Multi-layer perceptron based neural network model predicting maximum severity of *Spodoptera litura* **(Fabricius) on groundnut in relation to climate for Dharwad region of Karnataka (India)**

GIRISH K. JHA, GAJAB SINGH, S. VENNILA*, M. HEGDE**, M. S. RAO*** and H. PANWAR*

ICAR - Indian Agricultural Research Institute, New Delhi – 110 012, India

**ICAR - National Research Centre for Integrated Pest Management, New Delhi – 110 012, India*

***University of Agricultural Sciences, Dharwad – 580 005, Karnataka, India*

****ICAR - Central Research Institute for Dryland Agriculture (CRIDA), Hyderabad, India*

(*Received 24 February 2015, Accepted 27 February 2017*)

e mail : girish.stat@gmail.com

सार – धारवाड़ (कर्नाटक) में खरीफ ऋतु (26 - 44 मानक मौसम विज्ञानिक सप्ताह (SMW) वर्ष 1920-2013 मृंगफली की कृषि प्रणाली में स्पोडोप्टेरा लिटुरा (फैब्रीशियस), कीट भक्षक की संख्या के पूर्वानुमान के लिए, एक मल्टी लेयर परसेप्ट्रोन (MLP) तंत्रिका नेटवर्क तकनीक विकसित कर उसका परीक्षण किया गया। कीट की अधिकतम तीव्रता चौंतीसवें मानक मौसम विज्ञान सप्ताह के दौरान पायी गई। अधिकतम और न्यूनतम तापमान (°सें), वर्षा (मि. मी.) तथा सुबह और दोपहर की सापेक्ष आर्द्रता (%) जैसे मौसम प्राचलों का उपयोग करके 34 SMW में अधिकतम प्रचंडता के दौरान प्रेक्षित एस लिट्रा के साप्ताहिक नर कीट प्रेक्षित किए गए। कीट की अधिकतम तीव्रता से दो सप्ताह पूर्व (32 व 33) SMW) के मौसम प्राचलो, अधिकतम तथा न्यूनतम तापमान, वर्षा, प्रात:कालीन और सायकालीन सापेक्ष आद्रेत आँकड़ों का प्रयोग कर एस. लिट्रा कीट भक्षक की साप्ताहिक अधिकतम तीव्रता के अन्**मान के लिए मॉडल विकसित** किया गया। इस कीट के लिए पूर्वानुमान मॉडल विकसित करने के क्रम में दो सप्ताह पूर्वे सांख्यिकीय पद्धति "सिद्धांत घटक विश्लेषण" का उपयोग किया गया। इस पद्धति से पाँच घटक स्कोरों को प्राप्त किया गया जो कि सम्पूर्ण मौसम के आँकड़ों में 90 प्रतिशत परिवर्तनशीलता को दर्शाते हैं। सिद्धांत घटक विश्लेषण से प्राप्त इन घटक स्कोरों को समाश्रयण के रूप में लेते हुए मल्टी लेयर परसेप्ट्रोन-तंत्रिका नेटवर्क (MLP) तकनीक से कीट के पूर्वानुमान का पता लगाने का प्रयास किया गया। इस प्रक्रिया मे पाया गया कि एक हिंडिन लेयर एवम ग्यारह हिंडिन नोड्स से संयोजित मल्टी लेयर परसेप्ट्रोन-तंत्रिका नेटवर्क एस. लिट्रा कीट की संख्या में गतिशीलता के पूर्वान्**मान** के लिए सबसे उपयुक्त था। वर्ष (1990-2009) तक के मौसम तथा कीट आँकड़ों को मॉडल विकसित करने के प्रयोग में लाया गया तथा शेष चार वर्षों (2010-2013) के आँकड़ों को मॉडल के परीक्षण और पुष्टि के लिए रखा गया। वगे माध्य मूल त्र्टि (RMSE) तथा निरपेक्ष माध्य प्रतिशत त्र्टि (MAPE) के दवारा विकसित मॉडल का मूल्यांकन किया गया। सत्यापित परिणाम स्पष्ट रूप से दर्शाते हैं कि MLP - तंत्रिका नेटवर्क आधारित मॉडल, मूंगफली की खेती में एस. लिट्रा कीट के यादृच्छिक एव गतिशील व्यवहार के पूर्वानुमान में प्रभावी है।

ABSTRACT. A multi-layer perceptron (MLP) neural network model for predicting adult moth population of tobacco caterpillar (*Spodoptera litura* (Fabricius) in groundnut cropping system of Dharwad (Karnataka) was developed and tested using the long term (24 years : 1990-2013) trap catches of the pest and weather data of *Kharif* season [26 to 44 standard meteorological weeks (SMW)]. The weekly male moth catches of *S. litura* during maximum severity observed at 34 SMW was modelled using the weather parameters *viz.,* maximum temperature (°C), minimum temperature (°C), rainfall (mm) and morning and afternoon relative humidity (%) lagged by two weeks. The principle component analysis was performed using meteorological data of preceding two weeks (32 and 33 SMW) in order to create fewer linearly independent factors. Five principal component scores which together accounted for 90 per cent of variations in data were used as input variables for neural network model. A MLP neural network with five input nodes and one hidden layer consisting of eleven hidden nodes was found to be suitable in terms of adequacy measures for modelling the population dynamics of *S. litura*. While data sets of 1990-2009 were used for developing the model, data of four seasons (2010- 2013) were used for testing and validation. The performance of the model was assessed in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE). The validation results clearly showed that the neural network based model is effective in dealing with the apparently random behaviour of the *S. litura* dynamics on groundnut.

Key words – *Spodoptera litura*, Principal component analysis, Neural network, Groundnut.

1. Introduction

Groundnut also known as peanut (*Arachis hypogaea*) grown throughout the world is considered as one of the important oilseed crops with its economic and nutritional importance known worldwide. India is the second largest groundnut producing country (7.04 mt) after China (16.66 mt) with its cultivation throughout the year in three seasons *viz., Kharif, Rabi* and *Summer* contributing to one fifth of the total oilseed production in the country (Govt. of India, MoA, 2013 and FAOSTAT, 2013). Fifteen per cent out of Rs 8, 63, 884 million losses per annum due to insect pests in Indian agriculture is accounted for groundnut crop (Dhaliwal *et al*., 2010). Among the insect pests of groundnut, the polyphagous tobacco caterpillar (*Spodoptera litura*, Fabricius) attacks groundnut crop throughout the world and has shown increasing incidence in the recent times with its outbreaks during 2005, 2008 and 2009 in India (Prasad *et al*., 2013). Both gregarious and solitary feeding by early and late *S. litura* larval stages reduce the leaf area for photosynthesis which in turn limits the yield potential of the crop. *S. litura* in cropping systems is commonly monitored by deploying pheromone traps @ 2-5/ha that gives information on initiation and population levels during the cropping season (Manual of Groundnut Pest Surveillance, 2011). Population levels of insects monitored regularly over a long period across many seasons aid in development of forecast models. It is well known that the crop and its pests are influenced by the prevailing weather during the cropping season in a given location. Weather is a key driver determining insectpest abundance and thus the damage to the crop. Rao *et al.* (1989) have discussed the thresholds and thermal requirement for the development of tobacco caterpillar that would aid in forecast of future population. Several researchers have done modelling using linear as well as non-linear relationships in forewarning pests of agriculture crops (Agrawal *et al.*, 1986; Chattopadhyay *et al*., 2005; Samui *et al*., 2005 & 2007).

Recently, artificial neural networks are used in the area of classification, modelling and prediction (Huang *et al*., 2010), where regression and other related statistical models have traditionally been employed. Dewolf and Francl (1997) demonstrated the applicability of neural network technology for plant disease forecasting. Jha and Sinha (2014) briefly explained the working nature of neural network modelling for forecasting purpose. Yang *et al.* (2009) have applied principal components of weather variables as input for the back propagation artificial neural network (BP-ANN) methods to find out a non-linear relation between the pest occurrence and the meteorological factors.

Current study attempted to model the population dynamics of *S. litura* moths caught in pheromone traps with weather parameters of the location using principal component based multiple linear regression and neural network models.

2. Materials and method

2.1. *Study location and data accrual*

The groundnut cropping system of Dharwad (15 \degree 2' N to 74 \degree 57' E) located in the State of Karnataka was selected for the study. Data on *S. litura* moth catches (nos/trap/week) in pheromone traps, and weather data *viz.,* maximum temperature (MaxT, °C), minimum temperature (MinT, °C), morning and afternoon relative humidity (RHI & RHII,%) and rainfall (RF, mm) for *Kharif* season spread across nineteen weeks between 26 and 44 SMW for 24 years (1990 to 2013) recorded at the University of Agricultural Sciences, Krishinagar, Dharwad were obtained.

2.2. *Model development*

As a first step towards model development, all the weather and pest data were first normalised. Normalisation of each parameter was carried out by subtracting its mean value and dividing by its standard deviation to have zero mean value and unity variance for all variables. Secondly, the mean and the maximum *S. litura* moth catches in respect of standard meteorological weeks (SMW) over the years were examined to determine the peak period of infestation *viz.,* the time of maximum severity during *Kharif* season. The Bartlett's sphericity test was applied to ensure the applicability of principal component analysis to the original data. Subsequently, the population occurrence of *S. litura* during its maximum severity was modelled using principal component regression and principal component based neural network models with the help of collected data pertaining to trap catches of *S. litura* and weather parameters of *Kharif* season. Correlation between the *S. litura* at its maximum severity (34 SMW) and five weather parameters for the corresponding (34 SMW), one lagged (33 SMW) and two lagged (32 SMW) weeks was examined considering the cases from 24 seasons.

2.3. *Principal component (PC) analysis*

Since the pest population of the current week depends on the weather factors of preceding weeks, the principal component (PC) analysis was performed on 10 weather parameters (five for each 32 and 33 SMW) to obtain principal component scores base on correlation matrix (Montgomery *et al*., 2003) in order to model the

Fig. 1. A typical ANN structure with a single hidden layer

moth catch of *S. litura* during 34 SMW. The principal components with eigen values higher than one were utilised in the regression model. The PCs were obtained by varimax rotation for better interpretation of factor loading.

2.4. *Artificial neural network (ANN)*

The general form of the ANN output produced by a network consisting of *p* input nodes, *q* hidden nodes and one output node can be written as

$$
y_{i} = g \left\{ \alpha_{0} + \sum_{k=1}^{q} \alpha_{k} f \left(\beta_{0k} + \sum_{j=1}^{p} \beta_{jk} x_{j} \right) \right\}
$$

$$
i = 1, 2, ..., n
$$
 (1)

where, y_i denotes the pest population in i^{th} year, f and g denote the activation function at hidden and output layer respectively. β_{ik} is the weight attached to the connection between jth input node and the kth node of hidden layer, β_{0k} is the bias at the input layer, α_k is the weight attached to the connection from k^{th} hidden node to the output node and α_0 is the bias at the hidden layer. Since the universal approximation theorem suggests that a neural network with a single hidden layer and sufficiently large number of neurons can well approximate any nonlinear function (Haykin, 1994), neural network model with single hidden layer was designed for this study. Fig. 1 shows a typical MLP neural network model. For the estimation of neural network models, the data was divided into training and testing sets. The training test is used for model development and the testing is employed for evaluating the prediction ability as well as validation of the model in view of small data set in the present study.

Fig. 2. Maximum severity of *S. litura* with its corresponding SMW of occurrence

Here, four out of 24 points were kept for testing and validation. While the training and selection of model was performed using data pertaining to the years between 1990 and 2009, testing and validation was done with the real observations of the years from 2010 to 2013.The network was trained using Levenberg-Marquardt algorithm considering its fast convergence and high efficiency (Haykin, 1994). The training process was implemented using STATISTICA Neural Network Version 6.1.

2.5. *Evaluation and validation of developed models*

The forecasting performance of both the fitted models namely PCR and MLP-ANN were evaluated on the basis of coefficient of determination (R^2) , root mean square error (RMSE) and mean absolute percentage error (MAPE) (Jha and Sinha, 2014). The PCR and MLP-ANN models were tested for their performance by matching the predicted with the observed maximum severity of *S. litura* over the four seasons of 2010-2013.

3. Results and discussion

3.1. *Temporal distribution of S. litura*

Fig. 2 presents the maximum severity pertaining to *S. litura* moth catches (nos/trap/week) corresponding to the individual *Kharif* seasons (26 SMW (second last week of June) to 44 SMW (last week of October)) over the 24 years (1990-2013). The SMW of maximum severity in respect of individual seasons is also furnished in Fig. 2. The levels of maximum severity of *S. litura* (nos/trap/week) ranged from 34 during 36 SMW of 2001 to 1376 during 34 SMW of 2009 indicating greater seasonal differences. The vast difference in abundance is possibly associated with the size of the initial population modulated by the biotic and abiotic factors prior to *Kharif* season, and of the variables of crop/cropping system and natural enemies in addition to variations in weather on the *S. litura* stages in a given season. Period of maximum trap catches in respect of individual seasons indicated highest frequency of occurrence during 34 SMW (six times)

TABLE 1

Correlation between *S. litura* **(nos/ trap/week) and weather variables**

*: significant at $p \le 0.05$ & **: significant at $p \le 0.01$

TABLE 2

First five principal component loadings for weather variables*

SMW	Variable	PC1	PC ₂	PC ₃	PC ₄	PC ₅
32	MaxT	-0.82	-0.02	-0.22	0.07	-0.35
	MinT	-0.15	0.87	-0.21	0.06	-0.04
	RHI	0.86	-0.13	-0.16	0.39	0.17
	RHII	0.87	-0.09	0.42	-0.07	0.01
	RF	0.44	-0.19	0.07	0.02	0.84
33	MaxT	-0.15	0.32	-0.86	-0.16	-0.19
	MinT	-0.04	0.86	-0.14	-0.04	-0.09
	RHI	0.10	0.01	0.32	0.93	0.00
	RHII	0.20	-0.23	0.86	0.32	-0.04
	RF	-0.13	-0.62	0.33	0.47	0.39
% variance		42.9	19.86	12.92	8.72	5.52

*** : *The values marked in bold indicate the high correlation between the weather variables and the corresponding PCs*

followed by 36 SMW (five times) and 35 SMW (two times). Although the period of maximum severity showed the shifting peaks between 27 SMW in 1996 and 40 SMW in 2013, the long term averages over growing seasons for each of the SMWs indicated a continuously increasing population between 30 and 36 SMWs and the peak at 34 SMW.

3.2. *Influence of weather on S. litura*

Location of Dharwad showed ranges of 26.4-29.9 °C, 18.1-21.1 °C, 82.9-92.3%, 58.8 to 81.4% and 6.5-40.9 mm of maximum temperature (MaxT), minimum temperature (MinT), morning relative humidity (RHI), afternoon relative humidity (RHII) and weekly total rainfall (RF), respectively during the *Kharif* cropping seasons over years with mean *S. litura* population (nos/trap/week) as low as 6 in 44 SMW to 345 in 34 SMW (Fig. 3).

Fig. 3. Seasonal mean weather variables *vis a vis* mean and maximum population of *S. litura* over 24 years (1990-2013)

Significant positive and negative impact of MaxT and RHII of the lagged weeks (32 and 33 SMW), respectively were seen through correlation analysis on *S. litura* of 34 SMW (Table 1), the period of maximum severity due to the influence of these weather variables on the pupal stage of *S. litura*, non-feeding spent in the soil before emergence. The pupal duration of 6-12 days (Prasad *et al.*, 2013) coincide with lagged weeks of weather that had influenced the emergence of moths and hence the maximum severity of *S. litura*. Early studies have also shown the significance of weather variables of lagged weeks on the development of pests of rice (Samui *et al.*, 2005 & 2007) at Pattambi, Kerala.

3.3. *Models predicting maximum severity of S. litura*

3.3.1. *Principal component regression (PCR) model*

Principal component (PC) analysis is concerned with explaining the variance covariance structure through a few linear combinations of the original variables, and is one of the most effective multivariate techniques for dimensionality reduction which constitute an important step for statistical pattern recognition. Principal component regression also avoids the problem of multicollinearity associated with the present data arising out of the correlated weather variables. A calculated value of chi-square as 140.87 (df. 45, $p<0.01$) through Bartlett's sphericity test implied the suitability of PC analysis to the present data. The PC analysis was done using the ten weather variables (five weather variables for each of 32 and 33 SMW) for feature extraction of the data. Sensitivity of weather variables was evaluated in terms of principal component loadings, which are nothing but the correlations among principal component scores and the meteorological factors. All ten weather variables (five each of 32 and 33 SMW) were included in the five selected PCs. Table 2 summarises the results of the PC

TABLE 3

Error statistics for prediction models

analysis with the amount of variance explained by each PC. Only a few variables showed the high loading within each PC.

PC1 accounted for 43 per cent of variations in the input data and had loadings more than 0.8 with a combination of MaxT, RHI and RHII of 32 SMW. While the PC2 accounted for 20 percent of total variation by extracting the information from MinT of 32 and 33 SMWs and RF of 33 SMW. PC3 depended on MaxT and RHII of 33 SMW. Fourth and fifth principal components exhibited higher loadings with RHI and RF of 33 and 32 SMW, respectively. Five principal components with eigen values higher than one, together explained 90 per cent variation in the data. The principal component score values were obtained by multiplying eigen vectors with the standardized weather variables in a linear fashion and the scores were used as the predictors of adult moth population of *S. litura* at 34 SMW. The stepwise regression procedure identified the first, third and fourth principal component scores to have the significant linear relationship towards the prediction of maximum severity of *S. litura*. The variance inflation factor for the PCR model indicated the absence of collinearity problem. The regression equation based on principal components cores is given as :

 $\hat{Y}_{PC} = -0.153 \, PCS_1 - 0.489 PCS_3 + 0.326 PCS_4 \tag{2}$

where, \hat{Y}_{PC} is the predicted moth population of *S*. *litura*, PCS_1 , PCS_3 and PCS_4 are the first, third and fourth principal component scores respectively. The PCR model could explain only 50 per cent of variations (Table 3) in *S. litura* peak severity at 34 SMW due to its linear structure.

Fig. 4. Observed and predicted maximum severity of *S. litura* during 34 SMW using MLP model

3.3.2. *Multi-layered perceptron (MLP) based artificial neural network (ANN)*

Considering the non-linear distribution pattern of adult moth population, artificial neural network model with one hidden layer and one output node was used to predict the peak occurrence of *S. litura*. In this study we varied the input nodes between one to five principal component scores of weather variables. The number of hidden nodes was varied from 2 to 14 with basic crossvalidation method. The resulting neural network model performed relatively better with six to eleven hidden nodes. In general, more hidden nodes in the model suggest complex interactions among the weather variables. A total of 65 models with different number of nodes in the input and hidden layers were tested for the data. Table 4 shows the goodness of fit for the prediction models developed through PCR and MLP. A model with smallest RMSE and MAPE and the largest R^2 is considered to be the best.

A MLP neural network with five input nodes and eleven nodes in hidden layer was found to be the most accurate for modelling the *S. litura* maximum severity. Principal component based MLP network performs better in comparison to principal component regression. The principal component regression models could capture only linear relation while neural network based model could explain non-linear relationship between weather variables and the pest population occurrence and thus better suited for predicting pest population dynamics. Fig. 4 shows the comparison between the model prediction and the actual observation of *S. litura* maximum severity at 34 SMW.

4. Conclusions

The dynamics of *S. litura* in the groundnut cropping system as measured through moth catches in the pheromone traps at Dharwad followed a nonlinear relation with weather variables. Long term data indicated 34 SMW to be the period of maximum severity with the maximum temperature and relative humidity (both morning and afternoon) of the 32 and 33 SMWs having significant

influence on the *S. litura* moth population. The ability of linear and nonlinear models to predict *S. litura* moth population at maximum severity compared using PCR and MLP neural network, respectively indicated the superiority of the later over the former. Principal component analysis reduced the dimensionality of input weather data and improved the prediction ability of the multilayer perceptron (MLP) neural network model. Validation clearly demonstrated the utility of the MLP model in predicting the maximum severity of *S. litura.* Applicability of the MLP model in predicting S*. litura* maximum severity for future years requires all the weather variables considered in this study pertaining to 32 and 33 SMW as inputs in addition to integration of neural network methodology into a standalone or web enabled tool for forecasting. Since the peak population of *S. litura* largely determined its scenario and severity for a given season and the fact that adult moth population levels are directly related to the size of the previous week's larval population inflicting damage to the crop. Thus the weather-based forewarning using neural network based model would be highly useful for issuing pest advisory leading to be in readiness for pest management.

Acknowledgement

Authors gratefully acknowledge the funding of National Initiative on Climate Resilient Agriculture (NICRA) for undertaking the current study. As the long term data on pest and weather involves many unknown contributors over time, their efforts are placed on record with sincere appreciations. Authors are also thankful to the anonymous referee for his valuable comments and suggestions.

References

- Agrawal, R., Jain, R. C. and Jha, M. P., 1986, "Models for studying rice crop weather relationship", *Mausam*, **37,** 1, 67-70.
- Chattopadhyay, C., Agrawal, R., Kumar, A., Singh, Y. P., Roy, S. K., Khan, S. A., Bhar, L. M. Chakravarthy, N. V. K., Srivastava, A., Patel, B. S., Srivastava, B., Singh, C. P. and Mehta S. C., 2005, "Forecasting of *Lipaphis erysimi*on oilseed Brassicas in India - A case study", *Crop Protection*, **24**, 1042-1053.
- De Wolf, E. D. and Francl, L. J., 1997, "Neural network that distinguish infection periods of wheat tan spot in an outdoor environment", *Phytopath*, **87**, 83-87.
- Dhaliwal, G. S., Jindal, V. and Dhawan, A. K., 2010, "Insect Pest Problems and Crop Losses: Changing Trends", *Indian J. of Ecol.*, **37,** 1, 1-7.
- FAOSTAT, 2013, ["http://faostat.fao.org/site/567/d](http://faostat.fao.org/site/567/)efault.aspx", Food and Agriculture Organization of the United Nations.
- GOI, 2013, "Agricultural statistics at a glance", Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India.
- Haykin, S., 1994, "Neural networks: A comprehensive foundation", McMillan College Publishing Company, New York.
- Huang, Y., Lan, Y., Thomson, S. J., Fang, A., Hoffmann W. C. and Lacey, R. E., 2010, "Development of soft computing and application in agricultural and biological engineering", *Comput. Electron. Agric*., **71**, 107-127.
- Jha, G. K. and Sinha, K., 2014, "Time-delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India", *Neural Comput. Applic*., **24**, 563-571.
- Montgomery, D. C., Peck, E. A. and Vining, G. G., 2003, "Introduction to linear regression analysis", John Wiley and Sons Inc., New York.
- NCIPM, NICRA, 2011, "Manual for Groundnut Pest Surveillance", National Centre for Integrated Pest Management, New Delhi, Directorate of Groundnut Research, Gujarat & Central Research Institute for Dryland Agriculture, Hyderabad.
- Prasad, Y. G., Gayathri, M., Prabhakar, M., Jeyakumar, P., Vennila, S., Rao, A. V. M. S., Rao, I. B., Rao, K. V., Rao, G. R. and Rao, V. U. M., 2013, "Population dynamics of *Spodoptera litura* outbreak on soybean *vis a vis* rainfall events", *J. Agric. Meteorol.*, **15**, 37-40.
- Ranga Rao, G. V., Wightman, J. A. and Ranga Rao, D. V., 1989, **"**Threshold temperatures and thermal requirements for the development of Spodoptera litura *(Lepidoptera*: *Noctuidae)*", *Environ. Entomol*., **18,** 4, 548-551. ISSN 0046-225X.
- Samui, R. P., Chattopadhyay, N., Sable, J. P., Karthikeyan, K. and Balachandran, P. V., 2005, "Weather based forewarning of green jassid attack on kharif rice and operational crop protection at Pattambi, Kerala", *Mausam*, **56**, 2, 405-416.
- Samui, R. P., Chattopadhyay, N., Sabale, J. P., Gopinathan, P. B. and Jayprakesh, P., 2007, "Weather based forewarning of leaf folder attack on kharif rice & operational crop protection at Pattambi, Kerala", *Mausam*, **58**, 4, 525.
- Yang, L. N., Peng, L., Zhang, L. M., Zhang, L. L. and Yang, S. S., 2009, "A prediction model for population occurrence of paddy stem borer (*Scirpophaga incertulas*), based on back propagation artificial neural network and principal components analysis", *Comput. Electron. Agric.*, **68**, 200-206.