

# Comparing Neural Network and Multiple Regressions Models to Estimate Monthly Rainfall Data

Satyvan Yashwant<sup>1</sup>, S. L. Sananse<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Statistics Dr. B.A.M.U Aurangabad, Maharashtra, India

<sup>2</sup>Professor, Department of Statistics Dr. B.A.M.U Aurangabad, Maharashtra, India

**Abstract:** *Marathada region situated between 170 -35 N and 200- 40 N latitude and 740 - 40 E and 780 – 15 E longitudes in Maharashtra state. In this region consist of seven districts such as, Beed, Hingoli, Jalna, Latur, Nanded, Osmanabad and Parbhani. The monthly data of rainfall, maximum & minimum temperature, minimum & maximum humidity, wind speed, wind direction & total could cover (octa) of five years (2 009 to2014) . For analysis of data, statistical methods based on multiple regression and artificial neural networks have been used. According to results of this paper, a comparison between multiple regression and neural network models, using the SPSS& Mat lab software. Two statistical parameter multiple correlation coefficient (R) and Mean Square Errors (MSE) is used for selection best fit model .The model have maximum correlation coefficient (R) and minimum Mean Square Errors (MSE) that model is best fit to estimate the monthly rainfall of Marathwda region. Result show that Aurangabad station value of the Multiple correlation coefficient (R) of Aurangabad satiation 0.915 for multiple regression model while it was 0.949 in neural network model, secondly, the mean square error(MSE) for multiple regression model is 2.41 which is higher than that of the mean square error(MSE) obtained through neural network model as 0.945 on comparing of these result finally concluded the Aurangabad station have Artificial Neural Network (ANN) is best model for estimation of monthly rainfall data. Finally on the basis of maximum value of Multiple correlation coefficient (R ) and minimum value of mean square error(MSE) .it was concluded that the selected metrological station neural network model is better than multiple regression model for estimation of monthly rainfall data.*

**Keywords:** Multiple Linear Regression (MLR), Artificial Neural Network (ANN)

## 1. Introduction

Rainfall information is important for food production plan, water resource management and all activity plans in the nature. The occurrence of dry period or heavy rain at the critical stages of the growth and development may lead to significant reduce crop yield. Regression analysis is a statistical technique that utilizes the relation between two or more quantitative variables on observational database so that outcome variable can be predicted from the others. One of the purposes of a regression model is to find out to what extent the outcome (dependent variable) can be predicted by the independent variables. Rainfall forecast methods are employed in weather forecasting at regional and national levels. Fundamentally, there are two approaches to predict rainfall. They are Empirical method and dynamical methods. The empirical approach is based on analysis of historical data of the rainfall and its relationship to variety of atmospheric and oceanic variables over different parts of the world. The most widely use empirical approaches used for climate prediction are regression, artificial neural network, fuzzy logic and group method of data handling. The dynamical approach, predictions are generated by physical models based on system of equations that predict the future Rainfall. The forecasting of weather by computer using equations is known as numerical weather prediction. To predict the weather by numeric means, meteorologist has developed atmospheric models that approximate the change in temperature, pressure etc using mathematical equations.in this study we have use monthly data of rainfall, maximum & minimum temperature, minimum & maximum humidity, wind

speed, wind direction & total could cover (octa) of five years (2000 to2005) .these variable were used for analysis Artificial Neural Network (ANN) model and Multi Linear Regression (MLR). In this study we have use the best suited architecture of Feed Forward Neural Network Model (FFNN) for monthly rainfall data (june to oct). In the proposed neural network model three layers are considered, the input environment with the significant variables, two hidden layers and an output layer. The  $(7 \times 10 \times 1)$  configuration is considered with 7 neurons in the first hidden layer and 10 neurons in the second hidden layer and only one neuron in the output layer. On the basis of these result we have compare best fit model by using of multiple correlation coefficient (R) and Mean Square Errors (MSE) for estimation of monthly rainfall data of Marathwada Region .this region is less rainfall affected region in Maharashtra .the economy of this region depend on agriculture .Rainfall is directed affected to crop yield currently last five years dry spells increasing in Marathwada region. This study is help farmers in making decision concerning with their crop & regional and local plan they depend on rainfall of Marathwada region.

## 2. Related Work

Many researchers have use study the rainfall predication by using multiple regressions& neural network analysis such as.Chattopadhyay developed a three layer neural model which used the data ranging from 1950 to 1995 to predict the average monsoon rainfall in India .Sable et al. (2007) developed models for various locations of Maharashtra State, in which the group matching technique was used. From the

matching group weighted mean was calculated to predict the seasonal rainfall of different locations of Maharashtra State. French et al in 1992 created a neural network to forecast two-dimensional rainfall in one hour advance. Hu and since then, artificial neural networks have been extended to various environment forecasting tasks. N. Sen has presented long-range summer monsoon rainfall forecast model based on power regression technique with the use of El Nino, Eurasian snow cover, north west Europe temperature, Europe pressure gradient, 50 h Pa Wind pattern, Arabian sea SST, east Asia pressure and south Indian ocean temperature in previous year. The experimental results showed that the model error was 4%. Chhaya Sonar, S L Sananse, etc. Impact of monsoon and statistical forecasting of production of important oil seeds in Aurangabad division. Hitherto almost seven different methods for situation (without limitation) based on initial training artificial neural networks have been developed and each artificial neural network of an input layer, one or one can be used in a different analysis rather than more hidden layers and an output layer has been classical statistical methods. Winn Thida Zaw has developed a prediction model for determining Rainfall over Myanmar using multiple Linear regression where 15 predictors has been used. As a results of several experiments, the predicted rainfall amount is close to actual value. El-Shafie et al. (2011) developed feed forward neural network FFNN model and implemented to predict the rainfall on yearly and monthly basis. The analysis was made on the basis of four parameters viz. root mean square error (RMSE), mean absolute error (MAE), coefficient of Correlation (CC) and mean error (BIAS) which suggested that ANN provide better results than the MLR model. Recently, Zaefizadeh et al. (2011) recommended ANN approach as better predictor of yield in Barley than in multiple linear regressions these are related work .

### 3. Data and Methodology

Marathwada region consist of eight districts such as, Beed, Hingoli, Jalna, Latur, Nanded, Osmanabad and Parbhani. This In this study we have used 5 years (2009 to2014) of time series data in rainfall, minimum &maximum

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + \epsilon$$

Y = rainfall (mm)

X1= Maximum temperature (°C)

X2 = Minimum temperature (°C)

X3= Total cloud cover (octa)

X4= Maximum Relative humidity (%)

X5= Minimum Relative humidity (%)

X6= Wind speed (kmph)

X7= Wind direction (deg)

a= intercept

bi= regression coefficient of i th independent variables

(i=1,2,...7)

ε = error term

#### 2] Neural network model.

The term neural network applies to a loosely related family of models, characterized by a large Parameter space and

temperature, minimum &maximum humidity, wind speed, wind direction & total could cover(octa) in eight metrological station of monthly data(june to oct) in Marathwada region .In this region consist of eight districts such as, Beed, Hingoli, Jalna, Latur, Nanded, Osmanabad and Parbhani. This region is situated between 170 -35 N and 200- 40 N latitude and 740 - 40 E and 780 – 15 E longitudes. The data of seven districts from (2009 to2014) were obtained by IMD & Maharashtra agriculture department.

#### I. Regression Analysis

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modelling and analysing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. In this study use have used seven variables such as Rainfall(Y),Max temp(X1) ,Min temp(X2) ,Total could cover(X3) ,Max humidity(X4), Min humidity(X5) , Wind speed(X6) , Wind direction(X6) rainfall is dependent variable. In that study more than two variable present in data we have use multiple regression

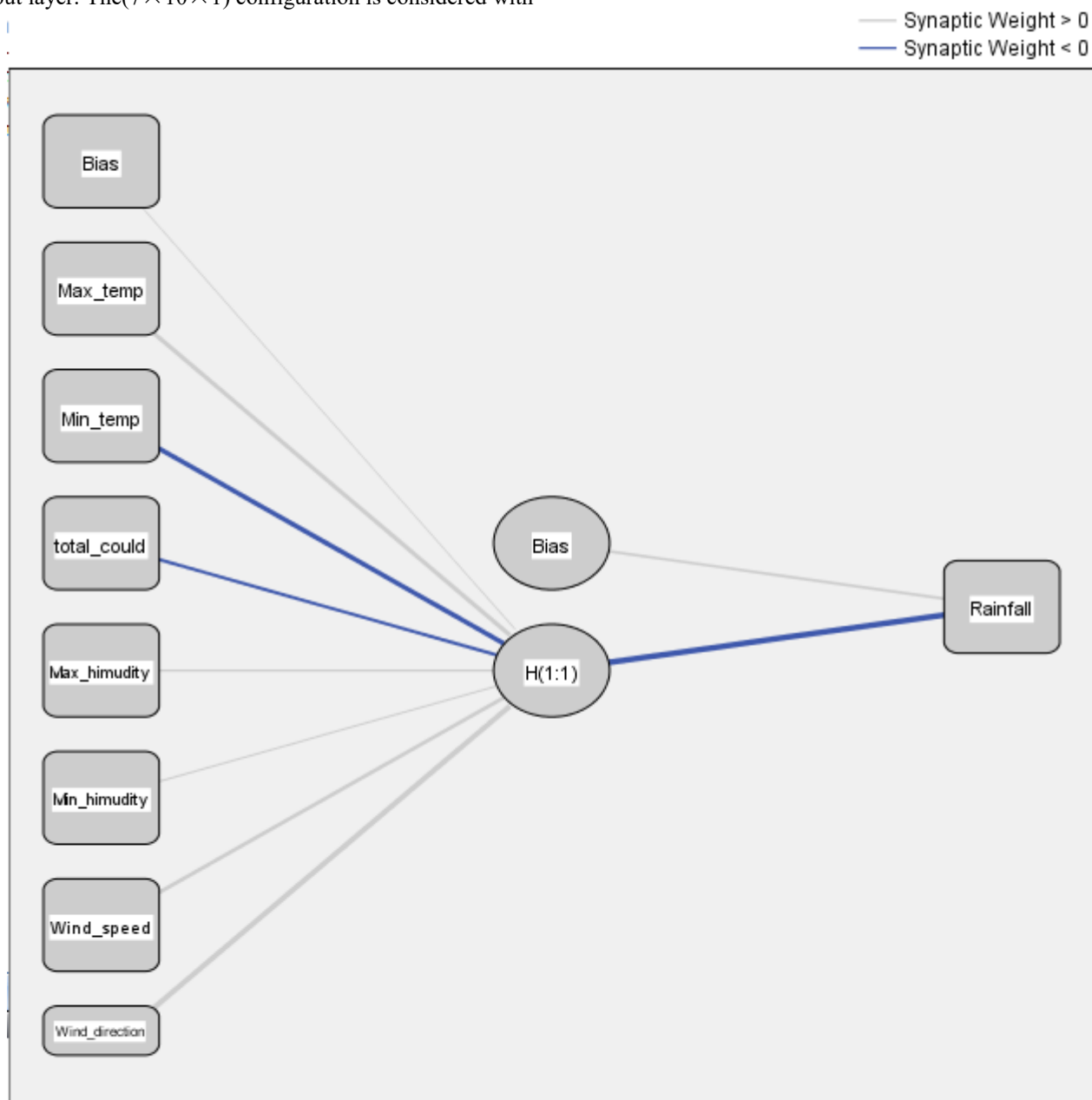
#### 1] Multiple Regression analysis.

Multiple regressions fit a model to predict a dependent (Y) variable from two or more independent (X) variables. Multiple linear regression models are often used as approximating functions. That is, true functional relationship between y and x1,x 2 ,x3 ... , is unknown, but over certain ranges of the repressor variables the linear regression model is an adequate approximation to the true unknown function. In the present study rainfall was treated as dependent variable and maximum & minimum temperature, minimum &maximum humidity ,wind speed, wind direction & total could cover(octa) as independent variable. The form of the multiple linear regression equation fitted to the weekly average weather parameters is given below.

flexible structure, descending from studies of brain functioning. As the family grew, most of the new models were designed for no biological applications, though much of the associated terminology reflects its origin.. In this paper, artificial neural networks have been used to create different models which predict the monthly rainfall of Marathwadal region in Maharashtra. The input data set comprises of parameters affecting rainfall such as minimum temperature, maximum temperature, minimum &maximum humidity ,wind speed, wind direction & total could cover(octa).The results obtained from all the models were then compared to each other using the Mean Square Error (MSE) and sub sequentially the conclusions were draw. The combination of neurons and synapses produce a value which is dependent on the values presented to them at their inputs. There are various types of neuron which generate their output based on

different processes. In this process we have used hidden (sigmoid) neuron. In this study we have used the best suited architecture of Feed Forward Neural Network Model (FFNN) for monthly rainfall data. In the proposed neural network model three layers are considered, the input environment with the significant variables, two hidden layers and an output layer. The  $(7 \times 10 \times 1)$  configuration is considered with

7 neurons in the first hidden layer and 10 neurons in the second hidden layer and only one neuron in the output layer. Fig 1 shows The  $(7 \times 10 \times 1)$  configuration is considered with 7 neurons in the first hidden layer and 10 neurons in the second hidden layer and only one neuron in the output layer of used Feed Forward Neural Network Model (FFNN).



Hidden layer activation function: Hyperbolic tangent

Output layer activation function: Identity

**Figure 1:** Feed Forward Neural Network

**3] Comparative study for using performance statistics.**

Comparison among the multiple regression model, and artificial neural network was made to identify the adequacy of the fitted models using the multiple correlation coefficients (R) and Mean Square Errors (MSE). The following two methods used for comparison selection for best fit model.

**i] Mean Square Error(MSE).**

In the dataset value  $y_i$  is associated with model value of  $\hat{y}_i$  are called predicted value and  $y_i$  is observed values of rainfall and model. In regression analysis, the term mean squared error is sometimes used to refer to the unbiased estimate of error variance. Mean square error is used for

comparison of model. the criteria for selecting best model is minimum mean square error .

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Where  $y_i$ =observed value of rainfall,  $\hat{y}_i$  = predicted values of rainfall data, n= Total no of day of rainfall

**ii) Multiple correlation coefficient (R)**

The correlation between  $\hat{y}_i$  and the actual  $y_i$  value is also called the multiple correlation coefficients,  $R_{y.12...k}$ , or simply R. Thus, R provides a measure of how well Y can be predicted from the set of X scores. The strength of prediction from a multiple regression equation is nicely measured by multiple correlation coefficient,(R). In the case of only two predictors,(R) can be found by using the formula

$$R = \sqrt{\frac{r_{Y1}^2 + r_{Y2}^2 - 2(r_{Y1})(r_{Y2})(r_{12})}{1 - r_{12}^2}}$$

**Table 1:** Descriptive statistics of selected Metrological stations of Marathwada region

Station	Variable	Mean	StDev	Minimum	Maximum	Skewnes	Kurtosis
Aurangabad	rainfall(mm)	3.8	5.235	0	17	1.49	1.34
	MaxTemp(°C)	5.096	8.399	0	56	3.34	14.81
	Min Temp(°C)	30.719	1.948	25	34	-0.64	0.74
	Total could cover(octa)	22.632	0.865	20	25	-0.37	0.75
	Max Rel humidity(%)	5.553	2.166	0	8	-0.77	-0.36
	Min Rel humidity(%)	79.851	7.604	53	93	-0.88	1.17
	Wind speed(kmph)	52.614	9.986	29	88	0.4	0.88
	Wind direction(deg)	13.825	5.553	3	28	0.53	-0.18
Latur	rainfall(mm)	4.991	10.543	0	93	5.69	43.44
	MaxTemp(°C)	31.851	1.741	28	35	-0.23	-0.58
	Min Temp(°C)	22.711	0.975	20	25	-0.2	0.28
	Total could cover(octa)	6.202	1.815	1	8	-1.14	0.66
	Max Rel humidity(%)	77.623	7.948	60	94	-0.03	-0.74
	Min Rel humidity(%)	51.991	9.798	32	91	0.68	1.46
	Wind speed(kmph)	15.482	5.805	3	30	0.15	-0.45
Beed	rainfall(mm)	5.214	7.796	0	54	2.93	12.45
	MaxTemp(°C)	34.057	3.451	25	42	0.3	-0.49
	Min Temp(°C)	24.071	1.966	20	29	-0.06	0.36
	Total could cover(octa)	6.036	1.875	0	8	-0.94	0.25
	Max Rel humidity(%)	70.34	12.65	34	94	-0.12	-0.21
	Min Rel humidity(%)	44.1	14.64	13	80	0.12	-0.54
	Wind speed(kmph)	17.393	5.019	2	28	-0.82	0.4
	Wind direction(deg)	262.76	31.78	102	311	-2.52	9.96
Osmanabad	rainfall(mm)	3.36	6.746	0	53	4.46	27.29
	MaxTemp(°C)	31.465	1.705	27	35	0	-0.84
	Min Temp(°C)	21.035	1.744	18	25	0.45	-0.94
	Total could cover(octa)	6.193	1.789	1	8	-0.98	0.19
	Max Rel humidity(%)	77.395	7.58	61	93	-0.05	-0.71
	Min Rel humidity(%)	50.947	9.359	27	86	0.78	2.15
	Wind speed(kmph)	15.561	5.635	3	31	0.48	-0.03
	Wind direction(deg)	261.9	46.21	86	328	-2.46	6.46
Jalna	rainfall(mm)	2.017	5.07	0	37	4.24	22.8
	MaxTemp(°C)	31.861	1.696	26	35	-0.65	0.69
	Min Temp(°C)	23.243	1.225	21	26	0.51	-0.76
	Total could cover(octa)	9.17	16.25	0	85	4.41	18.16
	Max Rel humidity(%)	77.678	9.059	53	92	-0.62	-0.1
	Min Rel humidity(%)	50.49	12.63	11	88	-0.42	1.71
	Wind speed(kmph)	27.33	56.14	2	299	4.46	18.5
	Wind direction(deg)	274.44	46.79	55	339	-2.39	8.85

Where rij= value of correlation coefficient rainfall, minimum &maximum temperature, minimum &maximum humidity, wind speed, wind direction & total could cover (octa)

**4. Result and Discussions**

The present mythology is applied to seven parameter, rainfall, maximum & minimum temperature, minimum &maximum humidity, wind speed, wind direction & total could cover (octa) of five years (2000 to2005) of Marathwada region. In this region five metrological stations have selected for study such as Aurangabad, Latur, Beed, Osmanabad, and Jalna. Multiple linear regression model we have use rainfall is dependent variable and other seven are independent variable. Table1.shows the descriptive statistics of each variable .Jalna mean rainfall value 2.01 mm is less than other selected station .it was less affected rainfall station in Marathwada region .in the table of value of skwnes indicate that continues less rainfall affected in this region.

**Multiple Linear Regression Model (MLR)**

In multiple regressions a common goal is to determine which independent variables contribute significantly to explaining the variability in the dependent variable. A goal in determining the best model is to minimize the mean square error (MSE), which would intern maximize the multiple correlation value (R).the three show the multiple regression model coefficient .The result indicated that using seven

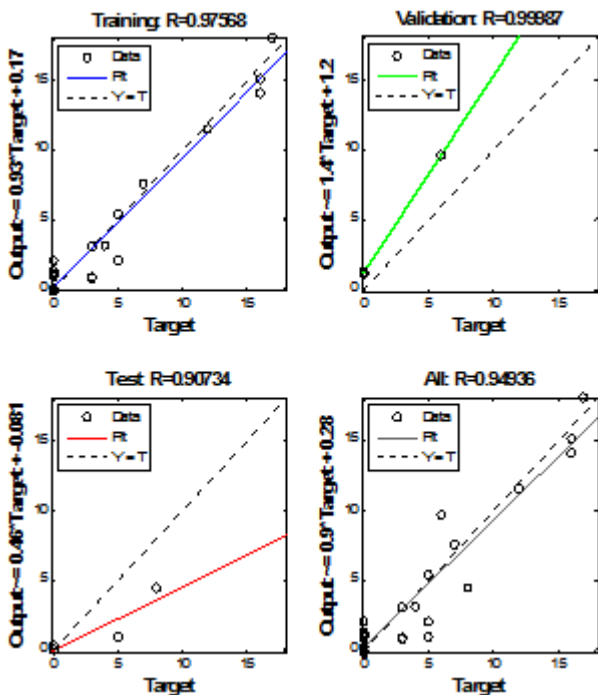
climatic variables under study minimum &maximum temperature , minimum &maximum humidity ,wind speed, wind direction & total could cover(octa) has positive impact on rainfall. The value of R-square Aurangabad station is 0.838in found to be significance at 5% level of significant indicates that 83.8% variation in the rainfall of explain by seven climate parameters under study.

**Table 2:** Multiple Regressions coefficient &constant of metrological station

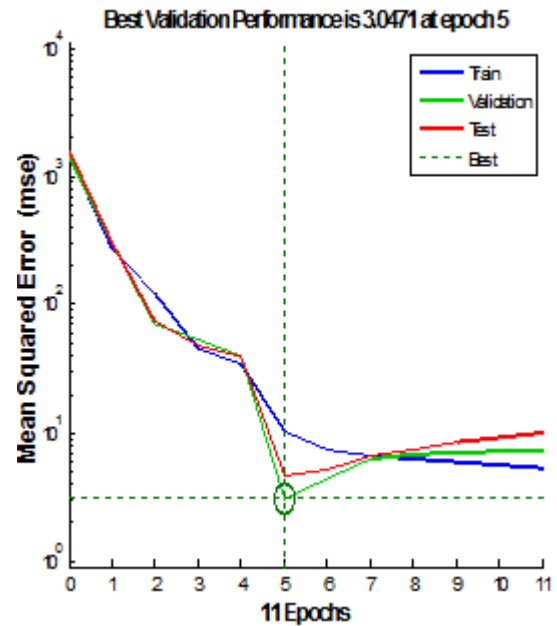
Station	constant	X1	X2	X3	X4	X5	X6	X7
Aurangabad	44.2	-0.680	-0.16	-0.711	0.285	0.522	0.09	0.213
Latur	-72	-0.722	2.02	1.38	0.904	-0.137	0.372	-0.0875
Beed	52.9	- 1.18	0.100	0.636	0.138	- 0.231	- 0.111	- 0.0426
Osmanbad	-47.7	-0.047	0.308	0.467	0.603	0.02	0.394	-0.0406
Jalna	-46.8	0.838	0.065	-0.041	0.061	0.336	0.035	-0.0072

**Feed Forward Neural Network Model (FFNN)**

In the forward neural network model have use parameter of rainfall such as rainfall, minimum &maximum temperature, minimum &maximum humidity ,wind speed, wind direction & total could by using the(7×10×1) configuration is considered with 7 neurons in the first hidden layer and 10 neurons in the second hidden layer and only one neuron in the output layer. Fig2. Show the correlation coefficient (R) of training &testing data of Artificial Neural Network (ANN) model.in this method we have use Select the Feed Forward Neural Network Model (FFNN) for monthly rainfall data. In the proposed neural network model three layers are considered, the input environment with the significantvariables.,fig2 show model performance of Aurangabad metrological station on the basis of mean square error (MSE)



**Figure 2:** Multiple correlation coefficient (R) Aurangabad



**Figure 3:** Model performance of Aurangabad Performance of Model on Basis of Selection Criteria

In this study we have use rainfall parameter such as as rainfall, minimum &maximum temperature, minimum &maximum humidity, wind speed, wind direction & total could. In multiple linear regression rainfall is dependent variable and other seven are independent variable .in forward neural network mode seven parameter are input variable &rainfall is output variable .The selection criteria for best performance of model is that model have maximum multiple correlation coefficient value (R) and minimum value of mean square error (MSE).

The results of statistical analyses for Multiple Regression (MLR) &Artificial Neuron Network (ANN) shown in table 3.The result shows that Aurangabad station value of the Multiple correlation coefficient (R ) of Aurangabad satiation 0.915 for multiple regression model while it was 0.949 in neural network model, secondly, the mean square error(MSE) for multiple regression model is 2.41 which is higher than that of the mean square error(MSE) obtained through neural network model as 0.945 on comparing of these result finally concluded the Aurangabad station have Artificial Neural Network (ANN) is best model for estimation of monthly rainfall data .

**Table 3:** Comparison of Multiple Linear Regression (MLR) & Artificial Neural Network (ANN)

Station	Multiple Linear Regression (MLR)		Artificial Neural Network (ANN)	
	MSE	R	MSE	R
Aurangabad	2.41	0.915	0.47	0.949
Latur	1.942	0.856	0.193	0.931
Beed	4.06	0.812	0.232	0.908
Osmanbad	1.792	0.792	0.332	0.905
Jalna	4.403	0.872	0.432	0.89

## 5. Conclusion

This aim this study to estimate monthly rainfall data of Marathwada region by using comparison of multiple linear regression & Artificial Neural Network method on the basis of comparison of statistical parameters such as minimum mean square error (MSE) & maximum is maximum multiple correlation value (R). In Multiple regression model we have use rainfall is dependent variable and other seven are independent variable. In neural network model we have use rainfall is output and other seven are input variable. Result show that Aurangabad station value of the Multiple correlation coefficient (R) of Aurangabad station 0.915 for multiple regression model while it was 0.949 in neural network model, secondly, the mean square error (MSE) for multiple regression model is 2.41 which is higher than that of the mean square error (MSE) obtained through neural network model as 0.47 on comparing of these result finally concluded the Aurangabad station have Artificial Neural Network (ANN) is best model for estimation of monthly rainfall data. Finally on the basis of maximum value of Multiple correlation coefficient (R) and minimum value of mean square error (MSE) it was concluded that the selected metrological station neural network model is better than multiple regression model for estimation of monthly rainfall data. This study is help farmers in making decision concerning with their crop & regional and local plan they depend on rainfall of Marathwada region.

## References

[1] Chattopadhyay, S." Multiplayer feed forward artificial neural network model to predict the average summer monsoon rainfall in India". Acta Geophysica 55(3): 369-382. 2007.

[2] Chhaya Sonar, S L Sananse, A Y Tayde, Ankush Ubale. "Impact of monsoon and statistical forecasting of production of important oil seeds in Aurangabad division" International Journal of Statistika and Mathematika May to July 13(1): 38-40 2015.

[3] Dahnal, V., K. Kuca and D. Jun, 2005. What are in Ecology. Ecological Modeling, 90: 39-52. Artificial neural networks and what they can do?. Fac.. Biomed. Papmed. Palacky olomuac Univ., Czech. Republic, 149: 221-224

[4] Kannan, M S., Prabhakaran, P., Ramachandran.). Rainfall forecasting sing data mining Technique.

International Journal of Engineering and Technology, 2:397-401. 2010

[5] S. Lee, S. Cho and P.M. Wong, "Rainfall prediction using Artificial Neural Networks," Journal of Geographic Information and Decision Analysis, Vol. 2, No. 2. 1998.

[6] Gato, S., Jayasuriya, N., Roberts, peter, "Temperature and rainfall thresholds for base urban water demand modeling." Journal of hydrology 337(2007): 364-376. 2007.

[7] S. Maxime, G. Hartmut, R. Lars, K. Nicole and O. Ricardo, "Statistical Downscaling of Precipitation and Temperature in North-Central Chile: An Assessment of Possible Climate Change Impacts in an Arid Andean Watershed," Hydrological Sciences Journal, Vol. 55, No. 1, 2010, pp. 41-57.

[8] T. Sohn, J. H. Lee, S. H. Lee, C. S. Ryu, "Statistical Prediction of Heavy Rain in South Korea", in Advances in Atmospheric Sciences, Vol. 22, No. 5, , pp.703-710. 2005

[9] Hu, Michael Jen-Chao." Application of the adaline system to weather forecasting". Diss. Department of Electrical Engineering, Stanford University. 1964.

[10] Sharma, M.A. and Singh, J.B. "comparative study of rainfall forecasting models "New York science journal 4(7):115-120. 2014

[11] Jyoti Upadhaya, Assam University."Climate Change and its impact on Rice productivity in Assam".

[12] Olaiya Folorunsho,"Application of Data mining Techniques in Weather Prediction and Climate change studies" 2012

[13] Wang, Zhi-L. and Sheng, Hui-h, Rainfall prediction using generalized regression neural network" case study Zhengzhou. Computational and Information Science (ICCIS) 1265-1268.2010.