.

[7] Leon Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent", *Proceedings of 19th International Conference on Computational Statistics*, pp. 177-186, 2010.

[8] Cort J. Willmott and Kenji Matsuura, "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance", *Climate Research*, Vol. 30, No. 1, pp. 79-82, 2005.

[9] James Douglas Hamilton, "*Time Series Analysis*", Princeton University Press, 1994.

[10] Fang Mei Tseng and Gwo Hshiung Tzeng. "A Fuzzy Seasonal ARIMA Model for Forecasting", *Fuzzy Sets and Systems"* Vol. 126, No. 3, pp. 367-376, 2002.

[11] Iain L. MacDonald and Walter Zucchini. "*Hidden Markov and other Models for Discrete-Valued Time Series*", CRC Press, 1997.

[12] Ajeet Singh, Asima Jillani and Pravendra Kumar, "Daily Rainfall Prediction using Artificial Neural Network (ANN) for Monsoon Season", *Trends in Bioscience*s, Vol. 8, No. 13, pp. 3302-3309, 2015.

[13] G. Peter Zhang and Min Qi, "Neural Network Forecasting for Seasonal and Trend Time Series", *European journal of Operational Research*, Vol. 160, No. 2, pp. 501-514, 2005.

[14] Philip Hans Franses and Gerrit Draisma, "Recognizing Changing Seasonal Patterns using Artificial Neural Networks", *Journal of Econometrics*, Vol. 81, No. 1, pp. 273-280, 1997.

[15] Assad Y Shamseldin, "Application of a Neural Network Technique to Rainfall-Runoff Modeling", *Journal of Hydrology*, Vol. 199, No. 3, pp. 272-294, 1997.

[16] Guoqiang Zhang, B. Eddy Patuwo and Michael Y. Hu, "Forecasting with Artificial Neural Networks: The State of the Art", *International Journal of Forecasting*, Vol. 14, No. 1, pp. 35-62, 1998.

[17] Amir F Atiya.et.al., "A comparison between Neural-Network Forecasting Techniques-Case Study: River Flow Forecasting", *IEEE Transactions on Neural Networks*, Vol. 10, No. 2, pp. 402-409, 1999.

[18] Holger R Maier and Graeme C. Dandy, "Neural Networks for the Prediction and Forecasting of Water Resources Variables: a Review of Modelling Issues and Applications", *Environmental Modelling and Software*, Vol. 15, No. 1, pp. 101-124, 2000.

[19] A.W. Minns and M.J. Hall, "Artificial Neural Networks as Rainfall-Runoff Models", *Hydrological Sciences Journal*, Vol. 41, No. 3, pp. 399-417, 1996.

[20] A. Sezin Tokar and Peggy A. Johnson, "Rainfall-Runoff Modeling using Artificial Neural networks", *Journal of Hydrologic Engineering*, Vol. 4,No. 3 ,pp. 232-239, 1999.

[21] Jason Smith and Robert N. Eli, "Neural-Network Models of Rainfall-Runoff Process", *Journal of Water Resources Planning and Management*, Vol. 121, No. 6, pp. 499-508, 1995.