Journal of Agrometeorology 19 (3) : 265-269 (September 2017)

Forecasting models for predicting pod damage of pigeonpea in Varanasi region

PRITY KUMARI[1]*, G.C.MISHRA[1] and C.P. SRIVASTAVA[2]

[1]Section of Agricultural Statistics, Departmentof Farm Engineering

[2] Departmentof Entomology and Agricultural Zoology

Institute of Agricultural Sciences, Banaras Hindu University, Varanasi-221005, India

*Email: psingh2506@aau.in

ABSTRACT

Present investigation considers comparison of time series statistical models like autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) with explanatory multiple linear regression model for predicting per cent pod damage in pigeonpea by pod borer for Varanasi region of Uttar Pradesh using 27 years of data (1985-86 to 2011-12).The evaluation of best suited model was assessed by root mean squared error (RMSE). Based on empirical studies, ANN was found to be best suited model with lowest RMSE having forecasted per cent pod damage in pigeonpea by pod borer during the year 2012-13 for Varanasi region.

Keywords: ANN ARIMA model, multiple regression, pigeonpea pod borer.

Pigeonpea (*Cajanuscajan*) is one of the important pulse crop and ranks sixth among pulses production in the world. Nutritionally, proteins and starch are the major constituents of pigeonpea and also a good source of dietary fiber, many vitamins and minerals.India is the largest producer of pigeonpea contributing 75 per cent of world production. Despite being the largest producer, it is also top most importer of this legumein the world, because of its rising demand within country. It is therefore become essential to reduce the gap between production and consumption. One of the major constraints for this gap is considerable damage in pods due to attack of major insect pests directly affecting the loss of yield. Pod borer (*Helicoverpa armigera*) is a key pest inflicting 80 to 90 per cent of loss (Kooner *et al*. 2006). Therefore, prevention against such losses needs an important consideration, so that timely control measures will be taken for future planning. The multiple regression models, autoregressive integrated moving average (ARIMA) model and artificial neural network (ANN) architecture have been widely used for forecasting yield as well as pests of different crops (Agarawal and Mehta, 2007;Kumari, *et. al*.,2013; Kumari, *et. al*., 2014; and Kumari, *et. al*., 2016).

Multiple linear regressions (MLR) are explanatory model and more suitable to short term or intermediate term forecasting (Varmola *et. al*.,2004;Chauhan *et. al*., 2009). ARIMA model (Box and Jenkins, 1970) is a forecasting technique that projects the future values of a series based entirely on its own inertia. However, ARIMA models work best when data exhibits a stable or consistent pattern over time with a minimum amount of outliers (Gorantiwar *et. al*.,2011; Kumar *et. al*.,2013).On the other side, ANN which is non-parametric model may be preferred over traditional parametric statistical models in the situations where input data do not meet the assumptions required by the parametric model.Neural Networks has capability to generalize the underlying pattern within a time series even when the underlyingsystem is too complex to describe (Mishra and Singh, 2013; Meena *et. al*.,2016).

In the present study, all the three approaches have been used and their performance is compared to forecast the per cent pod damage by pod borer in pigeonpea for Varanasi region of Uttar Pradesh.

MATERIALS AND METHODS

The time series secondary data on per cent pod damage by pod borer in pigeonpea for the period 1985-86 to 2011-12 were collected from All India Coordinated Research Project on Pigeonpea (Indian Council of Agricultural Research) and corresponding weekly weather data were collected from All India Coordinated Research Project on Dry Land Agriculture, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi. Five main weather variables maximum temperature (X_1), minimum temperature (X_2), rainfall (X_3), maximum relative humidity (X_4) and minimum relative humidity (X_5) were considered for building regression model.

Table 1: Estimate of the predictors of multiple regression model

Model		Coefficients		T	Sig.	R^2	Adjusted
		B	Std. Error				R^2
1	(Constant)	9.30	.47	19.66	.000	0.52	0.49
	Z351	.002	.000	4.88	.000		
2	(Constant)	7.59	.74	10.20	.000	0.65	0.62
	Z351	.002	.000	5.19	.000		
	T	.14	.05	2.77	.011		
3	(Constant)	18.34	3.80	4.82	.000	0.75	0.71
	Z351	.001	.02	4.34	.000		
	T	.16	.04	3.73	.001		
	Z141	.002	.01	2.87	.009		

Table 2: ARIMA model parameters and fit statistics value

Model	Parameter	Estimate	SE	T	Sig.	R^2	RMSE
ARIMA(0,0,1)	Constant	-560.48	162.33	-3.45	.002	.65	2.16
	MA (Lag 1)	-.59	.16	-3.53	.002		

Multiple linear regressions (MLR) model

The weekly weather data from July 1(25[th]SMW) to March 15(11[th]SMW) in each year from 1985-86 to 2011-12 were utilized for development of multiple regression models. Out of 27 years data, 24 years data were utilized for development of regression model and 3 years data were used to validate the forecasting ability of developed model. Agarwal and Mehta (2007) model was followed as given below:

$$Y = A_0 + \sum_{i=1}^{p}\sum_{j=0}^{1} a_{i,j} Z_{i,j} + \sum_{i \neq i'=1}^{p}\sum_{j=0}^{1} a_{i.i',j} Z_{i,i',j} + cT + e$$

Where, $Z_{i,j}$, $Z_{ii',j}$: weather indices ; i,i': 1, 2, …p; Y: Dependent variable; T: Year number; A_0: Intercept; p: Number of weather variables under study. 'e' *e*rror term, is normally distributed with mean zero and constant variance. Stepwise regression technique was used to select the important weather indices.

Autoregressive integrated moving average (ARIMA) model

ARIMA model analyzes and forecasts equally spaced univariate time series data, as a linear combination of its own past values, past errors (also called shocks). ARIMA model is defined as ARIMA (p, d, q) and is expressed in the form:

$$Y_t = \square_0 + \square_1 Y_{t-1} + \square_2 Y_{t-2} + \ldots + \square_p Y_{t-p} + e_t - \square_1 e_{t-1} - \square_2 e_{t-2} - \ldots - \square e_{t-q},$$

Where Y_t and e_t are the actual values and random error with mean zero and the constant variance $\square_e^2$ at time t, respectively, $\square_i$(i = 1,2,……,p) and $\square_j$(j = 1,2,……,q) are model parameters, p and q are referred to as orders of autoregressive and moving average polynomials respectively (Box and Jenkins, 1970).

During construction of best ARIMA model order of autoregressive (*p*), differencing (d) and moving average (*q*) parameters have to be effectively determined. The model having relatively small root mean squared error (RMSE), relatively high R^2 and adjusted R^2 was considered to be the best amongst all.

Artificial neural network (ANN)

For constructing ANN architechture, data is divided into three non-overlapping sets which are training, *validation* and *testing set*. The training set, consisting major portion of data, is used to teach the network in order to get the desired target function. Validation set is used to decide when to stop training process. The testing data set, which exposed to the unseen data, is used to measure performance of trained network by mean square error (MSE) or root mean square error (RMSE). In present study, neural network architectures were developed by using Levenberg Marquardt (LM) algorithm (Ranganathan, 2004;) a training algorithm of weight matrix.

RESULTS AND DISCUSSIONS

Multiple linear regression (MLR) model

The stepwise multiple linear regression analysis results (Table 1) showed that all the generated variables entered in three different models affected significantly pod

Table 3: ANN model parameters

Weights	H_1	H_2	Biases	Values
I_1	WI_1H_1= 1.20	WI_1H_2= -2.50	BH_1	-0.42
I_2	WI_2H_1= 1.47	WI_2H_2= -1.64	BH_2	0.85
O	WOH_1= -0.94	WOH_2= 0.88	B_O	-0.80

Note:

- I_i(i=1,2), H_j(j=1,2) and Oare two input nodes, two hidden nodes and one output node respectively.
- WI_1H_1, WI_2H_1, WI_1H_2&WI_2H_2 are weights among input & hidden neurons
- WOH_1&WOH_2 are weights among hidden & output neurons
- BH_1, BH_2, and $B_O$arebias values of two hidden nodes and one output node

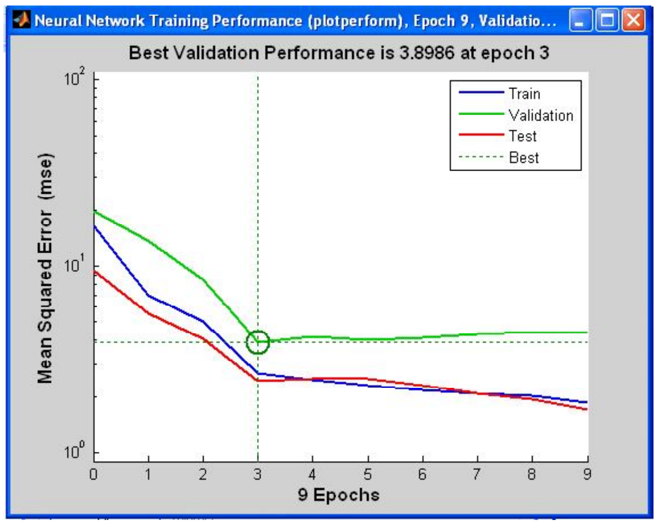

Fig. 1: Performance of LM algorithm

damage but Model 3 was considered better than the remaining two models because of greater value R^2/adjusted R^2 value. Model 3 was explained by the variables *viz.* constant, Z351, Z141 and T. The constituent of each of these generated variables is as follows:

$$Z_{3,5,1} = \sum_{w=1}^{37} r_{35w} X_{3w} X_{5w} \qquad Z_{1,4,1} = \sum_{w=1}^{37} r_{14w} X_{1w} X_{4w}$$

Where,

r_{35w}= Correlation coefficient between percent pod damage by pod borer (Y) and product of 3rd and 5th weather parameter (viz. total rainfall (X_3) and minimum relative humidity (X_5))

r_{14w} = Correlation coefficient between percent pod damage by pod borer (Y) and product of 1st and 4th weather parameter (viz. maximum temperature (X_1) and maximum relative humidity (X_4) respectively)

The estimates of the constants and independent variables entered in the Model 3, were 18.34, 0.001, 0.16 and

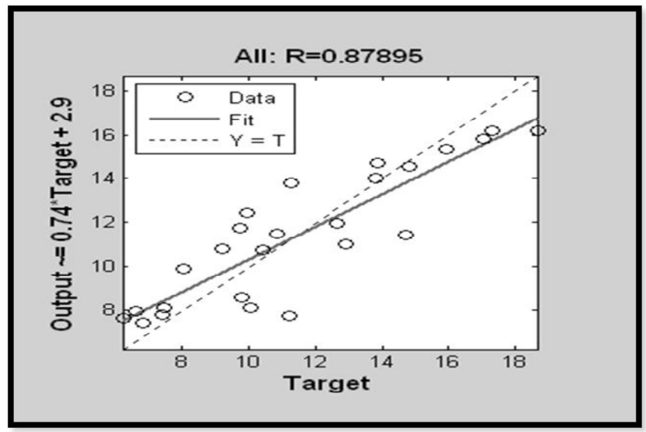

Fig. 2: Regression analysis of LM algorithm

0.002 with standard error of 3.80, 0.02, 0.04 and 0.01 respectively. Also they are statistically highly significant (Table 1). Since the models were developed only on the basis of 24 years data while three years data were taken as holdout in order to check the forecasting ability of the models by Mean squared error (MSE). MSE of the Model 3 was calculated on the basis of three years data which were used to explain the error in the forecasting model and its value for Model 3 is17.25 while R^2&adjusted R^2were 0.75 &0.71 respectively (Table 1). The forecasted value of percent pod damage by pod borer of pigeonpea during the year 2012-13 was obtained as 12.45 %.

ARIMA model

Out of various ARIMA models with different value of p, d and q, ARIMA (0,0,1) model was found to be the best. Table 2 represents value of parameters and model fit statistics.The parameters of ARIMA (0,0,1) *i.e.* constant term and moving average (MA) term at lag 1 was found to be statistically significant with an estimate of -560.48 and -0.59 respectively. $R^2$and MSE (RMSE) of this model were 0.65 and 4.65 (2.16) respectively. The forecasted value of per cent pod damage by pod borer of pigeonpea during the year

Table 4: Comparative performance of different models

Model accuracy and forecasted value	ANN	ARIMA	MLR
Forecast value (%)	16.38	15.70	12.45
RMSE	1.97	2.16	4.15
MSE	3.89	4.65	17.25
R square	0.77	0.65	0.75

2012-13 was 15.7 per cent. At the diagnostic checking stage residual were examined and their autocorrelation coefficients were found to be non significant which shows that the model fit was satisfactory.

Artificial neural network architecture

Neural network architecture was developed by using time series data of per cent pod damage by pod borer of pigeonpea where lag values are taken as independent variable and MATLAB neural network toolbox 2010 was used to develop these architectures. The network used was a two-layer feed-forward network.

Neural network architecture has following topology: a) two-layer feed-forward network (one input & one hidden layer), b) Input layer having two lag value of time series data as inputs, c) Hidden layer having two node with sigmoid activation function and d) Output layer having one node with linear activation function.Therefore, four weights for input to hidden neurons and two weights for hidden to output neurons and three bias values were chosen. For training 70 per cent, for each of validation and testing 15 per cent data were used by using random data division process.

The performance of the proposed network when trained with Levenberg-Marquardt (LM) algorithm was accessed by their mean squared error (MSE) value along with multiple correlation coefficient (R) between observed and predicted outputs. Here parameters of ANN model *i.e.* weights among different nodes and biases value of each node were mentioned in the Tables 3.From Fig. 1, it is observed that the best validation performance MSE (3.89) or RMSE (±1.97) was obtained at epoch 3. The regression analysis plot (Fig. 2) displayed a linear regression between network outputs and the corresponding targets with the R value as 0.88 (R^2 = 0.77) showing the fit was good for all data sets. The forecasted value of percent pod damage by pod borer of pigeonpea during the year 2012-13 was obtained as 16.38 per cent by this model.

Comparison of ANN, ARIMA and MLR model

Table 4 indicates that the forecasted value of per cent pod damage by pod borer of pigeonpea was best explained by ANN model during 2012-13 having relatively small value of RMSE (±1.97) and relatively high value of R^2 (0.77).

CONCLUSION

ANN was found to be more appropriate for forecasting percent pod damage by pod borer (16.38%) of pigeonpea in comparison to multiple linear regression (MLR), autoregressive integrated moving average (ARIMA). ANN therefore, can be recommended as appropriate forecasting model for the problem under study and will be helpful for farmers and policy makers for future planning in advance.

ACKNOWLEDGEMENT

The authors are thankful to the All India Coordinated Research Project on Pigeonpea (Indian Council of Agricultural Research) and All India Coordinated Research Project on Dry Land Agriculture, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, for providing data to carry out the present study.

REFERENCES

Agarawal, R. and Mehta, S.C. (2007). Weather based forecasting of crop yields, pest and diseases – IASRI Models. *J. Ind. Soc. Agril. Stat.*, 61 (2): 255-263.

Box G.E.P. and Jenkins G. (1970). Time series analysis, Forecasting and control, Holden-Day, San Francisco, CA.

Chauhan V.S., Shekh A.M., Dixit S.K., Mishra A.P. and Kumar Sanjay (2009), Yield prediction model of rice in Bulsar district of Gujarat, *J.Agrometeorol.*, 11 (2): 162-168.

Gorantiwar S.D., Meshram D.T. and Mittal H. K. (2011). Seasonal ARIMA model for generation and forecasting evapotranspirtion of Solapur district of Maharashtra, *J.Agrometeorol.,* 13 (2): 119-122.

Kooner, Bant Singh and Cheema, Harpreet Kaur (2006). Evaluation of pigeonpea genotypes for resistance to pod borer complex. *Ind. J. Crop Sci.*, 1: 194-196.

Kumar J. Ashok, Muralidhar M., Jayanthi M. and Kumaran M. (2013). Trend analysis of weather data in shrimp farming areas of Nagapattinam district of Tamil Nadu, *J.Agrometeorol.,* 15 (2): 129-134.

Kumari Prity, Mishra G.C. and Srivastava C.P. (2013). Forecasting of Productivity and Pod Damage by *Helicoverpaarmigera* using Artificial Neural Network

Model in Pigeonpea (*Cajanuscajan), Int. J. Agri., Env. & Biotech.*, 6(2): 187-193.

KumariPrity, Mishra G.C. and Srivastava C.P. (2014). Time series forecasting of losses due to pod borer, pod fly and productivity of pigeonpea (*CajanasCajan*) for North West Plain Zone (NWPZ) by Artificial Neurat Network (ANN), *Int. J. Agric. Stat. Sci.*, 10 (1): 15-21.

KumariPrity, Mishra G.C. and Srivastava C.P. (2016). Statistical models for forecasting pigeonpea yield in Varanasi region, *J.Agrometeorol.*, 18 (2): 306-310.

Meena P. K., Khare Deepak and Nema M. K. (2016). Constructing the downscale precipitation using ANN model over the Kshipra river basin, Madhya Pradesh, *J.Agrometeorol.*, 18(1): 113-119.

Mishra, G. C. and Singh, A. (2013). A study on forecasting price of groundnut oil in Delhi by ARIMA Methodology and Artificial Neural Network, AGRIS On-line Papers in Economics and Informatics, 5(3): 25-34.

RanganathanAnanth (2004). The Levenberg-Marquardt Algorithm [web page] http://www.ananth.in/Notes_files/lmtut.pdf.

Varmola S. L., Dixit S. K., Patel J. S. and Bhatt H. M. (2004). Forecasting of wheat yields on the basis of weather variables, *J.Agrometeorol.*, 6 (2): 223-228.

Received : April 2017 ; Accepted : July 2017