Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.1, No.3, 2011

www.iiste.org

Crop Yield Forecasting by Adaptive Neuro Fuzzy Inference System

Pankaj Kumar*

VCSG College of Horticulture, GBPUA&T, Pantnagar, Uttarakhand

*pankaj591@gmail.com

Abstract

Meteorological uncertainties affect crop yield portentously during different stages of crop growing seasons, therefore several studies have been carried out to forecast crop yield using climatic parameters with empirical statistical regression equations relating regional yield with predictor variables. In this study an attempt has been made to develop Crop Yield Forecasting models to map relation between climatic data and crop yield. Present study was undertaken for forecasting rice yield by adaptive neuro fuzzy inference system (ANFIS) technique based on time series data of 27 years, yield and weather data (w.e.f. 1981-82 to 2007-08) obtained from G. B. Pant University of Agriculture and Technology, Pantnagar, District Udham Singh Nagar, Uttarakhand, India.

Keywords: climate, crop yield, ANFIS, forecasting

1. Introduction

In India it is possible to cultivate large number of crops due to diverse climatic conditions. Among these crops Rice (*Oryza sativa* L.) is important food crops of the country. The total area under cultivation of Rice 43.6 (22.6% of gross cropped area) and the total production of Rice is 94.1 million tonnes during the year 2007-08 constituting about 42.9% of total food production (Economic Survey 2007-2008). Crop failure on account of drought or flood will have a severe repercussion not only on the country's economy but also on food security.

The forecasting of crop yield may be done by using three major objective methods (i) biometrical characteristics (ii) weather variables and (iii) agricultural inputs (Agrawal *et al*. 2001). These approaches can be used individually or in combination to give a composite model. Forecasting is an significant aid in effective and efficient planning and is more important aspect for a developing economy such as ours so that adequate planning exercise is undertaken for sustainable growth, overall development and poverty alleviation. In management and administrative situations the need for planning is great because the lead time for decision making ranges from several years (for capital investments) to a few days or hours (for transportation or production schedules) to a few seconds (for telecommunication routing or electrical utility loading). Forecast of crop yield is of immense utility to the government and planners in formulation and implementation of various policies relating to food procurement, storage, distribution, price, import-export etc.

Several studies have been carried out to forecast crop yield using weather parameters (Huda *et al.* (1975), Choudhary and Sarkar (1981), Kokate *et al.* (2000), etc.). However, such forecast studies based on statistical models need to be done on continuing basis and for different agro-climatic zones, due to visible effects of changing environmental conditions and weather shifts at different locations and areas. Therefore, there is a need to develop area specific forecasting models based on time series data to help the policy makers for taking effective decisions to counter adverse situations in food production.

In recent years soft computing techniques like, artificial neural network (ANN), fuzzy logic, genetic algorithm and chaos theory have been widely applied in the sphere of prediction and time series modelling. Adaptive neuro-fuzzy inference system (ANFIS) which is integration of neural networks and fuzzy logic has the potential to capture the benefits of both these fields in a single framework. ANFIS utilizes linguistic information from the fuzzy logic as well learning capability of an ANN. Thus the present study was carried out to develop forecasting model based on time series agricultural data related to yield of rice crops based on weather parameters.

2. Materials and Methods

2.1 Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

Adaptive Neuro Fuzzy Inference System (ANFIS) is a fuzzy mapping algorithm that is based on Tagaki-Sugeno-Kang (TSK) fuzzy inference system (Jang *et al.*, 1997 and Loukas, 2001).ANFIS is integration of neural networks and fuzzy logic and have the potential to capture the benefits of both these fields in a single framework. ANFIS utilizes linguistic information from the fuzzy logic as well learning capability of an ANN for automatic fuzzy if-then rule generation and parameter optimization.

A conceptual ANFIS consists of five components: inputs and output database, a Fuzzy system generator, a Fuzzy Inference System (FIS), and an Adaptive Neural Network. The Sugeno- type Fuzzy Inference System, (Takagi and Sugeno, 1985) which is the combination of a FIS and an Adaptive Neural Network, was used in this study for rainfall-runoff modeling. The optimization method used is hybrid learning algorithms.

For a first-order Sugeno model, a common rule set with two fuzzy if-then rules is as follows:

Rule 1: If x_1 *is* A_1 *and* x_2 *is* B_1*, then* $f_1 = a_1 x_1 + b_1 x_2 + c_1$.

Rule 2: If x_1 *is* A_2 *and* x_2 *is* B_2*, then* $f_2 = a_2 x_1 + b_2 x_2 + c_2$.

where, x_1 and x_2 are the crisp inputs to the node and A_1, B_1, A_2, B_2 are fuzzy sets, a_i, b_i and c_i ($i = 1, 2$) are the coefficients of the first-order polynomial linear functions. Structure of a two-input first-order Sugeno fuzzy model with two rules is shown in Figure 1 It is possible to assign a different weight to each rule based on the structure of the system, where, weights w_1 and w_2 are assigned to rules 1 and 2 respectively.

and f = weighted average

The ANFIS consists of five layers (Jang, 1993), shown in Figure 1. The five layers of model are as follows:

Layer1: Each node output in this layer is fuzzified by membership grade of a fuzzy set corresponding to each input.

$$O_{1,i} = \mu_{Ai}(x_1) \qquad i = 1, 2$$

or

$$O_{1,i} = \mu_{Bi-2}(x_2) \qquad i = 3, 4 \tag{1}$$

Where, x_1 and x_2 are the inputs to node i ($i = 1, 2$ for x_1 and $i = 3, 4$ for x_2) and x_1 (or x_2) is the input to the i^{th} node and A_i (or B_{i-2}) is a fuzzy label.

Layer 2: Each node output in this layer represents the firing strength of a rule, which performs fuzzy, AND operation. Each node in this layer, labeled Π, is a stable node which multiplies incoming signals and sends the product out.

$$O_{2,i} = W_i = \mu_{Ai}(x_1)\,\mu_{Bi}(x_2) \qquad i = 1, 2 \tag{2}$$

Layer 3: Each node output in this layer is the normalized value of layer 2, i.e., the normalized firing strengths.

$$O_{3i} = \overline{W}_i = \frac{W_i}{W_1 + W_2} \qquad i = 1, 2 \tag{3}$$

Layer 4: Each node output in this layer is the normalized value of each fuzzy rule. The nodes in this layer are adaptive. Here $\overline{W}_i$ is the output of layer 3, and $\{a_i, b_i, c_i\}$ are the parameter set. Parameters of this layer are referred to as consequence or output parameters.

$$O_{4i} = \overline{W}_i f_i = \overline{W}_i (a_i x_1 + b_i x_2 + c_i) \quad i = 1, 2 \tag{4}$$

Layer 5: The node output in this layer is the overall output of the system, which is the summation of all coming signals.

$$Y = \sum_1^2 \overline{W}_i f_i = \frac{\sum_1^2 W_i f_i}{\sum_1^2 W_i} \tag{5}$$

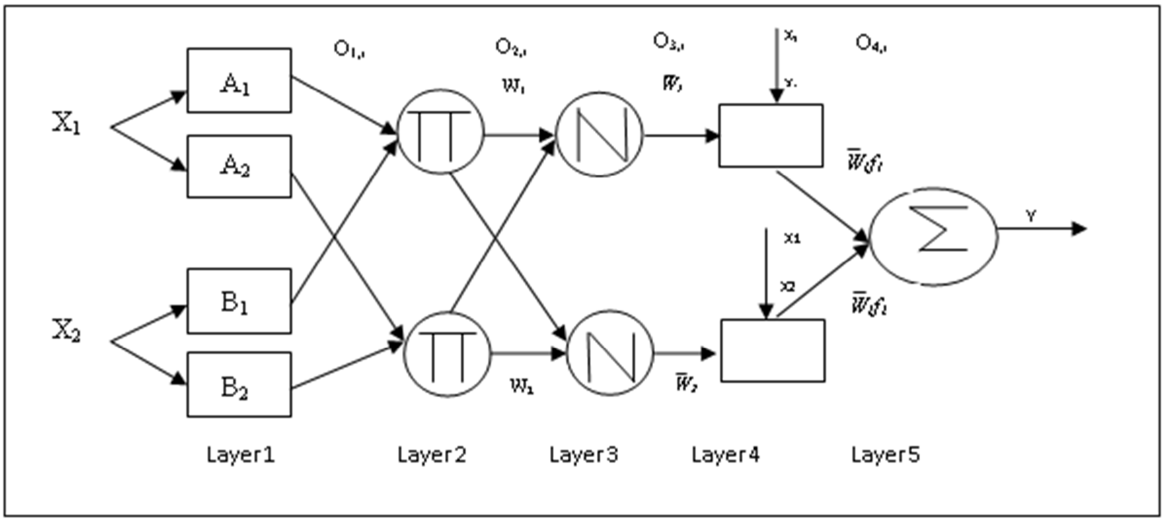

Figure 1.ANFIS architecture

In this way the input vector was fed through the network layer by layer. The two major phases for implementing the ANFIS for applications are the structure identification phase and the parameter identification phase. The structure identification phase involves finding a suitable number of fuzzy rules and fuzzy sets and a proper partition feature space. The parameter identification phase involves the adjustment of the premise and consequence parameters of the system.

Optimizing the values of the adaptive parameters is of vital importance for the performance of the adaptive system. Jang et al. (1997) developed a hybrid learning algorithm for ANFIS to approximate the precise value of the model parameters. The hybrid algorithm, which is a combination of gradient descent and the least-squares method, consists of two alternating phases: (1) in the backward pass, the error signals recursively propagated backwards and the premise parameters are updated by gradient descent, and (2) least squares method finds a proper set of consequent parameters (Jang *et al*., 1997). In premise parameters set for a given fixed values, the overall output can be expressed as a linear combination of the consequent parameters.

$$AX = B \tag{6}$$

Where, X is an unknown vector whose elements are the consequent parameters. A least squares estimator of X, namely X^*, is chosen to minimize the squared error $\|AX - B\|^2$. Sequential formulas are employed to compute the least squares estimator of X. For given fixed values of premise parameters, the estimated consequent parameters are known to be globally optimal.

2.2 Study Area and Model Application

2.2.1 Study area

The present study was carried out to develop forecasting models for predicting the yield of rice and wheat at Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, District Udham Singh Nagar, Uttarakhand, India, which is situated in the Shivalik range of Himalayan foot hills at approximately 243.89 meters above mean sea level. It is located at 29^0 N latitude and 79.3^0 E longitudes. This region is known as *Tarai* (meaning wet) and has a dry season from early October to mid-June and a wet season from June to early October. The soil of *Tarai* is mainly silt and loamy having good moisture storage capacity, about 7.3 pH and is highly productive. The annual rainfall of this region is about 1400 mm, which is subjected to large variation.

Yearly yield data of rice (q/ha) for 27 years (w.e.f. 1981-82 to 2007-08) was collected from the University Farm of G. B. Pant University of Agriculture and Technology, Pantnagar, which is a commercial well organized farm having 2714.43 ha land utilized for seed production, research purpose and commercial farming. The farm is well irrigated, fully mechanized and well equipped with all modern agricultural technologies and is governed with a high level management practices with the regular use of high yielding varieties, recommended doses of fertilizers and manures, proper plant protection measures including soil water management.

The time series weekly weather data of 27 years (from 1981-82 to 2007-08) were collected from Agro-meteorological observatory situated at the crop research centre of G. B. Pant University of Agriculture and Technology, Pantnagar, which is equipped with modern technology and instruments. Six weather parameters were included in this study; namely average weekly temperature (T°C), average weekly relative humidity (Rh%), average weekly sunshine hrs (S), average weekly total rainfall (P), average weekly number of rainfall days (n) and average weekly pan evaporation (E). However, weekly weather data related to kharif (the autumn harvest also known as the summer or monsoon crop in India) crop seasons starting from a fortnight before sowing up to one month before harvest were utilized for the development of models in the present study therefore, the weather data for rice crop (Kharif season), from May 21 (about a fortnight before sowing) to October 22 (one month before harvest) in each year were employed. Data for last one month of the crop season was excluded as to forecast crop yield at least one month before harvest.

2.3 Model Application

Total 27 year data are divided in two sets namely training (calibration) and testing (validation). Fig 2 and Fig3 indicates graphical representation of rice crop yield and weather variables are shown in.21 years data from 1981-82 to 2001-02 were used for calibration of model rest 6 years data from 2002-03 to 2007-08 were used for testing of model.

Among various inputs like average weekly temperature (T), average weekly relative humidity (Rh%), average weekly sunshine hrs (S), average weekly total rainfall (P), average weekly number of rainfall days (n) and average weekly pan evaporation (E) weather parameters best parameter to forecast rice crop yield were chosen by Gamma test. Gamma test is a technique for estimating the variance of the noise, or the mean square error, that can be achieved without over fitting (Jones, 2004). It is used for evaluation of the nonlinear correlation between random variables, namely, input and output pairs. Gamma test was applied on all possible combinations of weather variables. Gamma test concluded that only the combination of weekly pan evaporation (E) and average weekly temperature (T) are sufficient to predict rice crop yield. Including average weekly relative humidity (Rh%), average weekly sunshine hrs (S), average weekly total rainfall (P) and average weekly number of rainfall days (n) cause complexity in model and yields a poorer root mean square value between actual yield and predicted yield.

Combinations of weekly pan evaporation (E) and average weekly temperature (T) were considered as the inputs to the model, and Yield of the current year was considered as the output. Input space partitioning for model structure identification was done by grid partition. Hybrid learning algorithm was used to train the models for runoff prediction. The optimal learning parameters were determined by trial and error (Kim *et al*., 2002) for constant trapezoidal membership function. In order to choose better model among developed models root mean square error was computed.

3. Result and Discussions

The study revealed that the least value of root mean square error were obtained for model with two inputs, weather data of weekly pan evaporation and average weekly temperature for rice crop in Kharif season, from May 21 (about a fortnight before sowing) to October 22 (one month before harvest) and output as rice crop yield of the current year. Among triangular, bell shaped, trapezoidal and gaussian membership function, the trapezoidal membership function was found most suitable for this study. 2 trapezoidal and constant membership functions were used for 30 epochs. The observed and estimated crop yields are shown in Fig 5 and Fig 6 for training and testing periods respectively. Values of root mean square error is 3.485 in case of training period and 1.756 in case of testing.

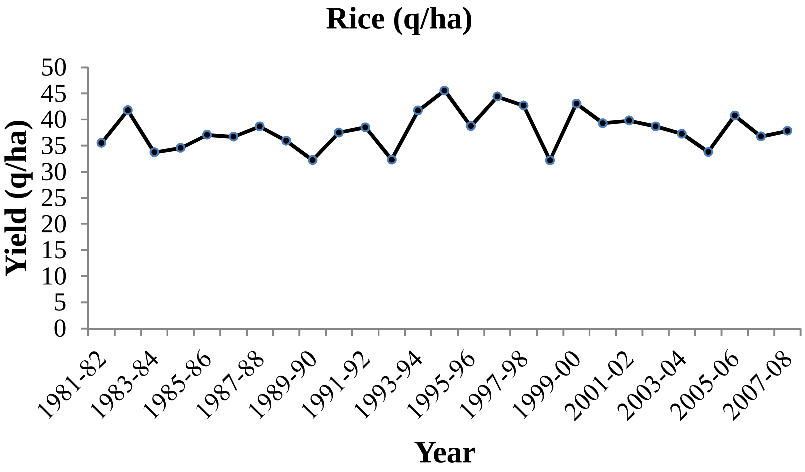

Figure 2. Rice yield (q/ha)

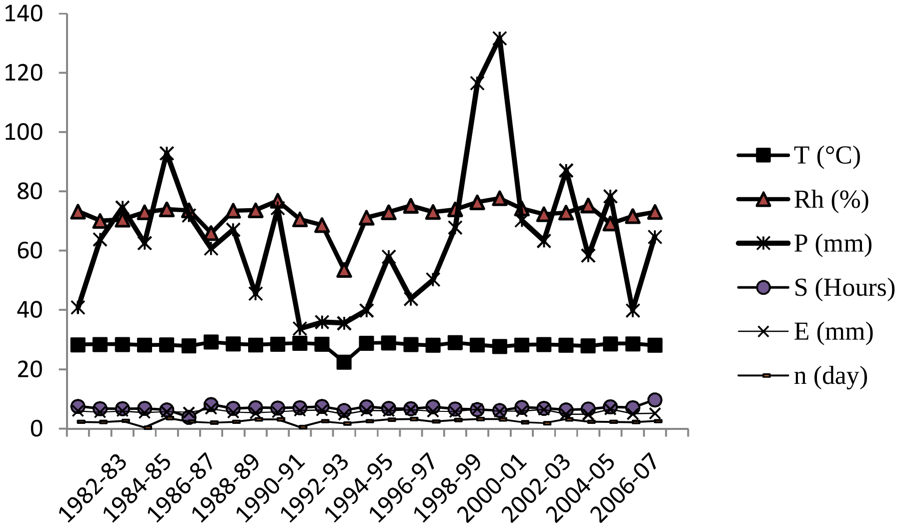

Figure 3. Weather variable

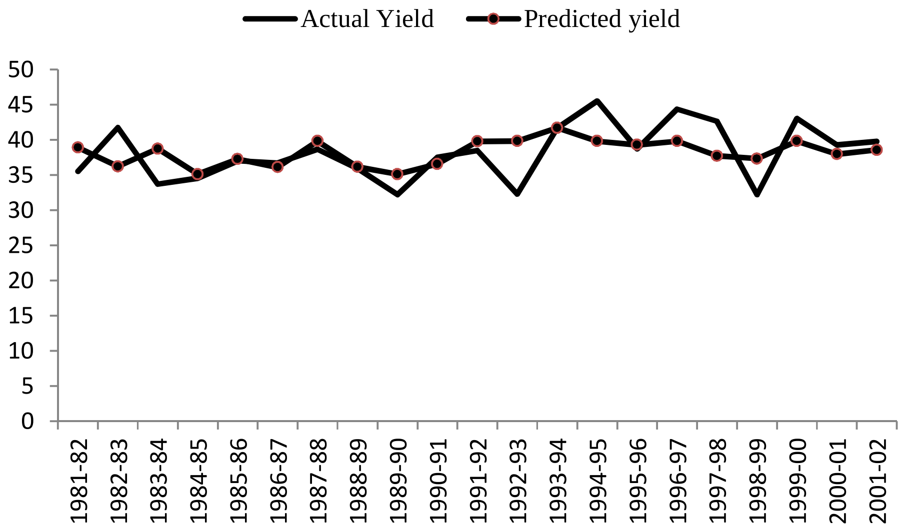

Figure 4.Observed and estimated rice crop yield during training period

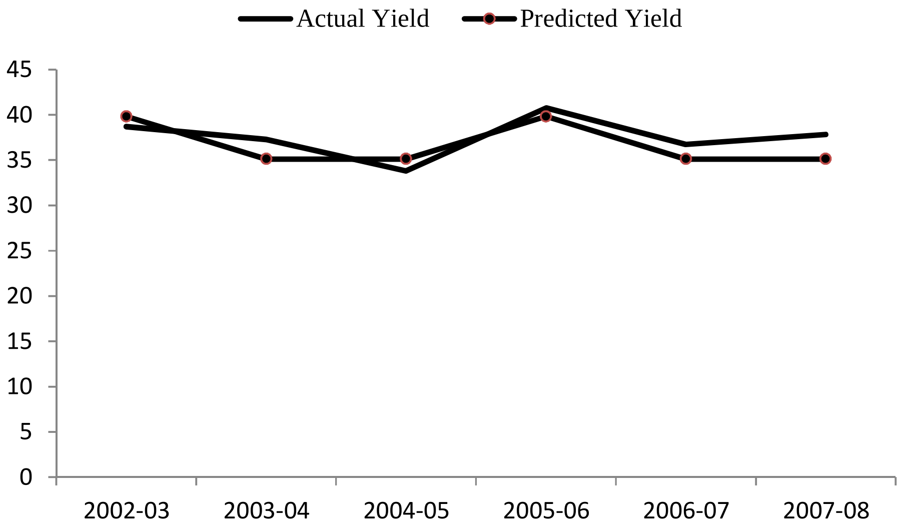

Figure 5. Observed and estimated rice crop yield during testing period

4. Conclusions

The present study discusses the application and usefulness of adaptive neuro fuzzy inference system based forecasting approach for forecasting of rice crop yield. The visual observation based on the graphical comparison between observed and predicted values and the qualitative performance assessment of the model indicates that ANFIS can be used effectively for crop yield forecasting.

References

Jang, J.-S. R. (1993) “ANFIS: adaptive network-based fuzzy inference systems”, *IEEE Trans. Syst., Man Cybern* **23**, 665–685.

Jang, J.S.R.; Sun, C.T. and Mizutani, E.(1997), “Neuro-Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligence”, Prentice Hall, NJ,USA ISBN 0-13-261066-3.

Kim, B.; Park, J.H.; Kim, B.S.(2002**),** “Fuzzy logic model of Langmuir probe discharge data”, *Comput Chem*,**26**(6), 573–581.

Loukas, Y.L (2001) “Adaptive neuro-fuzzy inference system: an instant and architecture-free predictor for improved QSAR studies”,*J Med Chem* **44**(17),2772–2783.

Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramasastri, K.S. (2004) “A neuro-fuzzy computing technique for modelling hydrological time series”, *J. Hydrology* **291**, 52-66.

Takagi, T., M. Sugeno.(1985) “Fuzzy identification of systems and its application to modeling and control”, *IEEE Transactions on Systems, Man, and Cybernetics,* **15,** 116–132.

Zadeh, L.A. (1965)”Fuzzy Sets”, *Information and Control* **8**, 338-353.

Agrawal; R.; Jain, R.C., Mehta, S.C. (2001) “Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis”, *Ind. J. Agri. Sci.* **71** (7), 487-490.

Huda, A.K.S.; B.P. Ghildyal; Tomar, V.S., Jain, R.C. (1975**)** “Contribution of climate variables in predicting rice yield”, *Agric. Met.* **15**,71-86.

Kokate, K.D.; Sananse, S.L. , . Kadam, J.R. (2000) “Pre-harvest forecasting of rice yield in Konkan region of Maharashtra” *J. Maha. Agric. Univ.* **25** (3),289-293.

Chowdhury, A., Sarkar, M.B. (1981) “Estimation of rice yield through weather factors in a dry sub-humid region” *Mausam* **32** (4), 393-396.

Jones, A. J.(2004) “New Tools in Non-linear Modeling and Prediction”, *Computational Management Science***1**(2), 109-149.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage:
http://www.iiste.org

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a fast manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar