Data Mining and Wireless Sensor Network for Groundnut Pest/Disease Interaction and Predictions - A Preliminary Study

A. K. Tripathy[1*], J. Adinarayana[1+], D. Sudharsan[1#],
K. Vijayalakshmi[2], S. N. Merchant[3], U. B. Desai[4]

[1] Centre of Studies in Resources Engineering (CSRE), Indian Institute of Technology Bombay,
Powai, Mumbai-400076, India
[1*] *aktripathy@iitb.ac.in*
[1+] *adi@iitb.ac.in*
[1#] *sudharsan@iitb.ac.in*

[2] Department of Entomology, Agricultural College, ANGR Agricultural University,
Rajendranagar, Hyderabad-500030, India
[2] *vijayalkorlipara@yahoo.com*

[3] Department of Electrical Engineering, Indian Institute of Technology Bombay,
Powai, Mumbai-400076, India
[3] *merchant@ee.iitb.ac.in*

[4] Indian Institute of Technology Hyderabad,
Yeddumailaram, Andhra Pradesh-502205, India
[4] *ubdesai@iith.ac.in*

Abstract: **Data driven precision agriculture aspects, particularly the pest/disease management, require a dynamic crop-weather data. An experiment was conducted in semi-arid region of India to understand the crop-weather-pest/disease relations using wireless sensory and field-level surveillance data on closely related and interdependent pest (Thrips) – disease (Bud Necrosis) dynamics of groundnut (peanut) crop. Various data mining techniques were used to turn the data into useful information/ knowledge/ relations/ trends and correlation of crop-weather-pest/disease continuum. These dynamics obtained from the data mining techniques and trained through mathematical models were validated with corresponding ground level surveillance data. It was found that Bud Necrosis viral disease infection is strongly influenced by Humidity, Maximum Temperature, prolonged duration of leaf wetness, age of the crop and propelled by a carrier pest Thrips. Results obtained from the four continuous agriculture seasons (monsoon & post monsoon) data has led to develop cumulative and non-cumulative prediction models, which can assist the user community to take respective ameliorative measures.**

Keywords: Data Mining, Knowledge Discovery, Wireless Sensor Network, Precision Farming and Pest/Disease Management.

I. Introduction

In recent years, there has been a large infection area of crop diseases and insect pests as well as the degree of its seriousness, which caused enormous economic losses to the farming community. Crop losses due to pests and diseases are quite considerable, particularly in the Indian semi-arid conditions [20]. Weather plays an important role in agricultural production. Oilseed crops are more prevalent in the weather based fragile semi-arid agriculture systems. Among the oilseed crops, groundnut (peanut) crop is more prone to attack by numerous pests/diseases to a much larger extent than any other crops. Significant crop losses by these diseases have been reported from Australia, India and the USA [3], [18], [22].

Among the pests, Thrips species occur as a complex, starting from vegetative stage till the harvest of the crop. It damages the chlorophyll content of the leaf terminals. Besides causing direct damage to the crop, Thrips are known to cause more indirect damage by attacking as vectors of viral disease viz, viz. groundnut Bud Necrosis Virus (BNV). In India, the disease occurs with the incidence ranging from 0-98% [13], [28]. BNV infection in the young stage will result in death of the plant due to severe necrosis. Forecasting systems are based on assumptions about the pathogen's interactions with the host environment and the disease triangle. The objective is to accurately predict when the three factors - host, environment, and pathogen - all interact in such a fashion that disease can occur and cause economic losses [9], [14] (Figure 1). Critical threshold of the meteorological elements for the incidence, spread and intensification of pests and disease determined in the laboratory conditions have little relevance to the field condition. Therefore, they have to be determined and monitored under field conditions through simultaneous observation of micrometeorological parameters and the pertinent data on pest and disease [23].

Sensor network technology (wired or wireless) is a one of

the potential information, communication and dissemination (ICD) systems suitable for collecting the real time data on different parameters pertaining to weather, crop, soil and environment, which in turn helps in developing open solutions for majority of the agricultural processes. The wireless sensors are cheap enough for wide spread deployment in the form of a mesh network and also it offers robust communication through redundant propagation paths. Wireless Sensor Network (WSN) allows faster deployment and installation of various types of sensors as the network provides self-organizing, self-configuring and self-diagnosing capabilities to the sensor nodes. It is a system comprised of radio frequency transceivers, sensors, microcontrollers and power sources. They are relatively low-cost, consumes low-power, small devices equipped with limited sensing, data processing and wireless communication capabilities, which perfectly suites to precision agriculture where decisions are to be made at micro-climatic level at right time/place/input [24].
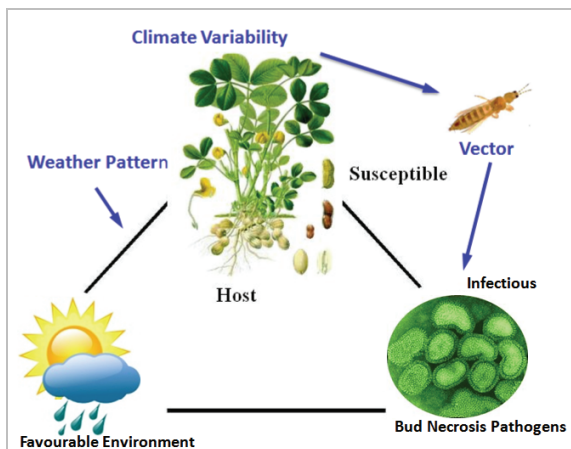

Figure 1. Pest/Disease Triangle modified from [19]

In the present scenario, agricultural data are virtually being harvested along with the crops and are being collected/ stored in databases. As the volume of the data increases, the gap between the amount of the data stored and the amount of the data analyzed increases. Such data can be used in productive decision making if appropriate data mining (DM) techniques are applied. DM allows to extract the most important information from such a vast data and to uncover previously unknown patterns and hidden relationships within the data that may be relevant to current dynamic agricultural problems [1]. With the ever-increasing amount of information about their farms, farmers are not only harvesting in terms of agriculture output but also a large amounts of data. These data should be used for optimization [10], [26].

Researches on utility of macroclimatic data on precision protection has been carried out at [17], [27], [28], however, there have been a few studies concerning use of WSN in pest/disease management. Mote based AgriSens were used to test the feasibility of capturing and analyzing data and facilitated global data accessibility from multiple wireless sensor pods to study the efficient irrigation as well disease forecasting for grape vineyard [24]. [17] Discussed through a WSN named COMMON-Sense Net that monitors several

environment parameters and is deployed in an Indian semi-arid region for the benefit of small and marginal farmers to provide better diagnosis for better crop management. 'U-Agri' from Centre for Development of Advanced Computing (C-DAC), Hyderabad has developed low cost sensor networks, which encompass the farm environment and provide macro and micro climate information on groundnut crop for a Decision Support System for groundnut pest Leaf Miner and disease Leaf Spot [27]. However, U-Agri does not address the hidden correlation of weather-pest/disease-crop and the vector pest (Thrips) and disease (BNV) interactions. It is essential that an efficient methodology should be capable of forecasting the pest and disease dynamics accurately. Thus, there is a need for development of a viable and functionally realistic model to correlate pest/disease with weather and surveillance data.

In the present study, micro-level weather data (Temperature, Humidity and Leaf Wetness) were obtained through Mote based AgriSens WSN distributed sensing system; and DM techniques and surveillance data have been used to understand and quantify hidden correlations in the crop-pest/disease-weather continuum. Subsequently, one week as well as a cumulative prediction models for (Thrips, BNV) have been developed with which one can develop a Decision Support System (DSS) with multi-season data.

In this work, an attempt has been made to develop a viable model for groundnut pest/disease dynamics using the state of the art data mining techniques to understand the hidden correlations (crop-pest and disease–meteorological continuum) and there by development of an empirical as well as Multivariate Regression Models. Initial experimental results revealed interesting crop-weather correlations that helped in generating a multivariate regression model for Thrips and an empirical model for BNV disease in association with carrier pest (Thrips). Results obtained from four consecutive agriculture seasons have been validated with repeated long-term ongoing experiments.

II. Material and Methods

In order to study the crop-weather-pest/disease interactions, a test bed for WSN experiments was chosen at Agriculture Research Institute (ARI) of A.N.G.R. Agricultural University, Hyderabad falling in semi-arid tropic region. The test bed, where a long term weather-based experiments are being carried out on groundnut crop, will provide a platform for validation of proposed model with ground level surveillance studies. This work is a part of Indo-Japan initiative to develop a real time decision support system called GeoSense [25], [26], integrating Geo-ICT and WSN for Precision Agriculture.

A. Standard Experimental Setup

A standard field experiment design was laid out in the test bed during Kharif (monsoon) (2009 and 2010) and Rabi (2009-10 and 2010-11) to study the crop-weather-pest/disease interactions (Figure 2). Four different dates of sowing (D1, D2, D3 and D4) were taken in to consideration (Table -1). These different dates will determine the impact of pest and disease incidence in order to observe dynamics in pre (D1) and post normal weeks of sowing (D4). D2 and D3 are normal

dates of sowing. Two protection (P) treatments were included in the experiment where, P1 stands for unprotected plot (a normal situation in farmer's field) and P2 stands for weather based protection plots. Apart from this, to have uniform and unbiased observation, surveillance data has been collected from each plot in randomly selected one square meter area locations of the plot (S1D1, S1D2, S1D3, S1D4, S2D1, S2D2, S2D3, S2D4,………) through flowering to harvesting phenological stages (Figure 2) in three replicated plots (R1, R2 and R3).

Table 1: Different Dates of Sowing groundnut crop in test bed

Sowing Date	D1	D2	D3	D4
Kharif 2009	07/07/2009	21/07/2009	07/08/2009	22/08/2009
Rabi 2009-10	12/10/2009	26/10/2009	10/11/2009	27/11/2009
Kharif 2010	19/06/2010	03/07/2010	19/07/2010	04/08/2010
Rabi 2010-11	14/10/2010	29/10/2010	16/11/2010	04/12/2010

B. Surveillance Data Collection

Thrips population dynamics (surveillance data) were obtained at every week from flowering to reproductive pheonological stages, where majority of pest and disease incidences occur, at various locations in the experimental site.
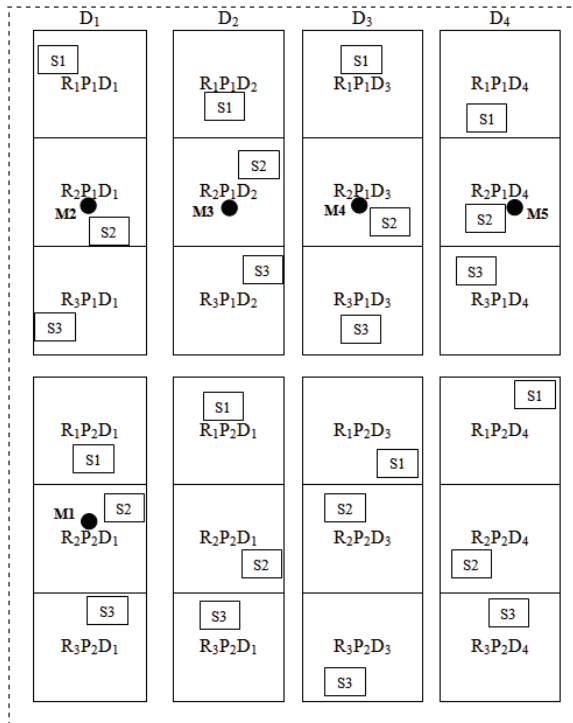

Figure 2. Experimental Layout

The surveillance data has been collected weekly instead of daily as there won't be any significant visible changes in pest/disease incidences. A total of 48 observations (12 X 4) have been made with respect to different dates of sowing in each season. Along with this, Groundnut crop age (that is at which stage of the crop the pest/disease attack takes place and their dynamics trends) was also recorded weekly to understand the infection dynamics of Thrips. The weeks in a year are mapped into integer values by considering the first week of January as first standard week.

Five sticks were placed in each one square meter and named 1, 2, 3, 4 and 5 (Figure 3). Subsequently data has been collected manually from the plants adjacent to each stick for both Kharif (2009 & 2010) and Rabi seasons (2009-10 and 2010-11). The surveillance data like number of leaf-lets that Thrips have punctured/infected/visited, how many plants infected in the one square meter area, date of flowering, date of recording the surveillance data, etc. The plants have been numbered in the form of ID like $R_1P_1D_1$-1, $R_1P_1D_1$-2, $R_1P_1D_2$-1, etc. so as to retrieve their respective data easily from data base. This is the one of the standard design practice that has been observed in the long term experiment in the test bed.

Table 2: Groundnut Thrips foliage damage Score 0-5 (modified from [22])

Score	Description	Damage Severity (%)
0	No damage	0
1	Negligible damage	1-10
2	Low level damage	11-30
3	Moderate damage	31-60
4	Severe damage	61-80
5	Heavy damage	81-100

Figure 3: Surveillance data collection from One square meter area at Flowering stage.

C. Sensory Data collection

AgriSens-based sensory data was collected from the field and transmitted through General Packet Radio Service (GPRS) to the remote server for data storing, analysis and mining. The data was stored in an OpenSource data base (PostgreSQL) [16] for further analysis.

The deployed WSN system consists of the battery-powered nodes equipped with sensors for continuously monitoring agricultural/weather parameters such as temperature, relative humidity, soil temperature and leaf wetness [24]. Figure 4 shows the schematics of wireless sensor network with agricultural / environment sensors deployed in the field. Each node was able to transmit/receive data packets to/from other nodes every 15 minute over a transmission range of 25 meter. Data collected by the sensors were wirelessly transferred in a multi-hop manner to a base station node (stargate) connected with embedded gateway for data logging. In a WSN, when the transmission range of a sensor node is not sufficient, it uses multi-hop communication to reach the destination node or

sink node. This data forwarding mechanism continues till it reaches the sink node.

The base station has a GPRS connectivity through which it routes data to the GeoSense server setup at Agro-Informatics Lab at CSRE, IIT Bombay and collect all the sensory information. The sensory data coming to the server through GPRS is raw and has been converted in to a usable format in real time through appropriate conversion formula (by using open-source server-side scripting language PHP) in the server end [20]. Both raw data and the real-time data have been stored in different database for further analysis and mining. Other related weather data (sunshine hours SH, wind speed WS, rainfall RF and evapotransipiration ET) were obtained from the weather station situated within the vicinity of the test bed (Figure 5).
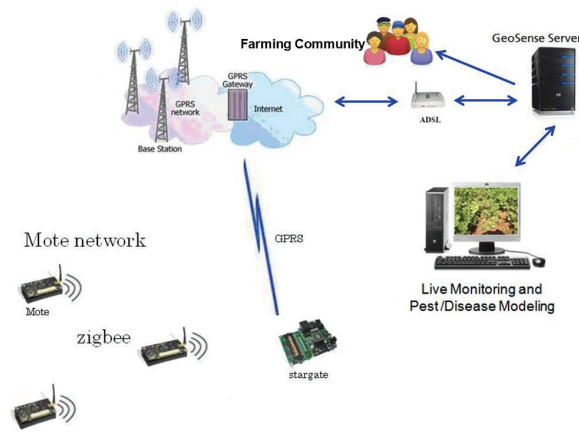

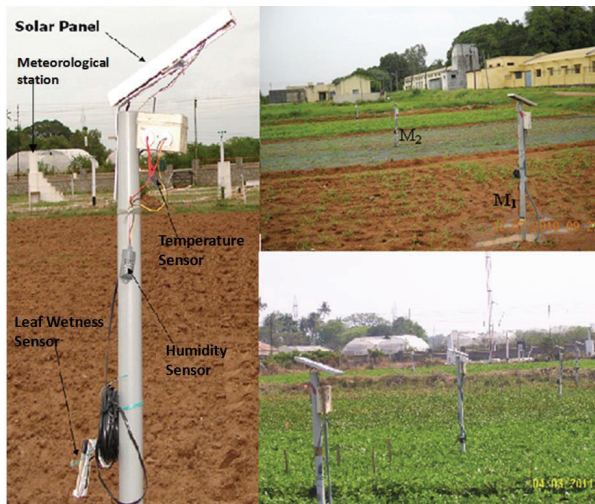**Figure 4**: WSN architecture in the experimental

Figure 5. Agrisens sensor network deployed in the field

Sensory data was collected during the both Kharif (monsoon) seasons as well as Rabi (post monsoon) seasons with multiple sensor nodes, i.e. M1, M2, M3 and M4 for four different dates of sowing (D1, D2, D3 and D4) in four different test plots, respectively in groundnut field (Figure 5).

D. Data Mining and Statistical Models

Various data mining techniques and a few algorithms were used/developed to understand the pest dynamics and the general processing flow is depicted in Figure 6.

Raw sensory data obtained from the experimental field is not uniform in its collection. Owing to the climatic conditions

or non-function of field sensors or due to network errors, there have been a few breaks in collecting continuous sensory data that may lead to biased outcomes while developing the model. Expectation–Maximization (EM) algorithm was used to deal with missing data [11]. Relative data from nearest sensor node was used to fill the missing data with EM algorithm. The data set is provided in daily and weekly means wherever is required. Quality data is accomplished by performing satisfactory data pre-processing (data selection, data reduction and elimination of null values or other noise values). Though the real-time data was collected at 15 minute interval, all such data were not used for the current experiment. For example, the temperature has been used by computing maximum and minimum temperature of respective day; Relative Humidity has been taken as RH1 and RH2 having recorded data at morning 7:30 am and afternoon 3:00 pm (which is a standard practice in Indian agriculture [3]. Leaf Wetness (LW) data has been used in the scale of 1 to 10 as Leaf Wetness Index, and the Leaf Wetness Index value above 5 has been taken as leaf that is wet for favourable pest/disease conditions and, hence, computed for wetness period [24].
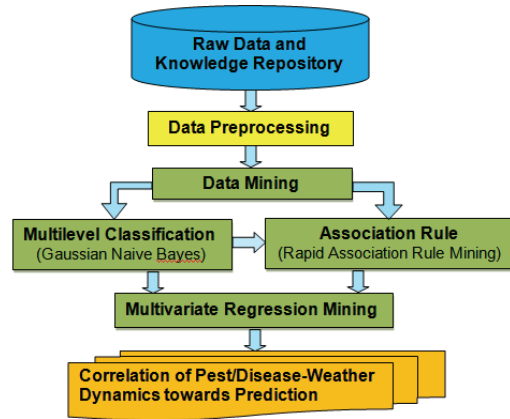

Figure 6. DM Processing flow for Pest/Disease Dynamics[25]

As groundnut crop was infested with multiple pests and diseases, multi-level classification modules were developed in the model, which classifies crop pests & diseases based on the severity. Gaussian Naive Bayes classification [8], [15] was used in the experiment.

Bayesian network principle was used to model uncertainty by combining experimental knowledge and observational evidences. Gaussian Naive Bayes (NB) classifier (NB classifier with Gaussian distribution), which is a term in Bayesian statistics dealing with a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence, was used to assume the presence (or absence) of a particular feature of a class unrelated to the presence (or absence) of any other feature [15].

Rapid association rule mining has been used in association with classification techniques to find out correlation of multiple weather parameters with respect to Thrips/BNV. This phenomenon is to identify signature patterns as well to discover their presence/dependency [1], [4] with other related pest and/or weather parameters. This algorithm helps in

discovering effects of pest/disease over the other pest/disease with respect to weather parameters. For example, how the Thrips incidence occurred due to presence/absence of BNV or vice versa. The outcomes are in the form of correlation indexes scaling from -1 to 1.

Regression Mining is a data mining (machine learning) technique used for developing multivariate equation for the training dataset. Following are the multivariate regression equations developed by using XLminer (a data mining tool) [30].

Thrips-

$$Y_{TH} = -4.84 + 1.23 *Tmax- 0.78 *Tmin-0.11 * RH1+ 0.25 * RH2 - 5.38 * RF - 2.05 * SH - 0.59 * WS - 0.28 * ET + 1.56E-02 * AC \quad (1)$$

Where, Tmax is the Maximum Temperature, Tmin is the Minimum Temperature, RH1 and RH2 are the Relative Humidity in morning and afternoon, RF is Rainfall, RD is Rainy days, SH is Sunshine Hours, WS is Wind Speed, ET is Evapotranspiration and AC is the age of crop.

An empirical model [29] has been adapted and modified by taking Temperature, Humidity and Leaf wetness factors. Later, in association with the carrier pest Thrips, the infection index of BNV was calculated:

$$I = f (W, T, RH1, RH1) = a (1- exp \{-[b (W- c)]^d\}) \quad (2)$$

Where, W = Cumulative Wetness duration (obtained from Leaf Wetness Sensor), T= Temperature, b = characterize the intrinsic rate of increase (0 to 1) of Incidence with respect to W, d = the rate of acceleration, c = characterizes the lag period (0 to ∞) before the response of I to W begins, I = Infection index for BNV. RH1 & RH2 = Relative Humidity at 7:30 am and 2:30 pm respectively, a = scale of response to W and varies with temperature and RH2

$$a = f(T, RH2) = \frac{e'\{\exp[f(T-g)/RH_2 +1]}{\{1 + \exp[f(T-g)]\}} \quad (3)$$

where, g = Optimum Temperature for GBN.
f = a parameter which is the intrinsic rate of change of Temperature with respect to optimal temperature (0 to 1)

$$e' = e\frac{[RH1 - RH2]/[T_{max} - T_{opt}]}{RH2} \quad (4)$$

in which e characterizes the scale of the response to T and RH. The BNV virus attack infection in association with pest Thrips can be modeled as

I' = f(RH, Leaf Wetness, Crop Age, Thrips)

$$I' = \frac{dI}{dI_{th}} + I + \frac{dI_{th}}{dW_{th}} + AGDD \quad (5)$$

Where, I' = Infection Index, I = GBN Incidence, I_{th} = Thrips Incidence, W_{nd} = Wind Speed and AGDD is the accumulated growing degree days (Experimental Temperature – Lower Development Threshold) * Development Time in days + GDD).

Empirical as well as multivariate prediction model, thus developed, has been used by taking historical data and was used for one week prediction. For example, one week prediction of Kharif 2010 season Leaf Spot incidence level has been predicted for one week prediction by using Kharif 2009 data and similarly for Rabi 2010-11 seasons one week prediction was carried out with 2009-10 Rabi historical parameters.

In addition, Complex Polynomial Cumulative model [2] was adapted and modified for pest/disease forecasting by including various aspects viz. maximum pest population/disease severity, time of first appearance, time of maximum pest population/disease severity as well as life cycle, season, weather parameter, crop phenological stages, incidence at flowering stage, growing degree days, previous year record, correlation with other pest/disease, previous season Crop:

$$Y = a_0 + \sum_{i=1}^{p} \sum_{j=0}^{1} a_{ij}Z_{ij} + \sum_{i\neq1}^{p} \sum_{j=0}^{1} b_{ii'j}Z_{ii'j} + e \quad (6)$$

Where $Z_{ij} = \sum_{w=n1}^{n2} r_{iw}^j X_{iw}$ and $Z_{ii'j} = \sum_{w=n1}^{n2} r_{ii'w}^j X_{iw}X_{i'w}$

Z_i's and Z_{ij}'s are the independent variables which are functions of the basic weather variables like maximum temperature, relative humidity, leaf wetness, etc., Y= variable to forecast, X_{iw} = Value of ith weather variable in wth week, r_{iw} = Correlation coefficient between Y and i[th] weather variable in wth week, $r_{ii'w}$ = Correlation coefficient between Y and product of X_i and $X_{ii'}$ in wth week, n1= initial incidence and n2 = fist Peak population week.

Pest/Disease life cycle plays an important role for their prediction. For example, the first BNV incidence in a particular date of sowing and then cumulative affect with respect to Thrips population increase or decrease has carried out by taking the life cycle of Thrips vector. Though the Thrips population increases rapidly, its impact (on BNV incidence) will be seen two weeks later, i.e. after going through three stages i.e. larva (picking the BNV virus), pupa (an inactive stage around one week) and then adult Thrips transmit to the plant. Hence, the high BNV infection value could be the cumulative result of the virus acquired two week prior to the present infection level. Moreover, if BNV infection has occurred in the initial stage of the plant then the plant will die if proper care has not been taken. However, if the infestation is in the later half then the plant survives but with low yields.

III. Results and Discussions

The different life cycles of plant pathogen diverse groups of organisms and their different interactions with host plants produce a wide range of responses to environmental and climatic drivers. Viruses may be present in hosts while symptom expression is dependent on temperature [6], [7]. Some relationships between climate/micro-climate and disease risk are obvious such as some pathogens' inability to infect without sufficient moisture (i.e. dew or rain droplets or rain fall) [12] or other pathogens' or vectors (e.g. Thrips)' ability to carry when temperatures go above a critical level. For example, a given pathogen (e.g. groundnut bud necrosis virus) may only be able to infect its host(s) when the plants are in certain (flowering to reproductive) developmental stages.

Climatic/meteorological features such as temperature, humidity and leaf wetness are important drivers of disease, and inappropriate levels of these features for a particular disease may be the limiting factor in disease risk [12], [7].

The crop-weather-pest/disease interaction results discussed here are from the D2 only, as it is a normal practice and coincide with farmers conditions, where immediate solutions/predictions would help the farming community for strategic decisions.

Correlation of Thrips and BNV with the help of weather and surveillance data (including crop age), from all four (continuous) seasons, were discovered and quantified using DM techniques in the test bed.
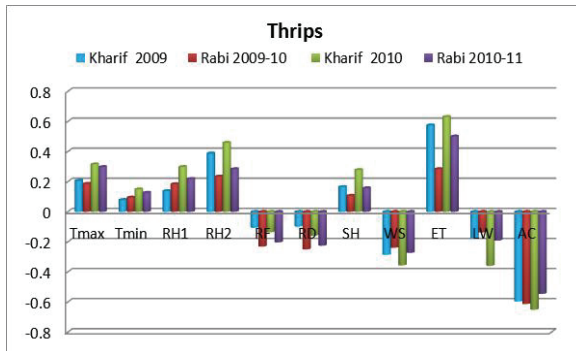

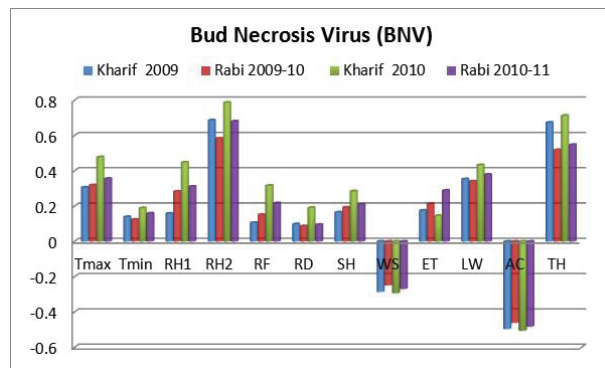**Figure 7.** Correlation Index Values for Thrips

Figure 8. Correlation Index Values for BNV

Where,
Tmax = Maximum Temperature
Tmin = Minimum Temperature
RH1 = Relative Humidity at 7:30 am.
RH2 = Relative Humidity at 3pm.
RF = Rain Fall
RD=Rain Days
WS=Wind Speed
LW = Leaf Wetness
SH = Sunshine Hours
ET = Evapotranspiration
AC= Age of the Crop
TH= Thrips

A correlation Index matrix was obtained by using Rapid Association Rule Mining algorithm, in which correlation index with of pest/disease with weather parameter were quantified in the range of -1 to +1. The correlation values (both positive and negative) of predictor (T_{max}, RH1, ET, etc.) versus target Thrips and BNV infection index were obtained from various datasets (sensory, weather-station and surveillance) during flowering to harvesting stages of all four seasons are shown in Figure 7 and 8, respectively.

Table 3 shows overall interpretation of correlation (Thrips/BNV/Weather/Crop Age) values with negligible, moderate and strong level drawn from all four seasons. Based on concept of infection index with respect to pest/disease risk model [5], [21], correlation index greater than 0.5 in the scale of -1 to +1 has been considered as strong +ve, whereas -0.5 to more for strong negative correlation.

Table 3. Overall Interpretation of Correlation index values (modified from [5], [21])

Correlation value (- 1 to +1)	Correlation Levels (e.g. Thrips with RH1)	Thrips	BNV
0.0 to 0.1 1.0 (+ve) 0.0 to - 0.1 (-ve)	Negligible or No correlation	Tmin RD RF	Tmin
> 0.1 to 0.5	Moderate (+ve)	Tmax, RH2, SH, RH1	RH1,SH, Tmax, ET, LW, RF, RD
< - 0.1 to - 0.5	Moderate (-ve)	WS, LW	WS
> 0.5 to 1 (+ve)	Strong (Good Correlation)	**ET**	RH2
< -0.5 to – 1 (-ve)	Strong (Good Correlation)	**AC**	**AC**

It was found that Relative Humidity has strong positive correlation and AC has strong negative correlation with BNV. In case of Thrips, ET and AC found to be strongly correlated. Apart from this, it was also observed that there is a strong correlation with pest Thrips and disease BNV with an index value of 0.674 (for Kharif 2009), 0.519 (for Rabi 2009-10), 0.713 (for Kharif 2010), 0.548 (for Rabi 2010-11), respectively in the D2 sowing plot (Figure 8). The higher pest/disease correlation index value has seen in Kharif as compared with Rabi season is due to high humidity that causes the BNV infestation.

With correlation studies revealing the crop-weather-pest/ disease relationship/ interactions, there is a possibility of developing an early warning models (Cumulative and non-cumulative) on pest/disease infestations. Preliminary prediction computations have been carried out and presented in graphical format for Kharif 2010 (Figure 9 for Thrips and Figure 10 for BNV) and Rabi 2010-11 (Figure 11 and 12 for Thrips and BNV) from D2 date of sowing. These computations have been carried out for near equal to peak period only as rest of the prediction has a less pest/disease management significance if the crop is already infected during these peak stages. MVR model is the graph obtained by empirical computation from above mentioned weather parameters. ARI graph is the obtained from the ground level surveillance data. It is observed that from Figures 9 to 12 that MVR model appears to be close to the ground based

observations. Thrips pest incidence increase with increase in age of crop up to the reproductive stage and then decreases. Rabi season found to be more Thrips incidences (24.11%) in comparison with Kharif season (18.29%). However, in case of BNV, Kharif season found to be more prone to the disease incidence due to high humidity, which plays a major role on this disease. Rapid BNV incidences were observed with increase in humidity beyond 80% RH2.

Figure 9. Thrips Incidence Prediction for *Kharif* 2010

Figure 10. BNV Incidence Prediction for *Kharif* 2010

Figure 11. Thrips Incidence Prediction for *Rabi* 2010-11

It has been observed in all the cases that one week (1wk) prediction is very closer to the regression/empirical model (MVR), where as the cumulative (CWK) method is probably a preferable prediction strategy as it is closer to the ground level data. The impact of carrier pest Thrips on BNV was more prone after two weeks (starting from the first BNV

initiation) as the virus acquired first week of the incidence from an infected plant and multiplies in the vector in 2nd week. Then the transition happen consequently when it come in contact with the normal plant. For example, 17.75% BNV incidence on 17/8/2010 comes up with 49.34% on 8/9/2010 (Figure 10).
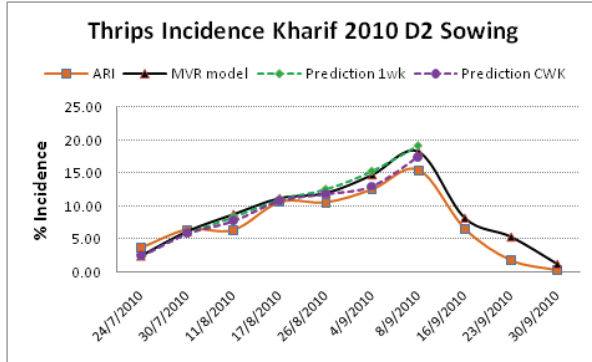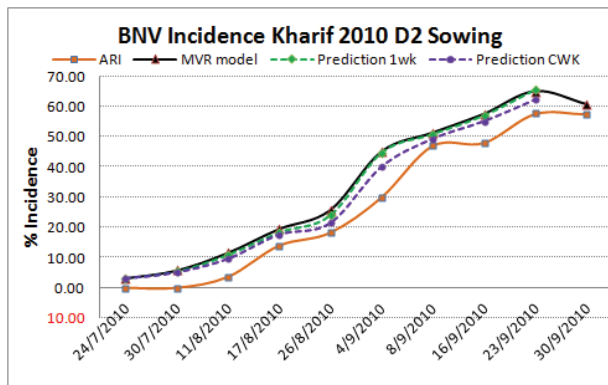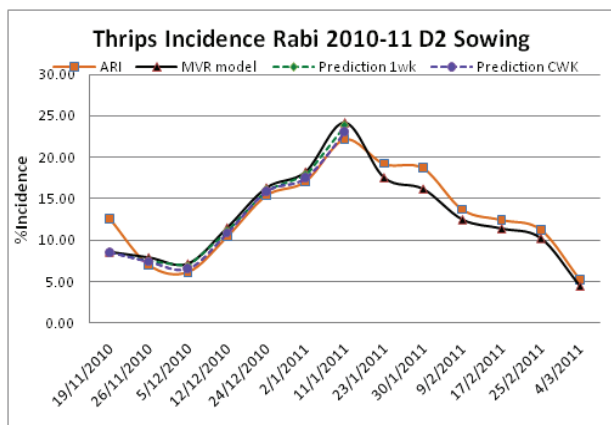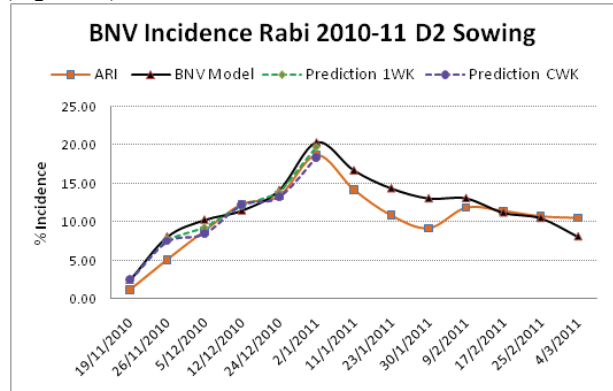

Figure 12. Thrips Incidence Prediction for *Rabi* 2010-11

These pest/disease interactions indicate that if Thrips pest are controlled on 17/8/2010, there is a possibility to counter BNV disease and its yield loss during normal (D2) sowing date. Overall, the CWK value has been observed to be in the range of 2 to 5% increased value with respect to ground level data. However, in case of 1wk prediction approach, an increase of 5 to 10% in prediction value was observed as compared to ground level data. Thus, the preliminary results show that the CWK model appears to be a better prediction method.

Figure 13. Thrips pest (in microscope) and its Incidence in the test bed

Figure 14. Different stages of BNV Incidence in the test bed

Similar trends were also observed during 2010-11 Rabi season, and proper care should be taken after appearance of

first peak incidence of Thrips. Figure 13 and 14 depicts the field photos of Thrips and BNV incidences, respectively. Four different stages of BNV incidence was captured and shown in Figure 14.

IV. Conclusion

An attempt has been made to understand the hidden relationships between most prevailed and interrelated disease (BNV) / pest (Thrips) and weather parameters of Groundnut crop. WSN was established in the test bed to obtain real-time weather parameters (Temperature, Humidity and Leaf wetness) at micro-climatic level and a few other related weather parameters were taken from the nearby weather station. The crop-weather-pest/disease dynamics and hidden relations were obtained and quantified using DM techniques. The statistical approach together with regression mining based correlations helped in developing multivariate regression model that have been used to develop an empirical prediction model (non-cumulative) to issue the forecast for population buildup, initiation & severity of pest /disease. However, this is a preliminary investigation limited to one date of sowing and this has to be experimented/ validated continuously with different dates of sowing in assertion with long term experiments. Apart from this, a cumulative prediction model has been developed; this will lead to the development of Decision Support System for Pest/Disease Prediction. This will help to take strategic decisions so as to save the crop from pest/disease affects and improve the crop yields and environmental conditions.

Acknowledgment

References

[1]   Agrawal, R., Imielinski, T., Swami, A.N. "Mining Association Rules between Sets of Items in Large Databases". *SIGMOD,* June, 22(2): pp. 207-16, 1993.

[2]   Agrawal R. and Mehta S.C., Weather Based Forecasting of Crop Yields, Pests and Diseases - IASRI Models, *Journal of Indian Society of Agricultural Statistics*, 61(2): pp. 255-263, 2007.

[3]   AgroMet-Cell: *Annual Report of Entomology Work 2008-09*. AgroMet-Cell, Agriculture Research Institute, ANGRAU, Hyderabad, pp. 57-69, 2009.

[4]   Das A., Ng W. K. and Woon Y. K. "Rapid association rule mining", *Proceedings of the tenth international conference on Information and knowledge management*

[5]   Das, I, Naveen, C.P.R.G., Yadav S S., Kodilkar, A A., Shah, N.G., Merchant, S.N. and Desai, U.B. "WSN Monitoring of Weather and Crop Parameters for Possible Disease Risk Evaluation for Grape Farms - Sula Vineyards- A Case Study", *Proceedings of Conference on Geospatial Technologies and Applications, Geomatrix- 2009,* IIT Bombay, Mumbai, 27-29, 2009.

[6]   DeBokx J., Piron P. Potato Res. 20 pp.207–213, 1977. (Accessed on 24 May, 2011)
      www.k-state.edu/pdecology/GarrettNita2009.pdf

[7]   Garrett, K.A., Nita, M., Wolf, E. D. D., Gomez, L. and Sparks, A.H. "Plant pathogens as indicators of climate change", *Climate and Global Change: Observed Impacts on Planet Earth*. T. Letcher, editor. Elsevier. Pp. 425-437, 2009.

[8]   George H. J. and Pat L. "Estimating Continuous Distributions in Bayesian Classifiers". *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Mateo. pp.338-345, 1995.

[9]   George N. A. "Plant Pathology", *Elsevier Academic Press. 5th ed*, ISBN 978-0120445653, 2005.

[10] Georg R., Rudolf K., Martin S., Peter W. "Data Mining with Neural Networks for Wheat Yield Prediction". P. Perner (ed.), *Springer LNAI*, Vol. 5077, pp. 47-56, 2008.

[11] Gupta M. R. and Chen, Y. "Theory and Use of the EM Algorithm", *Foundations and Trends in Signal Processing*", Vol. 4, No. 3, pp. 223–296, 2010.

[12] Huber L., Gillespie T.J., Modeling leaf wetness in relation to plant disease epidemiology, *Annual Review of Phytopathol*, 30 (1992) 553–577.

[13] Kendre M. S., Patange N. R., Neharkar P. S. and Telang S.M. "Occurrence of Thrips plami, a vector of bud necrosis disease of groundnut in Marathwada region", *Journal of Soils and Crops*, 10(2), pp. 226-230, 2010.

[14] Madden L., Gareth H, Frank V, Den B. "Study of Plant Disease Epidemics". *American Phytopathological Society*. ISBN 978-0890543542, 2007.

[15] Mitchell T. M. "Generatvie and Discruninative Classifier: Naive Bayes and Logistic Regression", *In Machine Learning, McGraw Hill* (Draft of January 19, 2010), pp. 1-17, 2010.

[16] PostgreSQL, PostgreSQL Global Development Group, (Accessed on 08 July 2009) http://www.postgresql.org/

[17] Prabhakar T.V, Jamadagni H. S., Sahu A, Prasad R. V. "Lessons from the Sparse Sensor Network Deployment in Rural India", *Proceeding of 11th International Conference on Distributed Computing and Networking (ICDCN' 10)*.Springer LNCS, Vol- 5935, pp. 104-115, 2010.

[18] Prasad R.D.V.J. "Thrips Transmitted Viral Diseases in Groundnut and Management", *Proceeding of workshop Changing Insect Pest Scenario and Management strategies in Different Crops*, Hydearbad, pp. 89-97, 2009.

[19] Olatinwo, R., Paz J. and Hoogenboom G. "Application of climate and weather information in peanut disease management", *Biological and Agricultural Engineering*

(CKIM'01), November 5-10, 2001, Atlanta, GA, USA, pp. 474-481, 2001.

Department University of Georgia, Griffin. BAE seminar 10 February 2009, (Accessed on 27-Janauary 2012). www.griffin.uga.edu/ageng/Rabiu.pdf

[20] RajiReddy, D. and Ravindra, B. "Groundnut Insect Pests, Diseases and Nutritional Disorders", *Acharya N.G. Ranga Agricultural University, Information Bulletin* No.4, pp. 20-36, 2008.

[21] Royle, D.J., Burr, D.J. "The Place of Multiple Regression Analysis in Modern Approaches to Disease Control", *EPPO Bulletin*, 9(3), 155-163, 2008.

[22] Singh F. and Oswalt D.L. "Major Diseases of Groundnut", *ICRISAT-skill development series*, No.6, pp. 50-36, 1992.

[23] Shekh, A.M., Patel, H.R., and Pandey, V. "Predicting the outbreak and spread of pests and diseases using weather parameters", *Dept. of Agri. Meteorology, Gujarat Agriculture University Anand, Annual Research Report*, Vol.2, pp. 249-255, 1999.

[24] SPANN Lab. WSN project report 2009, *Electrical Engineering Department, IIT Bombay,* Available at, Http://www.ee.iitb.ac.in/spann/agrisens/ (Accessed on 23 March 2009).

[25] Sudharsan D., Adinarayana J., Tripathy A. K., Desai U. B., Merchant S. N., Ninomiya S.,Hirafuji M., Kiura T. and Tanaka K. "GeoSense: Geo-ICT and Wireless Sensor Network based Dynamic Real-time System for Precision Agriculture", *Proceeding of CIGR 2011 Conference*, September 19-23, 2011, Tokyo, Japan, 2011.

[26] Tripathy, A.K., Adinarayana, J., Sudharsan, D., Merchant, S.N., Desai, U.B., Vijayalakshmi, K., Reddy, D.R., Sreenivas, G., Ninomiya, S., Hirafuji, M., Kiura, T., Tanaka, K. "Data Mining & Wireless Sensor Network for Agriculture Pest/Disease Preiction", *Proceedings of 2011 World Congress on Information and Communication Technologies (WICT 2011)*, 11-14 December 2011, Mumbai, India. Pp: 1233-1238. @IEEE, DOI: 10.1109/WICT.2011.6141424, 2011.

[27] U-Agri, Ubiquitous Agriculture, *CDAC, Hyderabad Project Report* (accessed on 21 February 2009) http://www.ubicomp.in/uagri

[28] Vijayalakshmi K., Reddy Raji D., Varma N.R.G. and Pranuthi G. "Weather based pest and disease forwarning models in groundnut in the context of climate change", *Workshop Proceedings: Impact of Climate Change on Agriculture, ISPRS Archives XXXVIII-8/W3*: pp. 48-50, 2009.

[29] Wu, L., Damicone, J. P., Duthie, J. A., and Melouk, H. A. "Effects of temperature and wetness duration on infection of peanut cultivars by *Cercospora arachidicola*", *Jounal of Phytopathology*, 89: pp. 653-659,1999.

[30] XLminer- *a Cytel Software*, (accessed on 17 Sept. 2009) http://www.resample.com/xlminer/index.shtml

Author Biographies

A. K. Tripathy is a Ph.D. scholar at Centre of Studies in Resource Engineering (CSRE), India Institute of Technology Bombay, Mumbai, India, working on Data Mining & Wireless Sensor Network in Agriculture Pest/Disease Dynamics and Prediction. He has received M.Tech. degree in Computer Science & Engineering from Motilal Nehru National Institute of Technology, Allahabad, India in 2004. He has been working on precision protection aspects as a part of Indo-Japan Initiative research project GeoSense: GeoICT and Sensor Network based Decision Support System in Agriculture and Environment Assessment since 2009. He is a member of ACM. His research interests include Data Mining, Machine Learning, Sensor Networks, ICT for Rural Developments.

J. Adinarayana is an Associate Professor at Centre of Studies in Resources Engineering (CSRE), Indian Institute of Technology Bombay, India. He has received Ph.D. from Banaras Hindu University, Varanasi, India in 1984. He is the Vice President of the Asian Federation for Information Technology in Agriculture (AFITA) for 2010-12 and is the Vice-Chair of the Asia-Pacific Advanced Network Agriculture Working Group for 2012-14. His main area of research includes Agro-informatics and Rural Developments, Sensor Network in Precision Agriculture, Watershed Management Information System, etc. and has published extensively in these areas. He serves on editorial boards of various journals and numerous conference program committees.

D. Sudharsan is a Ph.D. scholar at Centre of Studies in Resource Engineering (CSRE), Indian Institute of Technology Bombay, with extensive experience and management skills in Geographical Information and Communication Technology and Wireless Sensor Network System technologies. Mr. Sudharsan has M.Sc. in Geology from the Presidency College, Chennai, Tamilnadu, India (2004) and M.Tech in Geoinformatics from Bharathidasan University, Trichy, Tamilnadu, India (2006). He developed a cost effective dynamic real time sensory network/ communication/ database management based decision support system as a part of doctoral research.

K. Vijayalakshmi is an Associate Professor in Department of Entomology, College of Agriculture, A.N.G.R. Agricultural University, Hyderabad, India. She has received Ph.D. from A.P. Agricultural University, Hyderabad, India in 1995 under CSIR and ICRISAT fellowship. She has served as an Entomologist in a project from University of Wageningen, Netherlands for a year during 1993-94 on screening of genotypes of groundnut upon peanut bud necrosis virus at ICRISAT. She has served as Senior Scientist (Entomology) at Agromet-Cell, A.N.G.R. Agricultural University, Hyderabad and worked on impact of climate change on pests/diseases. Her broad areas of interest include Entomology, Climate Change in Pest/Disease, Forewarning Models, etc. and have published more than 50 publications extensively in these areas.

S. N. Merchant is a Professor in Department of Electrical Engineering, IIT Bombay. He has received his B. Tech, M. Tech, and PhD degrees all from Department of Electrical Engineering, Indian Institute of Technology Bombay, India. He has more than 25 years of experience in teaching and research. Dr. Merchant has made significant contributions in the field of signal processing and its applications. His noteworthy contributions have

been in solving state of the art signal and image processing problems faced by Indian defence. His broad area of research interests are signal and image processing, multimedia communication, wireless sensor networks and wireless communications, and has published extensively in these areas. He has been a chief investigator for a number of sponsored and consultancy projects. He is a Fellow of IETE. He is a recipient of 10th IETE Prof. S. V. C. Aiya Memorial Award for his contribution in the field of detection and tracking. He is also a recipient of 9th IETE SVC Aiya Memorial Award for 'Excellence in Telecom Education'.

U. B. Desai is the Director of IIT of Hyderabad. He received the B. Tech. degree from Indian Institute of Technology, Kanpur in 1974, M.S. degree from the State University of New York, Buffalo, in 1976, and the Ph.D. degree from The Johns Hopkins University, Baltimore, U.S.A., in 1979, all in Electrical Engineering. Prior to being the Director of IIT Hyderabad, he was a Professor in Electrical Engineering Department at IIT Bombay, India. His research interest includes wireless communication, sensor networks, statistical signal processing, etc. He has a special interest in the development of communication and other technologies that can be exploited for the betterment of rural India. Dr. Desai is a senior member of IEEE. He is a Fellow of INSA and INAE. He is on Board of Tata Communications Ltd. and on the Technology Advisory Board of Microsoft Research Lab. India.