WEATHER BASED FOREWARNING MODELS FOR PESTS & DISEASES AND YIELD LOSS ASSESSMENT

Amrender Kumar
Indian Agricultural Statistics Research Institute, New Delhi-11012
akjha@iasri.res.in

Pests and diseases are one of the major causes of reduction in crop yield. Timely application of remedial measures may reduce the yield loss. For application of these measures one must have prior knowledge of the time and severity of the outbreak of pests and diseases. Forecasting system can help in this direction.

In pests and diseases forewarning system, the variables of interest may be maximum pest population / disease severity, pest population / disease severity at most damaging stage of the crop, pest population / disease severity at different stages of crop growth or at various standard weeks, time of first appearance of pests / diseases, time of maximum pest population / disease severity, time of pest population / disease severity crossing threshold limit, extent of damage, weekly monitoring of pests and diseases progress, occurrence / non-occurrence of pests and diseases.

If data are available at periodic interval for 15-20 years, the detailed study can be carried out for different variables of interest. However, depending upon the data availability, different types of models can be utilized for developing forewarning system. The models could be of two types, 'Between year model' and 'Within year model'.

Between year models

These models are developed using previous years' data. An assumption is made that the present year is a part of the composite population of the previous years and accordingly the relationships developed on the basis of previous years' data will be applicable for the present year. The forecast for pests and diseases are obtained by substituting the current year data into the model developed upon the previous years. Various methods have been attempted when data are available in quantitative form. Some of the important techniques are discussed below :

Thumb rule

This approach is the most common and extensively used. It is a simple system which describes the forecasting of the pests and diseases based on past experience. For example for potato late blight, a day is favorable if

- the 5 day temperature average is $< 25.5^0$ C
- the total rainfall for the last 10 days is > 3.0 cm
- the minimum temperature on that day is $> 7.2^0$ C

When this situation arises, there is a possibility of potato late blight appearance.

Regression Model

The regression model taking pest / disease variable as dependent and suitable independent variables such as weather variables, crop stages, population of natural enemies/predators etc. is used. The form of the model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots + \beta_p X_p + e$$

where β_0, β_1, β_2,…….β_p are regression coefficients, X_1, X_2, ……,X_p are independent variables and e is error term. These variables are used in original scale or on a suitable transformed scale such as cos, log, exponential etc. (Coakley et al 1985; Trivedi et al. 1999).

Fuzzy regression

In regression analysis, the unfitted errors between a regression model and observed data are generally assumed as observation error that is a random variable having a normal distribution, constant variance, and a zero mean. In fuzzy regression analysis, the same unfitted errors are viewed as the fuzziness. Fuzzy regression can be quite useful in estimating the relationship among variables where the availability data are imprecise and fuzzy.
Fuzzy regression analysis gives a fuzzy functional relationship between dependent and independent variables where vagueness is present in some form. There are three situations where the fuzzy analysis can be viewed *viz.* Crisp parameters and fuzzy data, Fuzzy parameters and crisp data and Fuzzy parameters and fuzzy data. Fuzzy regression method is based on minimizing fuzziness as an optimal criterion, which can be achieved by linear programming procedures.

Growing Degree Day Approach

This method is based on the assumption that the pest becomes inactive below a certain temperature known as base temperature. Growing degree day is worked out as

$$GDD = \Sigma \text{ (mean temp. – base temp.)}$$

GDD is used in the model as explanatory variable. This method requires proper knowledge of base temperature and initial time from which accumulation is to start.

Model based on weather indices

In this approach, using weekly and fortnightly weather variables suitable indices are worked out which are used as regressors in the model. The model is of the form

$$Y = a_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq i'}^{p} \sum_{j=0}^{1} b_{ii'j} Z_{ii'j} + e$$

where

$$Z_{ij} = \sum_{w=n_1}^{n_2} r_{iw}^{j} X_{iw} \quad \& \quad Z_{ii'j} = \sum_{w=n_1}^{n_2} r_{ii'w}^{j} X_{iw} X_{i'w}$$

Y variable to forecast; X_{iw} is value of i^{th} weather variable in w^{th} period; r_{iw} / $r_{ii'w}$ is suitable weight given to i^{th} weather variable / product of i^{th} and i'^{th} weather variable in w^{th} period; p is number of weather variables considered; n_1 and n_2 are the initial & final periods for which weather data were included in the model and e is error term.

If information on favourable weather conditions is known, subjective weights based on this information can be used for constructing weather indices. In absence of such information correlation coefficients between Y and respective weather variable/product of weather variables can be used [Agrawal *et al.* (2004), Chattopadhyay *et al.* (2005-a), Chattopadhyay *et al.* (2005-b), Desai *et al.* (2004) and Dhar *et al.* (2007)]

Principal component regression

Forewarning models can be developed using the principal component technique as normally relevant weather variables are large in number and are expected to be highly correlated among themselves. Using the first few principal components of weather variables as independent variables forecast models can be developed.

Discriminant function analysis

Forewarning models of pests and diseases based on time series data on weather variables can be developed using the discriminant function analysis. For this analysis, a series of data for 25-30 years are required. Based on the pest and diseases variables, data can be divided into different groups – low, medium and high etc. and using weather data in these groups, linear or quadratic discriminant functions can be fitted which can be used to find discriminant scores. Considering these discriminant scores as independent variables and diseases / pest as a dependent variable, regression analysis can be performed. Johnson et al (1996) used discriminant analysis for forecasting potato late blight.

Complex polynomial [Group Method of Data Handling (GMDH)]

It provides complex polynomial in independent variables. It selects the structure of the model itself without prior information about relationship. Form of the model :

$$Y = a + \sum_{i=1}^{m} b_i X_i + \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} X_i X_j + \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} d_{ijk} X_i X_j X_k + \ldots.$$

The technique involves fitting of quadratic equations for all pairs of independent variables and identifying a few best performers in terms of predictive ability (using appropriate statistics); converting entire set of independent variables (called zero generation variables) to new variables (first generation variables) which are obtained as predicted values from these selected quadratic equations (of zero generation variables). The process of fitting and identifying best quadratic equations is repeated using first generation variables and second generation variables are obtained. The whole process is repeated with every new generation of variables till appropriate model is obtained (using certain criteria). At final stage, one best quadratic equation is selected as the final model. (Bahuguna et al 1992; Trivedi et al 1999).

Machine Learning Techniques

Machine learning techniques offer many methodologies like decision tree induction algorithms, genetic algorithms, neural networks, rough sets, fuzzy sets as well as many hybridized strategies for the classification ad prediction (Han and Kamber, 2001; Pujari, 2000; Komorowski *et al.*, 1999; Witten and Frank, 1999). Decision tree induction represents a simple and powerful method of classification that generates a tree and a set of rules, representing the model of different classes, from a given dataset. Decision Tree (DT) is a flow chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node represents the class. The top most node in a tree is the root node. For decision tree ID3 algorithm and its successor C4.5 algorithm by Quinlan (1993) are widely used. One of the strengths of decision trees compared to other methods of induction is the ease with which they can be used for numeric as well as nonnumeric domains. Another advantage of decision tree is that it can be easily mapped to rules.

Artificial Neural Networks (ANNs) is another attractive tool under machine learning techniques for forecasting and classification purposes. ANNs are data driven self-adaptive methods in that there are few apriori assumptions about the models for problems under study. These learn from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown or hard to describe. After learning from the available data, ANNs can often correctly infer the unseen part of a population even if data contains noisy information. As forecasting is performed via prediction of future behaviour (unseen part) from examples of past behaviour, it is an ideal application area for ANNs, at least in principle. (Agrawal et al. 2004; Dewolf et al. 1997, 2000; Kumar, et al. 2010). However, the technique requires a large data base.

Deviation Method

This method can be utilized when periodical data at different intervals during the crop season are available for only 5-6 years. The pest population at a given point of crop stage is assumed to be due to two reasons – natural cycle of the pest and weather. To identify the natural cycle, data at different intervals is averaged over years and a suitable model is fitted to these averaged data points. Then the entire data is converted as deviations from the predicted natural cycle. Appropriate model is fitted using these deviations as dependent and weather as independent variables. [Mehta *et al.*(2001)]

Ordinal logistic model – model for qualitative data

The timely control measures to prevent pest / disease outbreak can be taken even if the information on the extent of severity is not available but merely the epidemic status is accessible. This information could be obtained through modeling qualitative data. Such models have added advantage that these could be obtained even if the detailed and exact information on pest count / disease severity is not available but only the qualitative status such as epidemic or no epidemic / low, medium or high is known. Such a situation arises quite often in pest / disease data. In such cases, the data are classified as 0/1 (2 categories); 0,1,2 (three categories). The logistic regression is used for obtaining probabilities of different categories. For example, for two categories, the model is of the form :

$$P(E=1) = \frac{1}{1 + \exp(-z)} + e$$

where z is a function of weather variables.

Forecast / Prediction rule :

If $P \geq .5$ more chance of occurrence of epidemic
If $P < .5$ probability of occurrence of epidemic is minimum

(Mehta et al. 2001; Mishra et al. 2004; Johnson et al. 1996; Agrawal, et al. 2004)

Within year model

Sometimes, past data on pests and diseases are not available but the pests and diseases status at different points of time during the crop season are available for the current season only. In such situations, within years growth model can be used for forewarning maximum disease severity / pest population, provided there are 10-12 data points between time of first appearance of pest / disease and maximum or most damaging stage.
The methodology consists in fitting appropriate growth pattern to the pests and diseases data based on partial data and using this growth curve for forecasting the maximum value of variable of interest. A number of growth models such as logistic, Gompertz etc. can be used for this purpose (Agrawal et al. 2004). Prajneshu (1998) developed a non linear statistical model for describing the dynamic population growth.

Crop yield loses due to pests and diseases

Crop loss models are essential for estimation of loss due to the pests / diseases. Accurate information concerning losses are needed by growers or plant protection specialists to develop decision thresholds for determining when the cost effective control measures should be deployed. Yield loss due to pests / diseases needs to represent a dynamic interaction between pathogen and host. Using proper relationship for loss modeling, the general procedure is to identify a dependent variable e.g. yield and quantify the manner in which this changes with independent variable(s). This independent variable(s) commonly represents pests or disease severity / pests or disease index. Based on this model we can predict the yield when there are pests and diseases incidences. When there is no pest or disease incidence (by putting zero for pest and diseases variables), we can get the maximum potential yield. Then loss can be calculated as

$$\text{Yield loss(\%)} = \frac{(Y_0 - Y_1)}{Y_0} \text{ X } 100$$

Y_0 = yield when there are no pests and diseases
Y_1 = yield when there are of pests and diseases

Stynes (1975) used the multivariate statistical techniques such as component analysis and canonical correlation in crop loss due to pests and diseases. Similarly loss due to other factors such as weeds, flood, drought, etc. can also be obtained.

References

Agrawal, Ranjana, Mehta, S.C., Bhar, L.M. and Kumar, Amrender (2004). Development of weather based forewarning system for crop pests and diseases - Mission mode project under NATP.

Bahuguna, G.N. and Chandrahas (1992). Models for forecasting aphid pest of mustard crop. (IASRI Publication).

Chattopadhyay, C, Agrawal, R., Kumar, A., Bhar, L.M., Meena, P.D., Meena, R.L., Khan, S.A., Chattopad3hyay, A.K., Awasthi, R.P., Singh, S.N., Chakravarthy, N.V.K., Kumar, A., Singh, R.B. and Bhunia, C.K. (2005-a). Epidemiology and forecasting of Alternaria blight of oilseed Brassica in India – a case study. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz (Journal of Plant Diseases and Protection)*, 112(4), 351-365.

Chattopadhyay, C., Agrawal, R., Kumar, Amrender, 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-b). Forecasting of *Lipaphis erysimi* on oilseed Brassicas in India – a case study. *Crop Protection*, 24, 1042-1053.

Coakley, S.M., Mcdaniel, L.R. and Shaner, G. (1985). Models for predicting severity of septoria tritici blotch on winter wheat. *Phytopathalogy*, 75(11) : 1245-51.

Desai, A. G., Chattopadhyay, C., Agrawal, Ranjana, Kumar, A., Meena, R.L., Meena, P.D., Sharma, K.C., Srinivasa Rao, M., Prasad, Y.G. and Ramakrishna, Y.S. (2004). *Brassica juncea* powdery mildew epidemiology and weather- based forecasting models for India – a case study. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz (Journal of plant Diseases and Protection)* 111(5), 429-438.

Dewolf, E.D. and Francl, L.J. (1997). Neural network that distinguish in period of wheat tan spot in an outdoor environment. Phytopathalogy, 87(1) : 83-7.

Dewolf, E.D. and Francl, L.J. (2000). Neural network classification of tan spot and stagonespore blotch infection period in wheat field environment. Phytopathalogy, 20(2) :108-13.

Dhar, Vishwa, Singh, S.K., Kumar, M., Agrawal, R. and Kumar Amrender (2007). 'Prediction of pod borer (Helicoverpa armigera) infestation in short duration pigeonpea (Cajanus cajan) in central Uttar Pradesh', *Indian Journal of Agricultural Sciences*, 77(10), 701-04.

Han, J. and Kamber, M. (2001). *Data Mining Concepts and Techniques.* Morgan Kaufmann Publisher.

Johnson, D.A., Alldredge, J.R. and Vakoch, D.L. (1996) Potato late blight forecasting models for the semi-arid environment of South-Central Washington. *Phytopathalogy*, 86(5) : 480-84.

Komorowski, J., Pawlak, Z., Polkowki, L. and Skowron, A.(1999). Rough Sets: A Tutorial. In : *Rough Fuzzy Hybridization, Pal S.K. and Skowron, A.(eds.),* Springer, 3-99.

Kumar, Amrender, Ramasubramianian, V. and Agrawal, Ranjana (2010). Neural Network Based Forecast Modelling in Crops. IASRI Publication.

Mehta, S.C., Agrawal, Ranjana, Shukla, R.P. and Sharma, Shashi (2001). A Statistical model for prediction of mango–fruit–fly outbreak. Presentation in National Seminar on Agro-meteorological research for Sustainable Agricultural Production organized by Gujarat Agricultural University, Anand.

Mishra, A.K., Prakash, O. and Ramasubramanian, V. (2004). Forewarning powdery mildew caused by *Oidinm mangiferae* in mango (*Mangifera indica*) using logistic regression models. Indian Journal of Agricultural Sciences, 74(2) : 84-7.

Prajneshu (1998). A non-linear statistical model for aphid population growth. Jour. Ind. Soc. Agril. Stat., 51(1), 81-93.

Pujari, A.K. (2000). *Data Mining Techniques.* Universities Press.

Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kauffman.

Stynes (1980). Crop loss assessment. *Mics. Publ. Univ. Minn. Agric. Exp. Stn*. **7.**

Trivedi, T.P., Paul Khurana, S.M., Jain, R.C., Mehta, S.C. and Bhar, L.M. (1999), Development of Forewarning system for aphids (*Myzus persicae*) on potato, Annual Report NCIPM, N. Delhi.

Witten, I.H. and Frank E. (1999). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers.