Introduction
Reliable and timely forecasts provide important and useful input for proper, foresighted and informed planning, more so, in agriculture which is full of uncertainties. Agriculture now-a-days has become highly input and cost intensive. Under the changed scenario today, forecasting of various aspects relating to agriculture are becoming essential. But in-spite of strong need for reliable and timely forecasts, the current status is far from satisfactory. For most of the sectors, there is no organized system of forecasting. The official forecasts (advance estimates) of major cereal and commercial crops are issued by Directorate of Economics and Statistics. Some work has been attempted in fisheries, livestock, market projections, etc. by various organizations but these are mostly at research stage. Details of forecasting techniques in crops is discussed here. In crops, production and attack of pests and diseases are the two major aspects which need attention. Forecasts of crop production before harvest are required for various policy decisions relating to storage, distribution, pricing, marketing, import-export, etc. 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.

1. Forecasting Techniques in Crop Production
The advance estimates provided by Directorate of Economics and Statistics are subjective. There is, thus, a need to develop sound objective forecasts of crop production.

Various organisations in India and abroad are engaged in developing methodology for pre-harvest forecast of crop yields / production using various approaches. The main factors affecting crop yield / production are inputs and weather. Use of these factors forms one approach for forecasting crop production. The other approach uses plant vigour measured through plant characters. It can be assumed that plant characters are integrated effects of all the factors affecting production. Yet another approach is measurement of crop vigour through remotely sensed data. Since area under the crop is available in advance of harvest through Timely Reporting Scheme (TRS) and remotely sensed data, the problem mainly reduces to forecast of yield rate.

1.1 Yield forecast using weather parameters
Weather affects crop differently during different stages of crop growth. Thus extent of weather influence on crop yield depends not only on the magnitude of weather variables but also on the distribution pattern of weather over the crop season which, as such, calls for the necessity of dividing the whole crop season into fine intervals. This will increase number of variables in the model and in turn a large number of parameters will have to be evaluated from the data. This will require a long series of data for precise estimation of the parameters which may not be available in practice. Thus, a technique based on relatively smaller number
of manageable parameters and at the same time taking care of entire weather distribution may solve the problem.

Fisher (1924) has suggested technique which requires small number of parameters to be estimated while taking care of distribution pattern of weather over the crop season. He assumed that the effect of change in weather variable on crop in successive weeks would not be an abrupt or erratic change but an orderly one that follows some mathematical law. He assumed that these effects are composed of the terms of a polynomial function of time. Further, the value of weather variable in w-th week, \( X_w \) was also expressed in terms of orthogonal functions of time. Substituting these in usual regression equation

\[
Y = A_0 + A_1X_1 + A_2X_2 + \ldots + A_nX_n
\]

(here \( Y \) denoted yield and \( X_w \) rainfall in w-th week \( w = 1,2,\ldots,n \)) and utilising the properties of orthogonal and normalised functions, he obtained

\[
Y = A_0 + a_0\rho_0 + a_1\rho_1 + a_2\rho_2 + \ldots + a_k\rho_k
\]

where \( A_0, a_0, a_1, a_2, \ldots, a_k \) are constants to be determined and \( \rho_i \) \( (i = 1, \ldots, k) \) are distribution constants of \( X_w \). Fisher has suggested to use \( k = 5 \) for most of the practical situations. In fitting this equation for \( k = 5 \), the number of constants to be evaluated will remain 7, no matter how finely growing season is divided. This model was used by Fisher for studying the influence of rainfall on the yield of wheat.

Hendricks and Scholl (1943) have modified Fisher's technique. They divided the crop season into \( n \) weekly intervals and have assumed that a second degree polynomial in week number would be sufficiently flexible to express the relationship. Under this assumption, the model was obtained as

\[
Y = A_0 + a_0\sum wX_w + a_1\sum wX_w + a_2\sum w^2X_w
\]

In this model number of constants to be determined reduces to 4, irrespective of \( n \). This model was extended for two weather variables to study joint effects. Since the data for such studies extended over a long period of years, an additional variate \( T \) representing the year was included to make allowance for time trend.

Another important contribution in this field is by Baier (1977). He has classified the crop-weather models in three basic types.

1. Crop growth simulation models
2. Crop-weather Analysis models
3. Empirical statistical models

The most commonly used models in crop forecasting are Empirical Statistical models. In this approach, one or several variables (representing weather or climate, soil characteristics or a time trend) are related to crop responses such as yield. The weighting coefficients in these equations are by necessity obtained in an empirical manner using standard statistical procedures, such as multi-variable regression analysis. Several Empirical Statistical models were developed all over the world. The independent variables included weather variables, agrometeorological variables, soil characteristics or some suitably derived indices of these variables. Water
Requirement Satisfaction Index (WRSI), Thermal Interception Rate Index (TIR), Growing Degree Days (GDD) are some agroclimatic indices used in models. Southern Oscillation Index (SOI) has also been used with other weather variables to forecast crop yield (Ramakrishna et al. 2003). To account for the technological changes year variable or some suitable function of time trend was used in the models. Some workers have also used two time trends. Moving averages of yield were also used to depict the technological changes.

In contrast to empirical regression models, the Joint Agricultural Weather Information Centre employs the crop weather analysis models that simulate accumulated crop responses to selected agrometeorological variables as a function of crop phenology. Observed weather data and derived agrometeorological variables are used as input data.

USDA and FAO are the two organisations that systematically forecast world agricultural production and global crop information based on weather. Daily monitoring of satellite weather images and meteorological data provides the framework for agricultural weather analysis. Daily, weekly and seasonal summaries are processed and merged with historical weather and crop data for evaluation of the crop–yield potential.

FAO has also carried out number of studies using agro–meteorological models. The methodology consists of developing an index depending on water deficit / water surplus in successive periods of crop growth. These models have good potential for early crop yield assessment for rainfed crops. (Frere and Popov. 1979).

In India, major organisations involved in developing methodology for forecasting crop yield based on weather parameters are IMD and IASRI. The methodology adopted by IMD involves identification of significant correlations between yield and weather factors during successive overlapping periods of 7 to 60 days of the crop growing season. By analysing the correlation coefficients for statistical and phenological significance, the critical periods when the weather parameters have significant effect on yield are identified. The weather parameters in critical periods along with trend variables are used through multiple regression analysis to obtain forecast equations. Using this methodology models were developed for principal crops on meteorological subdivisions basis. Data from various locations are averaged to get the figures for meteorological sub-divisions and these are utilised to develop the forecast model. Monthly forecasts are issued from these models by taking the actual data up to time of forecast and normal for the remaining period. In some models yield Moisture Index, Generalised Monsoon Index, Moisture Stress, aridity anomaly Index are also used (Sarwade, 1988; Sarkar, 2002).

At IASRI, the model suggested by Hendricks and Scholl has been modified by expressing effects of changes in weather variables on yield in the w-th week as function of respective correlation coefficients between yield and weather variables. This will explain the relationship in a better way as it gives appropriate weightage to different periods. Under this assumption, the models were developed for studying the effects of weather variables on yield using complete crop season data whereas forecast model utilised partial crop season data. These models were found to be better than the one suggested by Hendricks and Scholl.

The forecast model finally recommended was of the form
Forecasting Techniques in Crops

\[ Y = A_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij} Z_{ij} + \sum_{i \neq i'}^{p} \sum_{j=0}^{1} a_{i'ij} Z_{i'j} + cT + e \]

Where \( Z_{ij} = \sum_{w=1}^{m} r_{iw}^j X_{iw} \) and \( Z_{i'ij} = \sum_{w=1}^{m} r_{i'iw}^j X_{i'w} X_{iw} \)

Here \( Y \) is yield, \( r_{iw}/r_{i'w} \) is correlation coefficient of yield (adjusted for trend effect) with \( i \)-th weather variable \( (X_{iw}) \)/product of \( i \)-th and \( i' \)-th weather \( (X_{iw}X_{i'w}) \) variables in \( w \)-th week, \( m \) is week of forecast, \( p \) is number of weather variables used and \( e \) is error term.

*Models were successfully used for forecasting yields of various crops at district level as well as agroclimatic zone level. (Agrawal et al 1980; 1983, 1986, 2001; Jain et al 1980; Mehta, et al. 2000).*

These models were used to forecast yield of paddy and wheat in different situations, viz (i) rainfed area having deficient rainfall (paddy), (ii) rainfed area having adequate rainfall (paddy) and (iii) irrigated area (wheat). The results revealed that reliable forecasts can be obtained using this approach when the crops are 10-12 weeks old. This approach was also used to develop forecast model for sugarcane at district level (Mehta, et al. 2000). However, these studies were carried out at district level and required a long series data of 25-30 years which are not available for most of the locations. Therefore, the study has been undertaken to develop the model on agro-climatic zone basis by combining the data of various districts within the zone so that a long series could be obtained in a relatively shorter period. Previous years yield, moving averages of yield and agricultural inputs were taken as the variables taking care of variation between districts within the zone. Year variable was included to take care of technological changes. Different strategies for pooling district level data for the zone were adopted. Results revealed that reliable forecasts can be obtained using this methodology at 12 weeks after sowing i.e. about 2 months before harvest. The data requirement reduced to 10-15 years as against 25-30 years for district level models. The approach has been successfully used for forecasting yields of rice, wheat and sugarcane for Uttar Pradesh. (Agrawal et al. 2001).

At district level, model based on time series data on weather parameters has also been developed using technique of discriminant function analysis. The long series of 25–30 years has been classified into three groups – congenial, normal and adverse with respect to crop yields. Using weather data of these groups, linear / quadratic discriminant functions were fitted. These functions were used to find weather scores for each year at different phases of crop growth and were used as regressors in forecast model. (Rai, et al. 2000).

In another approach based on water balanced technique, models for rainfed crops using weighted stress indices have been developed. In this approach, water deficit / surplus has been worked out at different phases of crop growth and using suitable weights, accumulated weighted stress index has been developed for each year which was used as regressor in the forecast model. (Saksena et al. 2001).
1.2 Yield forecast based on plant characters
Effects of weather and inputs are manifested through crop stand, number of tillers, leaf area, number of earheads etc. which ultimately determine crop yield. As such, plant characters can be taken as the integrated effects of various weather parameters and crop inputs. Thus the other approach to forecast crop yield is to use plant characters.

In USDA, the net yield per acre for each sample plot is computed as (Vogel, 1985).

\[ y_i = (F_i \times C_i \times W_i) - L_i \]

where
- \( F_i \) = Number of fruits harvested or forecast to be harvested in the i-th sample plot.
- \( C_i \) = Conversion factor using the row space measurement to inflate the plot counts to a per acre basis.
- \( W_i \) = Average weight of fruit harvested or forecast to be harvested.
- \( L_i \) = Harvest loss as measured from post-harvest gleanings (the historic average is used during the forecast season).

\[ \bar{y}_i = \frac{\sum y_i}{n} \]

for the n sample plots.

Separate models are used to forecast the number of fruits (\( F_i \)) to be harvested and the final head weight (\( W_i \)).

This method cannot be followed in India/tropical countries as time period from head emergence to maturity is hardly one to two months for most of the crops whereas in USA this takes two to three months. Forecast of head weight at maturity therefore cannot be obtained much in advance in India, as such this will not be useful for obtaining early forecast in such countries.

In India, yield is directly regressed on plant counts and yield contributing characters for obtaining forecast model. Considerable work has been done at IASRI using this approach. The data are collected at different periodic intervals through suitable sampling design for 3 to 4 years from farmers’ fields. Two types of approaches have been attempted - Between year model and Within year model.

1.2.1 Between year models
These models are developed taking previous year(s) data. Objective yield forecasts are obtained by substituting the current year plant data into a model developed from the previous year(s). An assumption is made that the present year is a part of the composite population of these (previous) years. Most commonly used models are based on regression approach.

Different Between year Models
Linear regression models

\[ Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + e \]

where \( Y \) and \( X_i \) are yield and plant characters respectively. These may be used in original scale or some suitably transformed variables of these can be used. \( \beta_0 \) and \( \beta_i \) are constants to be estimated and \( e \) is random error. These models utilise data at one point of time only during the crop growth. (Sardana et al. 1972, Singh et al. 1976, Jha et al. 1981, Singh et al. 1988).
These models were improved taking regressors as principal components of plant characters (Jain et al. 1984) or growth indices based on plant characters observed on two or more points of time during the crop growth (Jain et al. 1985). The growth indices are obtained as weighted accumulations of observations on plant characters in different periods, weights being respective correlation coefficients between yield and plant characters. The model can be written as

\[ Y = \beta_0 + \sum \beta_i G_i + e \]

where

\[ G_i = \sum_{w=n_1}^{n_2} r_{iw} X_{iw} \]

\( G_i \) is the index of the i-th character, \( w \) is period identification, \( n_1 \) & \( n_2 \) are the initial and final periods considered in developing the index of the character, \( r_{iw} \) is simple/partial correlation coefficient between yield and i-th character in w-th period \( (X_{iw}) \).

**Probability model**

Multiple regression technique has been extensively used in developing models for crop yield forecasting. Least squares technique is used for estimating the parameters of the regression model. The optimality properties of these estimates are described in an ideal setting which is not often realized in practice. It has been observed that regression based on different subsets of data produce very different results, raising questions of model stability. To overcome some of the drawbacks of regression model probability model for forecasting crop yield using Markov Chain theory has been developed. This method, being completely model free, does not require any assumption about independent and dependent variables. Markov Chain method has the advantage of providing non-parametric interval estimates and is robust against outliers/extreme values.

In this method, growth process of the crop is divided into s phenological stages. A Markov chain model is constructed by defining a set of states, which describe the condition of an individual plant (or average condition of a group of plants) at specified time within the phenological stages. Individual states are defined on the basis of available qualitative and quantitative information to describe plant condition. Let \( n_i \) for \( i = 1, 2, ..., s \) denote the number of states at the commencement of stage \( i \). Let \( A_{i,i+1} \) for \( i = 1, 2, ..., (s-1) \) denote the \( (n_i x n_{i+1}) \) transition matrix which gives the transition probabilities of a plant (or group of plants) moving from any possible state of stage \( i \) to any possible state of stage \( i+1 \). As a property of transition matrices, each row of an \( A_{i,i+1} \) matrix sums to 1.

Let \( F \) denote the matrix of transition probabilities from each of the \( (n-n_s) \) states of \( (s-1) \) intermediate stages to each of the \( n_s \) states, the last (harvest) stage. The \( n_s \) states are defined as quantitative intervals of yield. \( F \) matrix can be obtained as

\[
F = \begin{bmatrix}
\pi_{1,1} & \pi_{1,2} & \cdots & \pi_{1,s-1} & \pi_{1,s}
\\
\pi_{2,1} & \pi_{2,2} & \cdots & \pi_{2,s-1} & \pi_{2,s}
\\
\vdots & \vdots & \ddots & \vdots & \vdots
\\
\pi_{s-1,1} & \pi_{s-1,2} & \cdots & \pi_{s-1,s-1} & \pi_{s-1,s}
\\
\pi_{s,1} & \pi_{s,2} & \cdots & \pi_{s,s-1} & \pi_{s,s}
\end{bmatrix}
\]
F matrix can be used to forecast crop yields. Each row of F represents a crop condition (state) at a certain crop stage. The \( n_s \) states of the final stage are defined as quantitative intervals of yield. Each column of F represents a different yield interval. The values of each row of F are the estimated probabilities of the crop producing a final yield within each of the \( n_s \) intervals. Thus, each row of F is a predicted yield distribution for a given stage and state. Each of the \((n-n_s)\) forecast yield distributions in the F matrix may be analysed to get mean and standard error of the forecast. In particular, transition probability matrix \( A_{s-1,s} \) will give mean and standard error of forecast at stage(s-1).

This method was applied to forecast yield of corn and cotton by USDA (Matis et al. 1985, 1989) and sugarcane (Jain and Agrawal 1992(a); Agrawal and Jain 1996). Models using higher order markov chain and using principal components and growth indices of plant characters in markov chain approach were also developed. (Jain and Ramasubramanian,1998 ; Ramasubramanian and Jain 1999; Ramasubramanian, et al. 2004).

### 1.2.2 Within year models

The 'between year models' while performing satisfactorily in typical years may falter in atypical years. A model which uses data from the current growing season only may be beneficial in improving forecasts during a year with atypical growing conditions. These models are developed to provide forecasts of pertinent components of crop yield relying entirely on growth data collected from plant observations during the current growing season. A logistic model having some yield components as dependent variable and an independent 'time' variable generally fits well to the growth process of crop yield components like dry matter accumulation etc. The model uses repeated observations from the current year to estimate the parameters needed to forecast the dependent variable at maturity. The model is

\[
Y_i = \frac{\alpha}{(1+\beta \rho^i)} + e_i \quad i = 1,2,\ldots,n; \quad \alpha > 0, \beta > 0, 0 < \rho < 1
\]

where
- \( Y_i \) = dependent growth variable
- \( t_i \) = independent time variable
- \( e_i \) = error term

Partial crop season data are utilised to fit the curve and the value at harvest is predicted through this curve which in turn is used to forecast yield. (Nealon 1976, House 1977, Larsen 1978, Jain et al. 1992(b)).

The parameter \( \alpha \) is the most important parameter to be estimated as it gives average amount of yield component (eg. dry matter) at maturity. It is likely to be over estimated when partial crop season data based on small data points that too falling on the lower side of the curve where the growth has steep rise are used to fit the model to forecast the yield component at maturity. This may need suitable modification in the model so as to capture \( \alpha \) (dry matter at maturity) from partial crop season data. The modified logistic model (Jain et al 1992(b)) is as follows:

\[
Y_i = \frac{t_m}{\sqrt[\alpha}{1+\beta \rho^i} + e
\]

where \( t_m \) is time of maturity and \( t_f \) is the time of forecast.
1.3 Models using spectral data
Since the approach using plant characters requires collection of data from farmers' fields, the data can be used on characters which can be measured easily without involving much expertise, cost and sophisticated instruments. Some characters contributing significantly towards yield may not find place in the model due to these limitations. This calls for the necessity of including some other variables in the model along with biometrical characters which could take care of such variables indirectly.

During the last three decades, considerable work has been carried out in India in the spectral response and yield relationships of different crops at Space Applications Centre, Ahmedabad, under the remote sensing applications mission called ‘Crop Acreage and Production Estimation’ (CAPE). Spectral indices such as ratio of infra red (IR)/Red(R) and Normalised difference (ND) = (IR-R) / (IR+R) are calculated from remotely sensed data and are used as regressors in the model (SAC report 1990).

The scheme needed further improvement. Project has been formulated to integrate Agrometeorology and Land-based observations along with remote sensing data. Project title is “Forecasting Agricultural outputs using Space, Agrometeorology and Land-based observations” (FASAL).

The experience in this context is that remote sensing can supplement the existing data collection system but never completely replace it. The two data collection systems must be integrated through rigorous statistical methodology. At Space Application Centre, methodology has been developed which provides multiple forecasts for rice and wheat using remotely sensed data for acreage forecast whereas forecasts for productivity are obtained using meteorological and agrometeorological indices. (Patel et al. 2004)

1.4 Models using Farmers' Appraisal
Farmer is the best judge of the likely production in the field. Farmers' appraisal, therefore, could serve as a good input for forecasting the yield and replace some of the characters requiring expertise or use of sophisticated instruments for their measurements and thus reducing the cost on data collection. A study has been carried out to study the feasibility of using farmers' appraisal in the forecast model for sugarcane. (Agrawal and Jain, 1996). The results revealed that a reliable forecast could be obtained using plant population and farmers’ appraisal.

Another methodology based on farmers’ appraisal data has been developed using Bayesian approach. The study has been carried out for wheat in Muzaffarnagar district. Expert opinion data were collected in a number of rounds in a year by interviewing the selected farmers regarding their assessment about the likely crop production and chance of occurrences in various yield classes. From these responses average prior probabilities were computed. Actual harvest yield and farmers’ appraisal data on yield for previous year(s) were taken into account to obtain posterior probabilities which were then used for obtaining Bayesian forecast of crop yield for current year. (Chandrahas et al. 2001)
1.5 Integrated approach
Models using data on plant characters along with agricultural inputs were found to be better than models based on plant characters alone in jowar and apple (Jain et al. 1985; Chandrahas and Narain, 1992).

Often it is not possible to include all the variables in a single model. In such situations, composite forecast can be obtained as a suitable combination of forecasts obtained from different models. Various strategies for combining forecasts have been suggested under different situations. (S.C.Mehta, 2000).

2. Forewarning System for Crop Pests and Diseases
In pests and diseases forewarning system, the variables of interest could 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 parameters. However, depending upon the data availability, different types of models can be utilized for developing forewarning system. As in the case of yield forecasts, for pest and disease forewarning also the models could be of two types, 'Between year model' and 'Within year model'.

2.1 Between year models
If data are available in quantitative form, the different types of models used are:

**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°C
- the total rainfall for the last 10 days is > 3.0 cm
- the minimum temperature on that day is > 7.2°C

**Regression Model:** The regression model taking pest / disease parameter as dependent variable and suitable independent variables such as weather variables, crop stages, population of natural enemies/predators etc. is used. 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).

**Model based on weather indices:** This is a model similar to the one used in yield forecasting. The recommended models is of the form

\[ Y = a_0 + \sum_{i=1}^{P} a_i Z_i + \sum_{i \neq i'} b_{ii'} Z_{ii'} + e \]

where \( Z_i = \sum_{w=n_1}^{n_2} r_{iw} X_{iw} \); \( Z_{ii'} = \sum_{w=n_1}^{n_2} r_{ii'} X_{iw} X_{iw} \)
Y is variable to forecast; $X_{iw}$ is value of $i$–th weather parameter in $w$–th period; $r_{iw} / r_{i'w}$ is value of correlation coefficient between $Y$ and $i$–th weather parameter/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 and final periods for which weather parameters are included in the model; $e$ is error term. (Agrawal, et al. 2004).

If information on favourable weather conditions is known, subjective weights based on this information can be used for constructing weather indices.

**Principal component regression:** Forewarning models can be developed using the principal component techniques 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:** The methodology is similar to the one used for yield forecasting, replacing yield by the character depicting pests and diseases. (Johnson et al 1996)

**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:

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

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).

**Artificial Neural Network (ANN):** ANN provides an attractive alternative tool for forecasting purposes. ANNs are data driven self-adaptive methods in that there are few apriori assumptions about the models for problems under study. They learn from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown or hard to describe. After learning the data presented to them, 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. (Dewolf et al. 1997, 2000; Agrawal et al. 2004). 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 adjusted by taking deviations from this natural cycle. Appropriate model is the fitted using these deviations form natural cycle as dependent and weather as independent variables.

**Model for Qualitative data:** Sometimes quantitative information on pests and diseases is not available but is available in qualitative form such as occurrence / non-occurrence, low / medium / high. 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:

\[
P(E = 1) = \frac{1}{1 + \exp(-z)}
\]

where \(Z\) is a function of weather variables.

**Forecast / Prediction rule :**
- If \(P \geq 0.5\) more chance of occurrence of epidemic
- If \(P < 0.5\) probability of occurrence of epidemic is minimum


### 2.2 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. 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 is similar to yield forecast model as given in section 1.2.2 (Agrawal et al., 2004).

**References**


Agrawal, Ranjana, Jain, R.C. and Jha, M.P. 1983 : Joint effects of weather variables on rice yield, Mausam, 34 (2) :189-94.


Sarkar, J. (2002). Forecasting rice and wheat yield over different meteorological sub–divisions of India using statistical models. 56th Annual Conference of ISAS.