Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 69(1) 2015 65-82

Prediction of Area and Crop Production for Summer Rice and Maize of Upper Brahmaputra Valley Zone of Assam using ANN

Raju Prasad Paswan[1] and Shahin Ara Begum[2]

[1]*Assam Agricultural University, Jorhat, Assam*

[2]*Assam University, Silchar, Assam*

Received 30 December 2013; Revised 03 April 2015; Accepted 06 April 2015

SUMMARY

The information related to crop area and crop production have been playing a vital role in planning and allocations of resources for the development of agriculture sector. In this paper, Artificial Neural Networks (ANN) has been used to predict the crop area and crop production for the crops Summer Rice and Maize of Upper Brahmaputra Valley (UBV) Zone of Assam as an alternative statistical tool. Multilayer Perceptron (MLP) with single hidden layer and Radial Basis Function (RBF) network have been trained with the secondary data of the crop area, crop production and meteorological data. The appropriate configuration for each of the network model is identified. The performance of the constructed ANN models has been measured using Root Mean Squared Errors (RMSE) and Correlation Coefficients (CC). The predictive accuracy of the developed ANN models has been compared with Multiple Linear Regression (MLR) Model. The performance comparison show that the constructed ANN-MLP and ANN-RBF models outperform MLR model. Sensitivity analysis has been performed for Prediction of Summer Rice and Maize production and results show that technology index is the most sensitive parameter followed by rainfall index for production of Summer Rice Production while temperature (maximum) is the most sensitive parameter for prediction of the crop Maize production followed by technology index for the UBV Zone of Assam with the considered secondary data.

Keywords: ANN, MLP, RBF, Crop Area, Crop Production.

1. INTRODUCTION

Amongst several states of India, Assam is one of the important states where agriculture is the main stay of the state economy and at the national level; the state contributes over 5 per cent of rice area and 4 per cent of rice production (Barah *et al*. 2001). An overview of the economy reveals that over 75% of the people of the state of Assam are directly or indirectly dependent on the agricultural sector. In terms of the state domestic product (SDP), the agriculture sector contributed over 28 percent of the state income in 2010-11 (Assam Portal, www.assam.org.node/2371).

Information on crop area statistics is backbone of agricultural statistical system. Reliable and timely information on crop area is of great importance to planners and policy makers for efficient and timely agricultural development and making important decisions with respect to procurement, storage, public distribution, export, import and other related issues. The agricultural production information is very important for planning and allocation of resources to different sectors of agriculture. Advance estimates of crop production are needed much before the actual harvest of crops for making various decisions such as pricing, distribution, export and import etc. Therefore, there is

Corresponding author: Raju Prasad Paswan
E-mail address: raju.pas66@gmail.com

a great need for developing suitable and reliable models using information from different sources like agricultural inputs, meteorological data, soil parameters and remote sensing data for providing the reliable and timely forecast of agricultural crop production. The goal of this study is to develop crop production prediction models with readily available data that could be easily applied by an end user. With this objective, the present study has been carried out to 1) investigate the applicability of ANN models *viz.* MLP and RBF for effective prediction of area and crop production for the crops Summer Rice and Maize of Upper Brahmaputra Valley Zone (UBVZ) of Assam under considered climatic conditions; 2) evaluate the performance of the constructed ANN models using RMSE and CC; 3) compare the performance of the developed ANN models with Multiple Linear Regression (MLR) model and finally 4) carry out sensitivity analysis to identify the influencing factor in relation to crop production.

This paper focuses on Prediction of crop Area and crop Production for the crops Rice and Maize using ANN models for the UBV Zone of Assam and sensitivity analysis has been performed to find the influencing parameter for prediction of crop production with the aforesaid objectives. In the rest of the paper Summer Rice is referred to as Rice.

The rest of the paper is organized as follows: Section 2 provides a discussion on related literature for prediction of crop production and other forecasting problem using ANN as well as multiple linear regression (MLR). Section 3 gives an overview of Multilayer Perceptron (MLP) and Radial Basis Function (RBF) network which have used in the present study. Section 4 presents Materials and Methods adopted for the study. The MLP, RBF and Statistical model developed for prediction of Area and Productions of the crops Rice and Maize of UBV Zone of Assam are presented in section 5. Section 6 presents experimental results discussion. Section 7 concludes the paper.

2. RELATED WORK

In recent years, ANNs have become popular for prediction and forecasting in a wide variety of applications where statistical methods are traditionally employed. The problems which were solved by classical statistical methods, such as discriminant analysis, logistic regression, Bayesian analysis, multiple regression models are now tackled by ANNs as an alternate statistical tool. Hence, in recent years the ANN approach for forecasting task has gained a great deal of interest from researchers in many fields of study including agriculture. ANNs have proved to be a more powerful and self adaptive method of yield estimation as compared to the traditional methods (Simpson 1994, Baret *et al*. 1995, Jiang 2004). Batchelor *et al.* (1997) showed that ANNs produced better results than traditional statistics methods when predicting soyabean rust epidemics. Starrett *et al.* (1997), reported that an ANN performed better than a regression model when predicting applied-nitrogen leaking below the root zone of turf grass. Gaudart *et al.* (2004) compared the performance of Multilayer perceptron (MLP) and that of linear regression for epidemiological data with regard to quality of prediction and robustness to deviation from underlying assumptions of normality, homoscedasticity and independence of errors and it was found that MLP performed better than linear regression. Several researchers have reported the attempt to apply the ANN approach in the forecast of agricultural yield (Haskett *et al*. 1998, Camago *et al*. 1999, Jackson and Looney 1999, Shearer *et al*. 1999). In a review paper, Paswan *et al*. (2013) have summarized some of the study of comparison of ANNs and Statistical models in literature for prediction of agricultural crop production.

3. ARTIFICIAL NEURAL NETWORKS

Haykin (1999), defined ANN as a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: First, knowledge is acquired by the network from its environment through a learning process; Second, interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge. A neural network model is composed of many neurons, which are basic units that work like computer processors. The units are connected by communication paths (connections) with weights. The units operate only on the inputs that they receive via the connections, and then send the outputs to the next layer of units. The neuron may receive the exciting signals and inhibitory signals at the same time, and then integrate them. The output is sent through an outlet to the other neurons. A neural network model has "training" rule whereby the weights of connections are adjusted on the basis of training set data. Thus, neural networks "learn" from input examples and exhibit some capability for generalization beyond the example data (Sarle 1999).

There are several classes of neural network architectures, classified according to their learning mechanisms and learning algorithms. Three fundamentally different classes of networks architectures are identified as single layer feedforward network, multilayer feedforward network and recurrent network. A feedforward network is one where units in one layer are connected only to units in the next layer, and not to units in a preceding layer or units in the same layer. Networks where the units are connected to other units in the same layer, to units in the preceding layer, or even to themselves are termed recurrent networks. Feedforward networks can be viewed as a special case of recurrent networks. In the present study, multilayer feedforward network (also called MLP) and Radial Basis Function (RBF) has been employed for the prediction of crop production.

The most obvious statistical interpretation of multilayer perceptron (MLP) is that it provides non-linear regression functions that are estimated by optimizing some measure of fit to the training data. Relatively less computer memory is required for the MLP network as compared to other ANN models and is generally fast (Lawrence 1994). Data move through the layers in one direction from the input through the hidden to the output layers. Generally an MLP architecture can have a number of hidden layers and different number of hidden units per layer. During training of the network, it produces its own output and tries to minimize the error between its own output and the target output. The minimization of the error is done by the weight adjustment during the learning process. Backpropagation is the most commonly used method for training MLP or multilayered feedforward networks. Backpropagation (Hertz *et al*. 1991, Warner *et al*. 1996, Singh 2008) is a form of supervised learning where the error rate is sent back through the network to change the weights to improve prediction and decrease error.

A RBF network is an artificial neural network that uses radial basis functions as activation function. They are used in function approximation, time series prediction and control. It can be used for approximating functions and recognizing pattern. RBF is a popular alternative network to MLP. RBF neural network is based on supervised learning. It has similar structure of neural network with MLP. The main difference is that RBF has a hidden layer which contains nodes called RBF unit. The hidden unit measures the distance between an input data vector and the centre of its RBF. In the present study, the RBF network consisting of an input layer, a hidden layer and an output layer has been employed. The input layer is made up of source nodes that connect the network to its environment. The hidden layer has sufficient dimensions, which applies a non-linear transformation of the input domain to a higher dimensional domain such that the training patterns can be linearly separated. The output layer is linear which provides the response of the network to the input patterns applied to the source nodes (Kim *et al.* 2005).

4. MATERIALS AND METHODS

4.1 Study Region

Based on rainfall, terrain and soil characteristics, Assam State has been broadly divided into the following six agro-climatic zones *viz.* North Bank Plain Zone (NBP Zone), Upper Brahmaputra Valley Zone (UBV Zone), Central Brahmaputra Valley Zone (CBV Zone), Lower Brahmaputra Valley Zone (LBV Zone), Barak Valley Zone (BV Zone), and Hills Zone (H Zone), each zone consisting of certain number of districts.

(**Source**: National Agricultural Research Project (NARP), www.aau.ac.in/dee/index.php)

Out of the six agro-climatic zones of Assam only the UBV Zone comprising of five districts has been considered for the present study and the developed ANN models may similarly be applied to other zones for the same crops.

4.2 Data Collection and Pre-processing

The crops considered in this study are Rice and Maize which comes under kharif crops and are cultivated during May to September every year. The study is based on secondary data for the last 30 years for the period 1981-1982 to 2010-2011 of the said crops. Two types of data considered in this study are - Crop data and Meteorological data. Crop-wise Area in hectare, Production in tonnes and Technology Index (Fertilizer consumption, High yielding varieties) have been considered in Crop data while Monthly Total Rainfall in mm for each year, Monthly Mean Maximum Temperature, Monthly Mean Minimum Temperature in degree Celsius (^{o}C) and Monthly Total Sunshine in

Hours (from May to September) have been considered in Meteorological data.

District-wise data for the zone of crop Area and Production of the crops Rice and Maize for the period 1981-1982 to 2010-2011 were obtained from the Directorate of Economics and Statistics, Directorate of Agriculture, Govt. of Assam, Guwahati and the Data on Technology index (Fertilizer consumption and High yielding varieties) have been collected from the Directorate of Agro-Economic Research Centre, Assam Agricultural University, Jorhat and from the Directorate of Agriculture, Govt. of Assam, Guwahati. District-wise Meteorological data have been obtained from National Data Centre, India Meteorological Department (IMD), Pune; Regional Meteorological Centre, IMD, Borjhar, Guwahati and from the Department of Agrometeorology, Assam Agricultural University, Jorhat, Assam.

The year-wise crop data i.e. data on Area and Production of Rice and Maize of all the districts (5 districts) of UBVZ have been summed up separately (crop-wise) to find the total crop Area under (in hectare) and crop Production (in tonnes). District-wise Area data of a crop have been added to find total crop Area for a single year of the zone and same procedure has been repeated to find the 30 years crop Area data of the considered zone. Similarly, district-wise, yearly crop Production data have been added to find total crop Production for a year of the zone and the same procedure is repeated up to 30 years to find yearly Production of the crop of the zone. Technology index for both the crops have been calculated as per expression in subsection 4.4. Since, these considered crops are produced during the period May to September each year, the Meteorological data such as Monthly Total Rainfall during May to September for all the districts of the zone are added year wise and average is calculated for each year which is actual rainfall for a year and from this actual rainfall, rainfall index for a particular year for 30 years has been calculated by using the formula in equation (1). Similarly, in case of Temperature, the Monthly Mean Maximum and Monthly Mean Minimum temperatures from May to September each year of the considered zone are summed up separately and average for the both has been evaluated separately and this is followed up to 30 years. In the same way, from the Monthly Total Sunshine Hours, average for each year of the zone, total sunshine hours has been calculated.

It is assumed here that farmers would decide to cover the area under certain crop depending upon the agricultural technology and rainfall (RF) available during the period of crop. Here, the term technology refers broadly to all the necessary inputs like HYV, fertilizer, irrigation, machinery etc. To use the information on each of these agricultural input as such in the model, it was observed that as these input were not recorded regularly and uniformly over the time and different places of a zone. Under such situation, instead of using the information on every agricultural inputs, an agricultural index depending on the available data has been evolved. Similarly, as the availability of irrigation is far from below, the farmers have to depend upon the RF during the period of the crop. Moreover, the records of RF over the different places of a zone was not uniformly recorded in the data collected for the present study. In some areas, the RF being excess than the necessary caused damage to the crops. To take account of the variation in the RF behavior a simple RF index has been used to represent RF over the entire zone.

4.3 Formulation of Rainfall Index

Equation (1) is used to calculate the rainfall index for a particular crop at a particular time.

$$r_i = (\alpha_j / n_j) \times 100 \tag{1}$$

where r_i, α_j and n_j represents rainfall index for the i^{th} crop, actual rainfall in the j^{th} zone and normal rainfall in j^{th} zone respectively. Although, the result drawn on the basis of may suffer slightly from robustness. It is expected to serve better as it is taking care of the wide range of rainfall with the zone and its normal rainfall. Normal rainfall is defined as the average seasonal rainfall for last 80 years from different zones of Assam.

4.4 Formulation of Technology Index

The Technology Index (P_i) is obtained as the proportion of High Yielding Varieties (HYV) area in the particular year and total area under the kharif crop *i.e.*

$$P_i = \frac{\text{HYV area}}{\text{Total area under kharif crop}}$$

The proportions (P_i) lies between 0 and 1.00. These are categorized into four different ranges and the scores provided to different ranges are tabulated in Table 1. The ratio P_i was considered for assigning a score according to the intensity of adoption of HYV.

The index P_i is defined as the proportion of HYV area to total area so as to divide the whole range *i.e.* 0-1 into four equal parts from minimum '0' to maximum '1'.

Table 1. Scores of proportions of HYV area

Sl. No.	Proportion of HYV Area	Scores
1	0.00 - 0.25	1
2	0.26 - 0.50	2
3	0.51 - 0.75	3
4	0.76 - 1.00	4

Calculation of Fertilizer Consumption/ha of area (it is given):

$$P = \frac{\text{Total fertilizer consumption}}{\text{Total area}}$$

Fertilizer consumption in the particular crop in the year is calculated as:

$$C_i = P \times \text{area under particular crop}$$

Proportion of fertilizer consumption for a year in a particular crop is calculated as:

$$C_j = \frac{C_i}{\sum C_i}$$

ΣC_i is the total consumption for all the crops. The C_j for particular crop also lies between 0 to 1.00. Since the information on fertilizer consumption (*FC*) for each crop separately is not available. As there is large variation in *FC* per year, between the districts of the zone and also in the area under a particular crop, the value of C_j can be expected to lie between 0-1 but not uniformly close to "1". The scores obtained for particular crop for fertilizer consumption are then added to scores obtained from the HYV areas to get different technology index for different crops in particular zone for different years.

5. ANN AND MLR MODELS

ANN-MLP with Back Propagation (BP) algorithm and ANN-RBF network along with MLR have been implemented in the present study. The constructed models have been trained with the secondary data of Area and Production and Meteorological data collected from various sources and an appropriate model is identified. This section provides the architecture of the constructed MLP and RBF models for the prediction of crop production. The MLR model used to compare the predictive accuracies is also presented here.

5.1 Multilayer Perceptron Network (MLP)

In the present study, backpropagation learning algorithm has been used to train the MLP. The structure of the MLP network depends on the number of dependent and independent variables present. In the present study, there are two predictor variables available, one is Area and another is Production. Area has two (functional) inputs - Technology Index (T), Rainfall Index (R) and Production has six inputs - Predicted Area (PA), T, R, Temperature Maximum (TMX), Temperature Minimum (TMN) and Sunshine Hour (SH). Therefore, the MLP structure for Prediction of Area and Production, are presented in Fig. 1 and Fig. 2 respectively. Both the models consist of three layers - an input layer, a hidden layer and an output layer. The

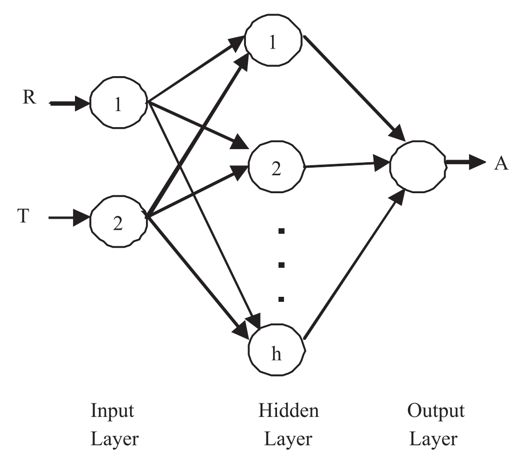

Fig. 1. MLP for Prediction of Area under Rice/Maize

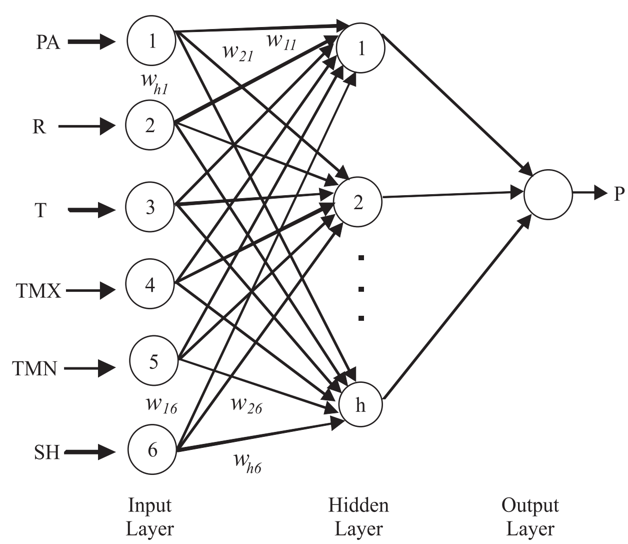

Fig. 2. MLP for Prediction of Rice/Maize Production

input and out layers contain nodes that represent input and output variables respectively. The output (v_j) of the j^{th} hidden node is given by equation (2)

$$v_j = f\left(\sum_{i=1}^{n} w_{ji} x_i\right) \tag{2}$$

where, x_i is the number of input, n is the number of input nodes, is the weight of connection between j^{th} hidden node and i^{th} input node and f is the transfer function is associated with j^{th} hidden node. The outputs of the networks for Area (A) and Production (P) are derived from the equation (3).

$$d = f\left(\sum_{i=1}^{h} w_{ji} v_i\right) \tag{3}$$

where, d is the desired output for Area and Production and h is the number of hidden nodes.

The most commonly used transfer function is a sigmoid function for hidden and output layer and linear transfer function is commonly used for the input layer. In the present study, during ANN model development log-sigmoid transfer function given in equation (4) is used in both hidden layer and output layer.

$$\log.sig(x) = \frac{1}{(1 + exp^{(-x)})} \tag{4}$$

The sigmoid function introduces non-linearity in the network. Sigmoid function is continuous and differentiable, its derivative is very fast to compute (as opposed to the derivative of tanh, which has similar properties) and has a limited range (from 0 to 1, exclusive). Simple neural networks can represent a wide variety of interesting functions when given appropriate parameters. One of the first versions of the theorem proved by Cybenko (1989) in 1989 for sigmoid activation functions states that arbitrary decision regions can be arbitrarily well approximated by continuous feedforward neural networks with only a single internal, hidden layer and any continuous sigmoidal nonlinearity. Hornik *et. al* (1989) showed that it is not the specific choice of the activation function, but rather the multilayer feed forward architecture itself which gives neural networks the potential of being universal approximators.

5.2 Radial Basis Function (RBF)

The number of hidden neuron is a crucial parameter in RBF. To determine the number of hidden neuron the same approach as that of back-propagation algorithm is followed. The forward selection method has been applied to determine the optimal number of nodes in the hidden layer. Then the radial basis function is trained and tested. This process continues until about the performance improvement of the neural network is no longer significant (Huang *et al*. 2009). Gaussian transfer functions are used in the hidden layer neurons whose outputs are inversely proportional to the distance from the center of the neuron. The output of the j^{th} hidden node is derived from equation (5).

$$\phi_j(x) = exp\left(-\frac{\| x - \mu_j \|^2}{2\sigma_j^2}\right) \tag{5}$$

where, σ_j and μ_j are the width and centre of the j^{th} hidden unit respectively and norm is the Euclidean distance and d_p is the desired output of the network. The transfer function of output node is linear. The output of the network is given by equation (6) and the RBF structure for Prediction of Rice and Maize Production is presented in Fig. 3.

$$d_p(x) = \sum_{j=1}^{h} w_j \phi_j(x) \tag{6}$$

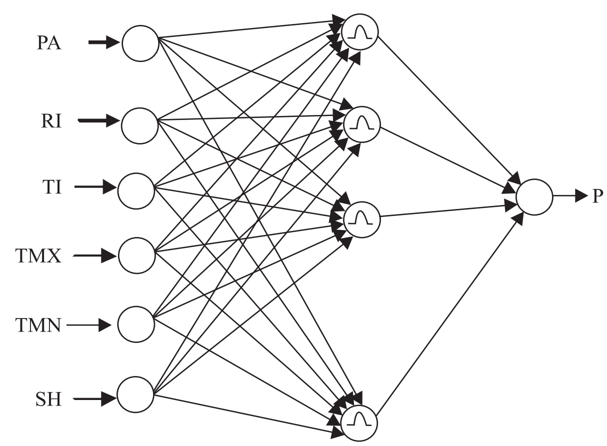

Fig. 3. RBF for Prediction of Rice/Maize Production

5.3 Multiple Linear Regression (MLR)

MLR technique is generally applied to investigate the relationship between a dependent variable and several independent variables. In this study, two models-one for predicting Area based on Rainfall Index and Technology Index, another models for predicting Crop Production based on predicted Area, Rainfall Index and Technology Index, Monthly mean maximum temperature, Monthly mean minimum temperature and, Monthly total sunshine hours (Goswami *et al.* 2002) have been constructed. While choosing variables and form it was assumed that the production of crop can be approximated fairly with a multiple linear model of the form equation (8) considering different types of agricultural and climatic factor involved with it. However, as the model was fitted with the data of a long period of past, the non availability of information relevant to all the factors occurred uniformly over the years and places posing problem. Hence, the factors for which information are available uniformly (over the places) are counted for the study. Similarly, for agricultural input the situation has been taken care of with the help of technology index.

The MLR models for prediction of crop area and production are represented in equations (7) and (8) respectively.

$$A = a + bT_j + cR_j \tag{7}$$

where, A = Area under crops; T_j = Technology index for crops; R_j = Rainfall Index for crops; a is constant and b, c are the regression coefficients.

$$P = a + bA + cT_j + dR_j + eTMX_j + gTMN_j + hSH_j \tag{8}$$

where, P = Production of crops; T_j = Technology index for crops; R_j = Rainfall Index for crops; TMX_j = Maximum temperature; TMN_j = Minimum Temperature; SH_j = Sunshine hour and a is constant and b, c, d, e, g, h are the regression coefficients whose values are estimated using matrix form of solving multiple linear regression.

If **Y** be the vector of observations Y, **X** be the matrix of independent variables, β be the vector of parameters to be estimated and **E** the vector of errors, thus for equation

$$\mathbf{Y}_i = \beta_0 + \beta_1\mathbf{X}_i + \beta_2\mathbf{X}_i + \ldots + \beta_n\mathbf{X}_n + \mathrm{E}_i$$

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \text{ and } E = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix}$$

The values for the parameters are obtained by substituting these values in equations (7) and (8), then the values for Area and Crop Production is estimated (Bhattacharyya 1999).

5.4 Determination of Input and Output Variables and Data Normalization

This study is carried out to predict Area using Rainfall index and Technology index and based on this predicted Area, Production of the crop is predicted. Therefore, for prediction of Area two inputs i.e. Rainfall index and Technology index have been considered while for prediction of Production, six inputs *viz.* Predicted area, Rainfall index, Technology index, Maximum temperature, Minimum temperature and Sunshine hours have been taken. As, all these input variables used have different dimensions, and there were major differences among the values, hence input and output variables are normalized within the range 0.1 - 0.9 using equation (9) in order to enhance the training speed and the precision of the neural network.

$$x_N = \frac{0.9 - 0.1}{x_{max} - x_{min}}(x - x_{min}) + 0.1 \tag{9}$$

where, x_N is the normalized value of x, x_{max} and x_{min} are maximum and minimum value of each parameter in the original data (input and the output data).

5.5 MLP for Prediction of Area and Production for the Crops Rice and Maize

The model to be established for Prediction of Area and Production for the crops Rice and Maize should both meet precision requirements and reduce learning time as much as possible. It is established from the literature that three-layer BP neural network using Sigmoid as activation function can approximate any multivariate function (Sun *et al*. 2004). Therefore, in the present study, one hidden layer in the network has been considered.

Determining the number of neurons in hidden layer is one important step to establish BP neural network. On one hand, if the number of neurons was too small, the network might be not trained enough; on the other hand, if the number of neurons was too large, that number of connection weights increased might make the network over parameterized (Sudheer *et al*. 2002). The method of trial and error is applied here. That is, a hidden layer of less neuron is selected and trained at first and gradually the number of neurons is increased till the convergence of the results.

5.6 Weight Initialization Methods

Weight initialization has been found as the most effective way to make the training of neural networks speed up. Experimentation is carried out with different weight initialization methods for determining the optimal initial weights of neural networks to increase the rate of convergence of networks. It is worth noting that the time required for the initialization process is negligible when compared with the training process. The different weight initialization methods used for determining the optimal initial weights of neural networks in this study *viz*. random weight initialization, Nguyen-Widrow and Hassoun method for MLP network and pseudo-inverse method for RBF network are discussed in brief in the next sub-sections.

5.6.1 Random weight initialization method

One of the most trivial methods to initialize the weights is to assign them to random values. Generally, weights are initialized with uniform random values within the interval [–1, 1]. This method is included to provide a reference (Kumar 2004).

5.6.2 Nguyen-Widrow method

This method was proposed by Derrick Nguyen and Bernard Widrow (Nguyen *et al*. 1990). Nguyen-Widrow method generates initial weights and bias values for a layer, approximately evenly over the input space. Small random values as initial weights of the neural network is considered in this method. The weights are altered in such a way that the area of interest is divided into small intervals. It is logical to consider speeding up the training process by setting the initial weights of the first layer so that each node is assigned its own interval at the start of training. To implement Nguyen-Widrow method, the neural network is first initialized with random weight values (w_{t+1}) in a specific range. Next, the value of beta is calculated using equation (10).

$$\beta = 0.7h^{1/i} \tag{10}$$

where, h and i represent the number of hidden neurons and number of input neurons respectively. Next, the Euclidean distance for all the weights on a layer is calculated using the equation (11).

$$n = \sqrt{\sum_{i=0}^{i<w_{\max}} w_{i2}} \tag{11}$$

Once the beta and distance values have been calculated, the random numbers can be adjusted. The equation (12) shows how weights are adjusted using the earlier calculation.

$$w_{t+1} = \frac{\beta w_{i+1}}{n} \tag{12}$$

5.6.3 Hassoun method

With Hassoun method, uniform random weights can be generated within the range $[-3/\sqrt{f_i} + 3/\sqrt{f_i}]$. where, f_i is the number of inputs (fan-in) for unit i. Here, it is assumed that zero mean random weights are in the range $[-r, + r]$ and inputs are normalized in the range [0, 1], random weighted sum net_i has zero mean and standard deviation $\sigma_{net_i} = (r/3) + \sqrt{f_i}$ (Hassoun 2002).

5.6.4 Pseudo-inverse method

In the present study, the pseudo-inverse method for linear weight in the output layer of the RBF network is used.

5.7 Training and Testing of Neural Network

In the present study, 30 years data for the period 1981-82 to 2010-11 for Area and Crop Production have been considered. The data set for prediction of Area and prediction of Production of crops considered (Rice and Maize), have been divided randomly into a training set and testing set. The training set consists of 24 years data (80%) while testing set consists of 6 years data (20%). This has been done after normalizing both the data sets separately by the normalization expression in equation (9). The network has been trained with the normalized data, and the weights (w) are determined so as to minimize the cost function given by equation (13).

$$E = \frac{1}{2}\sum_{p=1}^{N}(t^p - y^p)^2 \tag{13}$$

where, t^p, y^p are the target and network output for p^{th} training pattern and N is the total number of training patterns.

The optimum neural networks configurations are determined by trial and error method with an objective to minimize the difference among the network predicted values and the target values. Coefficient of correlation between the measured and predicted values is a good indicator to check the prediction performance of the model. The performance of each configuration was evaluated based on the root mean square error (RMSE) and correlation coefficient (CC) given by equations (14) and (15) respectively.

$$RMSE = \sqrt{\frac{1}{N}\sum_{p=1}^{N}(t^p - y^p)^2} \tag{14}$$

$$CC = \frac{\sum_{p=1}^{n}(y^p - \overline{y})(t^p - \overline{t})}{\sqrt{\sum_{p=1}^{n}(y^p - \overline{y})^2 (t^p - \overline{t})^2}} \tag{15}$$

where, $\overline{y}$ and $\overline{t}$ are average over network and target outputs.

The optimal configuration, based upon minimizing the difference between the neural networks predicted outputs and the target (desired) outputs, corresponds to the minimum value of RMSE and the maximum value of CC (Drummond *et al.* 2003).

6. EXPERIMENTAL RESULTS AND DISCUSSION

Experiments have been carried out to identify the model fit amongst ANN models (MLP, RBF) and MLR model to predict crop Area and crop Production for the crops Rice and Maize of UBV Zone of Assam. Same data set has been used in all test cases for ANN and statistical models to evaluate the performance of the models. Some of the experimental results of crop Area and crop Production for both the crops have been tabulated and from these results the best cases have been plotted. The best case result of all the three models have been compared to find the model fit. Finally, sensitivity analysis has been carried out to find the most influencing parameter for crop (rice and maize) Production for UBV Zone of Assam.

Matlab 7.9 environment has been used for the implementation of MLP with BP algorithm and RBF models. The results obtained from training and testing are used to construct an ANN model that can estimate crop Area and crop Production for the crops Rice and Maize. The reliability of the predicted values for both Area and crop Production depends on the ANN structure and the input data. To get the better results, the input data needs to be reliable. In the present study, the input data used for training and testing the ANN models have been collected from the sources stated in section 4 which are assumed to be reliable. The training dataset is used to train the network and testing dataset is used to check the generalization capability of the network.

Next task in the study is the selection of the optimum network architecture. There are a number of different parameters that must be decided upon when designing an ANN. Among these the ANN-MLP network parameters are Hidden neuron (HN) - HN is the number of neurons in the hidden layer, Learning rate (LR) - LR is the training parameter that controls the size of weight and bias changes in learning of the training algorithm, Momentum (M) - the momentum parameter is used to prevent the system from converging to a local minimum or saddle point and the number of epochs represents the maximum number of iterations in the algorithm. The ANN-RBF network parameters are HN and Spread constant (SC) - the spread constant determines the width of an area in the input space to which each neuron responds. Large SC makes the network function smoother and it results in better generalization. The individual test cases have been ranked according to the magnitude of RMSE and CC, which have been calculated using equations (14) and (15) respectively to develop the optimum network architecture. The model having the minimum RMSE and maximum CC is selected as the best.

The training and the testing results of the MLP model amongst 310 tested cases for Area under Rice and 286 tested cases for Rice Production is selected and tabulated in Table 2 and Table 3 respectively. The corresponding best case result of MLP for prediction of Rice Area are plotted in Fig. 4 and Fig. 5 and for Rice Production are plotted in Fig. 6 and Fig. 7 respectively. The training and the testing results of the RBF model amongst 198 tested cases for Area under Rice and 214 tested cases for Rice Production is selected and tabulated in Table 4 and Table 5 and corresponding best case result for Area under Rice are plotted in Fig. 8 and Fig. 9, while for Rice Production these are plotted in Fig. 10 and Fig. 11 respectively.

For the crop Maize, the training and the testing results of the MLP model amongst 270 tested cases of Area under Maize and amongst 230 tested cases of Maize Production is selected and are tabulated in Table 6 and Table 7 respectively. The corresponding best case result of MLP for prediction of Area under Maize are

Table 2. MLP Training and Testing results for Area under Rice

Weight initialization method	Parameters of Network				Training		Testing	
	HN	LR	M	Epoch	RMSE	CC	RMSE	CC
Random	3	.3	.5	20000	0.0802	0.9078	0.0962	0.8020
		.5	.3	50000	0.0847	0.9261	0.0537	0.7737
		.5	.7	50000	0.0854	0.9247	0.0580	0.7813
		.7	.7	30000	0.0898	0.9167	0.0733	0.7561
		.9	.5	50000	0.0939	0.9160	0.0649	0.7687
	5	.3	.7	20000	0.0683	0.9397	0.0688	0.9359
		.5	.3	20000	0.0717	0.9332	0.0702	0.9340
		.5	.5	10000	0.0749	0.9310	0.0137	0.9305
		.5	**.7**	**10000**	**0.0752**	**0.9371**	**0.0052**	**0.9496**
		.7	.3	50000	0.0755	0.9278	0.0062	0.9383
		.7	.5	50000	0.0765	0.9297	0.0060	0.9493
	7	.3	.3	10000	0.0971	0.9006	0.0296	0.8436
		.3	.7	10000	0.0805	0.9514	0.0228	0.7877
		.5	.7	50000	0.0700	0.9372	0.0054	0.9281
	9	.3	.7	50000	0.0650	0.9437	0.0082	0.9248
		.5	.5	20000	0.0709	0.9345	0.0258	0.9414
		.5	.7	50000	0.0709	0.9340	0.0706	0.9388
		.7	.5	20000	0.0751	0.9296	0.0055	0.9410
	11	.7	.3	50000	0.0755	0.9279	0.0091	0.9408

Table 3. MLP Training and Testing results for Rice Production

Weight initialization method	Parameters of Network				Training		Testing	
	HN	LR	M	Epoch	RMSE	CC	RMSE	CC
Random	7	.3	.3	20000	0.0020	0.9999	0.0843	0.6148
		.3	.5	10000	0.0075	0.9991	0.0417	0.5101
		.5	.3	20000	0.0004	0.9999	0.0563	0.3542
		.5	**.5**	**30000**	**0.0004**	**0.9999**	**0.0364**	**0.8290**
		.7	.5	50000	0.0004	0.9999	0.0558	0.4488
	9	.3	.3	30000	0.0009	0.9999	0.0497	0.5424
		.3	.5	10000	0.0028	0.9998	0.0437	0.2807
		.5	.3	15000	0.0020	0.9999	0.0744	0.4971
		.7	.5	50000	0.0004	0.9999	0.0462	0.4918
	11	.3	.5	10000	0.0039	0.9997	0.0566	0.4208
		.7	.5	10000	0.0039	0.9997	0.0744	0.6162

Table 4. RBF Training and Testing results for Area under Rice

Spread of Gaussian RBF	RBF Center	Parameters of Network		Training		Testing	
		HN	SC	RMSE	CC	RMSE	CC
Normal-ization	Fixed Center	9	-	0.1503	0.6059	0.0122	0.7201
		15	-	0.1037	0.8357	0.0129	0.5798
		19	-	0.0165	0.9961	0.0098	0.7346
		22	-	0.0114	0.9981	0.0195	0.5649
	K-means	3	-	0.1575	0.5517	0.0182	0.6498
		5	-	0.1447	0.6430	0.0109	0.7534
		7	-	0.1135	0.7994	0.0129	0.5260
		15	-	0.0909	0.8766	0.0069	0.7693
		17	-	0.0527	0.9602	0.0132	0.6232
		19	-	0.0720	0.9244	0.0076	0.6852
Hassoun	Fixed Center	3	1.2	0.0631	0.5043	0.0138	0.8945
			1.4	0.1629	0.5063	0.0118	0.9748
			1.5	0.1657	0.4802	0.0116	0.9791
		5	1.3	0.1454	0.6380	0.0188	0.7742
		7	1.1	0.1508	0.6019	0.0146	0.9224
			1.3	0.1394	0.6745	0.0168	0.6103
			1.4	0.1509	0.6014	0.0201	0.7886
			1.5	0.1370	0.6885	0.0232	0.5170
	***K*-means**	**3**	1.1	0.1646	0.4907	0.0115	0.8836
			1.3	**0.0139**	**0.9906**	**0.0058**	**0.9906**
			1.4	0.1589	0.5409	0.0146	0.9454
			1.5	0.1635	0.5011	0.0122	0.9039
		5	1.4	0.0508	0.6022	0.0173	0.8646
		7	1.1	0.1510	0.6007	0.0195	0.8015

Table 5. RBF Training and Testing results for Rice Production

Spread of Gaussian RBF	RBF Center	Parameters of Network		Training		Testing	
		HN	SC	RMSE	CC	RMSE	CC
Normal-zation	Fixed Center	9	-	0.1633	0.4889	0.0404	0.6392
		17	-	0.0387	0.9783	0.0410	0.5275
		19	-	0.0437	0.9723	0.0445	0.5500
		20	-	0.0471	0.9678	0.0470	0.4803
		21	-	0.0030	0.9998	0.0620	0.3901
		22	-	0.0088	0.9988	0.0450	0.5317
	K-means	11	-	0.1356	0.6897	0.0345	0.6267
		13	-	0.0478	0.9667	0.0485	0.4915
		15	-	0.0598	0.9476	0.0555	0.4198
		17	-	0.0639	0.9399	0.0630	0.5192
		22	-	0.0009	0.9999	0.0475	0.5274
Hassoun	Fixed Center	7	1.3	0.1244	0.7473	0.0317	0.5681
		13	1.1	0.0669	0.9339	0.0448	0.4753
		17	1.1	0.0657	0.9362	0.0457	0.4834
		17	1.5	0.0828	0.8968	0.0415	0.5610
		20	1.3	0.0374	0.9798	0.0319	0.6019
		21	1.1	0.0316	0.9856	0.0285	0.6263
		21	1.4	0.0274	0.9892	0.0280	0.6558
	K-means	15	1.2	0.0548	0.9561	0.0380	0.5747
		15	1.5	0.0662	0.9353	0.0324	0.6388
		17	1.4	0.0602	0.9468	0.0339	0.6239
		22	**1.1**	**0.0151**	**0.9967**	**0.0394**	**0.6847**

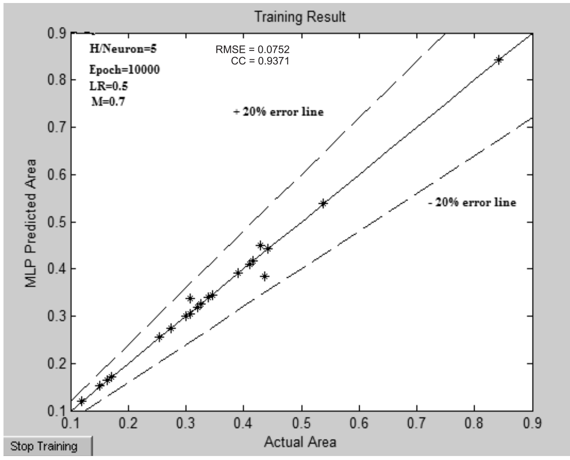

Fig. 4. Comparison of Actual and MLP Predicted Area under Rice (Training)

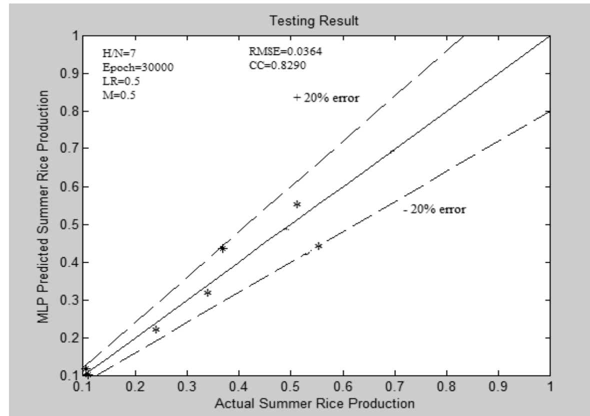

Fig. 7. Comparison of actual and MLP Predicted Rice Production (Testing)

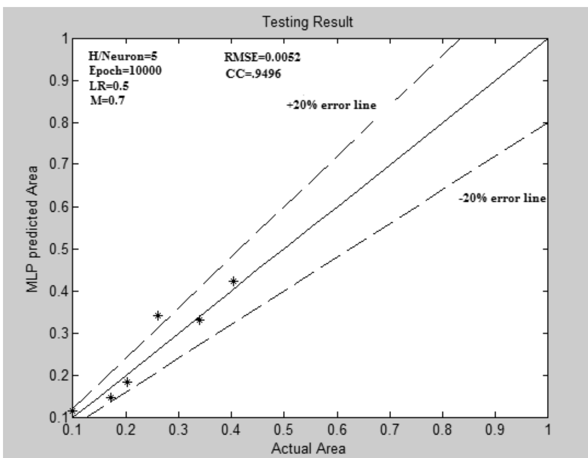

Fig. 5. Comparison of actual and MLP Predicted Area under Rice (Testing)

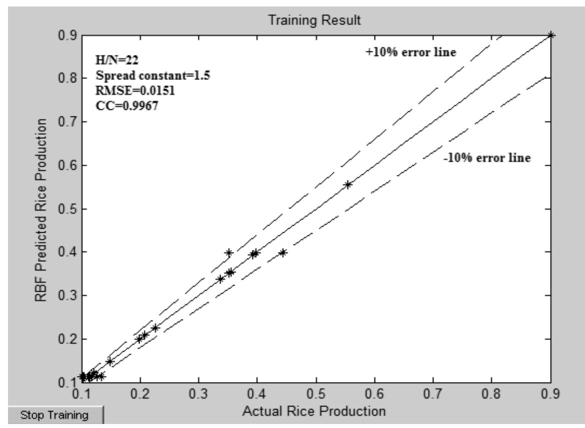

Fig. 8. Comparison of Actual and RBF Predicted Area under Rice (Training)

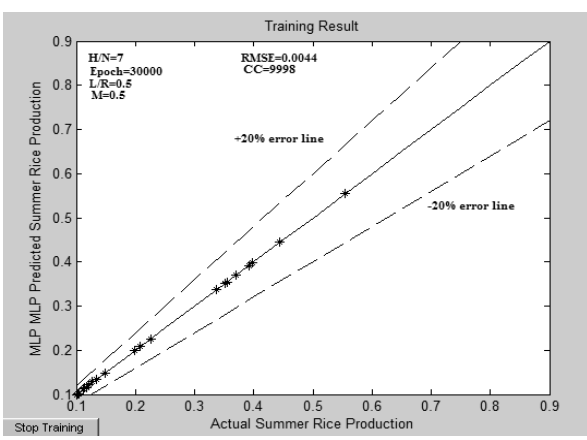

Fig. 6. Comparison of actual and MLP Predicted Rice Production (Training)

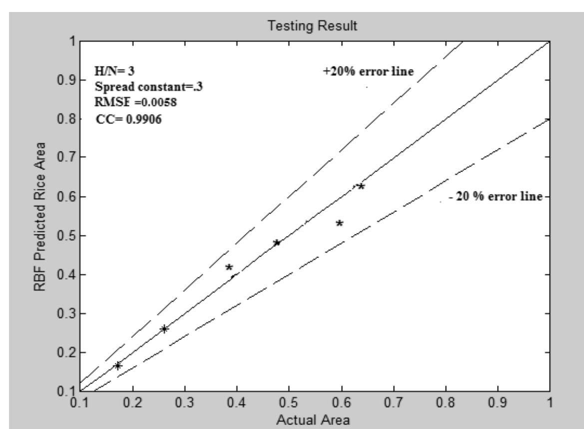

Fig. 9. Comparison of Actual and RBF Predicted Area under Rice (Testing)

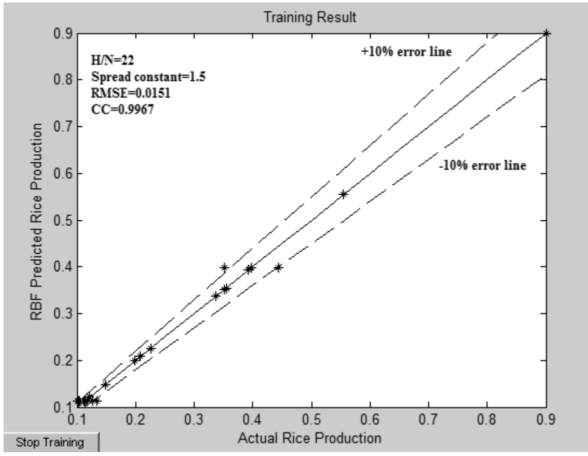

Fig. 10. Comparison of Actual and RBF Predicted Rice Production (Training)

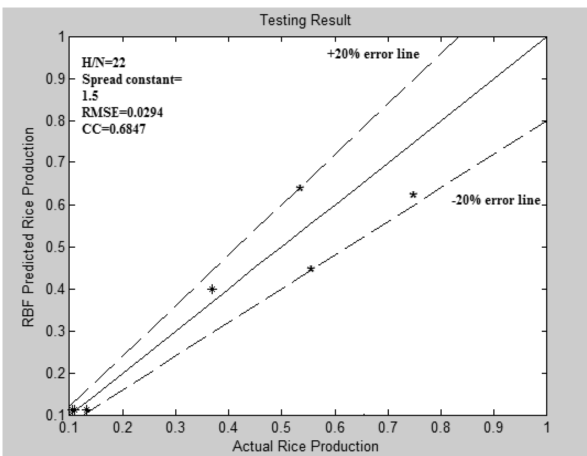

Fig. 11. Comparison of Actual and RBF Predicted Rice Production (Testing)

plotted in Fig. 12 and Fig. 13 and for Maize Production it is plotted in Fig. 14 and Fig. 15 respectively. Similarly, the training and the testing results of the RBF model amongst 210 tested cases of Area under Maize are tabulated in Table 8 and training and testing results are plotted in Fig. 16 and Fig. 17. The training and the testing results of the RBF model among 199 tested cases of Production for the crop Maize are Tabulated in Table 9 and corresponding best case result for Production for the crop Maize are plotted in Fig. 18 and Fig. 19 respectively.

On comparison of the experimental results obtained with MLP and RBF models for prediction of Area under Rice, the MLP model with random weight initialization method, with RMSE value of 0.0052, CC value of 0.9496 and the network configuration with one hidden layer having hidden neurons 5, learning rate of

Table 6. MLP Training and Testing results (best case) of Area under Maize

Weight initialization method	Parameters of Network				Training		Testing	
	HN	LR	M	Epoch	RMSE	CC	RMSE	CC
Random	5	.3	.3	20000	0.1436	0.6138	0.0867	0.6806
	5	.3	.5	20000	0.1362	0.6482	0.0624	0.6322
	5	.3	.7	20000	0.1337	0.6405	0.0743	0.7088
	3	.3	.5	30000	0.1343	0.6364	0.0739	0.6869
				70000	0.1222	0.7194	0.0164	0.7146
	3	.7	.5	20000	0.1379	0.6681	0.0846	0.6202
	7	.3	.3	20000	0.1339	0.6389	0.0826	0.6809
	9	.3	.3	20000	0.1349	0.6335	0.0587	0.6669
				20000	0.1341	0.6396	0.0641	0.6379
Hassoun	5	.3	.3	15000	0.1344	0.6385	0.0685	0.6806
	5	.3	.7	15000	0.1317	0.6434	0.0839	0.6289
	5	.3	.5	50000	0.1523	0.6062	0.0832	0.6024
Nguyen -Widrow	**5**	**.3**	**.5**	**1000**	**0.0561**	**0.7733**	**0.0417**	**0.9915**
	5	.3	.3	10000	0.1492	0.6014	0.0792	0.4957
				20000	0.1330	0.6480	0.0607	0.6724
	5	.3	.5	10000	0.1375	0.6352	0.0637	0.6219

Table 7. MLP Training and Testing results (best case) for Production (Maize)

Weight initialization method	Parameters of Network				Training		Testing	
	HN	LR	M	Epoch	RMSE	CC	RMSE	CC
Random	3	.3	.7	10000	0.0291	0.9951	0.0989	0.9693
Hassoun	**3**	**.3**	**.3**	**10000**	**0.0231**	**0.9968**	**0.0354**	**0.8726**
	5	.3	.5	10000	0.0016	0.9999	0.0856	0.1409
	9	.3	.3	10000	0.0024	0.9990	0.0456	0.8293
	9	.3	.7	10000	0.0044	0.9997	0.0608	0.5692
Nguyen-Widrow	16	.5	.7	5000	0.0267	0.9862	0.1764	0.9977
	17	.5	.7	10000	0.0014	0.9999	0.0940	0.2893
	11	.3	.3	10000	0.0091	0.9996	0.0978	0.8143
	11	.3	.7	10000	0.0004	0.9999	0.0496	0.7020
	13	.3	.3	10000	0.0048	0.9997	0.0628	0.6127
	14	.2	.4	10000	0.0064	0.9996	0.0557	0.7668
	5	.3	.7	10000	0.0040	0.9998	0.0361	0.4632

Table 8. RBF training and testing (best case) of Area under Maize

Spread of Gaussian RBF	RBF Center	Parameters of Network		Training		Testing	
		HN	SC	RMSE	CC	RMSE	CC
Normal-ization	Fixed Center	17	-	0.1046	0.7527	0.0484	0.7670
	K-means	3	-	0.1444	0.4178	0.0538	0.8355
Hassoun	Fixed Center	3	1.3	0.1526	0.2789	0.0591	0.9266
		9	1.3	0.1282	0.5908	0.0614	0.7712
			1.4	0.1213	0.6460	0.0495	0.8868
			1.5	0.1280	0.5925	0.0537	0.8572
		11	1.2	0.1293	0.5819	0.0542	0.9146
		13	**1.3**	**0.0213**	**0.6461**	**0.0419**	**0.9286**
			1.4	0.1099	0.7225	0.0667	0.6539
		15	1.1	0.1090	0.7280	0.0754	0.6019
			1.2	0.1064	0.7427	0.0820	0.7230
		17	1.1	0.1063	0.7433	0.0716	0.7062
			1.2	0.1020	0.7668	0.0847	0.6845
	K-means	7	1.3	0.1085	0.7306	0.0688	0.7624
			1.4	0.1288	0.5862	0.0635	0.7873
			1.2	0.1307	0.5686	0.0693	0.7531
			1.3	0.1275	0.5967	0.0553	0.8241
			1.5	0.1231	0.6327	0.0599	0.7932
		11	1.1	0.1247	0.6200	0.0552	0.8135
			1.2	0.1271	0.6006	0.0542	0.9181
			1.3	0.1266	0.6043	0.0598	0.8230
		13	1.2	0.1134	0.7009	0.0714	0.7737
		15	1.1	0.1083	0.7319	0.0672	0.7523

Table 9. RBF training and testing (best case) of Maize Production

Spread of Gaussian RBF	RBF Center	Parameters of Network		Training		Testing	
		HN	SC	RMSE	CC	RMSE	CC
Norma-lization	Fixed Center	13	-	0.0663	0.7494	0.0752	0.7931
Hassoun	Fixed Center	5	1.3	0.0938	0.3524	0.0745	0.6904
			1.5	0.0951	0.3146	0.0761	0.5523
		7	1.3	0.0891	0.4572	0.0702	0.5938
		9	1.1	0.0720	0.6959	0.0689	0.5074
			1.5	0.0632	0.7759	0.0890	0.9935
		13	1.2	0.0623	0.7834	0.0524	0.8783
		17	1.1	0.0542	0.8411	0.0568	0.7006
			1.3	0.0356	0.9345	0.0594	0.5772
		19	1.2	0.0317	0.9485	0.0548	0.6902
			1.3	0.0218	0.9759	0.0496	0.7594
		21	1.1	0.0107	0.9941	0.0467	0.7954
			1.3	**0.0073**	**0.9972**	**0.0371**	**0.8634**
		22	1.1	0.0009	0.9999	0.0424	0.7972
			1.4	0.0019	0.9998	0.0511	0.7199
	K-means	7	1.3	0.0786	0.6208	0.0705	0.5138
		9	1.1	0.0824	0.5689	0.0680	0.6965
		11	1.5	0.0727	0.6880	0.0652	0.6024
		15	1.3	0.0608	0.7951	0.0594	0.6232
		22	1.3	0.0007	0.9999	0.0379	0.8330
			1.4	0.0083	0.9965	0.0494	0.7948

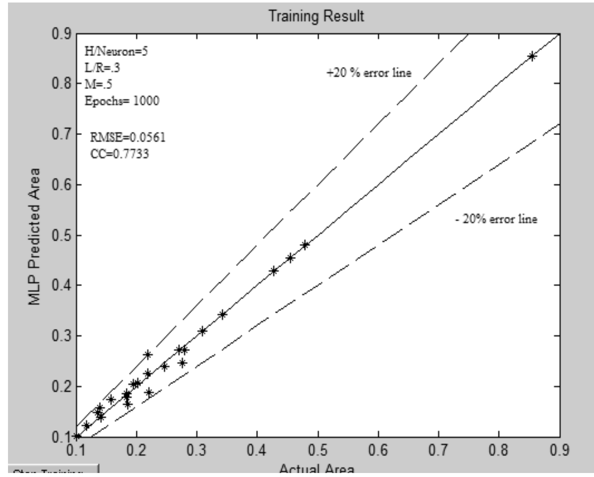

Fig. 12. Comparison of Actual and MLP Predicted Area under Maize (Training)

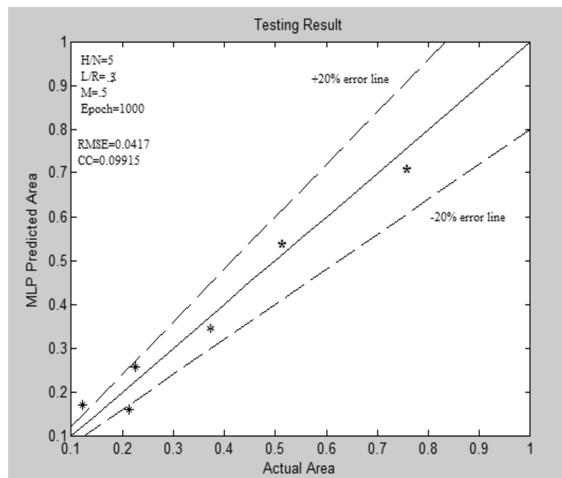

Fig. 13. Comparison of Actual and MLP Predicted Area under Maize (Testing)

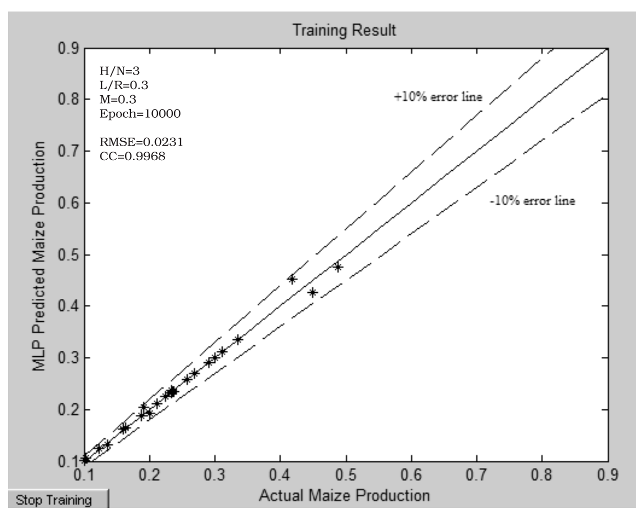

Fig. 14. Comparison of Actual and MLP Predicted Maize Production (Training)

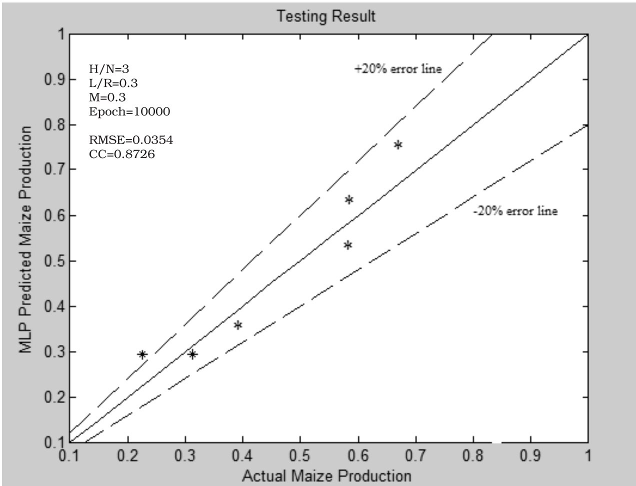

Fig. 15. Comparison of Actual and MLP Predicted Maize Production (Testing)

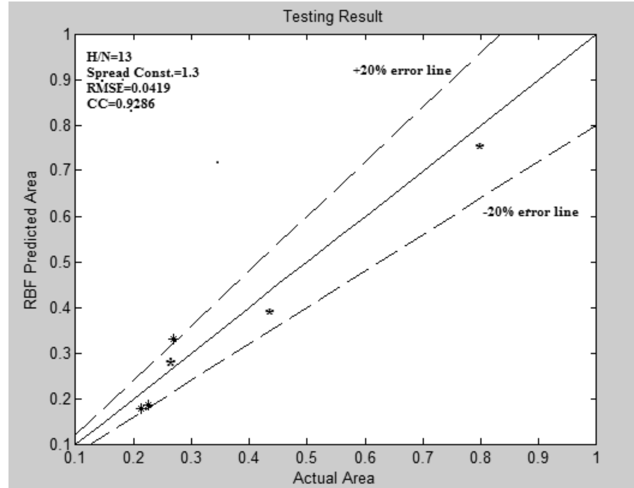

Fig. 17. Comparison of Actual and RBF Predicted Area under Maize (Testing)

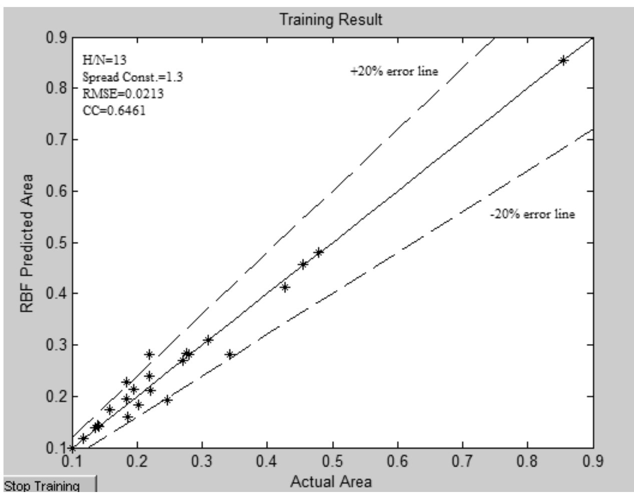

Fig. 16. Comparison of Actual and RBF Predicted Area under Maize (Training)

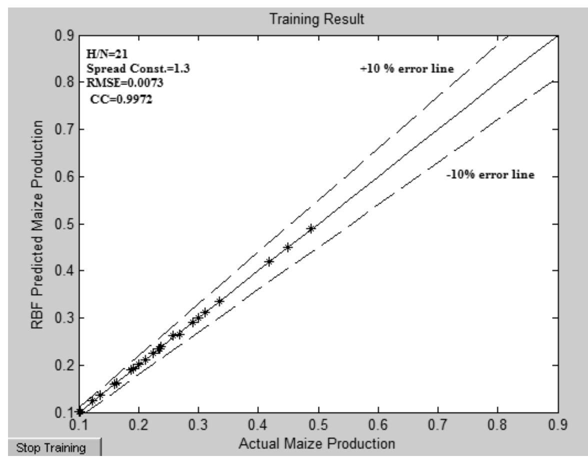

Fig. 18. Comparison of Actual and RBF Predicted Maize Production (Training)

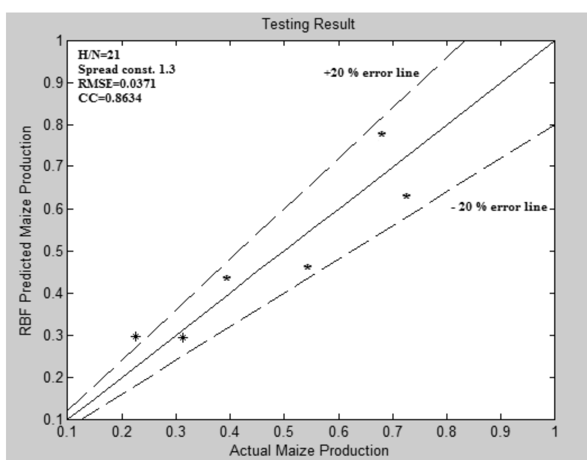

Fig. 19. Comparison of Actual and RBF Predicted Maize Production (Testing)

0.5, a momentum constant of 0.7 and epochs equal to 10000 is the best model and the entries in bold in Table 2 are the optimum values. For the prediction of Rice production, the MLP model with random weight initialization method, with RMSE value of 0.0364, CC value of 0.8290 and the network configuration with one hidden layer having 7 hidden neurons, learning rate of 0.5, a momentum constant of 0.5 and epochs equal to 30000 is found to be the optimum model. The entries shown in bold in Table 3 represent the optimum values.

Similarly, the experimental results of MLP and RBF models obtained for prediction of Area under Maize, the MLP model with Nguyen-Widrow weight initialization method, with RMSE value of 0.0417, CC value of 0.9915 and the network configuration with one hidden layer having 5 hidden neurons, learning rate of 0.3, a momentum constant of 0.5 and epochs equal to 1000 is found to be the best model. The entries in bold in Table 6 are the optimum values. For prediction of Maize Production, on comparison of the results obtained from experimentation with MLP and RBF models, the MLP model with Hassoun method of weight initialization, with RMSE value of 0.0354, CC value of 0.8726 and the network configuration with one hidden layer having 3 hidden neurons, learning rate of 0.3, a momentum constant of 0.3 and epochs equal to 10000 is found to be the best model and the optimum values are highlighted in bold in Table 7.

The outputs of the ANN-MLP and ANN-RBF models of crop Area and crop Production for both the crops have been compared with the results of regression equations used by (Goswami *et al*. 2002) using the same dataset and the result is tabulated in Table 10. From the Table 10, it is seen that, ANN-MLP model obtains the best results followed by ANN-RBF model for prediction of crop Area and Crop Production for both the crops i.e. Rice and Maize. The predictive accuracy of the ANN models is relatively higher than the MLR model for the test set of data. Comparatively low value of RMSE and high value of CC in case of ANN-MLP in comparison to MLR model shows that ANN model is a better alternative to the conventional regression model for prediction of crop Area and crop Production for Rice and Maize crops of UBV Zone of Assam. It may be noted here that the variation in predicted value between the methods occurs due to different underlying ANN structure, activation (transfer) function and the network parameter set during network training.

Although ANN models outperformed MLR model for prediction of crop Area and crop Production for Rice and Maize crops of UBV Zone of Assam, one problem that occur during neural network training is over-fitting. In Figs. 4-19 and the Tables 2-9, it can be seen that the error on the training set is driven to a small value, but when new data is presented to the network the error is relatively large as compared to the error on training set. The network has memorized the training examples, but it has not learned to generalize to new situations with the same accuracy as with the training set.

To determine the relative significance of each of the input parameters on Rice and Maize production, sensitivity analysis has been performed. Sensitivity analysis with the best neural network configuration shows the relative influence of various input parameters for prediction of Rice and Maize Production. Sensitivity analysis has been performed with the best ANN configurations.

To carry out the sensitivity analysis, one of the input parameters has been eliminated in each case, and the results obtained from MLP and RBF networks are tabulated in Table 11 for Rice Production and in Table 12 for Production of the crop Maize respectively**.** From Table 11 for Rice Production, it can be concluded that in both MLP and RBF networks, technology index is the most sensitive factor followed by the rainfall index while from Table 12 it can be concluded that, the

Table 10. Performance indices (RMSE and CC) for models

Technique	Rice				Maize			
	Area		Production		Area		Production	
	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
ANN-MLP	**0.0052**	**0.9496**	**0.0364**	**0.8290**	**0.0417**	**0.9915**	**0.0354**	**0.8726**
ANN-RBF	0.0058	0.9906	0.0394	0.6847	0.0419	0.9286	0.0371	0.8634
MLR	0.0857	0.8745	0.4975	0.3880	0.0664	0.3479	0.0758	0.4581

Table 11. Sensitivity analysis with ANN models for Rice production

Technique	Production		Technique	Production	
	RMSE	CC		RMSE	CC
ANN-MLP with PA, RI, TI, TMX, TMN,SH	0.0364	0.8290	ANN-RBF with PA, RI, TI, TMX, TMN, SH	0.0294	0.6847
ANN-MLP without RI	0.4538	0.4230	ANN-RBF without RI	0.5295	0.4163
MLP without TI	**0.5802**	**0.1671**	ANN-RBF without TI	**0.6456**	**0.2462**
ANN-MLP without TMX	0.0442	0.5066	ANN-RBF without TMX	0.0723	0.5144
ANN-MLP without TMN	0.0544	0.6431	ANN-RBF without TMN	0.0700	0.4462
ANN-MLP without SH	0.0482	0.5722	ANN-RBF without SH	0.0456	0.4062

Table 12. Sensitivity analysis with ANN models for the crop Maize production

Technique	Production		Technique	Production	
	RMSE	CC		RMSE	CC
ANN-MLP with PA, RI, TI, TMX, TMN, SH	0.0354	0.8726	ANN-RBF with PA, RI, TI, TMX, TMN,SH	0.0371	0.8634
ANN-MLP without TI	0.3273	0.4684	ANN-RBF without TI	0.1542	0.6034
ANN-MLP without RI	0.2519	- 0.1311	ANN-RBF without RI	0.0997	- 0.0636
ANN-MLP without TMX	**0.4144**	**0.4672**	ANN-RBF without TMX	**0.1649**	**0.4655**
ANN-MLP without TMN	0.2856	0.6616	ANN-RBF without TMN	0.2564	- 0.1419
ANN-MLP without SH	0.1518	0.4921	ANN-RBF without SH	0.0995	- 0.0316

temperature (maximum) is the most sensitive parameter followed by technology index for Maize production than other parameters for the Upper Brahmaputra Valley Zone of Assam.

7. CONCLUSION

Reliable and timely information on crop area and crop production is of great importance to planners and policy makers. In this paper, use of Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) models, for the prediction of crop Area and Crop Production for the crops Rice and Maize of UBV Zone of Assam, was described and compared. The performance comparison showed that the ANN-MLP is the best model for both the crops i.e. Rice and Maize for prediction of crop Area and Crop Production followed by the RBF model and it can be concluded that ANN models outperform MLR model for prediction of Area under the crop and crop Production for the crops Rice and Maize. Sensitivity analysis has been performed with the best neural network configurations (MLP and FBF) to find the most influencing factor on crop Production. From the results of sensitivity analysis, technology index has been observed to be the most sensitive parameter followed by rainfall index for the Rice Production and temperature (maximum) has been found to be the most sensitive parameter followed by technology index for prediction of Maize production in UBV Zone of Assam. From these results, it can be concluded that technology index has the effective impact on the production of rice. If the farmers adopt improved technology, they may be benefitted in terms of Rice production. For the crop Maize, it may be concluded that if the sufficient amount of Temperature is not received, it may affect the production of Maize crop.

Experimental result indicates the good predictive capabilities of ANN models for prediction of crop Area and crop production in Upper Brahmaputra Valley Zone of Assam and the experiments can further be extended to other five Agro-climatic zones of Assam. Although ANN models outperform MLR model for prediction of crop Area and crop Production for Rice and Maize crops due to the problem of overfitting, that occur during neural network training, it is seen that the error on the training set is driven to a small value, but when new data is presented to the network the error is relatively large. The network has memorized the training examples, but it has not learned to generalize to new situations with the same accuracy as with the training data.

It is established from the literature that neural networks and statistical models are not competing methodologies for data analysis. There is an overlap between the two fields. Neural networks include several models, such as MLPs that are useful for statistical applications. ANN can automatically approximate any nonlinear mathematical function. This aspect of neural networks is particularly useful when the relationship between the variables is not known or is complex and hence it is difficult to handle statistically. However, the determination of various parameters associated with neural networks *viz*. the number of hidden layers, number of nodes in the hidden layer etc. is not straightforward and finding the optimal configuration of neural networks is a time consuming process. Methods for improving generalization capability of the ANN model (MLP) are set as the goal for future research.

The present study gives an indication that ANN models can serve as an alternative to statistical model under non-linear or unknown relationship between the variables for increasing accuracy in prediction of crop area and crop production. In view of the importance of reliable estimates of crop production, there is a great need for further development of suitable and reliable prediction models using information from different sources like agricultural inputs, meteorological data, soil parameters and remote sensing data to improve the accuracy of the predictive models. The accuracy of the predictive models may be improved by application of intelligent methods like fuzzy logic and genetic algorithm in addition to statistical methods being used. Further experimentation needs to be carried out with other neural network models as well as hybrid models *viz.* neuro-fuzzy and neuro-genetic models for better prediction of Area and Production for the crops rice and maize of UBV Zone of Assam.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Directorate of Agriculture; The Directorate of Economics and Statistics; Govt. of Assam, Guwahati, Directorate of Agro-economic Research, Assam Agricultural University (AAU), Jorhat for providing the crop data and National Data Centre, India Meteorological Department (IMD), Pune, India; Regional Meteorological Centre, IMD, Guwahati and the Department of Agrometeorology, AAU, Jorhat for providing Meteorological data needed for the study. The author is grateful to Dr. A.C. Bora, Professor and Head (Retd.), Department of Agricultural Statistics, AAU, Jorhat for his valuable suggestions during the study of research. The authors would like to thank the reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES

Barah, B.C., Betne, R. and Bhowmick, B.C. (2001.) Status of rice production in Assam: A research perspective. "Abridged version of a paper presented at the 4th workshop of the NATP project RRPS2 on Socio-economic dynamics of rice production system in eastern India at the Central Rice Research Institute, Cuttack, 9-11.

Baret, F., Clevera, J.G. and Steven, M.D. (1995). The robustness of canopy gap fraction estimates from red and near-infrared reflectance: a comparison of approaches. *Remote Sensing Environ.*, **54,** 141-151.

Batchelor, W.D., Yang, X.B. and Tshanz, A.T. (1997). Development of a neural network for soybean rust epidemics. *Trans. ASAE,* **40,** 247-252.

Bhattacharyya, B.K. (1999). *Statistics in Applied Sciences.* Kalyani Publisher, ISBN 81-7663-265-1.

Camago, M., Ortolani, A.A., Pedro, Jr. M.J. and Rosa, S.M. (1999). Agrometeorological model for yield prediction of orange cultivar Valencia. *Bragantia,* **58(1)**, 171-178.

Cybenko, G., (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signals Sys.,* **2,** 303-314.

Drummond, S.T., Sudduth, K.A., Joshi, A., Birrell, S.J. and Kitchen, N.R. (2002). Statistical and neural methods for site–specific yield prediction. *ASAE,* USDA, University of Maryland.

Gaudart, J., Giusiano, B. and Huiart, L. (2004). Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data. *Comput. Statist. Data Analy.*, **44,** 547-570.

Goswami, K. (2002). *Prediction of area and production of selected kharif crops based on Rainfall indices of different agro-climatic zones of Assam.* M.Sc. (Agricultural Statistics) Thesis, Assam Agricultural University, Jorhat, Assam.

Haskett, J., Pachepski, Y. and Acock, B. (1998). Use of Artificial Neural Networks to simulate soybean crop yield as affected by global change. *Tektran,* United States Department of Agriculture, Agriculture Research Service.

Hassoun, M.H. (2002). *Fundamentals of Artificial Neural Networks*. Prentice Hall of India.

Haykins, S. (1994). *Neural Networks: A Comprehensive Foundation*. 2nd ed., Pearson, Prentice Hall.

Hertz, J., Krogh, A., and Palmer, R.G. (1991), Introduction to the Theory of Neural Computation. Santa Fe Institute Studies in the Sciences of Complexity (Vol. 1), Addison-Wesley, Redwood City, CA.

Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks,* **2**, 359-366

Hornik, Kurt. (1991). Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks*, **4(2)**, 251-257.

Huang, Y. (2009). Advances in artificial neural networks – methodological development and application. *Algorithms*, **2**, 973-1007; doi:10.3390/algor2030973.

Jackson, D.I. and Looney, N.E. (1999). *Temperate and Subtropical Fruit Production*. Second edition. *CAB International*, Wallingford, Oxon, UK, 332.

Jiang, D., Yang, X., Clinton, N. and Wang, N. (2004). An artificial neural network model for estimating crop yield using remotely sensed information. *Int. J. Remote Sensing,* **25(1)**, 1723-1732.

Kim, E.H., Hyun, K.H. and Kwak, Y.K. (2005). Optimal design of radial basis function using taguchi method. *IEEE Xplore*, 171-177.

Kumar, S. (2004). *Neural Networks: A Classroom Approach.* Prentice Hall of India.

Lawrence, J. (1994). *Introduction to Neural Networks.* California Scientific Software Press, Nevada City, CA.

Nguyen, D. and WIdrow, B. ((1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Proceedings of the International Joint Conference on Neural Networks,* **3**, 21-26.

Paswan, R.P. and Begum, S.A. (2013). Regression and neural networks models for prediction of crop production. *Intt. J. Sci. Engg. Res.*, Volume 4, Issue 9, September 2013, ISSN 2229-5518.

Sarle, W.S. (1994). Neural networks and statistical models. *Proceedings of the Nineteenth Annual SAS Users Group International Conference* pp. 1538-1550. SAS Institute.

Shearer, S.A., Burks, T.F., Fulton, S.F., Thomasson, J.A. and Samson, S. (1999). Yield prediction using a Neural Network classifier trained using soil landscape features and soil fertility data. *ASAE*. Paper No. 993041. St. Joseph, Michigan.

Simpson, G. (1994). Crop yield prediction using a CMAC neural network. *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, 2315, 160-171.

Singh, R.K. and Prajneshu. (2008). Artificial neural network methodology for modelling and forecasting maize crop yield. *Agric. Econ. Res. Rev.* **21**, 5-10.

Starrett, S.K. and Adams, G.L. (1997). Using artificial neural networks and regression to predict percentage of applied nitrogen leached under turf grass. *Comm. Soil Sci. Plant Analy.,* **28,** 497-507.

Sudheer, K.P., Gosain, A.K.D., Rangan, M. and Saheb, S.M. (2002). Modeling evaporation using an artificial neural network. *Hydrological Process*, **16,** 3189-3202.

Sun, T., Li, J.R. and Pan, S.B. (2004). Application study of Artificial Neural Network model in soybean crop yield as affected by global change. *Tektran*, United States Department of Agriculture, Agriculture Research Service.

Warner, B. and Misra, M. (1996). Understanding neural networks as statistical tools. *Amer. Statist.,* **50**, 284-293.