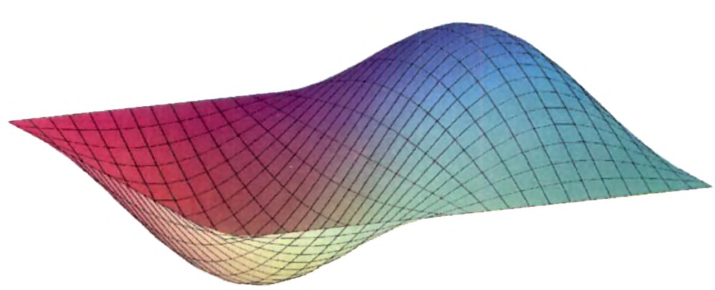
Chapter 5

PREDICTION
OF
ANNUAL RAINFALL
BY
DOUBLE FOURIER SERIES
AND
ARTIFICIAL NEURAL NETWORKS



5.1 INTRODUCTION

The problem of prediction of weather parameters like annual rainfall (ARF), soil temperatures (ST) etc. is the primary concern in the field of Meteorology Processes like rainfall in nature are chaotic in behavior (*Farmer et al* [30]). In such processes regularities like periodicities are mixed with noise. This results in large error for long time prediction.

The sum of daily rainfall during a year is known as the Annual Rainfall (ARF).

In this chapter we deal with the problem of prediction of annual rainfall (ARF) by using two different methods namely,

i) Double Fourier Series Method (DFS).

ii) Artificial Neural Network Method (ANN).

Case study is done for Anand station.

In DFS the prediction is based on a Double Fourier Series model of the process. In ANN, the process is being approximated by a feed -forward neural Network. We make use of relevant data of the Anand station during the past 48 years (1958-2006).

5.2 PROBLEM FORMULATION

Our objective is to predict the ARF of a year, say, 2006 using the historical data series (DS) from 1958 to 2005. The following graph (Figure 5.1) depicts the known Data Series (DS) of ARF from 1958 to 2005 are depicted in Fig 5.1. Figure shows that there is no vivid trend in the ARF series.

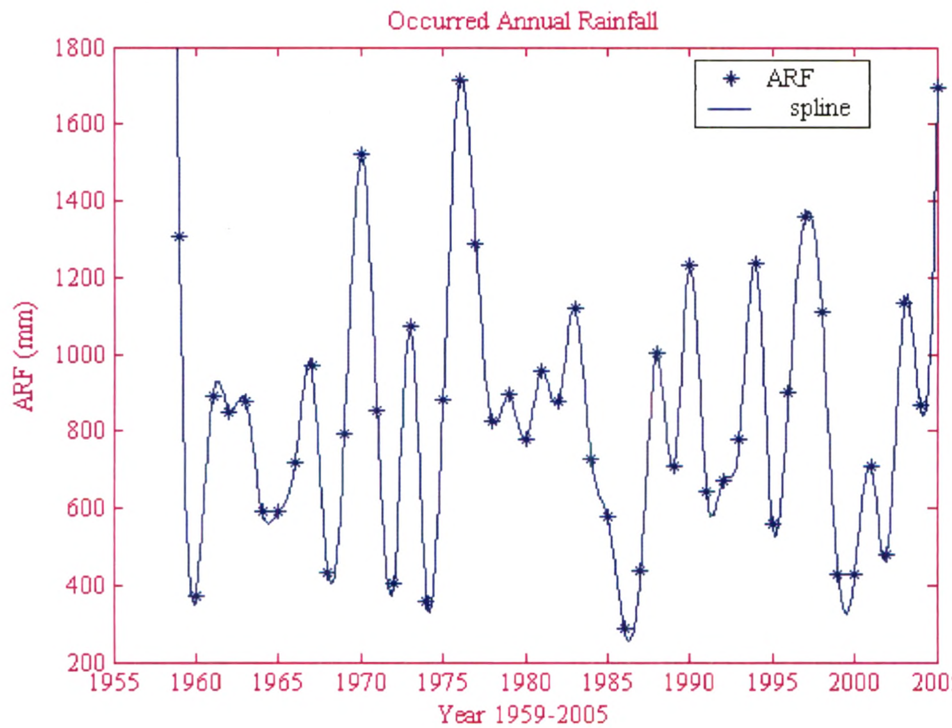


Fig. 5.1:

We now look for major variables, which affect the amount of ARF. Singh *et al.* [146] have studied the "Energy exchanges between ocean and atmosphere in relation to south-west monsoon over India" and provided the following conclusions:

(i) If higher evaporation occurs in the month of May over the southern belt of Indian seas (5°-15°N) then it is favorable for the ensuing southwest monsoon.

(ii) If a strong field of momentum flux occurs in the month of May over

Indian bay then it generates good monsoon.

Moreover, the May month's maximum air temperature (MAT) is the effect of Southern oscillation index (SOI) that is the difference of air pressure at Tahiti and Darwin Ireland situated at southeastern tropical Pacific regions (Pant *et al.*[118]; Walker [164]). This index is directly affecting ' rainfall of India (Parthsarthy *et al.* [120]; Shukla *et al.* [141])

To calculate the SOI index different methods are available. The method used by the Australian Bureau of Meteorology [6] is the Troup SOI, which is the standardized anomaly of the Mean Sea Level Pressure (MSLP) difference between Tahiti and Darwin Ireland. It is calculated as follows:

$$SOI = 10 * \left(\frac{\text{Pdiff - Pdiffav}}{\text{SD}\left(\text{Pdiff}\right)} \right)$$

where,

Pdiff = (average Tahiti MSLP for the month) - (average Darwin MSLP for the month),

Pdiffav = Long term average of Pdiff for the month,

and

SD (Pdiff) = Long term standard deviation of Pdiff for the month.

The effect of MAT on ARF is shows in Figure 5.2.

If we use curve-fitting methods to predict ARF as a function of MAT, we get the following results given in Table 5.1. Related Root Mean Square

Error (RMSE) and Percentage of Average Error (PAE) are found and given in the Table 5.1. RMSE and PAE are found by the following formulas.

Root Mean Square Error (RMSE)

$$= \sqrt{\sum_{i=1}^{n} \frac{(\text{Actual ARF - Predicted ARF})^2}{\text{n}}} \; ; \qquad (5.1)$$

where, n= total number of predictions

Percentage of Average Error

$$\text{PAE}= \frac{RMSE}{Average\ of\ Actual ARF} * 100 \qquad (5.2)$$

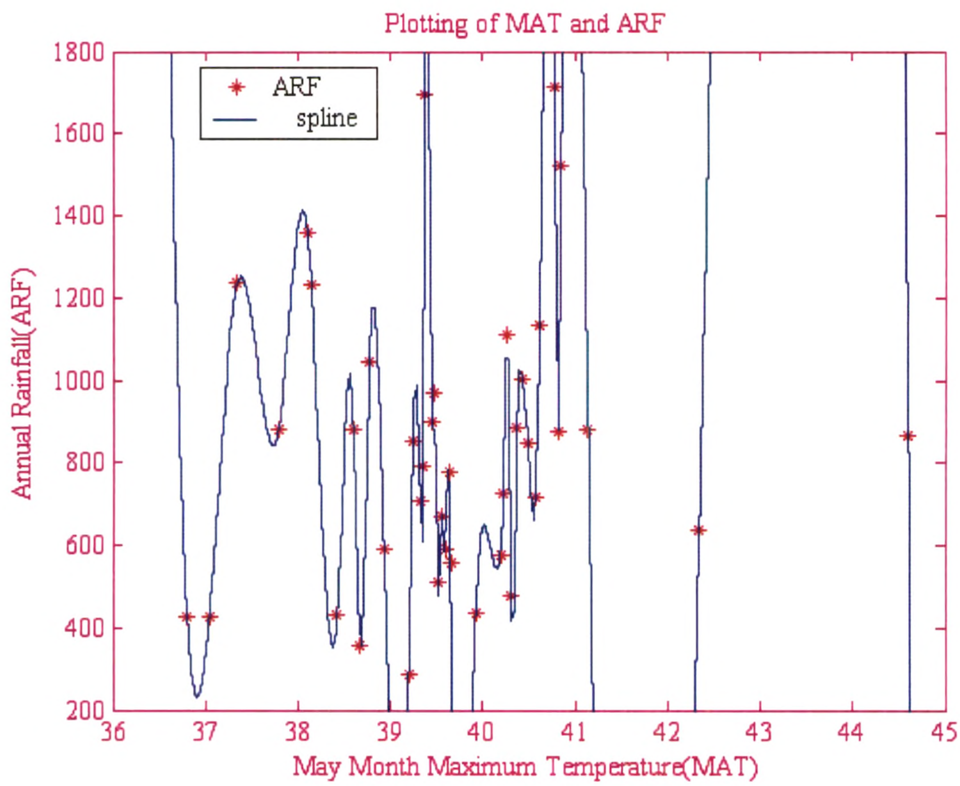

Plotting of MAT and ARF
ARF
spline
Annual Rainfall(ARF)
May Month Maximum Temperature(MAT)
1800
1600
1400
1200
1000
800
600
400
200
36
37
38
39
40
41
42
43
44
45

Fig 5.2

Table 5.1:
DETAILS OF THE CURVE FITTING WITH ONE PREDICTOR
'MAT' AND PREDICTAND 'ARF'

Sr. No	Type of the Curve Fit.	Predicted ARF (mm) Year	ARF	Actual ARF (mm)	RMSE	PAE (%)
1	Linear Function $y = P_1x + P_2$, where, P_1 =14 23 and P_2=274.74 (Fig: 5.3)	2002	855.73	479.0	346.41	35.5
		2003	865.17	1135.4		
		2004	986 75	866.0		
		2005	*827 39	1693.0		
		2006	913.62	1414.1		
2	Cubic Function $y = P_1x^3 + P_2x^2 + P_3 x + P4$ where, P_1= 1.28, P_2 =-161.44, P_3 = 6.77 and P_4=-9.37 (Fig · 5 4)	2002	865.36	479.0	337.47	34 66
		2003	878 93	1135.4		
		2004	848.42	866.0		
		2005	*820.63	1693 0		
		2006	924.16	1414.1		
3	Sixth Degree Poly. Function. $y = P_1x^6 + P_2x^5 + P_3x^4 + P_4x^3 + P_5x^2 + P_6x + P_7$ where, P_1 = 2.37, P_2 =-566 75, P_3 = 56336, P_4 = -2.9810^6, P_5 = 8.87x10^7, P_6=-1.41x10^9, P_7 = 9.28x10^9 (Fig : 5.5)	2002	6652	479.0	5883 0	526.51
		2003	7971	1135.4		
		2004	8231	866.0		
		2005	*1214.7	1693.0		
		2006	7227	1414.1		

*THIS VALUE IS NOT INCLUDED IN THE COMPUTATION OF RMSEs' AND PAE

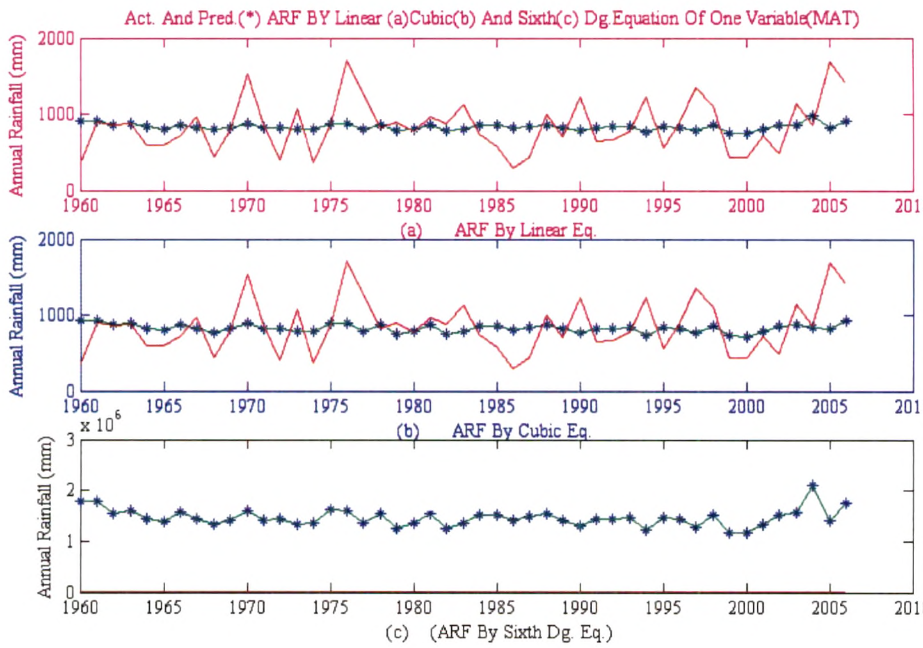


Fig. 5.3

Fig. 5.4

Fig. 5.5

Predicted ARF by a linear equation of one variable MAT that is

$$ARF = 14.23*(MAT) + 274.74 \qquad (5.3)$$

is depicted in the Figure 5.3. This figure shows very high difference between actual and predicted ARF.

Cubic equation fitted to predict the ARF as a function of one variable MAT is given by,

$$ARF = 1.28*(MAT)^3 + (-161.44)*(MAT)^2 + (6.77)*(MAT) - 9.37; \qquad (5.4)$$

Thus, prediction of ARF by linear, cubic and sixth degree polynomial of one variable MAT is not satisfactorily (Fig. 5.3, 5.4 and 5.5).

Rao *et al* [126] and Jagannathan *et al.* [74] did extensive study to obtain trends of annual rainfall at various stations taking very large number of historical data. Chowdhury *et al.* [19] have concluded that after the testing of the data series of 60 to 100 years for randomness, it is found that

i) There is no trend in yearly, 5 yearly, 10 yearly mean rainfalls for a majority of stations but at some places 2 or 3 years period is found.

ii) There is no short period of cycle in annual rainfall and in distribution of rainfall, particularly in arid and semi-arid regions of NW India.

iii) There is no increasing or decreasing trend but only oscillation from year to year.

Anand station lies in the semi arid region of India as per (ii) and Figure 5.1 it has no significant trend for annual rainfall but oscillation occur from year to year. So, we take the previous year ARF as another factor affecting the current year ARF.

Figure 5.6 shows the relation between previous year annual rainfall (PARF) and annual rainfall (ARF). PARF is plotted on the x-axis and ARF is on y-axis. This figure shows finite oscillations.

Curve fitting method is tried to derive the relation between PARF and ARF. Previous year Annual Rainfall (PARF) is used as predictor and linear, cubic and sixth degree polynomial equations are used to predict the ARF.

Table 5.3.shows these equations of the curve fit RMSEs (5.1) and related PAE (5.2) of prediction.

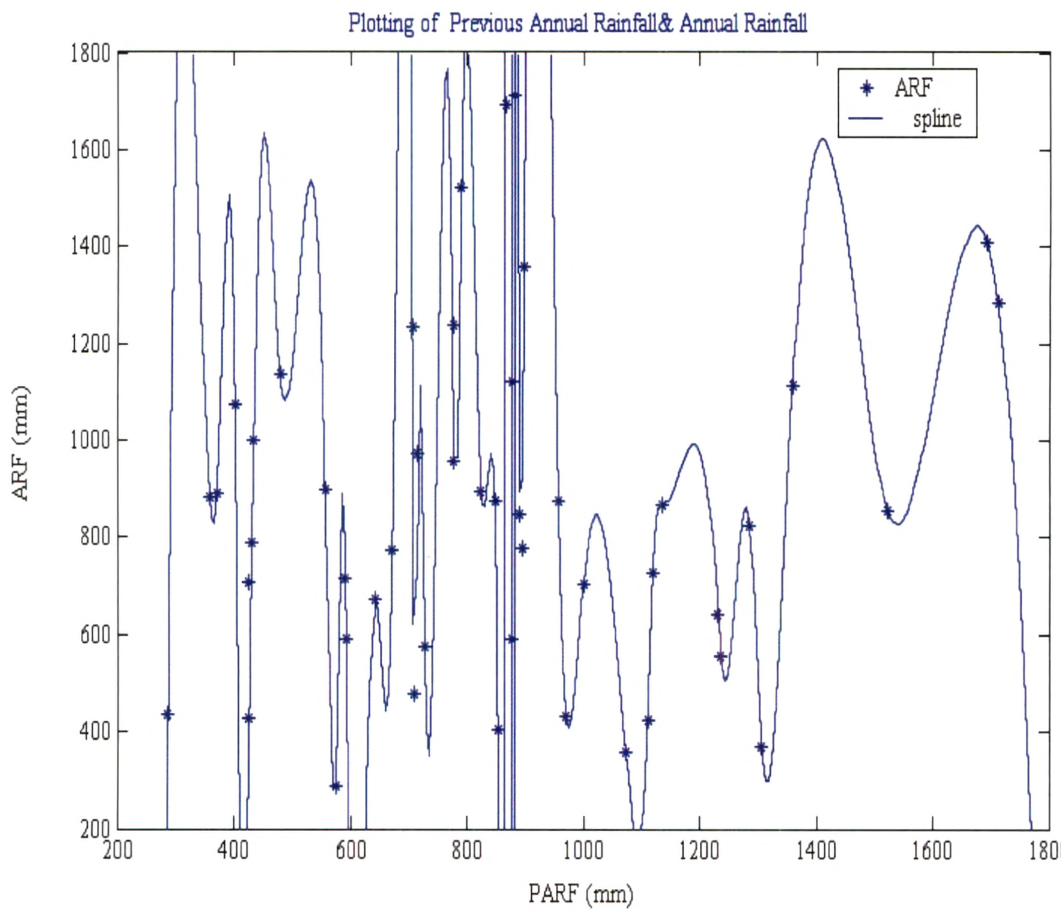


Fig: 5.6

The stared data in Table: 5.2 are not included for computation. As it is an outlier in the data series (DS).

Predicted ARF by linear, cubic and sixth degree equations are plotted in the Figure 5.7.

TABLE 5.2
DETAILS OF THE CURVE FITTING WITH ONE PREDICTOR '*PARF*' AND PREDICTAND '*ARF*'

Sr. No	Type of the Curve Fit.	Predicted ARF (mm) Year	ARF	Actual ARF (mm) (2002-2006)	RMSE	PAE (%)
1	Linear Function	2002	832 15	479.0		
	$y = P_1 x + P_2$,	2003	801.60	1135.4	331.76	34 08
	where, $P_1 = 0.13$ and	2004	888.76	866.0		
	$P_2 = 73.80$	2005	*853 03	*1693 0		
	(Fig:5.7)	2006	962.83	1414.1		
2	Cubic Function	2002	931.1	479.0		
	$y = P_1x^3 + P_2x^2 + P_3 10^3 + P_4$,	2003	852.4	1135.4		
		2004	734 4	866.0	281.70	28.93
	where, $P_1 = 0.00022276$	2005	*874.0	*1693.0		
	$P_2 = -0.064005$	2006	1289.1	1414.1		
	$P_3 = 5\ 5594$ $P_4 = -58.703$ (Fig:5.7)					
3	Sixth Degree Poly. Function.	2002	875.7	479.0		
	$y = P_1x^6 + P_2x^5 + P_3x^4 + P_4x^3$	2003	791 9	1135.4		
	$+ P_5x^2 + P_6x + P_7$ where,	2004	680.7	866.0	282.80	29.05
	$P_1 = (-2.938)10^{-9}$ $P_2 = (1.6684)10^{-6}$	2005	*954 1	*1693.0		
	$P_3 = -0.00037154$ $P_4 = 0.041326$ $P_5 = -2.4177$ $P_6 = 70.849$ $P_7 = -738.41$ (Fig.5.7)	2006	1313.0	1414 1		

*THIS VALUE IS NOT INCLUDED IN THE COMPUTATION OF RMSEs' AND PAE

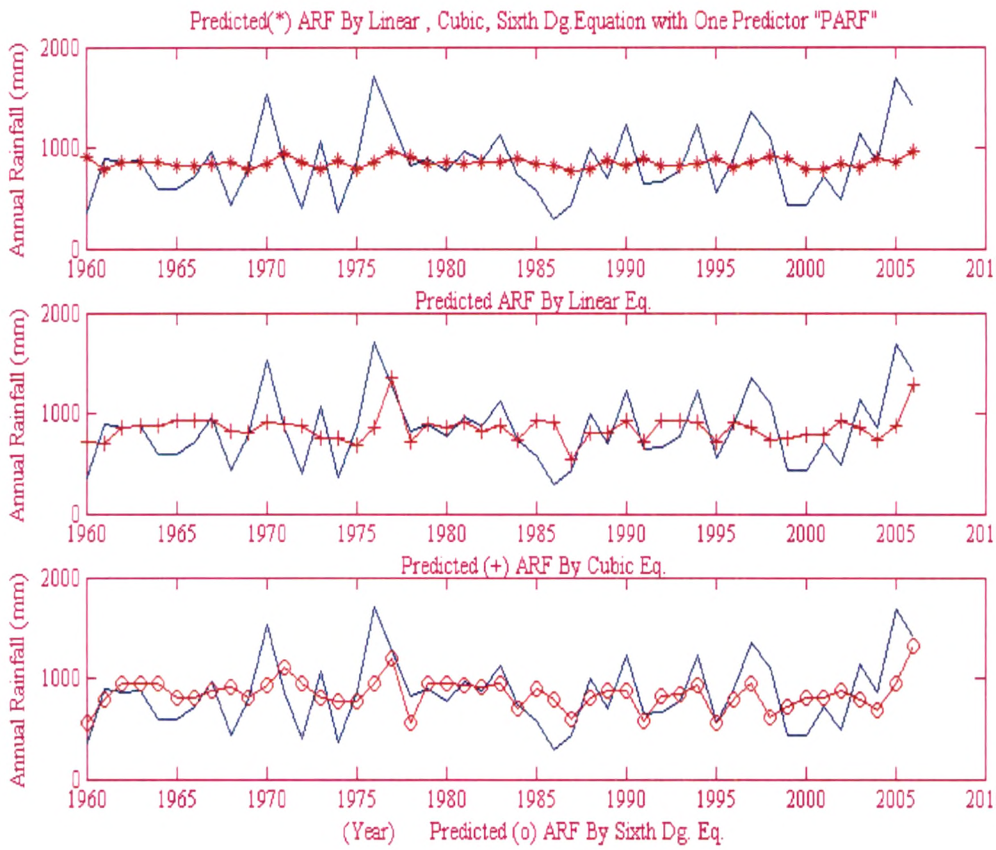


Fig: 5.7

Table 5.2 shows the RMSEs' are very high and Percentage of Average Error (PAE) of prediction is more than 10 %. Therefore, in both the cases ARF prediction by one variable at a time namely, MAT / PARF is found to be non-significant to actual ARF

Therefore, we take two predictors namely May month's Maximum Air Temperature (MAT) and Previous year Annual Rainfall (PARF) and current year Annual Rainfall (ARF) as the predictand.

It is found that the correlation between MAT & ARF and PARF & ARF are 11.5% and 12.8 % respectively. The relationship is not linear but oscillating.

First attempt to predict the ARF is to determine the best linear prediction equation (in the least square sense).

Let the multiple linear regression equation be of the form,

$$ARF= P_0 + P_1 * (MAT) + P_2 *(PARF); \quad (5.5)$$

where , P_1 & P_2 are slopes and P_0 is the intercept. Using the least square technique the slopes and intercept are found to be $P_0 = -596.5$ $P_1 = 34.5$ and P2 = 0 078. Thus we have

$$ARF= -596.5+ (34.5)*MAT + 0.078*PARF \quad (5\ 6)$$

The predicted ARF by using the linear regression equation is shown in Figure 5 8. We calculate the Root Mean Square Error and percentage of average error (PAE). These figures show that predictand (*) (that is ARF) is not in a good agreement with actual ARF. Details of the results are given in Table 5 3.

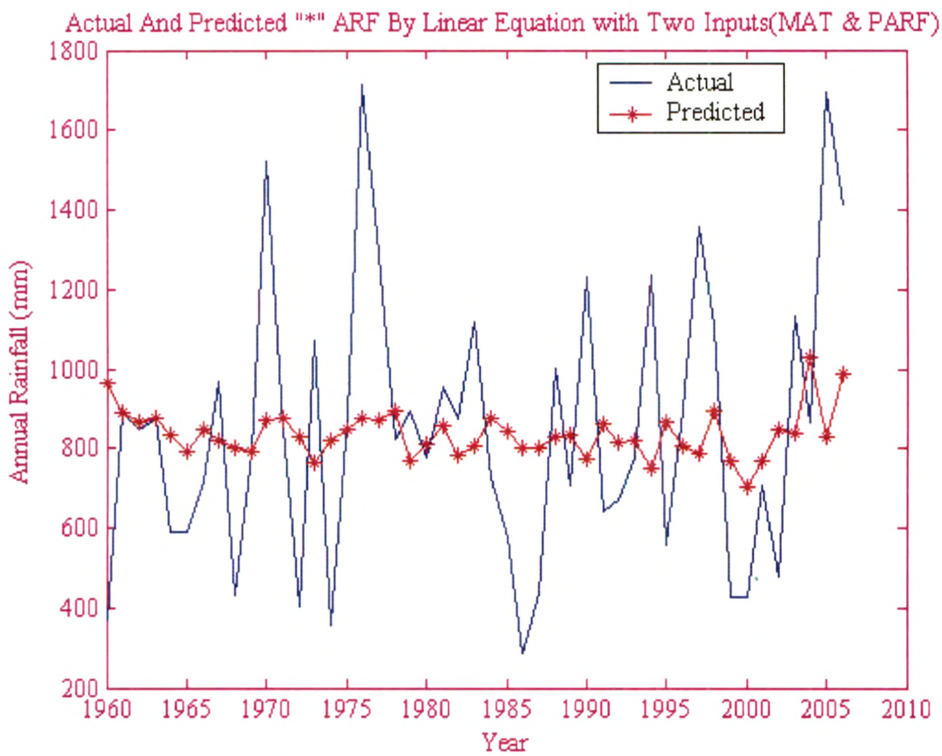


Fig: 5.8

TABLE 5.3

DETAILS OF THE RESULTS OBTAINED BY LINEAR EQUATION OF TWO VARIABLES MAT & PARF

Sr. No	Predicting Year	Actual ARF (mm) I	Predicted ARF (mm) II	Difference I-II (mm)	RMSE	Percentage of Average (PAE)Error (%)
1	2002	479.0	847.3	370.0	327.74	33.66
2	2003	1135.4	840.0	330.9		
3	2004	866.0	1028.7	156.8		
4	2005	*1693.0	827.5	867.2		
5	2006	1410.5	989.5	421.0		

* THIS VALUE IS NOT INCLUDED IN THE COMPUTATION OF RMSEs' AND PAE

Table 5.3 shows the high value of RMSE (327.74) and PAE (33.66 %) obtained , from formulae (5 1) and (5.2) for all the predictions of the year 2002 to 2006, excluding 2005.

Because of ARF's oscillatory nature (Figure 5.1) we have obtained high values of RMSE in both the cases.

Therefore, we look for non-linear models to represent the process. Here we try two methods, namely:

i) Double Fourier Series (DFS)

and

ii) Artificial Neural Network (ANN)

5.3 DATA

The two input variables used in the present problem of annual rainfall prediction are highest maximum May month's Air Temperature (MAT) and Previous year Annual Rainfall (PARF) (Table 5.4). Here ARF is predicted for the year 2002, 2003, 2004, 2005 and 2006 by the two methods. The estimation of ARF of the years 2002 to 2005 by DFS and ANN methods is helpful to select the parameters to get the good accuracy for next year prediction that is of the year 2006.

TABLE 5.4
USED INPUT (*MAT, PARF*) AND OUTPUT (*ARF*) DATA SERIES

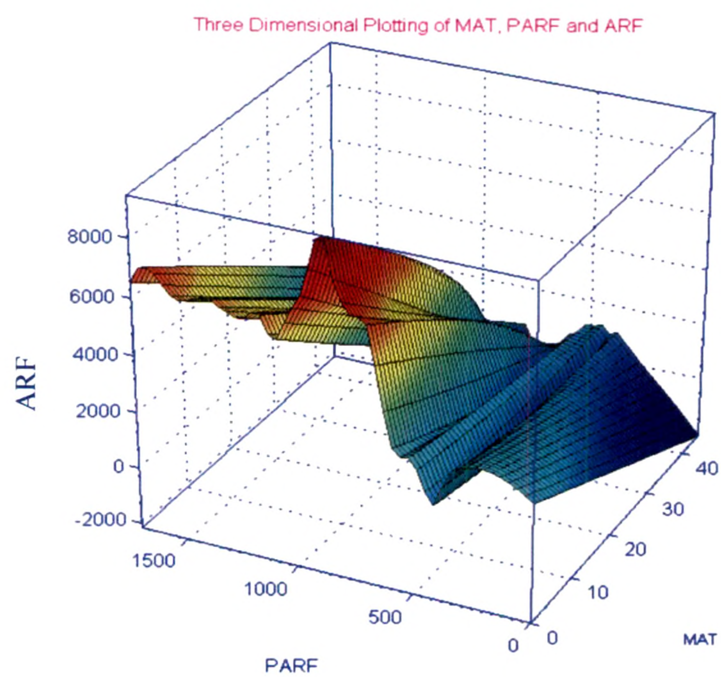
Sr. No.	Year	MAT (°C)	PARF (mm)	ARF (mm)	Sr. No.	Year	MAT (°C)	PARF (mm)	ARF (mm)
1	1959	42.05	1018.5	1306.2	25	1983	38.77	876.7	1119.6
2	1960	42.34	1306.2	370 5	26	1984	40.22	1119.6	726.7
3	1961	42 34	370.5	890.4	27	1985	40.19	726.7	574.5
4	1962	40.49	890.4	850 1	28	1986	39.2	574.5	286.9
5	1963	40.82	850.1	876.8	29	1987	39.92	286.9	434.0
6	1964	39.59	876.8	592.5	30	1988	40.43	434.0	1000.4
7	1965	38.93	592.5	591.4	31	1989	39 32	1000.4	706 0
8	1966	40.57	591.4	715.5	32	1990	38.14	706.0	1232.1
9	1967	39.48	715.5	971.5	33	1991	39.51	1232.1	643.1
10	1968	38.42	971.5	430.4	34	1992	39.55	643.1	672.0
11	1969	39.35	430.4	789 5	35	1993	39.64	672.0	775.5
12	1970	40.85	789.5	1522.6	36	1994	37.34	775.5	1236.7
13	1971	39.24	1522.6	854.2	37	1995	39.65	1236.7	557.4
14	1972	39.51	854.2	402.8	38	1996	39.46	557.4	897.9
15	1973	38 6	402.8	1074.0	39	1997	38.1	897.9	1359.2
16	1974	38.67	1074.0	357.9	40	1998	40.27	1359.2	1111.3
17	1975	41.14	357.9	882.8	41	1999	37.04	1111.3	425.2
18	1976	40.78	882.8	1712.5	42	2000	36.8	425.2	426.6
19	1977	38.77	1712.5	1286.0	43	2001	38.6	426.6	709.2
20	1978	40.37	1286.0	824.3	44	2001	40.3	709.2	479.2
21	1979	37.79	824.3	894.9	45	2003	40.61	479.2	1135.4
22	1980	38.77	894.9	778.0	46	2004	44.6	1135.4	866.0
23	1981	40.37	778.0	956.8	47	2005	39.37	866.0	1693.0
24	1982	37.79	956.8	876.7	48	2006	42.2	1693.0	1414.5

ANN model is trained and DFS coefficients are found for Anand station of Gujarat with historical data series available from 1959 to 2006 Training of the networks is done every time new, during the prediction of ARF for the years of 2002 to 2006.

Data series of the year 1959 to 2001 is used to train the network in ANN method and this data is also used to compute the Fourier Coefficients in

DFS method. In the second step ARF of the year 2002 is predicted by both the methods.

In the same way ARF for the year 2003, 2004, and 2005 are computed.



5.4 METHOD DOUBLE FOURIER SERIES

We attempt to make use of Double Fourier Series (DFS) to represent the natural process of annual rainfall (ARF).

As mentioned earlier Maximum May month's Air Temperature (MAT) and previous year annual rainfall (PARF) are the two major variables determining amount of annual rainfall of a year. Because of finite oscillatory nature of the annual rainfall process, we make use of Double Fourier Series

DFS as a Mapping Tool in Marine Cartography and it has been shown that DFS is a global model to map a projection process against coastal erosion.

Let x be the highest May month's Maximum Air Temperature (MAT), y be the Previous year Annual Rainfall (PARF) and z be the current year Annual Rainfall (ARF) Then by our assumption z is a function of x and y. That is, z=f(x,y)

5.4.1 DEFINITION OF DOUBLE FOURIER SERIES (DFS)

DFS contains terms of sine and cosine in combination of sine sine, sine cosine, cosine sine, and cosine cosine.

Let f(x,y) be a periodic function of two variables defined on a rectangle

$$K\ (\ (x,y).\ -P_1 < x < P_1,\ -P_2 < y < P_2\ ) \subset R^2,$$

The function is defined for all x and y with period $2P_1$ in x and period $2P_2$ in y This chosen rectangle can be converted in to $x \in (-\pi,\pi)$ *and* $y \in (-\pi,\pi)$ by making the substitutions

$X = \frac{\pi}{P_1} x$ *and* $Y = \frac{\pi}{P_2} y$; where, $-P_1 < x < P_1$ and $-P_2 < y < P_2$. Then the

function, $$f\left(\frac{P_1 X}{\pi}, \frac{P_1 Y}{\pi}\right) = \phi(X,Y)$$

has period 2π in both X and Y

Fourier series for this function f is given by z = ϕ (X,Y)=

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \lambda_{mn}[a_{mn} \cos mX \cos nY + b_{mn} \sin mX \cos nY + c_{mn} \cos mX \sin nY + d_{mn} \sin mX \sin nY] \tag{5.7}$$

where, $\lambda_{mn} = \begin{cases} \frac{1}{4} & for\ m = n = 0 \\ \frac{1}{2} & for\ m > 0,\ n = 0\ or\ m = 0,\ n > 0 \\ 1 & for\ m > 0,\ n > 0 \end{cases}$

the DFS co-efficient a_{mn}, b_{mn}, c_{mn} and d_{mn} are given by the following formulae.

$$a_{mn} = \frac{1}{\pi^2} \iint_K \phi(X,Y) \cos mX \cos nY \, dX \, dY,$$
$$b_{mn} = \frac{1}{\pi^2} \iint_K \phi(X,Y) \sin mX \cos nY \, dX \, dY,$$
$$c_{mn} = \frac{1}{\pi^2} \iint_K \phi(X,Y) \cos mX \sin nY \, dX \, dY$$
$$d_{mn} = \frac{1}{\pi^2} \iint_K \phi(X,Y) \sin mX \sin nY \, dX \, dY$$

The above series converges for a class of functions satisfying certain properties. We see this in the following theorem.

(II) THEOREM

Let f(x,y) be a continuous function defined on a square K, with bounded partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. Then the Fourier series of f(x,y) converges to f(x,y) at every interior point of K in a neighborhood of which the mixed partial derivative $\frac{\partial^2 f}{\partial x \partial y}$ exists. If f(x,y) is periodic with period 2π in x and y and has continuous partial derivatives $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ and $\frac{\partial^2 f}{\partial x \partial y}$, then the Fourier Series of f(x,y) converges to f(x,y) everywhere on K.

(II) FINDING DFS COEFFICIENTS

Let (x_i, y_i, z_i) be the observed Maximum May Month's Air Temperatures (MAT), Previous year Annual Rainfall (PARF) and Annual Rainfall (ARF) of the i^{th} year. Let the total number of observations be k. Define the observation vectors as,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ . \\ \\ x_k \end{bmatrix}, \; y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ \\ \\ y_k \end{bmatrix} \text{ and } z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ \\ . \\ z_k \end{bmatrix}$$

Thus, x is the highest May month's maximum air temperature data, y is the previous year annual rainfall data and z is the current year annual rainfall data.

We estimate the values of the DFS coefficients by the method of Least squares. We take finite values of m and n to estimate the DFS coefficients. Let c denote the coefficients vector to be determined and D be the data matrix and z be the observed output vector.

Therefore, the DFS coefficients is estimated by following system.

$$Dc=z$$

This system is solved in the Least square sense.

$$c= D^{+}z$$

where, D^{+} is the generalized inverse (Moore Penrose Inverse) of the data matrix D. This coefficient vector c is used for prediction of the ARF with required input vectors x and y.

5.4.2 METHOD IN PROBLEM SOLUTION:

To predict the ARF of the year 2002, DFS coefficients a_{mn}, b_{mn}, c_{mn} and d_{mn} are found by using the historical data namely MAT (1959 to 2001) and PARF (1958 to 2000). These coefficients are used in DFS method to predict the ARF of the year 2002.

In the same manner ARF of the year 2003 to 2006 are found (Table 5 6).

Computation is carried out in MATLAB and graphs are plotted.

5.4.3 RESULTS AND DISCUSSION

The ARF is predicted by using DFS, selecting suitable values of m and n. It has been found that the predicted ARF for the years 2002 to 2006 have non-significant difference to the actual one. However, the predicted ARF (750.9mm) of the year 2005 has large difference with occurred ARF (1693.0mm). This ARF is a rare event for the Anand station. Similar event was occurred in the year 1976

The ARF was 1712.0mm. This type of annual rainfall (ARF) is an outlier in the DS.

Therefore, the year 2005 prediction is eliminated from the computation of RMSE and PAE.

Fig 5.10 & Fig 5.11 show the prediction of the year 2002 and 2003. ARF of 2002 to 2004 are 423.1mm and 928.7 mm, respectively. These are one step ahead prediction in the DS (Table 5.6). Predicted ARF for the year 2004 and 2006 are also found significant with actual. Table 5.6 shows the values of Root Mean Square Error (RMSE) is 56.28 and Percentage of Average Error 5.78 %. This shows that all the four prediction by DFS are significant.

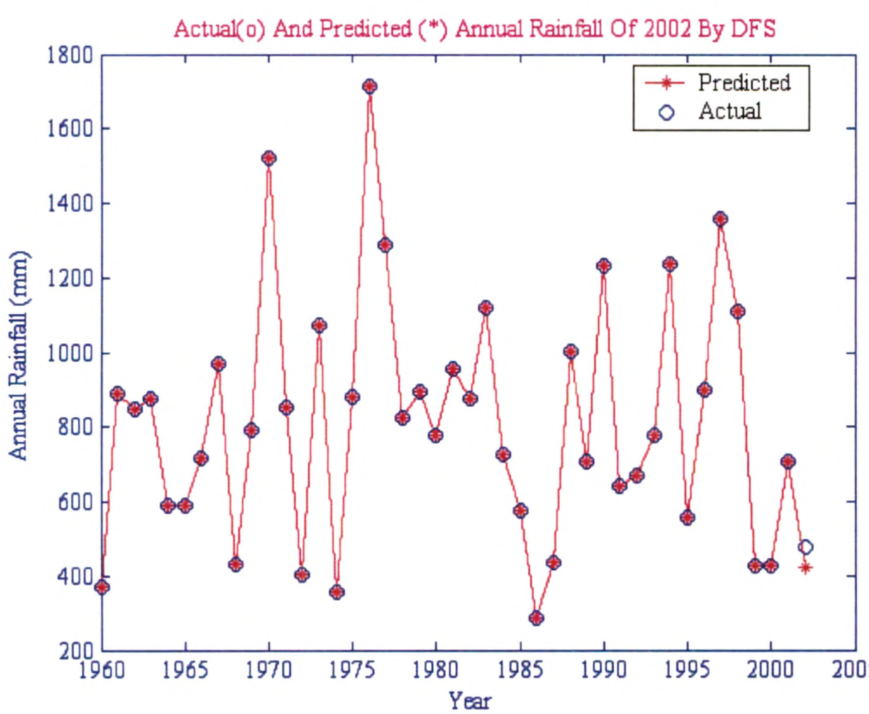


Fig: 5.10

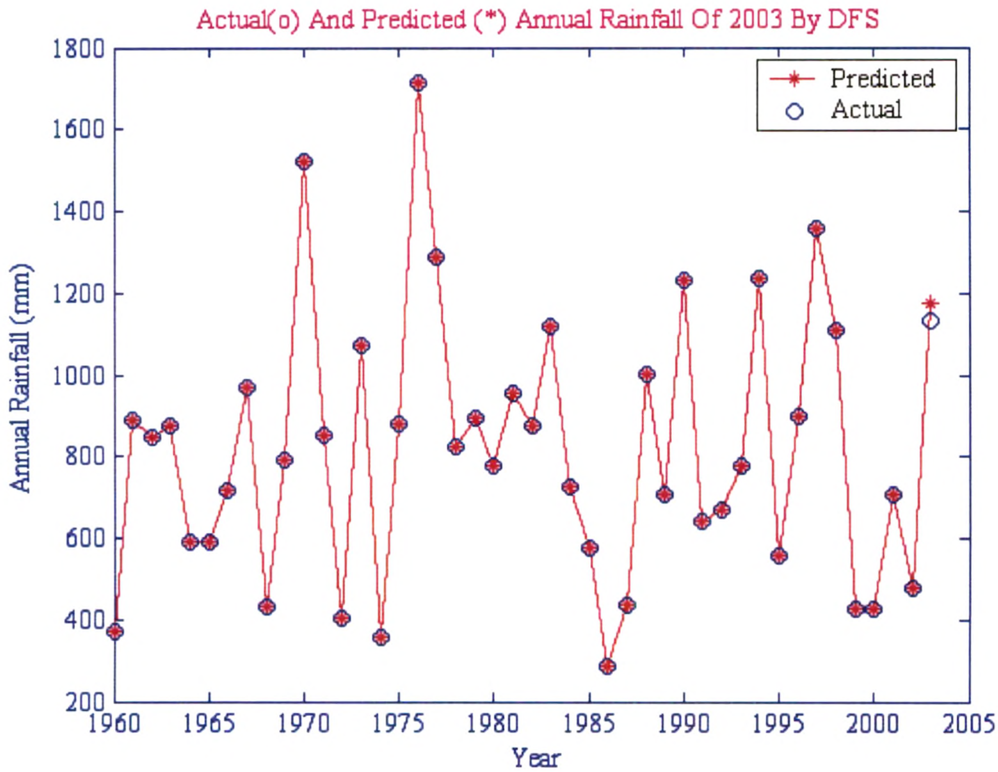


Fig: 5.11

Table 5.6
DETAILS OF THE PARAMETER VALUE USED IN DFS

Sr. No	Predicting Year	No. of Data Point	m and n values m	n	Predicted ARF P_i (mm)	Actual ARF A_i (mm)	$\lvert P_i-A_i \rvert$ (mm)	RMSE	PAE (%)
1	2002	43	21	12	423.0	479.2	56.2	56.28	5.78
2	2003	44	22	9	1175.4	1135.4	40.0		
3	2004	45	24	12	928.7	866.0	92.4		
4	2005	46	23	12	*750.9	1693.0	*720.7		
5	2006	47	23	15	1348.9	1428.7	79.8		

* Excluded from computation of RMSE and PAE.

5.4.4 CONCLUSION

i) Predicted annual rainfall for the years 2002 to 2006 (excluding the year 2005) is significant to the actual rainfall.

ii) Computation time is less in DFS approach.

5.5 ARTIFICIAL NEURAL NETWORK (ANN) APPROACH

Hall *et al.* [56] and Hsu *et al.* [65, 66] have applied artificial neural network for rainfall- runoff modeling. Goswami *et al.* [42, 43] have used ANN with three layers for experimental forecasts of all India Summer Monsoon Rainfall for 2002 and 2003.

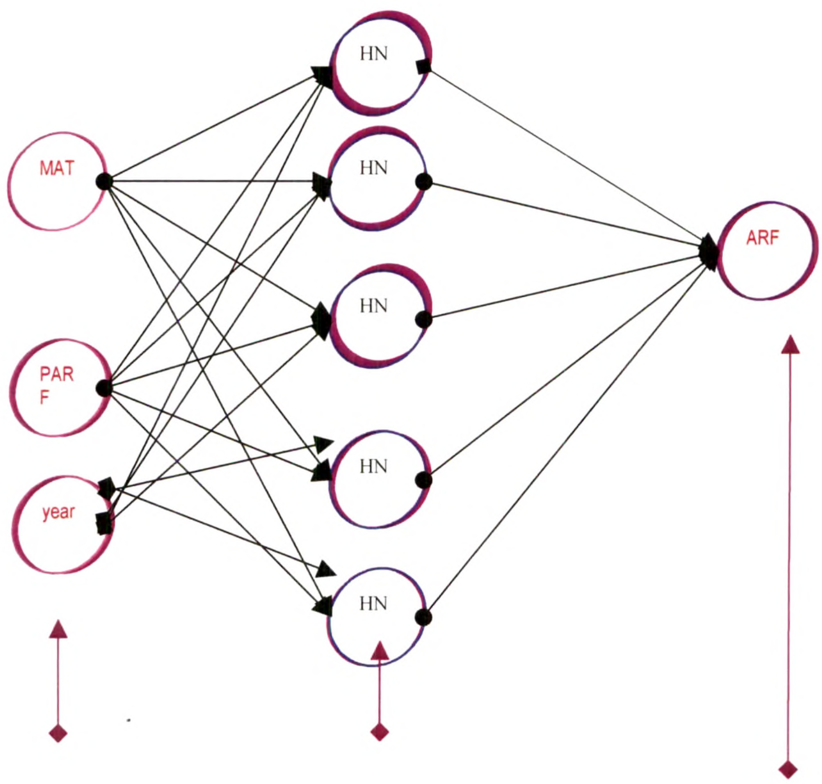
Here an attempt to represent the rainfall process in terms of a single - hidden layer feed forward neural network (Figure 5.12) is made.

5.5.1 NETWORK ARCHITECTURE:

We consider that an ANN that consists of an input layer of three neurons, one hidden layer and an output layer with one neuron. Each neuron is connected by feed forward network (Fig: 5.12).

Table 5.7
DETAILS OF THE PARAMETER VALUES USED IN ANN TRAINING

Sr. No.	Predicting Year	Number of epochs used	Learning rate	Momentum	No. of neurons	Error goal
1	2002	22230	0.001	0.3	147	0.0007
2	2003	237540	0.001	0.5	147	0.00001
3	2004	103684	0.001	0.3	147	0.00008
4	2005	18362	0.001	0.5	160	0.00089
5	2006	5942	0.001	0.5	147	0.0075



Input layer **Hidden layer** **Out put layer**
with 3 inputs, **with 1 output**
MAT , PARF and **ARF**
No.of the year

Fig: 5.12

It has three inputs, May month's Maximum Air Temperature (MAT) , Previous year Annual Rainfall (PARF) and number of the year. The output is the current year amount of Annual Rainfall (ARF).

Number of nodes in the input and output layer is equal to the number of variables of inputs and output, respectively. Figure 5.12 shows three nodes are in the input layer that has known values and one is in output layer to be predicted.

in the input layer that has known values and one is in output layer to be predicted.

In the present analysis numbers of neurons in input layer, hidden layer and output layer are 2, 147 and 1 , respectively (Table 5.7).

In ANN input nodes have known values and these values passes to the next layer (hidden layer) after multiplying with the weight of the connection. Hidden neuron get these weighted inputs and applies a sigmoidal function to determine other neuron fires or remains dormant. These neurons group determines the importance of that particular input to the overall prediction.

The sigmoid function is of S-shaped. Here, $F(x) = \frac{1}{1+\exp(-x)}$ is considered as activation function, where, x is the sum of all weighted inputs coming to the node, that is, x = netj

During the training, a learning algorithm namely, Back Propagation with momentum is used to iteratively modify the weights of the connections to minimize the total error in the approximation.

5.5.2 LEARNING OF THE NETWORK

Here learning is the type of Supervised Learning (Werner, [167]). In the network, supervisor is the observed output of ARF data series. Training is equally hard for larger as well as differently configured networks. The time required for training depends on the value of the parameters, like momentum, learning rate and error ratio. It also, depends on number of hidden neurons

(HN). If network is converging to actual output slowly then number of hidden neurons (HN) is required to increase. But simultaneously time of training is also increased. It also appears that some times larger nets that are with more number of hidden nodes (HN) do not give any significant change in accuracy [14]. Therefore, always it is desirable to decide by trial and error that how many numbers of hidden nodes are required. There is no standard method developed but Chu *et al.* [20] have proved that 'if the number of binary input cells is N (i.e. N-bit) for perception networks, the number of functional link cells that need to be generated to make the network learn successfully is at most 2^N -N-1'. In our analysis if we take number of hidden nodes 8 or 15 then it does not give successful learning. As per the need of the learning, 147 hidden nodes are chosen (Table 5.7).

Here to train the networks three data series are taken namely, the Highest Maximum May month's Air Temperature (MAT) for the period 1959 to May 2001, the Previous year Annual Rainfall (PARF) data series from 1958 to 2000 as a input series (Table 5.6) and the time point, that is, year numbers 1 to 43 are considered as a third input. Normalizations for all the inputs variable and actual output is done by dividing them by their norm. Training is done under the supervision of actual occurred ARF from 1960 to 2001. That means that training of the NN is done by minimizing the error which is the difference between actual ARF and output given by the NN at each epoch. At given minimum error the training is stopped. Output by this trained network is shown in the Figure 5.13. Here the error goal is 0.0007. This training required 22230 numbers of epochs (Table 5.7). To achieve the desired error goal, 147 numbers of hidden neurons is used. Less number of hidden nodes like 8, 10 or 15 do not give good performance to achieve the error goal and more than 147 numbers of hidden nodes took very large time to converge the aim or actual output. Other parameters values used in the training are given in the Table 5.7. Programmes are developed in MATLAB and given in the Appendix.

To predict the ARF, learning of the network becomes faster with three inputs namely time point, highest Maximum MAT and PARF instead of two namely MAT and PARF. Here, learning rate on the hidden layer is 0.001. Data with very high variations required learning rate less then 0.001.

5.5.4 RESULTS AND DISCUSSION

From the student's t-test for two tails at 95% of confidence interval the predicted ARF by ANN, is found to be significant to the actual rainfall except for the year 2005. For the outliers like 1693.0mm in the year of 2005 in the data series ANN is unable to predict the ARF accurately. All the predicted values are shown in the Fig 5.13 and Fig 5.14 and Table 5.8.

Predicted Annual Rainfall (ARF) of the year 2002 is 520.9mm and actual 479.2mm.Difference between these two ARF is 41.7mm.

RMSE and PAE are computed for the four years (Table 5.8).

Predicting ARF of the year 2005, training of the NN becomes difficult due to extreme ARF that is 1693.0mm. Here, error goal is 0.00089. If we increases the error goal or increase the numbers of hidden nodes, ANN is not convergent and predicting the non-significant ARF.

Predicted Annual Rainfall (ARF) of the year 2005 is 1319.5 mm . Actual Annual Rainfall (ARF) is 1693.0mm which is very high. Difference between these two ARF is 373.5 mm. This year's ARF is an outlier and therefore, ANN method can not predict this accurately.

From the predicted ARF found RMSE is 63.01 and PAE is 6.5 %.

For the year 2006 predicted ARF is 1485.1mm (Table 5.8; Fig: 5.14). Actual ARF is 1428.7 mm.

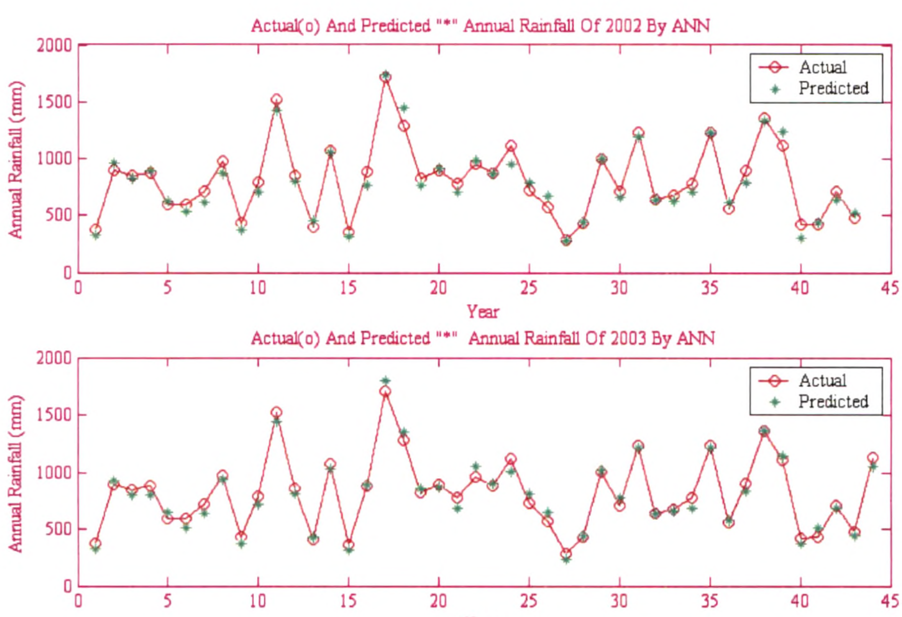


Fig: 5.13

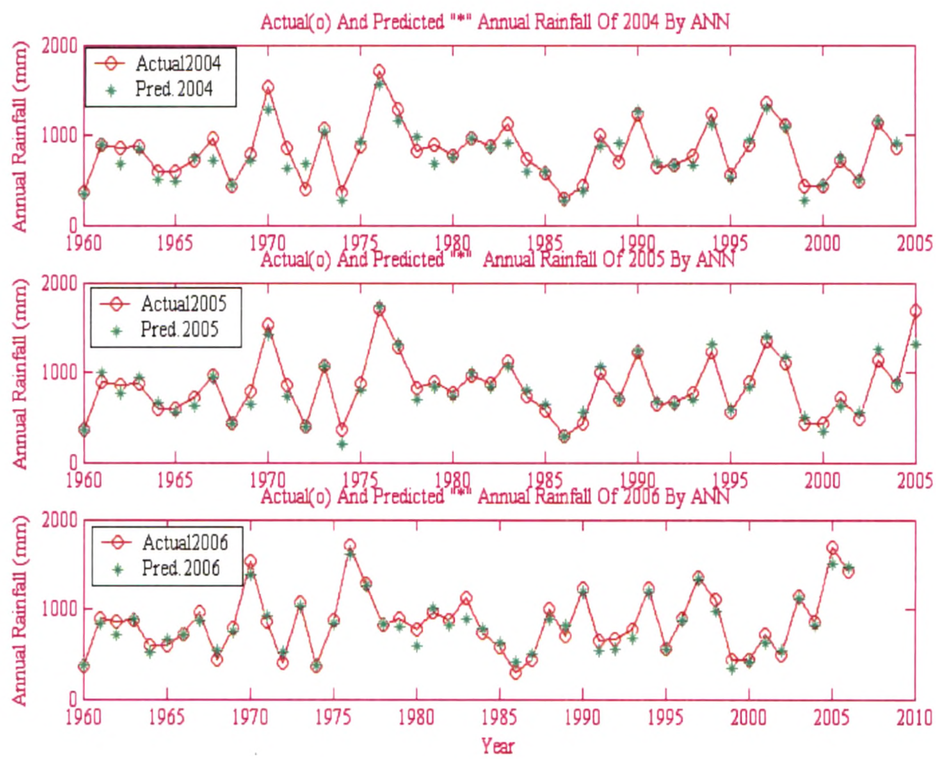


Fig: 5.14

TABLE 5.8

DETAILS OF THE ANN RESULTS

Sr. No	Predicting year (mm)	Predicted ARF I (mm)	Actual ARF II (mm)	Difference | I-II | (mm)	RMSE	PAE (%)
1	2002	0520.9	0479.2	41.7	63.01	6.5
2	2003	1054.0	1135.4	081.4		
3	2004	0816.4	0866.0	049.6		
4	2005	*1319.5	1693.0	*373.5		
5	2006	1485.1	1428.7	56.4		

* NOT INCLUDED IN THE COMPUTATION OF RMSE AND PAE.

5.5.5 CONCLUSION

i) The results show the applications of ANN to the rainfall analysis for predictions of ARF of the years 2002 to 2006 by ANN is significant to actual annual rainfall except for 2005.

ii) Prediction has been done for dependent random variable having non - linear relationships with predictors.

iii) ANN gives freedom to use multi-inputs and multi outputs without using any direct Mathematical or statistical model.

5.6 COMPARISON OF THE RESULTS BY TWO METHODS DFS AND ANN

Prediction of ARF by two methods, ANN and DFS are mentioned in the Table 5.9. Here comparison between two methods has been done by obtaining their Root Mean Square Error (RMSE). Percentage of Average Error (PAE) is found for both the methods.

Computed RMSE and PAE for ANN & DFS method are 63.011 & 56.28 and 6.5% & 5.78% (Table 5.9) respectively.

Comparing these two errors of two methods it is found that DFS model gives less error in comparison to ANN. But that difference is not very large.

TABLE 5.9
COMPARISON OF THE RESULTS

Sr. No	Year	Predicted ARF (mm)		Actual ARF (mm)	Difference with Actual ARF (mm)		RMSE (mm)		PAE	
		By ANN	By DFS		By ANN	By DFS	By ANN	By DFS	By ANN	By DFS
1.	2002	520.9	478.1	479.2	41.7	1.1	63.01	56.28	6.5	5.78
2.	2003	1054.0	1165.4	1135.4	81.4	30.0				
3.	2004	913.6	857.4	866.0	47.6	8.6				
4.	2005	*1319.5	*1152.4	1693.0	* 373.5	*540.6				
5.	2006	1485.1	1413.8	1428.7	56.4	14.9				

* NOT INCLUDED IN THE COMPUTATION OF RMSE AND PAE.

5.7 CONCLUSION

Prediction of annual rainfall (ARF) for 2002 to 2006 is found significant by both DFS and ANN methods except for outliers (2005). DFS method gives less PAE in comparison to ANN. However, the error difference is not very large. PAE in ANN is slightly more than that of DFS due to requirement of more efficient computer machine.

Obtained PAE by both the methods are less than 10 % and therefore, predictions are significant.

Further, predictions are also checked by Student t -test for two tails give non -significant difference with actual Annual Rainfall (ARF).

❖