Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 72(3) 2018 213–223

Seasonal Dynamics of Sterility Mosaic of Pigeonpea and its Prediction using Statistical Models for Banaskantha Region of Gujarat, India

Ranjit Kumar Paul[1], S. Vennila[2], Narendra Singh[3], Puran Chandra[2], S.K. Yadav[2], O.P. Sharma[2], V.K.Sharma[2], S. Nisar[2], M.N. Bhat[2], M.S. Rao[4] and M. Prabhakar[4]

[1]*ICAR- Indian Agricultural Statistics Research Institute, New Delhi*

[2]*ICAR- National Research Centre for Integrated Pest Management, New Delhi*

[3]*S.D. Agricultural University, Sardar Krushinagar*

[4]*ICAR- Central Research Institute for Dryland Agriculture, Hyderabad*

Received 17 July 2017; Revised 30 January 2018; Accepted 28 April 2018

SUMMARY

Sterility Mosaic Disease (SMD) is a major biotic stress limiting achievable yield levels of pigeonpea. Studies on field incidence of SMD carried out for four consecutive *kharif* seasons (2012-15) indicated the commencement of its infestation during second week of August with peak incidence during third week of October to November. Mean incidence of SMD was higher in 2013 (4.5%), 2014 (4.3%), 2012 (3.8%) with the least in 2015 (0.6%). Correlation analyses of SMD incidence with weather parameters of current, one and two weeks prior indicated significant and negative influence of evening humidity at current week; significant and positive influence of sunshine of current to two lagged weeks on mean SMD. Whereas, for maximum SMD, significant negative correlation is found with minimum temperature and evening humidity at current, one week and two week lags; significant positive correlation with sunshine at current to two week lags. Besides multiple regression model, advanced statistical models namely autoregressive integrated moving average model with exogenous variable (ARIMAX), support vector regression (SVR) model and artificial neural network (ANN) have been applied for predicting the mean and maximum SMD. A comparative performance of different models carried out in terms of root mean square error (RMSE) and mean square error (MSE) indicated that both MSE and RMSE of SVR model was less in comparison to regression, ARIMAX and ANN models for forecasting the incidence of sterility mosaic disease of pigeonpea.

Keywords: Weather, ARIMAX, SVR, ANN, Pigeonpea, Sterility mosaic disease.

1. INTRODUCTION

Pigeonpea [*Cajanus cajan (L.) Millsp*] is a multipurpose grain legume grown across varied agro climatic zones in India. Not only a source of protein in human diets but benefits soils by fixing up to 40 kg N/ha. India cultivates pigeonpea in 3.75 million hectares (mha) with production and productivity of 2.46 million tonnes (mt) and 656 kg/ha, respectively. Gujarat contributes 9.6% of area to all India during 2015-16 with an area, production and productivity of 0.23 mha, 0.24 mt and 1044 kg/ha, respectively. Productivity levels have fluctuated (Anonymous, 2016) between 751 kg/ha in 2006-07 to 1184 in 2012-13. Such fluctuations are attributed to various abiotic and biotic stresses encountered by the crop at different growth stages.

Amongst whole lot of insect pests and diseases attacking pigeonpea, sterility mosaic disease (SMD) transmitted by *Aceria cajani* Channabasavanna, (eriophyid mite) (Seth,1962 and Jones *et al.,* 2004) is an emerging concern at Banaskantha region of Gujarat. SMD is referred to as the “Green Plague” because at flowering time, affected plants appear green with excessive vegetative growth and have no flowers or seedpods and it spreads rapidly like a plague under congenial conditions leading to severe epidemics. Mosaic symptoms, reduction in leaf size, ring spots on leaflets and bushy appearance (phllody) of plants

Corresponding author: Ranjit Kumar Paul
E-mail address: ranjitstat@gmail.com

are common. Disease spread within fields in a season depends on proximity to source of inoculum, plant age, cultivars, climatic factors and mite population. Yield losses depend on plant growth stage at which infection occurs with estimates of loss up to 95% (Reddy and Nene, 1981; Kannaiyan *et al*., 1984., Ganapathy *et al.,* 2011). The incidence of SMD is more in Bihar (21.4%), followed by Uttar Pradesh (15.4%), Tamil Nadu (12.8%), Gujarat (12.2%) and Karnataka (9.8%) (Kannaiyan *et al*., 1981). Considering importance of crop and the disease at pigeonpea growing regions of Gujarat, study was carried out to establish status of the disease and influence of weather factors at Banaskantha belonging to the hot arid eco-region of North Gujarat agro-climatic zone.

Climate as an exogenous factor plays crucial role in determining severity of insects as well as diseases. However, only a very few theoretical frameworks are available to examine the effect of climate on population dynamics. Therefore, the quest for generalizations and development of adequate predictive process-based models of change remains difficult (Harrington *et al*. 2001). The common approach for analyzing the relationship between population size and climatic variables is by means of simple correlation or using the climate as an additive covariate in statistical models (Stenseth *et al*., 2002). Nevertheless, it has been shown that the influence of temperature (Huey and Berrigan, 2001) and humidity on population dynamics of ectotherms may not necessarily be additive, and more complex interactions could be involved (Royama, 1992). Development of weather based relations aid in forecasts that helps farmers in pest management programs. It is essential that a competent approach to analysis of pest weather relations is arrived at on priority to forecast the pest dynamics accurately. Applications of time series models in the field of agriculture such as forecasting agricultural prices using autoregressive integrated moving average (ARIMA) models (Paul and Das, 2010.,Paul *et al.,*2014a) and yield predictions based on ARIMA with exogenous variables (ARIMAX) (Paul *et al*., 2013.,Paul *et al*., 2014b) are many, but such applications for pest forecasting is scarce. Kim *et al*., (2014) have studied pest prediction using regression and machine learning techniques. Arya *et al*., (2015) applied ARIMAX model for predicting pest population. Vennila *et al*., (2018a) studied pigeonpea leaf webber damage dynamics in relation to weather. Vennila *et al*., (2018b) investigated the abundance, infestation and disease transmission by thrips on groundnut influenced by climatic variability at Kadiri, Andhra Pradesh. Kaundal *et al*., (2006) introduced a new prediction approach based on support vector machines for developing weather-based prediction models of plant diseases. Calyo *et al.* (2014) applied machine-learning techniques for prediction of sigatoka disease of banana and plantation crops in Central America. In the present investigation, an attempt has been made to apply time series models including weather variables to forecast the SMD severity for Banaskantha region of Gujarat.

2. MATERIALS AND METHODS

2.1 Study location and sampling of sterility mosaic disease

Field observations on SMD incidence were carried out on weekly basis among 10 villages located within 30 km radius of meteorological observatory of Sardar Krushinagar of Banaskantha district in Gujarat ($24^{0}19^{'}$:20N and $72^{0}18$'E) as a part of ICT based pest surveillance for study on pest dynamics in the context of climate change during *kharif* 2012-15. Twenty fields at the rate of two per village were used for pest surveillance. Major pigeonpea cultivars grown by farmers included Prabhat A-3, Sarda hybrid-3, Ankur, ICP-8863 and GT-101 with row and plant spacing of 90 x 30 cm and crop was raised as per standard practices of the region. Sampling area was approximately one-acre field and five spots/field were randomly selected with observations made on ten plants randomly selected per spot right from vegetative stage till crop harvest. Number of plants showing symptoms of phyllody out of 10 plants per spot was counted and percent incidence of SMD was calculated for each field. Mean as well as maximum incidence of SMD in respect of each period of observation across fields were considered for forecast purposes.

2.2 Meteorological observations

Data on weather variables *viz*., maximum and minimum temperature (MaxT & MinT in ^{0}C), morning and evening humidity (RHM & RHE in %), sunshine hour (SS in h/day), wind velocity (Wind in m/h), total rainfall (RF in mm), and rainy days (RD) recorded at meteorological observatory S.K. Nagar of Gujarat were gathered for study period (2012-2015). Data sets on standard meteorological week (SMW) basis were used to assess the influence of weather parameters on SMD.

2.3 Statistical Analyses

While seasonal dynamics of SMD across fields during each season (2012-15) was presented graphically, difference in mean incidence across seasons were compared one-way analysis of variance (ANOVA) using SAS 9.4®. Pearson correlation coefficients were worked out between SMD (% incidence) and weather variables *viz.* maximum and minimum temperature (MaxT & MinT) (^{0}C), morning and evening humidity (RHM & RHE) (%), sunshine hour (SS) (h/day), wind velocity (Wind) (km/h), total rainfall (RF) (mm), and rainy days (RD) lagged by one and two weeks using aggregate data over four (2012-15) seasons. Stepwise regressions, autoregressive integrated moving average model with exogenous variables (ARIMAX), nonlinear support vector regression (SVR) and artificial neural network (ANN) were employed for prediction of mean and maximum levels of incidence of SMD using weather variables lagged by one and two weeks together. SAS 9.4® and R 3.2® were used for all regressions and ANN. Brief description of models used is given hereunder:

2.3.1 Multiple Linear Regression Analysis

Multiple linear regression (MLR) is an extension of simple linear regression model. The data consist of N observations on a dependent or response variable Y and p predictor or explanatory variables, $X_1, X_2, \dots, X_p$. The relationship between Y and $X_1, X_2, \dots, X_p$ is formulated as a linear model

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon. \quad (1)$$

where $\beta_0, \beta_1, \dots \beta_p$ are constants referred to as the regression coefficients and ε is a random disturbance or error. It is assumed that Y is approximately a linear function of the X's, and ε measures the discrepancy in that approximation (Chatterjee and Hadi, 2012). Stepwise selection procedure for selecting the significant variable in the model was adopted.

2.3.2 Autoregressive Integrated Moving Average (ARIMA) Model

A generalization of ARMA model incorporates a wider class of non-stationary time-series obtained by introducing differencing. Non-stationary process that reduces to a stationary one after differencing is 'Random Walk'. A process $\{y_t\}$ is said to follow an Integrated ARMA model, denoted by ARIMA (*p*, *d*, *q*), if $\nabla^d y_t = (1-B)^d \varepsilon_t$ is ARMA (*p*, *q*). The model is written as

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \quad (2)$$

where $\varepsilon_t \sim WN(0, \sigma^2)$, WN indicating White Noise. $\phi(B)$ and $\theta(B)$ are the AR and MA polynomial of order p and q respectively. The integration parameter d is a non-negative integer. When $d = 0$, ARIMA (*p*, *d*, *q*) ≡ ARMA (*p*, *q*). The ARIMA methodology is carried out in three stages, *viz.* identification, estimation and diagnostic checking. Parameters of the tentatively selected ARIMA model at the identification stage are estimated and adequacy of tentatively selected model is tested. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for the time-series under consideration.

2.3.3. Autoregressive Integrated Moving Average Model with exogenous variable (ARIMAX)

The ARIMAX model (Bierens, 1987) is a generalization of the ARIMA model capable of incorporating an external input variable (X). Given a (k+1) time-series process $\{(y_t, x_t)\}$, where y_t and k-components of x_t are real valued random variables, the ARIMAX model assumes the form

$$\left(1-\sum_{s=1}^{p}\alpha_s L^s\right)\Delta y_t = \mu + \sum_{s=1}^{q}\beta_s' L^s x_t + \left(1+\sum_{s=1}^{r}\gamma_s L^s\right)e_t, \quad (3)$$

Where L is usual lag operator, i.e. $L^s y_t = y_{t-s}$, $\Delta y_t = y_t - y_{t-1}$, ì ∈ R, á$_s$ ∈ R, â$_s$ ∈ R^k and ã$_s$ ∈ R are unknown parameters and e_t, s are the errors, and p, q and r are natural numbers specified in advance.

2.3.4. Nonlinear Support Vector Regression (SVR) Model

For a data set $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in R^n$ is input vector, $y_i \in R$ is scalar output and N corresponds to size of data set, general form of

Nonlinear SVR estimating function is:

$$f(x) = w^T\varphi(x) + b, \quad (4)$$

Where $\varphi(.): R^n \to R^{n_h}$ is a nonlinear mapping function from original input space into a higher dimensional feature space, which can be infinitely dimensional, $w \in R^{n_h}$ is weight vector, b is bias term and superscript T indicates transpose. The coefficients w and b are estimated from data by minimizing the following regularized risk function:

$$R(\theta) = \frac{1}{2}||w||^2 + C\left[\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)\right]. \quad (5)$$

This regularized risk function minimizes both empirical error and regularized term simultaneously and implements Structural risk minimization (SRM) principle to avoid under and over fitting of training data. In Equation (5), first term $\frac{1}{2}||w||^2$ is called 'regularized term', which measures flatness of the function. Minimizing $\frac{1}{2}||w||^2$ will make a function as flat as possible. Second term $\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)$ called 'empirical error' is estimated by Vapnik ε-insensitive loss function representing radius of tube of accuracy located around the regression function given by:

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon; & |y_i - f(x_i)| \geq \varepsilon, \\ 0 & |y_i - f(x_i)| < \varepsilon, \end{cases} \quad (6)$$

where y_i is actual value and $f(x_i)$ is estimated value. In Equation(5), C referred to as regularized constant determines trade-off between empirical error and regularized term. The value ε is called as tube size equivalent to approximation accuracy in training data. Both C and ε areuser-determined hyper-parameters. Under above loss function (Figure 1), training points within the ε-tube have no loss and do not provide any information for decision. Only those data points located on or outside the ε-tube are penalized and will serve as support vectors. This property of sparseness algorithm results only from the ε-insensitive loss function and greatly simplifies computation of nonlinear SVR. Two positive slack variables ξ_i and ξ_i^* are introduced for representing the distance fromactual values to corresponding boundary values of the ε-tube.

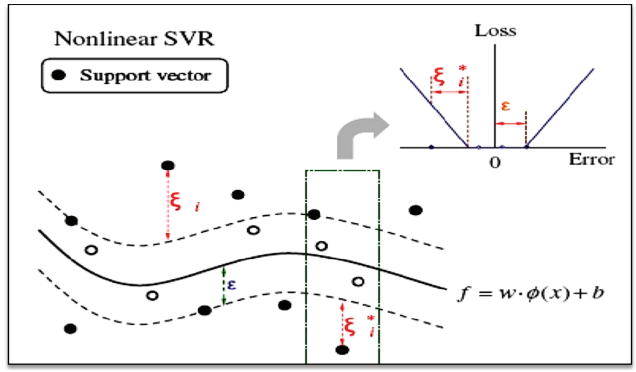

Fig. 1. A schematic representation of Vapnik ε-insensitive loss function and accuracy tube under nonlinear SVR model setup

These equal zeros when data points fall within ε-tube. Equation (5) is then reformulated into the following constrained QP problem in the primal space given by:

$$\text{Minimize: } R_P(w, b, \xi_i \xi_i^*) = \frac{1}{2}||w||^2 + C\left(\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right) \quad (7)$$

Subject to constraints

$$\begin{cases} w^T\varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^*, \\ y_i - w^T\varphi(x_i) - b \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0; \quad i = 1,2,\ldots,N. \end{cases} \quad (8)$$

A detailed description of above methodology can be found in Vapnik(2000) and Kecman(2001).

2.3.5. Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are nonlinear data driven self-adaptive approach and are powerful tools for modeling, especially when the underlying data relationship is unknown. A very important feature of these networks is their adaptive nature, where "learning by example" replaces "programming" in solving problems. A neural network consists of a set of connected cells (neurons). Neurons receive impulses from input cells or other neurons, perform some kind of transformation of input, and transmit outcome to other neurons to output cells. The neural networks are built from layers of neurons connected so that one layer receives input from the preceding layer of neurons and passes the output on to the subsequent layer.

Most commonly used ANN is multi-layer perceptron (MLP), a class of feed forward neural network. MLP consists of at least three layers of nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron, which distinguish data that is not linearly separable. A graphical presentation of MLP is given in Figure 2.

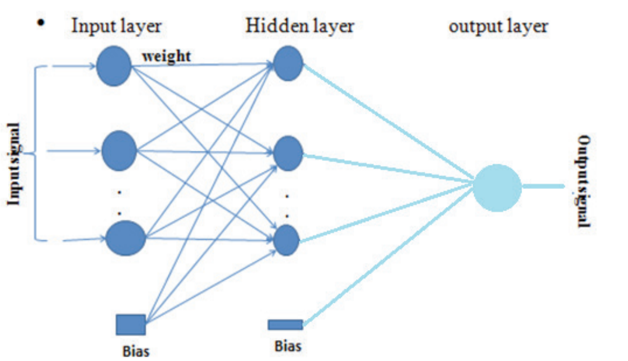

Fig. 2. A multilayer perceptron (MLP) architecture with one hidden layer

Mathematically, MLP network is a function consisting of compositions of weighted sums of the functions corresponding to the neurons. Let us consider following notations with p input and h hidden nodes: $x_i(i=1,2\ldots p)$ are network inputs; w_{ij}, refer the synaptic weight connection between neuron i and j; w_j refer the synaptic weight connection between j^{th} neuron of hidden node and output node. a_0, β_{0J} are bias term for output layer and hidden layer; φ is hidden output layer activation function, mainly logistic $\varphi(v_j)=\frac{1}{1+e^{-v_j}}$ and I as identity function. The output of a MLP with p input and h hidden nodes is expressed as

$$Y = I\left[a_0+\sum_{j=1}^{h} w_j\varphi_j[\beta_{0J}+\sum_{i=1}^{p} w_{ij}x_i]\right] = I\left[a_0+\sum_{j=1}^{h} w_j[v_j]\right]=I[y_j]$$

where

$$v_j=\varphi_j[\beta_{0J}+\sum_{i=1}^{p} w_{ij}x_i],$$

$$y_j=\left[a_0+\sum_{j=1}^{h} w_j[v_j]\right]$$

2.3.6. Validation

Out of total 62 observations across seasons relating to mean and maximum incidence of SMD, first 52 observations were used for estimation of parameters in models and remaining 10 observations were used for validation purpose. For validation of model, two statistics namely Mean square error (MSE) and Root mean square error (RMSE) have been used. Besides these, the residual diagnostics have also been carried out in order to ensure the adequacy of the fitted model.

3. RESULTS AND DISCUSSION

3.1 Seasonal dynamics and status of SMD

Earliest possible onset of SMD in fields of pigeonpea was in 38 SMW (September third week) during 2014 and 2015 with peak at 43 SMW as against 39 SMW (September fourth week) in 2012 and 2013 with peak incidence coinciding at 44 SMW. Progression of SMD was almost nil across years beyond 45 SMW (November second week) (Figure 3). Effective period of SMD occurrence at Banaskantha is between 38 and 45 SMW (September to November).

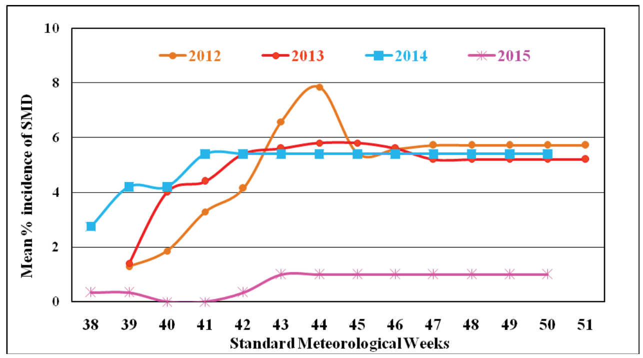

Fig. 3. Seasonal dynamics of SMD at Banaskantha (GJ) during 2012-15

Pairwise comparisons of means of SMD incidence for mean and maximum incidence levels across seasons using Duncan's Multiple Range Test (DMRT) (Table 1) indicated significantly reduced SMD during 2015 over 2012-2014 with later seasons being on par. Latest status report from Banaskantha for 2016 has indicated scarce to nil incidence of SMD in fields with mean and maximum incidence of 0.33 and 2%, respectively (NICRA database, 2016). Declining trend of SMD at Banaskantha region of Gujarat is obvious.

Table 1. Comparative analysis of SMD across the years

SMD incidence (%)	2012	2013	2014	2015
Mean incidence	3.81[a]	4.48[a]	4.26[a]	0.62[b]
Maximum incidence	9.76[a]	7.00[a]	7.00[a]	2.66[b]

**** Means in a row followed by the superscript of same at p<0.05 based on DMRT***

3.2 Summary statistics of response and regressor variables

Descriptive statistics for SMD as well as weather variables computed to understand variations in data sets for study attributes over four year study period (Table 2) revealed negative and positively skewed response variables of mean and maximum incidence, respectively. While maximum incidence was leptokurtic, mean incidence was paltykurtic. MaxT, MinT, RHM and SS were negatively skewed and all other variables (RHE, RF, Wind and RD) were positively skewed. The variability in SMD mean and maximum incidence measured in terms of coefficient of variation (CV) was found to be around 66 and 70%, respectively. Very high CV for rainfall and rainy days were observed possibly due to the unseasonal rains during 2013 between 39 and 41 SMWs (440 mm) in five rainy days (Table 2).

3.3 Correlation analysis

Correlation analysis considering the data sets of effective period of SMD incidence (38-45 SMWs) of all seasons indicated that mean and maximum incidence of SMD was significantly and positively influenced by sunshine (hours/day). Decreasing minimum temperature and evening relative humidity with increased hours of sunshine of current to previous two weeks had a significant role in triggering maximum incidence of SMD. Sunshine alone had significantly positive influence on both mean and maximum SMD incidence. Only evening relative humidity of current week showed negative association that was significant with mean and maximum SMD incidence. Rainy days in general had negative influence on SMD, however non-significant (Table 3). Reddy and Raju (1993) while studying factors contributing to increased incidence and seasonal variation of SMD for the period of 1980 to 1990 in peninsular India found that higher summer rainfall, humidity and lower temperature led to SMD outbreaks in the following season due to survival and multiplication of mites during off-season. Negative association of only maximum incidence of SMD with minimum temperature of present study and SMD outbreaks reported under low temperature condition by Reddy and Raju (1993) are similar. Padule *et al.* (1982) observed that SMD was highest (up to 95%) when the crop was irrigated or grown near other irrigated crops although it is unclear on its impact on micro or macroclimate of cropping system and on vector (mites).

Table 3. Correlation between mean and maximum incidence of SMD and weather parameters

	Mean incidence			Maximum incidence		
Weather parameters	**Current week**	**One week prior**	**Two week prior**	**Current week**	**One week prior**	**Two week prior**
MaxT	-0.04	0.03	0.06	-0.07	0.01	0.12
MinT	-0.33	-0.30	-0.14	-0.58**	-0.57**	-0.42*
RHM	-0.17	-0.15	-0.16	0.09	0.02	-0.10
RHE	-0.43*	-0.30	-0.22	-0.51**	-0.41*	-0.37*
RF	-0.16	-0.04	0.05	-0.19	0.01	-0.01
SS	0.42*	0.40*	0.52**	0.41*	0.40*	0.45**
Wind	0.15	0.13	-0.01	-0.12	-0.05	-0.09
RD	-0.24	-0.07	-0.08	-0.26	-0.13	-0.15

***: *Significant at* $p \leq 0.01$; *: *Significant at* $p \leq 0.05$

Table 2. Descriptive statistics of response and regressor variables

Statistical measures	Response variable		Regressor variable							
	Mean SMD	**Maximum SMD**	**MaxT.**	**MinT.**	**RHM.**	**RHE.**	**RF**	**SS**	**Wind**	**RD**
Mean	3.49	7.4	34.6	20.8	78.9	39.6	15.8	8.1	4.08	0.43
Median	4.17	8.0	34.4	21.4	77.5	36.6	0	9.1	3.71	0
Maximum	7.85	22	38.8	26.1	93.4	89.4	383.6	10.1	10	5.0
Minimum	0	0	27.1	13.3	61.1	18	0	0.43	1.71	0.0
Standard Deviation	2.32	5.2	2.5	3.43	7.92	15.4	70.2	2.52	1.98	1.10
CV	66.5	70.3	7.4	16.5	10.0	38.9	442.3	31.1	48.5	252.8
Skewness	-0.17	1.27	-0.78	-0.43	-0.12	1.4	5.31	-1.74	1.42	3.14
Kurtosis	-1.33	2.07	1.45	-0.45	-0.37	2.5	28.6	2.46	2.10	10.5

3.4 Multiple linear regression

The multiple linear regression model with stepwise selection procedure was applied for identification of significant variables influencing mean and maximum SMD incidence. Since the correlative analysis indicated significance of weather variables of preceding two weeks in addition to current week, 16 explanatory variables of weather were used. For mean incidence of SMD, $MaxT_{-1}$, RHE_{-1}, $Wind_{-1}$, SS_{-1}, RD_{-1} and RD_{-2}, were found significant. Significant variables influencing maximum incidence of SMD were RHM_{-1}, RHE_{-1}, $MaxT_{-2}$, SS_{-2} and RD_{-2}. The stepwise regression used eight iterations for Mean SMD and seven iterations for Maximum SMD. Significant parameter estimates of MLR model is presented in Table 4.

Table 4. Parameter estimates of multiple linear regression model using stepwise selection

Variables	Parameter Estimate	Standard Error	Type II SS	F Value	Pr * > F
SMD - mean incidence					
Intercept	-6.565	3.151	14.853	4.340	0.042
$MaxT_{-1}$	0.215	0.073	30.016	8.770	0.005
RHE_{-1}	-0.075	0.023	36.337	10.620	0.002
RD_{-1}	0.010	0.006	10.498	3.070	0.086
$Wind_{-1}$	0.164	0.093	10.656	3.120	0.083
SS_{-2}	0.558	0.155	44.362	12.970	0.001
RD_{-2}	1.516	0.343	66.758	19.520	<.0001
SMD -maximum incidence					
Intercept	-23.920	9.188	99.534	6.780	0.012
RHM_{-1}	0.107	0.067	37.300	2.540	0.117
RHE_{-1}	-0.125	0.039	147.013	10.010	0.003
$MaxT_{-2}$	0.574	0.168	172.325	11.730	0.001
SS_{-2}	0.890	0.316	116.280	7.920	0.007
RD_{-2}	2.082	0.722	122.020	8.310	0.006

*:probability values indicating significance of F in relation to parameter estimates

Many studies are available on the influence of weather on mite populations that transmit SMD and not SMD *per se*. Lower temperature and rainfall positively influencing mite populations and hence on manifestation of SMD was reported by Abhijit *et al.* (2014). Since there is always a period of acquisition, incubation and transmission of virus associated with mites in addition to manifestation of SMD on plants approach to use weather of lagged weeks is highly relevant and present study has attempted such an approach. The prediction models for SMD forecast have expressed 42% (R^2=0.42) and 39% (R^2=0.39) of variability in mean and maximum SMD incidence respectively.

3.5 ARIMA model

ARIMA model presented in Table 5 showed that autocorrelation function (ACF) decayed very slowly and significantly different from zero for both mean and maximum SMD indicating non-stationary nature of both series. Augmented Dickey Fuller (ADF) test applied to test for stationarity of the series confirmed that both the series are non-stationary. Therefore, one differencing has been done to make the series stationary. Four steps of ARIMA model building were followed *viz.,* identification, estimation, validation and forecasting. ARIMA (1, 1, 0) was selected as best fit model on the basis of minimum AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). The validation of the model was carried out based on MAPE (Mean Average Percentage Error) and by examining the residuals from fitted models.

Table 5. Parameter estimates of ARIMA model

Estimate	Standard Error	t Value	Approx. Pr > |t|
SMD - mean incidence			
0.020	0.137	0.15	0.883
-0.509	0.113	-4.51	<.0001
SMD - maximum incidence			
0.067	0.248	0.27	0.787
-0.442	0.117	-3.76	0.0004

3.6 ARIMAX model

ARIMAX model, an extended version of ARIMA model includes additionally other independent (predictor) variables. The model is also referred to as the dynamic regression model. The ARIMAX model is similar to a regression model, but allows to take advantage of autocorrelation that may be present in residuals of regression to improve the accuracy of a forecast. On the basis of minimum AIC values and considering the ACF and PACF of SMD (mean and max), ARIMAX model was applied on the differenced series to estimate the parameters using maximum likelihood estimation. Parameter constant for mean SMD was found not significant and hence dropped from the model. On the other hand, estimated parameter for SMD maximum severity was significant. All estimated parameters relating to weather variables with corresponding standard error based on ARIMAX model are given in Table 6.The best fitted model i.e. ARIMAX (1, 1, 1) for both the series (mean and max) was selected on the basis of minimum AIC and SBC criterion. While sunshine and relative morning humidity of previous two weeks had significant contribution in determining mean incidence of SMD, only the later weather variable showed up for maximum severity of SMD.

Table 6. Parameter estimates of ARIMAX model

Model parameters	Estimate	Standard Error	t value	Approx Pr > |t|
SMD - mean incidence				
SMD mean	1.927	4.861	0.40	0.69
RHE_{-1}	0.011	0.042	0.26	0.79
RF_{-1}	-0.032	0.034	-0.97	0.34
SS_{-1}	0.015	0.0063	2.46	0.02*
RHM_{-2}	0.200	0.086	2.32	0.02*
$Wind_{-2}$	0.163	0.178	0.92	0.36
SS_{-2}	0.011	0.006	1.87	0.07
SMD - maximum incidence				
SMDMax	1.000	0.040	24.43	<.0001**
RHE_{-1}	0.0011	0.031	0.04	0.97
RF_{-1}	-0.031	0.024	-1.28	0.21
SS_{-1}	-0.0083	0.005	-1.64	0.11
RHM_{-2}	-0.114	0.058	-1.95	0.05*
$Wind_{-2}$	0.094	0.133	0.71	0.48
SS_{-2}	-0.003	0.004	-0.78	0.44

It can be inferred that morning relative humidity is an important factor determining severity of SMD. It has to be mentioned that neither the correlative analysis nor the stepwise MLR models on prediction of mean and maximum severity had indicated the significance of morning relative humidity although sunshine was indicated across models including ARIMAX. Reddy *et al*. (1993) reported that shade and humidity encouraged mite multiplication and symptoms of SMD are suppressed during hot summer weather. Current study had revealed significant influence of morning relative humidity and sunshine of previous weeks on mean severity of SMD while the effect of RHM as significantly negative for maximum severity.

3.7 Support vector regression

Values of ε that lead to the lowest generalization error was experimentally determined. The value of ε defines a margin of tolerance where no penalty is given to errors. A good agreement with values predicted previously by a theoretical argument based on the asymptotic efficiency of a simplified model of support vector regression was found for various noise models and support vector parameter settings. The final estimate of the parameter εis found out to be 0.062 for both the series.

The performance of nonlinear SVR i.e. NLSVR model strongly depends on the kernel function and set of hyper-parameters. The radial basis function (RBF) in NLSVR requires optimization of two hyper-parameters, i.e. the regularisation parameter C, which balances the complexity and approximation accuracy of the model and the kernel bandwidth parameter, which defines the variance of RBF kernel function. These tuning parameters viz., C and are user defined parameters, these should be defined in such a way that training and testing error are minimum. The forecasting performance of NLSVR model in both training and testing data set are given in table 8.

3.8 Artificial neural network

Improving accuracy of time series forecasting is an important yet often difficult task. Combining multiple models or using ensemble methods can be an effective way to improve forecasting performance over a single best network model. In ANN, different strategies used to form neural network ensemble consists of neural network trained with different initial weights, neural network with different architecture and with different training data. Results showed that different strategies for neural network ensemble have different effects on forecasting (Table 7). Different combinations of sterility mosaic disease of pigeonpea both mean and maximum were used. It is concluded that (3,3) and (1,1) combination performed better than other combinations in case of field incidence of the cases of SMD (mean and maximum).

Among the 62 data points, 75% were used for training, 15% for validation and 15% for testing. The model has been examined at various delays with different number of hidden nodes as shown in table 8. Out of a total of 25 neural structures, a neural network model with 3 hidden nodes and 3 delays, 1 hidden nodes and 1 delays combination performed better than other combinations in case of field incidence of the cases of SMD (mean and maximum).

Table 7. Selection of ANN model based on RMSE

No. of Lag.	Hidden node	SMD - mean incidence		SMD – maximum incidence	
		MSE	RMSE	MSE	RMSE
1	1	27.58	5.25	**23.36**	**4.83**
1	2	20.98	4.58	62.89	7.93
1	3	20.57	4.54	70.23	8.38
1	4	19.30	4.39	108.61	10.42
1	5	20.99	4.58	104.40	10.22
2	1	17.31	4.16	41.04	6.41
2	2	17.39	4.17	138.39	11.76
2	3	15.89	3.99	79.84	8.94
2	4	15.39	3.92	89.80	9.48
2	5	16.86	4.11	85.46	9.24
3	1	13.38	3.66	111.77	10.57
3	2	18.10	4.25	110.13	10.49
3	3	**13.13**	**3.62**	119.60	10.94
3	4	16.91	4.11	59.47	7.71
3	5	15.70	3.96	77.08	8.78
4	1	16.61	4.08	120.50	10.98
4	2	16.56	4.07	147.87	12.16
4	3	17.61	4.20	140.47	11.85
4	4	16.87	4.11	120.95	11.00
4	5	17.27	4.16	131.42	11.46
5	1	19.42	4.41	156.81	12.52
5	2	18.14	4.26	65.93	8.12
5	3	18.76	4.33	92.41	9.61
5	4	17.23	4.15	55.39	7.44
5	5	17.35	4.17	48.59	6.97

3.9 Validation

Forecast for ten observations were obtained from estimated model equations of MLR, ARIMAX, SVR and ANN models and compared with original incidence of SMD. Performance of predictions of mean and maximum incidence of SMD through various models tested using the statistic of root mean square error (RMSE) and mean square error (MSE) (Table 8). Both the MSE and RMSE values of SVR models are less and are almost close to each other.MLR had the largest MSE and RMSE over other models and hence distantly precise over all other models. ARIMAX prediction for mean was better over maximum SMD incidence. ARIMAX model being stochastic in nature could be used successfully for modelling as well as for forecasting of sterility mosaic disease of pigeonpea in Gujarat. As the weather variables are of utmost important for studying the pest and disease incidence, in the present study ARIMA validation is not shown because it does not as such consider the exogenous variable in the model.

Table 8. Prediction performance of SMD (% incidence) using different models

S.No.	ORIGINAL		MLR		ARIMAX		SVR		ANN	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max
1	1.00	4.00	4.81	10.13	2.67	4.22	2.42	4.96	5.13	8.07
2	1.00	4.00	4.36	9.28	4.23	4.58	2.97	6.20	4.08	7.94
3	1.00	4.00	4.13	7.90	4.40	5.97	3.60	6.44	2.60	5.66
4	1.00	4.00	3.82	7.04	3.57	6.65	3.04	6.11	2.97	6.09
5	1.00	4.00	2.95	5.96	3.76	6.75	2.96	6.46	3.28	6.46
6	1.00	4.00	3.44	6.43	2.76	7.24	2.90	6.40	2.55	5.49
7	1.00	4.00	1.53	3.82	3.28	7.16	2.74	5.17	1.26	3.22
8	1.00	4.00	1.88	3.55	2.33	7.00	2.41	4.94	1.53	3.61
9	0.33	2.00	1.70	1.62	2.15	6.08	2.93	4.86	0.10	0.49
10	0.00	0.00	1.01	1.69	1.33	6.02	1.64	2.71	0.30	1.06
MSE			**5.72**	**10.29**	**5.41**	**10.13**	**3.88**	**4.58**	**4.13**	**5.16**
RMSE			**2.39**	**3.21**	**2.33**	**3.18**	**1.97**	**2.14**	**2.03**	**2.27**

SVR models are better over MLR, ARIMA, ARIMAX and ANN models for forecasting of sterility mosaic disease of pigeonpea. The case of SMD presented a situation of its seasonal dynamics over many seasons and of the associated weather that have influence on disease manifestation for application of ANN and the approach has proved useful. Nevertheless, each model has brought out biologically relevant inferences that could support or deviate from the field occurrence of SMD already reported.

4. CONCLUSION

Climate change has an adverse impact on pulse growing areas due to wide fluctuation in temperature and erraticrainfall patterns and assessment of seasonal dynamics of any pest in relation to weather variations is of significance. Present study revealed that sterility mosaic disease of pigeonpea at Banaskantha (GJ) is on the decline with 2015 having the lowest incidence over 2012-2014. Approaches to analyses of both mean and maximum incidence of SMD used indicated varying performances of models. Statistically, SVR models proved better over ARIMAX, ANN and MLR. MLR brought out significance of four variables *viz.,* MaxT and RHE of previous one week besides SS and RD of previous two weeks in reflecting mean SMD. Maximum SMD had similarity with mean SMD for RHE, SS and RD. Day temperature (MaxT) of previous and two weeks prior had significant effect on mean and maximum incidence of SMD, respectively. ARIMAX brought out importance of SS on mean SMD incidence and singled out RHM as the trigger for maximum SMD. In SVR and ANN all the weather variables have been taken into consideration to model both mean and maximum incidence of SMD. Inclusion of population of vector of SMD (mites) and analysis of vector (mite)-virus (SMD)-weather interactions are expected to yield improved models with higher prediction accuracy. Nevertheless current models could form a part of prediction for future seasons and for estimating scenario of SMD for projected period of climate change.

ACKNOWLEDGEMENTS

The Authors are thankful to Central Research Institute for Dryland Agriculture, Hyderabad (CRIDA), India, for financial support to undertake this study through National Innovations in Climate Resilient Agriculture (NICRA). The authors are thankful to the reviewers for providing comments for improving the quality of the manuscript.

REFERENCES

Abhijit, D., Fakrudin, B.,Shivarudrappa. B.B. Kavil, S.P., Narayana, Y.D., Muniswamy, Kaumar, A., Krishnaraj, P. U. Yerimani, A. and Khadi, B. M. (2014). Inheritance of pigeonpea sterility mosaic disease resistance in pigeonpea. *The Plant Pathol. J.*, **30(2)**, 188-194.

AICRP Report. (2013). Annual reports of All India Coordinated Research Project on Pigeonpea, IIPR, Kanpur. pp: **233**.

Anonymous. 2016. Agricultural Statistics At a Glance 2016, Department of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India.

Arya, P., Paul, R.K., Kumar, A., Singh, K.N., Sivaramne, N. and Chaudhary, P. (2015). Predicting pest population using weather variables: an ARIMAX time series framework. *Int. J. Agril. Stat. Sci*., **11(2),** 381-386.

Bierens, H.J. (1987). ARMAX model specification testing with an application to unemployment in the Netherlands. *J. Econom*., **35(1),** 161-90.

Chatterje, S. and Hadi, A.S. (2012). *Regression Analysis by Example*, John Wiley & Sons, Inc, New York.

Dash, P.K. (2012). Pigeonpea hybrid Pushkal. *Ag-biotech Digest*, **3(2)**:1-2.http://faostat.fao.org/site/567/default.aspx#ancor.

Daspute, A., Fakrudin, B., Bhairappanavar, S.B., Kavil, S.P., Narayana, Y.D., Muniswamy, Kaumar, A., Krishnaraj, P.U., Yerimani, A. and Khadi, B.M. (2014). Inheritance of Pigeon pea sterility mosaic disease resistance in Pigeonpea. *Plant Pathol. J*., **30(2)**:188-94.

Ganapathy, K.N., Gnanesh, B.N., Gowda, B.M., Venkatesha, S.C., Gomashe, S.S. and Channamallikarjuna, V. (2011). AFLP analysis in Pigeonpea (*Cajanus cajan* (L.) Millsp.) revealed close relationship of cultivated genotypes with some of its wild relatives. *Genet. Resour. Crop Evol.*, **58**: 837-847.

Harrington, R., Fleming, R. and Woiwod, I. (2001). Climate change impacts on insect management and conservation in temperate regions: can they be predicted? *Agric. Entomol*., **3,** 233-240.

Huey, R. and Berrigan, D. (2001). Temperature, demography, and ectotherm fitness. *Am. Nat*., **158,** 204-210.

Jones, A.T., Kumar, P.L., Saxena, K.B., Kulkarni, N.K., Muniyappa, V. and Waliyar, F. (2004). Sterility mosaic disease – the "green plague" of pigeonpea: advances in understanding the etiology, transmission and control of a major virus disease. *Plant Dis.*, **88**, 436-445.

Kannaiyan, J., Reddy, M.V., Nene, Y.L. and Raju, T.N. (1981). Prevalence of pigeonpea wilt and sterility mosaic in India (1975-80). *Intern. Pigeonpea Newslett*., **1,** 24-26.

Kannaiyan, J., Nene, Y.L., Reddy, M.V., Ryan, J.G. and Raju, T.N. (1984). Prevalence of Pigeonpea disease and associated crop losses in Asia, Africa and the Americas. *Trop Pest Manage*., **30**, 62-71.

Kaundal, R., Kapoor, A.S. and Raghava, P.S. (2006). Machine learning techniques in disease forecasting: a case study on rice blast prediction. *Bioinformatics*.,**7,** 485.

Kecman, V. (2001). Learning and Soft Computing: Support Vector Machines, Neural Networks,and Fuzzy Logic Models, MIT Press,Cambridge.

Kim, Y., Yoo, S., Gu, Y., Lim, J., Han, D. and Baik, S. (2014). Crop pests prediction method using Regression and machine learning technology: Survey, *IERI Procedia*., **6,** 52–56.

Kumar, P.L., Jones, A.T. and Reddy, D.V.R. (2003). A novel mite transmitted virus with a divided RNA genome closely associated with pigeonpea sterility mosaic disease. *Phytopath*., **93**, 81-91.

Padule, D.N., Newase, A.G. and Patil, B.P. (1982). Survey for incidence of *Fusarium* wilt and other diseases of pigeon pea in central parts of Western Maharashtra. *J. Maharashtra Agric. Univ.,* **7(2)**: 159-161.

Paul, R.K. and Das, M.K. (2010).Statistical modelling of inland fish production in India *J. Inl. Fish. Soc. Ind.*, **42(2),** 1-7.

Paul, R.K., Prajneshu. and Ghosh, H. (2013). Statistical modelling for forecasting of wheat yield based on weather variables. *Ind. J. Agr. Sci*., **83(2),** 180-183.

Paul, R.K., Alam, W. and Paul, A.K. (2014 a). Prospects of livestock and dairy production in India under time series framework. *Ind. J. Anim. Sci*., **84(4),** 130-134.

Paul, R.K., Ghosh, H. and Prajneshu (2014 b). Development of out-of-sample forecast formulae for ARIMAX-GARCH model and their application. *J. Ind. Soc. Agr. Stat.,* **68(1),** 85-92.

Reddy, M.V.and Raju, T.N. (1993). Some clues to increased incidence and seasonal variation of pigeonpea sterility mosaic in Peninsular India. *Intern. Pigeonpea Newslett*. **18**: 22-24.

Reddy, M.V., Nene, Y.L. (1981). Estimation of yield loss in pigeonpea due to sterility mosaic. *Proceedings of International Workshop on Pigeonpea*,15-19 December 1980, ICRISAT Center, Patancheru, AP, India, 305-312.

Reddy, M.V., Raju, T.N. and Lenne, J.M. (1998).T*he pathology of food and pasture legumes*. CAB International, Wallingford, Oxon, UK.

Reddy, M.V., Raju, T.N., Sharma, S.B., Nene, Y.L. and McDonald, D. (1993). Handbook of pigeonpea diseases. *Information bulletin*, International Crops Research Institute for Semi-Arid Tropics, A.P, India.

Royama, T. (1992). Analytical Population Dynamics. Chapman & Hall, New York, **371,** p.

Seth, M.L. (1962). Transmission of pigeonpea sterility by an eriophyid mite. *Indian Phytopath.,* **15,** 225-227.

Stenseth, N., Mysterud, A., Ottersen, G., Hurrell, J., Chan, K. and Lima M. (2002). Ecological effects of climate fluctuations. *Science*., **297,** 1292–1296

Valverde, L.A. Quesada, M.G. and Álvarez, J.A.G (2014). Considerations about application of machine learning to the prediction of sigatoka disease. *Proceedings of World Conference on Computers in Agriculture and Natural Resources*, 27-30 July 2014, San Jose CostaRica, 1-8.

Vapnik, V.N. (2000). *The Nature of Statistical Learning Theory.* Springer-Verlag, New York.

Vennila, S., Paul, R.K., Bhat, M.N., Yadav, S.K., Kumar, N.B., Rachappa, V., Yelshetty, S., Chandra, P. and Sharma, O.P. (2018a). Approach to study of pigeonpea leaf webber [Grapholita critica (Meyr.)] damage dynamics and its relation to weather. *Leg. Res. An Int. J.* DOI: 10.18805/LR-3937.

Vennila, S., Paul, R.K., Bhat, M.N., Yadav, S.K., Vemana, K., Chandrayudu, E., Nisar, S., Kumar, M., Tomar, A., Rao, M. S. and Prabhakar, M. (2018b). Abundance, infestation and disease transmission by thrips on groundnut as influenced by climatic variability at Kadiri, Andhra Pradesh. *J. Agromet.* **20 (3),** 227-233.