Understanding *Helicoverpa armigera* Pest Population Dynamics related to Chickpea Crop Using Neural Networks

Rajat Gupta
IIIT, Hyderabad India

BVL Narayana
IIIT, Hyderabad India

P.Krishna Reddy
IIIT, Hyderabad, India
E-mail:pkreddy@iiit.net

G.V.Ranga Rao
ICRISAT, Patancheru, India

CLL Gowda
ICRISAT, Patancheru, India

YVR Reddy
ICRISAT, Patancheru, India

G.Rama Murthy
IIIT, Hyderabad, India

Abstract

Insect pests are a major cause of crop loss globally. Pest management will be effective and efficient if we can predict the occurrence of peak activities of a given pest. Research efforts are going on to understand the pest dynamics by applying analytical and other techniques on pest surveillance data sets. In this study we make an effort to understand pest population dynamics using Neural Networks by analyzing pest surveillance data set of Helicoverpa armigera or Pod borer on chickpea (Cicer arietinum L.) crop. The results show that neural network method successfully predicts the pest attack incidences for one week in advance.

1. Introduction

Insect pests are well known as the major constraint to crop production. One of the problems in addressing pest management is inadequate knowledge about the factors influencing pest population dynamics. To understand pest dynamics, scientists collect pest surveillance data and related agricultural operations regarding crops, farming practices and other weather parameters. These databases contain details of pest incidence, climatic, soil, agricultural practices and serve as repositories of information. Correlations between some of these factors and pest incidence based on statistical models have been developed. However, a functionally viable model for pest forecast is still needed by farmers for efficient and effective pest management.

Pod borer, *Helicoverpa armigera* is one of the key pests causing severe yield losses, infesting several crops such as cereals, pulses, cotton, vegetables and fruit crops as well as wild hosts [5]. Ecological and Physiological features like high fecundity, multi-voltinism, ability to migrate long distances and diapause during unfavorable conditions contribute for it's severity in different situations. The climatic data follows a gradual seasonal pattern that repeats almost every year. The *Helicoverpa armigera* incidence, on the other hand, show a certain pattern in terms of population dynamics. However the peaks can change abruptly from one week to the other. In other words the overlapping generations of the pest lead to unpredictable biological events. This non-linear and complex nature of *Helicoverpa* population dynamics makes it difficult to predict population densities using traditional forecasting models.

In this paper, an effort has been made to understand the *Helicoverpa* population dynamics on the chickpea *(Cicer arietinum L.)* crop using data mining [4] techniques such as neural networks.

Literature Review:
A great deal of work on forecast models has been done especially on regression modeling and simulation models. The studies conducted by Trivedi et al., (1998) [8] have proposed a multinomial regression model to predict the impending attack of *Helicoverpa armigera.* However the model seems to be working well only when the pest population were moderate in years like 1992-1994. Whereas when there was an unusual spurt in the pest populations during 1995 the model outputs were not up to the expectations. Pimbert and Srivastava (1991) [6] analyzed the *Helicoverpa* larval counts, light trap data and related parameters over six years and showed that rainfall deficit year favor *Helicoverpa* population in Andhra Pradesh, India. Regression analysis techniques were used by Das et al., (2001) [1] to explore the relationship between rainfall and pest abundance in different years and the cumulative effect of drought on the abundance of *Helicoverpa.* Kruskal-Wallis [2] one way analysis of variance by ranks was used to compare the pest abundance in normal and rainfall deficient years. In their experiment they regressed rainfall versus larval count for a period of 9 years from 1983-1989. Their results hold good for most of the period between but fail for 87/88 where there is a departure in the usual behavior of *Helicoverpa* from the original trend. Zhao and Shen [9] discussed about building a Monte Carlo simulation model based on variance and did not use the deviations. They used the nonlinear least square regression for simulating insect stochastic population rather than methods in

Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)
0-7695-1978-4/03 $ 17.00 © 2003 IEEE

simulation of differential equations and estimating parameters of nonlinear equations. This innovation made simulation easy to use for plant protectionists. The results of simulation seemed to be the best till date and are better than regression models. However they are far from the required accuracy which necessitates the exploration of new techniques to address the pest problem.

Therefore, the present study was initiated to enhance the predictability of *Helicoverpa* population using neural networks. In the next section we explain about the data set. Next, we explain experimental procedure and discuss the results. The last section contains conclusions.

2. Data set

As a part of investigation to understand *Helicoverpa armigera* population dynamics, scientists at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) [3], Patancheru, near Hyderabad, India, have collected pest surveillance data. ICRISAT focuses on improving the productivity and production of the farming systems of the semi-arid tropical areas of the developing world and conducts research on the following crops: sorghum, pearl millet, groundnut, chickpea, and pigeonpea. Pest surveillance data collected at ICRISAT for chickpea crop on *Helicoverpa* pest contains a set of daily and weekly recordings about weather and pest incidence at various locations in the farm. We briefly describe these recordings.

- **Date and Week**
  - **Date**: The date of survey for a particular event. There were some missing values in the data, which were indicated as nulls in the database.
  - **Standard Week**: The weeks in a year are mapped to integer values by considering first week of January as first standard week.
- **Weather**
  - **Minimum and Maximum Temperature (Tmin and Tmax)**: The lowest and highest temperatures (^{o}c) recorded on ICRISAT campus on the date of survey respectively.
  - **Humidity:** The relative humidity recorded on ICRISAT campus on the date of survey.
  - **Rainfall:** The amount of rainfall (mm) recorded on ICRISAT campus on the date of survey.
- **Pest Incidence**
  - **Larvae/Plant:** The mean number of *Helicoverpa* larvae present per plant. Larval counts are based on 30-50 randomly picked plants per hectare.
  - **Eggs/Plant:** It is an estimate of the number of *Helicoverpa* eggs present per plant.
  - **Light and Pheromone Trap Catchs:** Number of *Helicoverpa* moths caught by Light trap and Pheromone traps.
- **Location**
  - **Zone:** ICRISAT farm was divided into various zones. So this attribute indicate the zone of the observation.
  - **Location:** Location of the observation in a particular zone.
  - **Season:** Two main seasons in Indian agriculture: Kharif (rainy) and Rabi (post-rainy).
  - **Area Surveyed:** The farm size surveyed in hectares.
  - **Plant Protection Type:** Different types of plant protection practices that are undertaken on farm.
  - **Observer:** The scout's name who has collected that particular information.

3. Pest attack prediction

A Neural network [7] is an interconnected set of input/output units where each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to able to predict the call label of input samples during testing phase. Neural networks posses high tolerance to noisy data as well as ability to identify patterns on which they were not trained. However, a deep problem in the use of neural network techniques involves regularization, complexity adjustment, or model selection, that is, selecting (or adjusting) the complexity of the network. Even though the number of inputs and outputs is given by the feature space, the total number of weights or parameters in the network is not known directly.

3.1 Data Preprocessing and neural network training

From the chickpea data set, weather parameters and larvae/plant information is selected and other information was ignored. We have decreased the granularity of the data set by taking weekly mean of Tmin, Tmax, rainfall, humidity and larvae/plant. The reasons for doing this are as follows: Daily data was having lots of gaps in it. Because, sometimes, it was not possible to collect the data on a particular day due to bad weather or absence of scout. Null values (about 2 %) have been ignored. We have used z-score normalization to scale the attribute data to fall within a small specified range [4].

Discrete Fourier transform (DFT) was used to convert the time domain periodic data into the frequency domain. DFT gives the set of complex numbers for the normalized data set. We have designed separate feed-forward neural networks for real values and imaginary values. Bayesian Regularization in combination with Levenberg-Marquardt algorithm (Gauss Newton) [7] is used for training. The typical performance function used by the feed-forward neural networks is the mean sum of squares of network

0-7695-1978-4/03 $ 17.00 © 2003 IEEE

errors. The neural network architecture is having two hidden layers. We have used hyperbolic tangent sigmoid function in both of these hidden layers. In the outer layer we have used linear transfer function.

3.2 Experiments and results

The pest surveillance data set for the chickpea crop was collected over a period of 11 years (1991-2001) and contains 2372 daily recordings. In this data set, eight years (1991-1998) data was selected for training and three years data (1999-2001) for testing. Each tuple in the data set is of the form <Tmin, Tmax, Humidity (H), Rainfall (RF), Larvae/plant (L)>, where each value represents the weekly mean. After reducing the data set into weekly means, the number of tuples comes to 380. Let us term this as a base data set. For prediction, we have generated four kinds of data sets from the base data set. Let the notation advance(x), where x = 0, 1,2, and 3, denote the data set. In the advance(0) data set, the L value is a function of corresponding Tmin, Tmax, H, R values of the same week, i.e., it is same as base data set. In advance(1), advance(2) and advance(3) data sets, the L value is a function of previous first, second and third week's Tmin, Tmax, H, and R values respectively. Experiments were conducted on advance(x) data set to predict Larvae for x-weeks advance.

Table 1. Results

Data set	Correl -ation	Hits	Misses	False hits
Advance(0)	0.91	27	4	6
Advance(1)	0.96	27	3	4
Advance(2)	0.91	27	2	11
Advance(3)	0.75	22	6	11

Each data set was transformed into complex domain through DFT. Two networks are being trained: one for predicting the real value, another for predicting the imaginary value. After the prediction, these predicted complex values are remapped back to corresponding values in time domain through inverse DFT. By starting with different initial values, the experiment was conducted fifteen times for each data set. Each neural network to predict both real and imaginary part of complex value consists of an input layer (four variables), two hidden layers and an output layer (one variable). For the neural network to predict the real value, first hidden layer consists of twelve neurons and second hidden layer consists of six neurons. And, for the neural network to predict complex value, first hidden layer consists of nine neurons and second hidden layer consists of six neurons.

Analysis: Figures 1, 2, 3 and 4 show the real and prediction curves for advance(0), advance(1), advance(2), and advance(3) data sets respectively. The standard week was plotted on X-axis and the corresponding mean L-values of fifteen experiments on Y-axis. In these figures the horizontal thick line indicates the threshold larvae/plant value for the pest emergence which is equal to 1.2. Whenever the actual value is greater than or equal to 1.2, we term it as a *real peak*. Whenever the predicted value is greater than or equal to 1.2, we term it as a *predicted peak*. The performance of neural network is measured by the number of real peaks it predicts.

Table 1 shows the correlation coefficients of actual and predicted curves, the number of hits, misses and false hits results corresponding to Figures 1-4. Here, a hit means, the network is able to predict the real peak. A miss means the network is unable to predict the peak. A false hit means, there is no real peak, but there is a predicted peak.

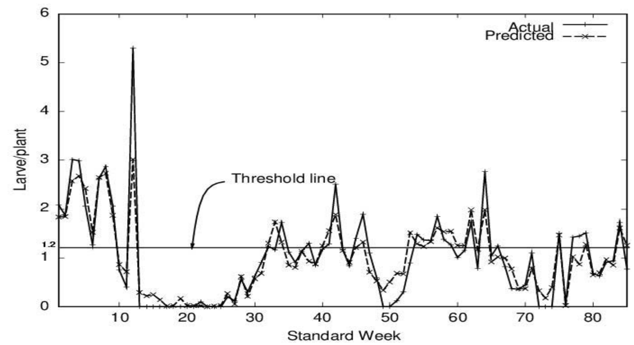

Figure 1. Advance(0) data set.

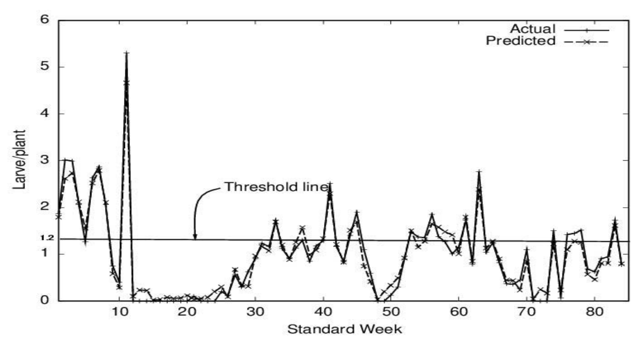

Figure 2. Advance(1) data set

Figure 1 shows the results for advance(0) data set. Given the weather parameters of a particular week, this experiment predicts the larvae/plant value of the same week. Out of 31 peaks, 27 peaks are being predicted and 4 peaks are being missed with 6 false hits.

Figure 2 shows the results for advance(1) data set. Given the weather parameters of a particular week, this experiment predicts the larvae/plant value of the next week. The correlation coefficient is 0.96. Surprisingly, it can be observed that we are getting an improved correlation over advance(0) data set (see Table 1). It means that the weather parameters of the current week are influencing more the larvae of the next week over the larvae of the current week. This is indeed a fact as it takes a four to five days for the eggs to hatch and convert into larvae. So this result agrees with the pattern of pest growth. Out of 30 peaks, 27 peaks are being predicted and 3 peaks are being missed with 4 false hits. This result shows that the neural network is able to predict the pest attack in one week advance with high accuracy.

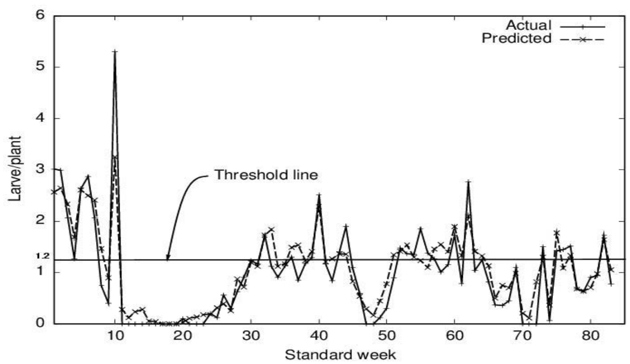

Figure 3. Advance(2) data set.

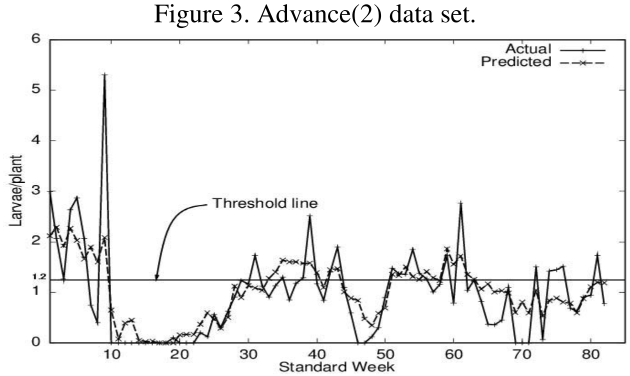

Figure 4. Advance(3) data set.

Figure 3 shows the results for advance(2) data set. Given the weather parameters of the particular week, this experiment predicts the larvae/plant value of the next two weeks. The correlation coefficient (0.91) is decreased over advance(1) data set. So the performance of neural network is decreasing with the delay. Out of 29 peaks, 27 peaks are being predicted and 2 peaks are being missed with 11 false hits. This result shows that the neural network is able to predict the pest attack two weeks advance with high accuracy, however with more number of false hits.

Figure 4 shows the results for advance(3) data set. Given the weather parameters of the particular week, this experiment predicted the larvae/plant value of the next three weeks. Here the correlation coefficient (0.75) is decreased significantly over advance(2) and advance(1) data sets. So the decrease in the performance continues with the increase in the delay as expected.

4. Summary and Conclusions

Experiments were conducted to predict pest attack by extracting pest dynamics patterns using climatic data and pest surveillance databases of *Helicoverpa armigera* pest dynamics on chickpea using neural network technique. The experimental results show that it is possible to predict the pest attack with high probability for one week in advance. These predictions would help the farmers in pest management programs by avoiding the crop losses with improved environment quality, as it can avoid unnecessary sprays of chemical pesticides.

Bibliography

1. D.K.Das, T.P.Trivedi and C.P.Srivastava. 2001. Simple rules to predict attack of *Helicoverpa armigera* on crops growing in Andhra Pradesh, Indian Journal of Agricultural Sciences 71: 421-423.

2. Dean W. Wichern and Richard Arnold Johnson, Applied Multivariat Statistical Analysis, Prentice Hall, 2002.

3. International Crop Research Institute for Semi-Arid tropics: ICRISAT, URL: www.icrisat.org, April 2003.

4. Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann, 2001.

5. T.M.Manjunath, V.S.Bhatnagar, C.S.Pawar, and S.Sithnantham. 1989. Economic importance of *Helicoverpa* spp. In India and an assessment of their natural enemies and host plants. In King EG, Jackson RD, eds. Proceeding of the workshop on Biological Control of *Helicoverpa:* Increasing the Effectiveness of natural Enemies, Far Eastern Regional Research Office, United States Department Agriculture, New Delhi. pp. 197-228.

6. M.P.Pimbert, C.P.Srivastava. 1991. The influence of rainfall deficits on the abundance of *Helicoverpa armegira* in Andhra Pradesh, India. Biological Agriculture and Horticulture, 8:153-176.

7. Simon Haykin, Neural Networks: A Comprehensive Foundation, Pearson Education, 2001.

8. T.P.Trivedi, D.K.Das, A.Dhandapani, and A.K.Kanojia, 2002. Models for Pests and Disease Forecasting, Resources Management in Plant Protection, Volume I, Plant Protection Association of India, Hyderabad, India, 2002.

9. Zhongua Zhao, Zuorui Shen. Theories and their applications of Stochastic Simulation Models for Insect population Dynamics. Department of Entomology, The China Agricultural University, Feb.2003.