Impact of Climatological Parameters on Yield of Wheat Using Neural Network Fitting

Parekh, F. P.[1], Suryanarayana, T. M. V.[2]

[1, 2] *(Water Resources Engineering and Management Institute, Faculty of Technology & Engineering, The Maharaja Sayajirao University of Baroda, Samiala, 391410, Gujarat, India)*

ABSTRACT : *The study was carried out to determine the predominance of various meteorological data on yield of wheat. The meteorological data for the Rabi season are collected and correlated with yield of wheat in Vallabh Vidyanagar for the period 1981- 1999 using neural network fitting. Then, the model is re-trained until the best coefficient of correlation is obtained and this corresponding model is considered as the best model and this is further used to validate the dataset from 2000 to 2006. This whole procedure is repeated for three different Alternatives. In Alternative 1, only maximum and minimum temperatures are correlated with yield data. In Alternative 2, maximum and minimum temperature and relative humidity are correlated with yield data. In Alternative 3, Maximum and minimum temperatures, sunshine hours and relative humidity are correlated with yield. The correlation between maximum and minimum temperatures & relative humidity and yield, the co-efficient of correlation for training, i.e. 1.00 validations, i.e. 0.97 and testing, i.e. 0.95 are best for 70%, whereas in 30% dataset, R comes out to be 0.62, which is good. It can be evidently concluded from the study that considering three variables, the correlation is achieved as the best. This reveals that yield of a crop is very much depended on maximum and minimum temperatures & relative humidity.*

Keywords: *Climatic variability, Coefficient of correlation, Crop yield, Meteorological data, Neural Fitting.*

I. INTRODUCTION

Climatic variability is the major factor influencing the agriculture productivity. Global climate change and its impacts on agriculture have becoming an important issue. Agriculture production is highly dependent on climate and it also adversely affected by increasing climatic variability. The aim is to develop the methodology for assessing this component of the total impact of climate variability on agricultural productivity. There is a need to quantify climatic variability to assess its effect on crop productivity. Among the major wheat growing states in India, Gujarat ranked 6th in production and having productivity of 2,377 kg/ha.

Agrometeorological models are defined as the product of two or more weather factors each representing functioning between yield and weather. These models do not require hypothesis of the plant and environment process. Thus the input requirement is less stringent but the output information is more dependent on the input data. Thus grometeorological models are a practical tool for the analysis of crop response to weather and estimating the yield. Khashei-siuki et al. (2011) studied the ability of Artificial Neural Network (ANN) technology and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) for the prediction of dryland wheat (Triticum aestivum) yield, based on the available daily weather and yearly agricultural data. Maqsood et al. (2004) presented the applicability of an ensemble of artificial neural networks (ANNs) and learning paradigms for weather forecasting in southern askatchewan, Canada. Parekh and Suryanarayana (2012) carried out the study to find the quantitative relationship between weather parameters and district level yield of wheat by seasonal trend analysis.

The present study was undertaken with a view to determine the predominance of various meteorological data on yield of wheat in Vallabh Vidhyanagar, Gujarat, India wherein the different weather parameters that are considered will be maximum & minimum temperatures, relative humidity, wind velocity and sunshine hours.

II. STUDY AREA

The entire Gujarat is divided into the various agro-climatic zones. Vallabh Vidyanagar is located in the Anand district and lies in middle Gujarat agro-climatic zone III of Gujarat state. Vallabh Vidyanagar is located at 22˚32' N latitude, 72˚54' E longitude at an altitude of 34 m above mean sea level. It is bounded on the north by the Kheda district and south by the Gulf of Khambhat, on the west by Ahmedabad district and, on the east by Vadodara district.

The climate of Vallabh Vidyanagar is semi-arid with fairly dry and hot summer. Winter is fairly cold and sets in, in the month of November and continues till the middle of February. Summer is hot and dry which commences from mid of February and ends by the month of June. May is the hottest month with mean maximum temperature around 40.08 ˚C. The average rainfall is 853 mm.

The soil of the region is popularly known as Goradu soil. It is alluvial in origin. The texture of the soil is sandy loam and black. The soil is deep enough to respond well to anuring and variety of crops of the tropical and sub-tropical regions. The soil is low in organic carbon and nitrogen, medium in available phosphorus and available sulphur. In this area paddy, tur, cotton and ground nut, til are grown in kharif season. In rabi season wheat, gram, and jowar are grown. Especially, in summer season the bajara and ground nut are grown. Tobacco is grown from August and harvested in March. In last few years, there is increase in amount of rainfall which facilitated in agriculture production and various irrigation scheme.

III. DATA COLLECTION

The data required for evaluation in this study are collected from India Meteorological Department, Pune and Krishi Bhavan, Gandhinagar. Long term meteorological daily data are collected from IMD (Indian

Meteorological Department), Pune for Vallabh Vidyanagar, town of Anand district of Gujarat.
The basic weekly meteorological data used comprises of Maximum and minimum temperature(˚C) , Relative humidity (%) and Sunshine hours (hours). The yield data of Wheat grown in Vallabh Vidyanagar are collected from the Krishi bhavan, Gandhinagar from year 1981-2006.

IV. METHODOLOGY

The weekly meteorological data viz. temperature, relative humidity, sunshine hours are converted into average monthly data and seasonal data. Wheat is cultivated in Rabi season.

Some data gaps or missing values are indentified in data. The missing values were found out with the SPSS 11.5 software. The linear trend at a point method is used to find the missing values of the data. To study the impact of climatological data on yield of wheat, Neural fitting tool of Artificial Neural Network (ANN) of MATLAB is used.

4.1 Artificial Neural Networks (ANN)

ANN was first introduced as a mathematical aid and was inspired by the neural structure of the brain. An input layer, which is used to present data to the network. An output layer, which is used to produce an appropriate response to the given input; and one or more intermediate layers, which are used to act as a collection of feature detectors. The ability of a neural network to process information is obtained through a learning process, which is the adaptation of link weights so that the network can produce an approximate output. In general, the learning process of an ANN will reward a correct response of the system to an input by increasing the strength of the current matrix of nodal weights.

There are several features in ANN that distinguish it from the empirical models. First, neural networks have flexible nonlinear function mapping capability which can approximate any continuous measurable function with arbitrarily desired accuracy, whereas most of the commonly used empirical model, do not have this property. Second, being non-parametric and data-driven neural networks impose few prior assumptions on the underlying process from which data are generated. Because of these properties, neural networks are less susceptible to model misspecification than most parametric nonlinear methods.

An ANN can be defined as data processing system consisting large number of simple highly interconnected processing elements (PEs or artificial neurons) in architecture analogous to cerebral cortex of brain. An ANN consists of input, hidden and output layers and each layer includes an array of artificial neurons. A typical neural network is fully connected, which means that there is a connection between each of the neurons in any given layer with each of the neuron in next layer. An artificial neuron is a model whose components are analogous to the components of actual neuron in next layer. An artificial neuron is a model whose components are analogous to the components of actual neuron. The array of input parameters is stored in the input layer and each input variable is represented by a neuron. Each of these inputs is modified by a weight sometimes called synaptic weight) whose function is analogous to that of the synaptic junction in a biological neuron. The neuron (processing element) consists of two parts. The first part simply aggregates the weighted inputs resulting in a quantity 1: the second part is essentially a nonlinear filter, usually cooed the transfer function or activation function. The activation function squashes or limits the values of the output of an artificial neuron to values between two asymptotes. The sigmoid function is the most commonly used activation function. It is a continuous function that varies gradually between two asymptotic values typically 0 and 1 or -1 and +1.

4.2 Neural Fitting Tool

In fitting problems, you want a neural network to map between a data set of numeric inputs and a set of numeric targets. Examples of this type of problem include estimating house prices from such input variables as tax rate, pupil/teacher ratio in local schools and crime rate (house_dataset); estimating engine emission levels based on measurements of fuel consumption and speed (engine_dataset); or predicting a patient's bodyfat level based on body measurements (bodyfat_dataset).

The Neural Network Fitting Tool will help you select data, create and train a network, and evaluate its performance using mean square error and regression analysis.

A two-layer feed-forward network with sigmoid hidden neurons and linear output neurons (newfit), can fit multi-dimensional mapping problems arbitrarily well, given consistent data and enough neurons in its hidden layer.

The network will be trained with Levenberg-Marquardt back propagation algorithm. (trainlm), unless there is not enough memory, in which case scaled conjugate gradient back propagation (trainscg) will be used. During past couple of years, the Levenberg-Marquardt (LM), a second order optimization technique is extensively employed in evapotranspiration modeling using neural networks.

In the present study, the meteorological data for the Rabi season are collected and correlated with yield of wheat for the period 1981- 1999 and validated for the 2000-2006 periods. The whole data is divided into 70% and 30% Datasets. The first part, i.e. 70% of the Dataset, is further divided into 70% for Training, 15% for Validation and 15% for Testing. For these datasets, correlations of input and output are observed using neural fitting tool. Then, the model is re-trained until the best coefficient of correlation is obtained and this corresponding model is considered as the best model and this is further used to validate the remaining 30% of the Dataset.

This whole procedure is repeated for three different Alternatives. In Alternative 1, only maximum and minimum temperatures are correlated with yield data. In Alternative 2, maximum and minimum temperature and relative humidity are correlated with yield data. In Alternative 3, Maximum and minimum temperatures, relative humidity and sunshine hours are correlated with yield.

4.3 Goodness of Fit Test Parameters

Coefficient of Correlation, R

Measure of the "goodness of fit" is the coefficient of correlation, R to explain the meaning of this measure, one has to define the standard deviation, which quantifies the spread of the data around the mean:

$$s_t = \sum_{i=1}^{n}\left(\bar{o} - o_i\right)^2$$

Where St is the standard deviation, o_i is the observed data points and p is the average of predicted data points and the average the observed data points given by,

$$\bar{o} = \frac{1}{n}\sum_{i-1}^{n} o_i$$

The quantity St considers the spread around a constant line (the mean) as opposed to the spread around the regression model. This is the uncertainty of the dependent variable prior to regression. One also defines the deviation from the fitting curve as

$$s_r = \sum_{i=1}^{n}\left(o_i - p_i\right)^2$$

Where Sr is the deviation from the fitting curve, p_i is the predicted data points.

$$R = \sqrt{\frac{s_t - s_r}{s_t}}$$

where R is defined as the coefficient of correlation. As the regression model starts improving describing the data, the correlation coefficient approaches unity. For a perfect fit, the standard error of the estimate will approach Sr = 0 and the correlation coefficient will approach R = 1.

Note the similarity of this expression to the standard error of the estimate; this quantity likewise measures the spread of the points around the fitting function. Thus, the improvement (or error reduction) due to describing the data in terms of a regression model can be quantified by subtracting the two quantities. Because the magnitude of the quantity is dependent on the scale of the data, this difference is normalized to yield .

V. RESULTS AND ANALYSIS

The predefined correlation coefficient, R is found for each stage (training, validation, testing) for each model and these values also were found for over all data in addition to an additional results for a randomly selected data for additional testing .Values of Correlation coefficient, R are plotted for (training, validation, testing) for each model and are given in Fig. 1, 2 and 3 for Alternatives 1, 2 and 3 respectively.

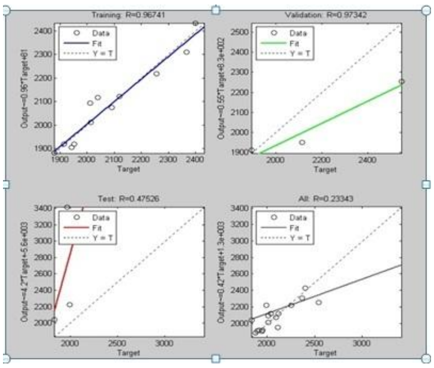

Fig.1 Correlation coefficient, R for Alternative 1

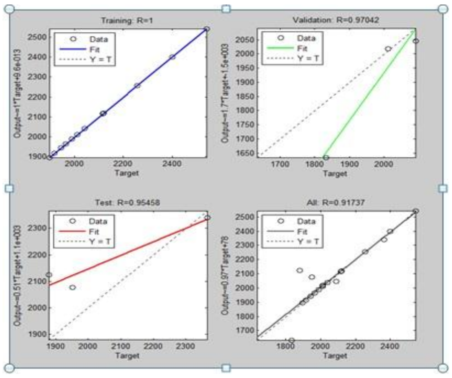

Fig.2 Correlation coefficient, R for Alternative 2

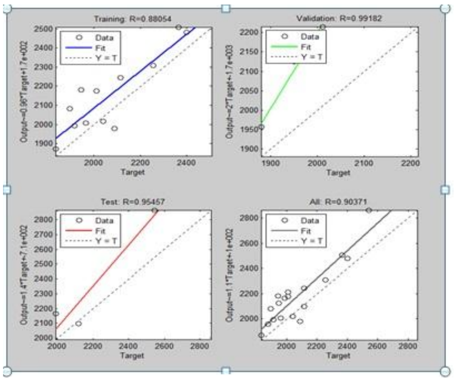

Fig.3 Correlation coefficient, R for Alternative 3

The Coefficient of correlation, R for 70 % and 30 % data set are given in Table 1.

Table 1 Coefficient of correlation, R for 70 % and 30 % data set

Alternative	Co- Efficient of Correlation, R			
	For 70 % of Dataset			For 30 % Dataset
	Training	Validation	Testing	Validation
1	0.96	0.97	0.47	0.39
2	1.00	0.97	0.95	0.62
3	0.88	0.99	0.95	0.66

Looking to the Table 1 and Figs. 1,2 and 3, as per Alternative 1, i.e. considering correlation between maximum and minimum temperatures and Yield, the co-efficient of correlation for training and validation are better compared to testing for 70%, whereas in 30% dataset, R comes out to be 0.39, which is comparatively low.

As per Alternative 2, i.e. considering correlation between maximum and minimum temperatures & Relative Humidity and Yield, the co-efficient of correlation for training and validation are better compared to testing for 70%, whereas in 30% dataset, R comes out to be 0.0.62, which is comparatively good. As per alternative 3, i.e. Considering correlation between maximum and minimum temperatures & relative Humidity & Sunshine Hour and Yield, the co-efficient of correlation for training, i.e. 0.88 for validation, i.e. 0.99 an for testing for 70%, i.e. 0.95, whereas in 30% dataset, R comes out to be 0.66, which is good. Here, one can observe that sunshine hours are not significantly affecting the yield of a crop compared to the effect of Maximum and Minimum Temperatures and relative humidity on the same yield of a crop.

Here, it is very much evident that considering maximum and minimum temperature and relative humidity variables, the correlation is achieved as the best, which reveals that yield of a crop is very much depended on maximum and minimum temperatures & relative humidity. Inclusion of Sunshine Hours does not play significant role.

VI. CONCLUSIONS

The correlation between maximum and minimum temperatures & relative Humidity and Yield, the co-efficient of correlation for training, i.e. 1.00 validation, i.e. 0.97 and testing, i.e. 0.95 are best for 70%, whereas in 30% dataset, R comes out to be 00.62, which is good.

The inclusion of sunshine hours with maximum and minimum temperatures and relative humidity does not show better correlation with yield of a crop compared to the consideration of only maximum and minimum temperatures and relative humidity.

It can be evidently concluded from the study that considering three variables, the correlation is achieved as the best. This reveals that yield of a crop is very much depended on maximum and minimum temperatures & relative humidity.

The use of Neural Network fitting is quite helpful in studying the predominance of any variables on one another.

REFERENCES

[1] Khashei-Siuki,A., Kouchakzadeh,M. and Ghahraman,B. "Predicting Dryland Wheat Yield from Meteorological Data Using Expert System", Khorasan Province, Iranî, *J. Agr. Sci. Tech. (2011) Vol. 13: 627-640*

[2] Imran Maqsood, Muhammad Riaz Khan and Ajith Abraham (2004). "An ensemble of neural networks for weather forecasting", *Neural Comput & Applic* (*2004) 13: 112ñ122.*

[3] Parekh, F.P. and Suryanarayana, T.M.V. (2012). "Climatic Variability and Crop Weather Model of Wheat Crop for Vallabh Vidyanagar", *The Indian Journal of Technical Education, (Special Issue for NCEVT'12), April 2012, 325-330.*