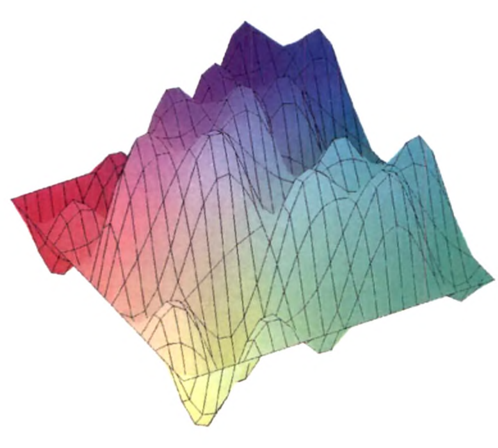
Chapter 1

INTRODUCTION



1.1 INTRODUCTION

The information on weather and climate is very important for agricultural planning and strategies. Growth of the plants and crop yield are highly influenced by weather. Weather can be defined as the instantaneous physical state of the atmosphere (gaseous layer surrounding air) at particular place (Ghadeker [39]). Air temperature, wind, cloud, visibility, evapotranspiration, atmospheric pressure, humidity, precipitation (rain and snow fall), radiation, pollution etc are the major weather elements.

Generalization of weather or statistical average of day-to-day weather conditions computed from long-term data for a given region or space is called climate. The complete description of the climate includes information on extreme weather also. Extreme weather contains information on highest and lowest temperatures, floods, destructive winds etc. Prediction of such extreme events helps us in preventing the unexpected losses.

The main objective of this thesis is to predict the weather elements especially, annual and weekly rainfall, soil temperature of Anand station of Gujarat state by using appropriate computational methods. In our analysis we employ tools of Applied Mathematics like, Artificial Neural Network, Harmonic analysis, Double Fourier Series, Curve fitting and Statistical distributions etc.

We confine our study for Anand station of Gujarat but it can be used for various other places too with proper choice of methods The present thesis consists of six chapters Chapter 1 is the introduction of whole thesis and

Chapter 2 gives necessary preliminaries on Weather, Artificial Neural Network, the required Mathematical Analysis and Fourier series etc. A brief description of the chapters 3 to 6 is as follows.

1.2 CHAPTER 3.: - PREDICTION OF WEEKLY SOIL TEMPERATURES

Chapter 3 deals with prediction of weekly soil temperatures at different depths by using following two different methods:

i) Artificial Neural Network (ANN).

ii) Harmonic Analysis (HA).

1.2.1 HARMONIC ANALYSIS (HA) APPROACH

The annual soil temperatures regime follows a wave pattern, which is entirely analogous to diurnal variation for a longer period. Wave amplitude for the annual pattern decreases less rapidly with depth. To describe the annual soil temperature cycle Harmonic Analysis (HA) is found to be a useful tool to describe the annual soil temperature cycle (Bocock [15] ; Bocock *et al.* [16] ; Dalrymple [26] ; Kulshrestha [90] ; Liakats [99]). The Fourier Theorem ensures that "Most physical functions that vary periodically with time with a frequency P can be expressed as a superposition of sinusoidal components of frequencies, P, 2P, 3P, 4P.. etc " That is, "A periodic function T of t, with period P, can be approximated as the following":

$$T(t) \approx T_0 + \sum_{k=1}^{N} a_k \cos(2\pi \frac{kt}{P}) + \sum_{k=1}^{N} b_k \sin(2\pi \frac{kt}{P}) \quad (1.1)$$

Here the Fourier coefficients are given by,

$$a_0 = \frac{2}{\pi}\int_0^P T(t)\ dt \qquad a_k = \frac{2}{\pi}\int_0^P T(t)\ \cos(2\pi\frac{kt}{P})dt\ ,\ and \qquad b_k = \frac{2}{\pi}\int_0^P T(t)\ \sin(2\pi\frac{kt}{P})dt\ , \qquad (1.2)$$

The Phase Angle form of Fourier series is

From this equation, $T(t) = T_0 + \sum_{k=1}^{no/2}\left(C_k\ *\sin\left(\frac{2\pi}{P}kt + A_k\right)\right)$ (1.3)

where, C_k is the Amplitude and A_k Phase Angle given by

$$C_k = \sqrt{a_k^{\ 2} + b_k^{\ 2}}\ , \quad A_k = \tan^{-1}(\frac{a_k}{b_k}) \qquad (1.4)$$

where a_k and b_k are the Fourier Coefficients, estimated by formula for discrete variable

$$b_k = \frac{2}{no}\sum[T(t)\ \sin(\frac{2\pi}{P}k\,t)]\ ; \qquad (1.5)$$

$$a_k = \frac{2}{no}\sum[T(t)\ \cos(\frac{2\pi}{P}k\,t)]\ ; \qquad (1\ 6)$$

Here, k = 1,2,3…N. N is the total number of Harmonics used, T_0 is the Mean Soil temperature of the data series, C_k is the Amplitude of the k^{th} Harmonic and t = 0,1,2,3. .51, P = 52 The values of Phase Angles A_k will be calculated according to procedure describe by (Panofsky [116]). The ratio of $C_k^2/2$ to total variance s^2 in observed soil temperature, which explains the fraction of variability accounted for by k^{th} Harmonic, is estimated.

In chapter 3, HA is used for prediction of soil temperature as a first method. In this analysis, soil temperatures data series for the period of 1982-2003 is used. Soil temperatures are recorded at 5, 10, and 20 cm depth at Anand station of Gujarat state. After finding Fourier coefficients, temperatures are predicted for the year 2004 and 2005. The actual weekly soil temperatures and computed temperatures are compared and tested by student t- test for the year 2004 and 2005. Accuracy of the results is checked in view of used number of data series. It is found that predicted temperatures are significant to the actual temperature with 26 harmonics.

It is found that Harmonic Analysis technique is a one of the good fit with fairly good accuracy, to estimate soil temperatures at different depths.

1.2.2 ARTIFICIAL NEURAL NETWORK (ANN) APPROACH

Second method in chapter 3 is Prediction of the Soil temperature by using Artificial Neural Network techniques Here we have taken Relative humidity, Wind speed and Air temperature as a Predictor and soil temperature as a Predictand. These data are recorded at land surface Process experiment (LASPEX) held at Anand during the year 1997. Three ways of prediction of soil temperature are considered.

i) Prediction of soil temperature for a given time point. That is time had been taken as input data and soil temperature as out put data. Here, a network with only one hidden layer had been considered These results are compared with the results found from HA. It is concluded that results obtained by ANN is more accurate to the actual temperatures.

ii) Two inputs that are predictors are air temperature at 1-meter height and time point. Out put that is Predictand is soil temperature Two hidden layers are used for the prediction of soil temperatures at the depth of 0-5 cm.

iii) A simple neural network having 3 McCulloch Type (McCulloch [105]) neurons is considered. This network has three input neurons and three output neurons, that is, n=3 and m=3. Three input parameters are relative humidity (%), wind speed and air temperature and three outputs are morning and afternoon soil temperatures and evaporation. We use the transfer function $f_i(x) = \tanh(x)$, for i= 1,2,3...... Lipschitz conditions and the monotonic conditions are satisfied and convergence of network is studied.

To train the network we have used 52 standard week's data of a year To model the weather system we consider a simple neural network having McCulloch Pitts Type neurons and use the generalized Widrow-Hoff algorithm to train the network. We give conditions on the learning rate and the transfer functions, which will guarantee the convergence of the generalized Widrow-Hoff algorithm. To prove the convergence we make use of Banach Fixed- point theorem. Our convergence theorem generalizes an earlier convergence theorem proved by Hui *et al* [67]

Soil temperatures predicted by ANN are more significant in comparison to HA from found RMSE.

1.3 CHAPTER 4: PREDICTION OF RAINFALL PROBABILITIES

Study of climatic variability is important for the designing and operating the water resource systems, which results in social benefits. Dams, reservoirs building constructions etc are highly affected by occurrence of rare events like heavy rainfall, storms etc. Therefore, to know the time of reoccurrence of such types of extreme rare events are essential. Thus ability to forecast hydrological effect of climate fluctuations would be a valuable asset to regional water management authorities. It is required to predict the time of reoccurrence of highest one-day maximum rainfall from one-day maximum rainfall historical data series. In this chapter, we investigate following two problems:

i) Return Period Analysis (RTPA) by Gumbel [53] and Fisher Tipett Type-II Distribution and ANN

ii) Rainfall Probability Analysis (RPA) by Gamma Distribution Model (GDM) and ANN.

1.3.1 RETURN PERIOD ANALYSIS (RTPA)

A return period (T) is a time of reoccurrence of some specified event. It is a statistical measure of how often an event of a certain size is likely to reoccur Here, this event is a highest, occurred rainfall from the one-day total, maximum rainfall data series. The return period has an inverse relationship with the probability.

In RTPA, we find the return period T of the rare event and its corresponding probability of reoccurence.

There have been several attempts to study such problems by many researchers by using some methods like Jenkon's (Rao [126] ; Upadhyay [158]), mixed of Gumbel and Fisher (Upadhyay *et al.* [159]), etc. Mukherjee *et al.*[109] reported that the Gumbel's distribution is not satisfactory in Bhuj (Gujarat) and Phalodi (Rajasthan) due to their belonging to a different population.

The aim of this research work is to apply Gumbel's and Fisher Tipett Type-II distribution to the above Return Period Analysis. Here, one-day maximum rainfall data series from 1901-1992 is used These distributions are applied to the eight-agro climatic zones of Gujarat state Total 58 stations have been used for calculation of RP values.

Fisher and Tipett Type-II distribution is actually Log Gumbel distribution, which minimizes error of Gumbel's distribution. In application of Fisher Tipett Type-II distribution, used data series (z = log (x)) is Natural logarithm of highest occurred rainfall. The distributions cumulative distribution functions (c.d.f) are given by

$$p(X \leq x) = \exp(-\exp(-\alpha\,(x-u))) \quad \text{and}$$

$$p(Z \leq z) = \exp(-\exp(-\alpha_{II}\,(z-u_{II}))) \;, \quad \text{respectively.}$$

Parameters are estimated by the method of Moments ([57], [136]). In this method, formulas to find parameters are as following

$$\alpha = \frac{s}{1.283},$$
$$u = Mean - 0.45s$$

where s is the variance. α_{II} and u_{II} is obtained by same formula from the log data series. Return period is given by $T = \frac{1}{1-p(x)}$. At chosen value of T, rainfall (x_T) is calculated by $x_T = u - (\ln(\ln(\frac{T}{T-1})))$.

Results of T will be presented and discussed in the chapter. Here, the obtained results show that T, calculated by Gumbel's distribution in comparison to Fisher and Tipett Type-II is very high (Kulshrestha *et al.* [90], Kulshrestha *et al.* [91]) in terms of years and not getting practical value This has been supported by Goel *et al.* [41] Authors in [41] have applied these distributions to annual maximum rainfall series over Krishna basin and suggested preferring the Fisher and Tipett-type II distribution for evaluating, RP of outliers (three times more then mean value) to Gumbel's distribution.

Conversely, rainfall amount is also computed for the given T that is 5,10,15,20,25 50 and 100 years

Return Period found for 14 districts are depicted in the figures in the chapter, with there occurred highest one-day maximum rainfall. Results obtained from both the distributions are compared. Standard Error (S.E.) for Fisher Tipett Type-II distribution is very less in comparison to Gumbel's distribution. Therefore, Fisher Tipett Type-II distribution is more appropriate than Gumbel's T values again predicted by ANN for the highest occurred one-day maximum rainfall from the data series of 1901 to 2005. The obtained results are compared by their tolerance with earlier results obtained from distributions

ANN has very negligible error of prediction therefore; ANN is best fit for prediction of return period for occurred highest, maximum rainfall.

Related programme is written in MATLAB.

1.3.2 RAINFALL PROBABILITY ANALYSIS (RPA)

Rainfall pattern is required to understand for crops sowing and its management (Khambhate *et al.* [84] ; Widrow *et al.* [170]). Cultural practices is also depends on rainfall pattern. The erosion studies of agricultural fields and moisture content are highly affected by precipitation. Hence, many scientists like Friedman *et al.* [34] and Barger [8] have used Gamma Distribution Model (GDM) (Government of India [46], Han [57]) to compute probability of assured rainfall In hydrology, rainfall variable has only positive value including zeros (Markovic [103]) and this is the big advantage of GDM as this model is defined on positive interval of x In meteorology, Thom ([154], [155], [156] has used, the GDM and fitted to the rainfall data, ranging from individual storms up to monthly and yearly distributions, on fairly large space and time scales.

Hence, the second problem is to predicted rainfall pattern by using 2-parameter GDM (Hafzullah Aksoy [55]) to standard weekly (22 to 42) rainfall data for the period of 1958-1998 of Anand station. Gamma density function is given by

$$p_x(x) = \frac{\lambda^{\eta}\ x^{\eta-1} e^{-\lambda x}}{\Gamma \eta} \quad , \ x, \lambda, \eta > 0$$

The mean and variance of the gamma distribution are defined as

$$E(X) = \frac{\eta}{\lambda} \quad \text{and } Var(X) = \frac{\eta}{\lambda^2} \ .$$

The cumulative gamma distribution (c.d.f) is given by

$$P_x(x) = \int_0^x \frac{\lambda^{\eta}\ x^{\eta-1} e^{-\lambda x}}{\Gamma \eta} \ dt$$

Applying this distribution, we obtained the probabilities, to occur particular intervals of rainfall amounts viz., 10-20 mm, 20-30 mm, 30-40 mm, 40-50 mm, 50-75 mm and 75-100 mm and visa versa also. Results and discussion will be given in chapter In this chapter week number having maximum rainfall amount, with its probability is pointed out. Further, the corresponding rainfall at different probability levels ranging from 10 to 90 percent were obtained and reported according to week numbers Rainfall amounts at different probability levels are shown in the figure in chapter. Relation of the Scale (λ) and Shape (η) parameters with rainfall amount is sketched in graph in chapter.

Same analysis is done by Artificial Neural Network and the result is compared by their RMSE.

In both the problems ANN method is found to be better than the used distributions. Related programme is written in MATLAB.

1.4 CHAPETR 5: PREDICTION OF ANNUAL RAINFALL

Chapter 5 deals with the prediction of Annual Rainfall by using the following two methods:

i) Artificial Neural Network, and

ii) Double Fourier Series

1.4.1 ARTIFICIAL NEURAL NETWORK (ANN) APPROACH

Rainfall prediction has always been a difficult and challenging task due to various factors like noise, discontinuity, and large number of zeros in the time series functions for rainfall Prediction of rainfall using different techniques and methodologies certainly require persistent efforts and multi- pronged strategy. Since the problem of predicting rainfall involves enormous quantity of data, many researchers have extensively used the Artificial Neural Network based computational algorithm (Zhang [174]). It has been found that this approach is more effective (Jacques *et al.* [72] ; Koekkoek [86]) because of availability of high-speed computers today.

An ANN is a parallel-distributed computational algorithm imitating the working of brain cells (neurons) It is also a function free estimator for the approximation of the functional relationship between a set of input data and its corresponding output data of given system. It learns from experience and it has the ability of generalization. Any continuous function mapping a compact set in n-dimensional Euclidean space to m dimensional Euclidean space could be approximated by means of a neural network having a single hidden layer. This result is the natural generalization of the Weirstrass Theorem "Any continuous function on compact set can be approximated by polynomials (Hecht-Nielsen [61] ; Kolmogrove [87])

ANN has been exploring and developing for several years ([30], [49]). A novel ANN, that is Composite Neuron Network (CN) has been used by Goswami ([42],[43] ,[44]) for prediction of mean rainfall of India, at CSIR, Center of Mathematical modeling and Computer Simulation (CCMMACS), Bangalore These predictions were more accurate than the forecast using parametric model.

Therefore, an attempt has been made in this thesis, to predict the annual rainfall of Anand at zonal level by ANN. The present analysis will also be applicable in the other region of Gujarat state.

There are various types of architectures for ANN (Werner [167]). One is Multi layered Feed Forward Network which is employed in present analysis. In this network, neurons are arranged in more than two layers viz. input layer, hidden layer and out put layer. In out put layer Back Propagation algorithm (BP) with a momentum term is used to correct the weights. Weight update rule in Back Propagation algorithm (BP) is given by ([59]

$$W_{ji}{}^{l}\,(n+1) = W_{ji}{}^{l}\,(n) + \alpha\,[W_{ji}{}^{l}\,(n-1)] + \eta\,\delta_j^{1}\,(n)\;y_j^{l-1}(n)$$

where α is the momentum and η the learning rate, l the layer index and $W_{ji}{}^{l}$ is the weights of the neuron.

Here, annual rainfall data series from 1958 to 2004 of Anand station of Gujarat is used Input data series are, one-day highest air temperature of month of May of the predicted year and previous years' annual rainfall. Details of the Training of the Network and algorithm will be given in chapter 2 Predicted annual rainfall for the year 1960 to 2005 is compared with the actual one Chi-Square (χ^2) test is applied for significance of the predicted rainfall. Predicted annual rainfall is significant according to Chi-square test (Panofsky [116]). The MATLAB based ANN with Back propagation algorithm is used in the present study.

1.4.2 DOUBLE FOURIER SERIES (DFS) APPROACH

Double Fourier series (DFS) is a global model (Vande, [161]). Author in [161] has used DFS as a mapping tool in marine cartography. DFS can be used for non-linear problems on globe like rainfall prediction.

Uniqueness for higher dimensional Trigonometric series is proved by Marshall [104]. Anita *et al.* [4] have studied the possibility of replacing the popular spectral transform method with the DFS methods.

Hence, to predict the annual rainfall, DFS may be employed. Fourier series with two variables is as under. Let f(x,y) be defined on square K = { (x,y): $-\pi < x < \pi$, $-\pi < y < \pi$] $\subset R^2$, then f(x,y)=

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \lambda_{mn}[a_{mn} \cos mx \cos ny + b_{mn} \sin mx \cos ny + c_{mn} \cos mx \sin ny + d_{mn} \sin mx \sin ny].$$

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

co-efficient a_{mn} b_{mn} c_{mn} and d_{mn} are found by the following formulas

$$a_{mn} = \frac{1}{\pi^2} \iint_K f(x,y) \cos mx \cos ny dx dy,$$
$$b_{mn} = \frac{1}{\pi^2} \iint_K f(x,y) \sin mx \cos ny dx dy,$$
$$c_{mn} = \frac{1}{\pi^2} \iint_K f(x,y) \cos mx \sin ny dx dy,$$
$$d_{mn} = \frac{1}{\pi^2} \iint_K f(x,y) \sin mx \sin ny dx dy$$

Here, we use May month's highest air temperature (x) and previous year's annual rainfall (y) as the inputs. Thus, in this analysis, we have two inputs and rainfall as one output.

Here, m, n=1, 2, 3. ....value of m and n are chosen by requirement of the analysis f(x,y) is to be predicted rainfall of the coming year

Root Mean Square Error and Percentage of Average Error obtained during the application of DFS are less than ANN. However, difference is very less Details of the methods and results are given in the thesis.

Programmes are developed in MATLAB.

1.4 CHAPTER 6 : PREDICTION OF HOURLY AIR TEMPERATURES

Chapter 6 deals with prediction of hourly air temperatures. The following two different methods are employed.

Our aim in this chapter is to predict the hourly air temperature (HAT). We estimate the hourly air temperatures in two situations namely,

(i) When input variables are daily Maximum Air Temperature (MaxAT) and Minimum Air temperatures (MinAT).

(ii) When Input variables are Hourly Maximum Air Temperature (MaxHAT) and Hourly Minimum Air temperatures (MinHAT).

In case (i), two methods are used:

(a) William and Logan Model [121] and

(b) Average of Daily Extreme Air Temperatures

In case (ii) four methods are used:

(a) William and Logan Model [121]

(b) Double Fourier Series

(c) Artificial Neural Network and

(d) Average of Hourly Extreme Air Temperatures

1.4.1 WILLIAM AND LOGAN MATHEMATICAL MODEL

Air temperatures exert a marked influence on plant growth and development. Temperature has direct effect on the rate of growth of plants. The diurnal and seasonal temperature patterns affect the structure of indigenous vegetation. The two most frequently used techniques for simulating diurnal variations in air temperatures are empirical models and energy-budget models Myrup, [111], Lemon *et al.* [98] and Goudriaan *et al.* [45] have used different methods to calculate the energy budget at the plant canopy and soil surface and thus determine soil and air temperatures.

A Mathematical model for diurnal variation in soil and air temperature was developed by Parton *et al.* [121] Model is useful for prediction of hourly soil or air temperatures. The model uses a truncated sine wave to predict variation of daytime temperatures and an exponential function to predict nighttime temperatures. The use of sine wave during the daytime is very similar to the approach used in the sinusoidal model proposed by Johnson *et al.* [78].

To find out the Day time temperature at a specified hour

$(T_i)_d = (T_x - T_N) \sin [(3.14\, m/(y+2a))] + T_N$; where y is day length in hour, $(T_i)_d$ is the temperature at the specified hour i in a day and T_x and T_N are the maximum and minimum temperatures of the specified Julian day (Table: A-3.1) respectively.

To find out the nighttime temperature at a specified hour i, the exponential function used is given as follows,

$$(T_i)_N = (T_N) + (T_{se} - T_N) \exp(-bn/Z).$$

where, T_{se} is the temperature at the time of the sunset, which will be calculated from the Model of Day time temperature, T_N is the minimum temperature of the specified Julian day and b is the nighttime temperature co-efficient.

Parameter estimation of 'a', 'b' and 'c' and calculation of y (Day length), z (night length) and m is based on i^{th} hour.

The important assumptions in this model are

i) Maximum temperature will occur sometime during the daylight hours before sunset.

ii) Minimum temperature will occur during the early morning hours

and

iii) Temperature variation during the day is described by a truncated sine wave and nighttime temperatures are described by an exponential function. The validity of these assumptions was tested with observed air and soil temperature.

One component of chapter 6 is to develop a similar model of Parton The mean temperatures recorded in the automatic weather station (Campbell USA) installed at the Agro-meteorological Observatory of the department of Agricultural Meteorology, B A College of Agriculture, Anand Agricultural University, Anand (22.33 ^{0} N, 72.55 ^{0}E and msl45m) for the year 1992 were used to derive the constants in the model. Manually recorded maximum and minimum temperature in the Agro meteorology observatory are also used for the same period.

The variations between two types of hourly air temperatures are found. Here, two different air temperatures are

i) Mean air temperature (D_{21}) found from maximum and minimum temperatures from the data series [(Max + Min)/2]

and

ii) Hourly air temperatures, obtained by model (D_1).

The difference of D_{21} and D_1 for Julian days has shown by graph.

The trend would be checked for all the Julian days. Thus, one could estimate the true daily mean temperature by calculating hourly temperatures using this Mathematical model.

Details of the Model, Parameter estimation and results are given in chapter. Results are tested by t- test and presented in chapter. Obtained results are significant to the actual hourly temperatures which shown by the graph in the thesis.

Related Root Mean Square Error (RMSE) and Percentage of Average Error (PAE) are found for all the used methods and compared with each other.

1.5.2 ARTIFICIAL NEURAL NETWORK (ANN) APPROACH

Artificial Neural Network is applied to find out the above hourly air temperatures of some Julian days for the year 2006. Back propagation algorithm is used for training of the network Number of neurons, learning rate, momentum and error of tolerance is chosen for the fast convergence of the algorithm Outcomes are compared with earlier one of William and Logan Model. From RMSE it is concluded that results found by ANN are more accurate.

1.6 CONCLUSION

This thesis, addresses the problem of prediction of weather Parameters like,

i) Prediction of annual rainfall.

ii) Prediction of return period of occurred highest one-day maximum rainfall.

iii) Prediction of weekly rainfall probabilities.

iv) Prediction of Hourly air temperatures for a day.

v) Prediction of soil temperatures at 5- 20 cm. Depths,

for Anand station by invoking various methods like Artificial Neural Network, Double Fourier Series , Return Period Analysis, Gamma Distribution Model, William and Logan Model and Harmonic Analysis. The above methods are first time used for prediction of weather parameters in the Gujarat Double Fourier Series approach is first time applied to the problems of prediction in Meteorology.

Results of these methods are compared with Artificial Neural Network technique by finding Root Mean Squared Error (RMSE). It turned out that ANN is the best-fit algorithm for prediction of Weather Parameters. We have also obtained a convergence result for multi input, multi output McCulloch Pitts type Neural Network by using fixed-point theorem.

❖