Int. J. Agricult. Stat. Sci. Vol. 11, No. 2, pp. 381-386, 2015 ***ISSN : 0973-1903***
ORIGINAL ARTICLE

PREDICTING PEST POPULATION USING WEATHER VARIABLES : AN ARIMAX TIME SERIES FRAMEWORK

Prawin Arya*, Ranjit Kumar Paul, Anil Kumar, K. N. Singh, N. Sivaramne[1] and Pradeep Chaudhary[2]

Indian Agricultural Statistics Research Institute, New Delhi - 110 012, India.
[1]NAARM, Hyderabad - 500 030, India.
[2]Department of Statistics, C.C.S. University, Meerut - 250 004, India.
E-mail: prawin_arya@yahoo.co.in

Abstract : Farmers are encountering several issues in endeavour to increase crop productivity. Despite several successful new agricultural technologies related with crop cultivation, India is unable to attain world average mark in productivity. One of the main reasons for this is climatic conditions and abundance of insects and pests. To mitigate the loss due to pest attacks and for better yield, forecasting of pest population based on historical data and pertinent external climatic information is considered. Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) time-series model is applied for modelling and forecasting the pest population after testing for stationarity. Primary weekly data (2008-2012) for three pests namely Jassids, Whitefly and Thrips in Guntur and Faridkot Districts along with weekly maximum temperature, minimum temperature, rainfall, maximum RH and minimum RH have been used for model development. Evaluation of forecasting is carried out with relative mean absolute prediction error (RMAPE). Diagnostic test were applied and results showed that maximum temperature and minimum temperature along with maximum relative humidity have a significant role for Whitefly and Thrips at Guntur district respectively. Rainfall was found to be significant at Faridkot district in case of Thrips. The fitted models along with the data points are also presented. A perusal of figures indicates that in both districts, the population of Whitefly is best predicted followed by Jassids and Thrips.

Key words : ARIMAX Model, Forecasting, Pest Population, Weather Variables.

1. Introduction

A significant part of Indian population living in the rural areas earns their living directly or indirectly from farming and agriculture and therefore, agriculture plays a significant role in Indian economy. Indian farming is encountering several issues in the endeavour to increase crop productivity. Despite the number of successful researches on new agricultural technologies related with crop cultivation majority of the farmers are still not able to produce upper-bound yield owing to several other reasons. One main reason is the abundance of insects and pests.

The last assessment report from the Intergovernmental Panel on Climate Change (IPCC) predicts an increment in mean temperature from 1.1 to 5.4°C towards the year 2100 [Meehl *et al.* (2007)]. An increment of this magnitude is expected to affect global agriculture significantly [Cannon (1998)]. In addition, such changes in climatic conditions could profoundly affect the population dynamics and the status of insect pests of crops [Porter *et al.* (1991), Cammell and Knight (1992), Woiwod (1997)]. These effects could either be direct, through the influence that weather may have on the insects' physiology and behaviour [Hagstrum and Milliken (1988), Porter *et al.* (1991), Harrington *et al.* (2001), Huey and Berrigan (2001), Bale *et al.* (2002), Samways (2005), Parmesan (2007), Merrill *et al.* (2008)], or may be mediated by host plants, competitors or natural enemies [Cammell and Knight (1992), Harrington *et al.* (2001), Bale *et al.* (2002)].

The climate as an exogenous factor plays a crucial role in determining abundance and distribution of insect and pest population. But even today, a very few theoretical frameworks are available to examine the effect of climate on population dynamics. Because of this, the quest for generalizations and for developing

**Author for correspondence* *Received December 11, 2014* *Revised April 19, 2015* *Accepted May 16, 2015*

adequate predictive process-based models of change [Harrington *et al.* (2001)] remains difficult. The common approach for analysing 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)].

Research works are intended to understand pest dynamics with the use of analytical and related procedures on pest surveillance data sets. These forecasts would help the farmers in the pest management programs as they are capable of avoiding unwarranted chemical pesticides being sprayed on the crop, thereby stopping crop losses and also bringing about improved environment quality. Normally, the farmers come across severe losses on their crops due to pests. To protect the farmers from the losses and to assist them in better production, successful pest prediction methodologies are expected. It is essential that a competent pest control methodology would be able to forecast the pest dynamics accurately. Thus, the need for creating an applicable and functionally viable model to forecast the pest incidence is the current requirement. Paul *et al.* (2010, 2014) applied autoregressive moving average (ARIMA) models for forecasting agricultural prices. Paul *et al.* (2013) studied the ARIMA with exogenous variables (ARIMAX) model for forecasting wheat yield. In the present investigation, time series weekly data on number of pest per three leaves is used. Time series models including weather variables to forecast the pest population in two districts namely Guntur and Faridkot have been attempted.

2. Materials and Methods

Transformation

We first perform transformations to improve the additivity and homoscedasticity of the time-series. Various forms of transformation could be used, including a square root transformation, a logarithmic transformation and, more generally, the Box–Cox transformation. Among these, a logarithmic transformation is usually preferred when analyzing population dynamics for the following reasons. Most populations change by multiplicative factors such as the mortality and birth rates. The logarithmic transformation makes a multiplicative factor into an additive factor. Therefore, if we use logarithmic transformation, we are able to use an additive model, which is analytically more tractable. Simultaneously, homoscedasticity arises in most cases because the multiplicative error factor becomes additive by a logarithmic transformation. Thus, we use a logarithmic transformation in the following analyses. To solve the problem that arises from the discreteness of the number of individuals, we use $\log(x+0.5)$ where x is the number of individuals, although most people traditionally use a transformation of the form $\log(x+1)$, which is less preferable [Yamamura (1999)]. We use a common logarithm $\log_{10}(x+0.5)$ instead of a natural logarithm $\log_e(x+0.5)$, so that we are able to easily back-transform the variable using mental calculations.

ARIMAX Model

The ARIMAX model [Bierens (1987)] is a generalization of the ARIMA model, which is 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

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

Where, L is the usual lag operator, *i.e.* $L^s y_t = y_{t-s}$, $\Delta y_t = y_t - y_{t-1}$, $\mu \in R$, $\alpha_s \in R$, $\beta_s \in R^k$ and $\gamma_s \in R$ are the unknown parameters and e_t's are the errors and p, q and r are natural numbers specified in advance.

The first step in building an ARIMAX model consists of identifying a suitable ARIMA model for the endogenous variable. The ARIMAX model concept requires testing of stationarity of exogenous variable before modelling. The transformed variable is added to the ARIMA model in the second step, in which the lag length r is also estimated. Nonlinear least squares estimation procedure is employed to estimate the parameters of ARIMAX model [Bierens (1987)]. Fortunately, the ARIMAX model can be fitted to data by using a software package, like SAS, MATLAB, EViews and R. In the present investigation, SAS, Version 9.3 is used for data analysis.

Table 1 : Pearson's correlation coefficient between pest population and weather variables.

	Guntur			Faridkot		
	Jassids	Thrips	Whitefly	Jassids	Thrips	Whitefly
Max temperature	-0.039	0.294	-0.319**	0.372**	0.228*	0.450**
Min Temperature	0.117	0.478**	-0.177	0.499**	0.444**	0.568**
Rainfall	-0.149	-0.038	-0.093	0.083	0.232*	0.115
Max RH	-0.024	-0.325**	0.339**	0.281**	0.194	0.280**
Min RH	0.027	0.075	0.051	0.305**	0.523**	0.362**

** and * denote level of significance at 1% and 5%, respectively.

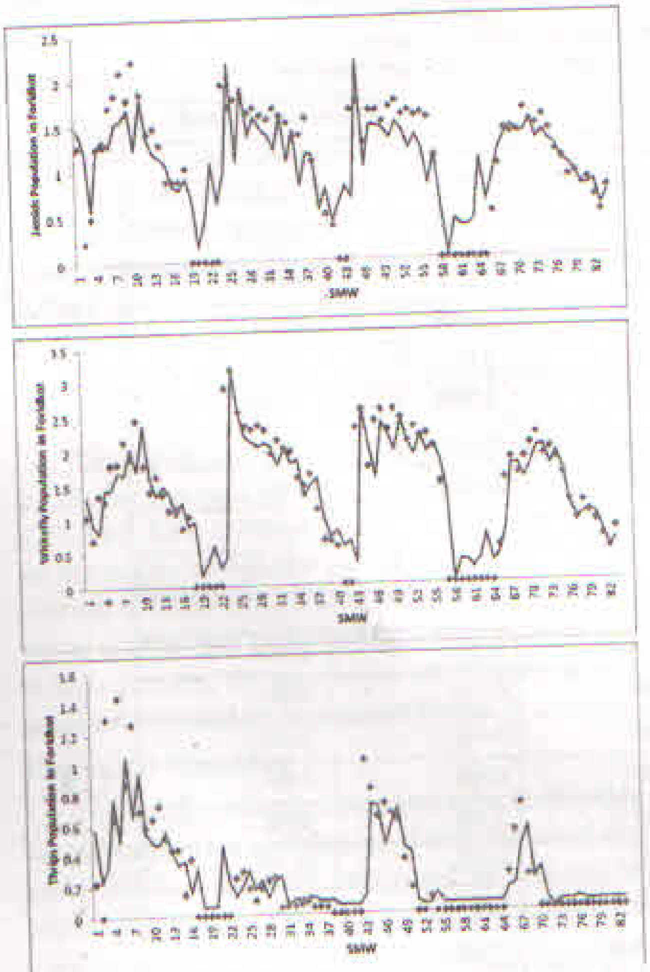

Fig. 2 : Observed vs predicted pest population in Foridkot district.

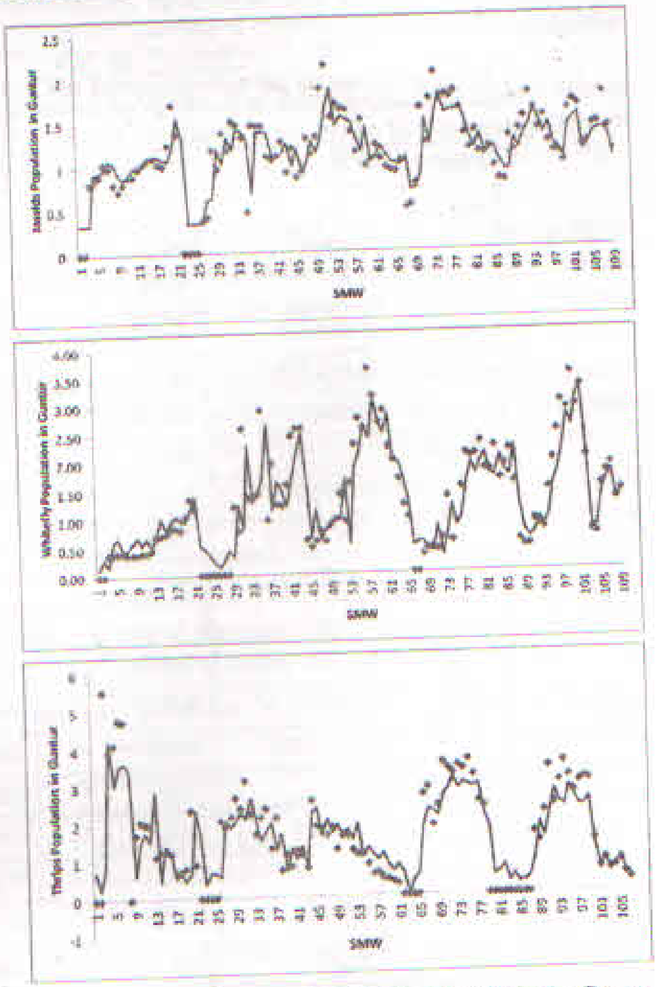

Fig. 3 : Observed vs predicted pest population in Guntur district.

3. Results and Discussion

For the present investigation, weekly data on number of pest per three leaves in Guntur and Faridkot district along with weekly maximum temperature, minimum temperature, rainfall, maximum RH and minimum RH have been used. Out of total 110 weeks' data (2008-09 to 2012-13), the first 96 observations were used for model building and the remaining 14 data points were used for validating the model. The correlation coefficient between number of pest per three leaves and maximum temperature, minimum temperature and minimum RH were found to be significant (Table 1). So, only these variables were used for subsequent model development.

Fitting of ARIMAX model

Evidently, the data set for pest population is not stationary. In order to select the order of the ARIMA model, Augmented Dickey Fuller (ADF) unit root test

Table 2 : ADF test for stationarity testing.

	Guntur	Faridkot
Jassids	-3.840	-3.062
Thrips	-4.502	-2.939
Whitefly	-3.675	-2.967
Max temperature	-3.542	-4.080
Min Temperature	-2.943	-2.915
Rainfall	-4.312	-4.713
Max RH	-2.912	-2.928
Min RH	-3.607	2.930

Tabulated value of ADF test statistic at 5% level of significance is -2.8889

Table 3 : Parameter estimate for Guntur and Faridkot districts.

Guntur					
Pest	Variable	Coefficient	Std. Error	t-Statistic	Prob.
Jassids	C	1.174	0.102	11.546	<0.001
	AR(1)	0.714	0.064	11.216	<0.001
Whitefly	C	2.618	1.555	1.684	0.095
	MAXT	-0.065	0.030	-2.189	0.031
	MAXRH	0.020	0.009	2.192	0.032
	AR(1)	0.779	0.060	12.943	<0.001
Thrips	C	0.953	1.631	0.584	0.560
	MINT	0.111	0.038	2.920	0.004
	MAXRH	-0.06	0.014	-4.334	<0.001
	AR(1)	0.648	0.074	8.760	<0.001
Faridkot					
Jassids	C	-0.843	0.504	-1.672	0.098
	MINT	0.033	0.009	3.614	0.001
	MAXRH	0.014	0.005	2.641	0.010
Whitefly	C	0.307	0.655	0.468	0.641
	MAXT	0.043	0.019	2.248	0.036
	AR(1)	0.677	0.103	6.551	<0.001
	MA(1)	0.276	0.135	2.034	0.045
	MINT	0.111	0.038	2.920	0.004
Thrips	C	0.184	0.086	2.152	0.034
	RAIN	0.002	0.001	2.757	0.007
	AR(1)	0.688	0.079	8.696	<0.001

proposed by Dickey and Fuller (1979) was applied for parameter ρ in the auxiliary regression

$$\Delta y_t = \rho y_{t-1} + \alpha_1 \Delta y_{t-1} + \varepsilon_t \qquad (2)$$

Where, $\Delta y_t = y_t - y_{t-1}$. The relevant null hypothesis is $H_0 : \rho = 0$ and the alternative is $H_1 : \rho < 0$. ADF test was conducted and the result is reported in Table 2. A perusal of table indicates that H_0 is rejected at 5% level. Thus, there is no presence of unit root and so differencing was not required.

On the other hand, the time-series data of weekly maximum temperature, minimum temperature, rainfall, maximum RH and minimum RH were found to be stationary. On the basis of minimum Akaike information criterion (AIC) and Schwartz-Bayesian criterion (BIC) values, best ARIMAX model was selected. The parameter estimates of ARIMAX model along with their significance level for Guntur and Faridkot districts are presented in Table 3. A perusal of Table 3 revealed that in Guntur district, for Jassids, no exogenous variable has entered in the model rather AR(1) model was found to be the best for this pest; for Whitefly, along with AR(1) coefficient, maximum temperature and maximum RH also have significant role in modelling as well as forecasting; for Thrips, along with AR(1) coefficient, minimum temperature and maximum RH have also entered into the model. Similarly, for Faridkot district, for Jassids, minimum temperature and maximum RH have taken part in model development; for Whitefly, along with the ARMA(1,1) coefficients, maximum temperature and minimum temperature also played significant role in forecasting; for Thrips, Rainfall and AR(1) coefficient are part of the model. From Table 3, it is also interesting to note that all the coefficients are significant at 5% level of significance except constant in the model. The fitted model along with the data points are displayed in Figs. 1-2. A perusal of figures indicates that the fitted model is a good fit for the data under consideration.

Validation of models for hold-out data

One-step ahead forecasts of pest population for the last 14 weeks data in respect of above fitted models were computed. A comparative study of forecasts by these models was carried out on the basis of Relative Mean Absolute Prediction Error (RMAPE) values defined as

$$\text{RMAPE} = 1/14\sum_{i=1}^{14}\left\{\left|y_{t+i} - \hat{y}_{t+i}\right|/y_{t+i}\right\} \times 100 \qquad (3)$$

The RMAPE values for Jassids, Whitefly and Thrps for Guntur and Faridkot districts are found to be 8.28, 6.71, 8.67, 8.91, 7.76 and 9.25%, respectively. It may be noted that models for all the three pests in two districts perform quite satisfactorily as the RMAPE values are less than 10%. Further, it is to be noted that in both districts, the population of Whitefly is best predicted followed by Jassids and Thrips.

Diagnostic Checking

The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen ARIMAX model, which has been done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, ACF and PACF up to 16 lags were computed. It is also found that none of these autocorrelations is significantly different from zero at any reasonable level. This proved that the selected ARIMAX model is an appropriate model for forecasting pest population, which also indicated the 'good fit' of the model.

Acknowledgements

Authors thankfully acknowledge the learned reviewer for his fruitful comments to improve the quality of this paper.

References

Bale, J., G. Masters, I. Hodkinson, C. Awmack, T. Bezemer, V. Brown, J. Butterfield, A. Buse, J. Coulson, J. Farrar and others (2002). Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Glob. Chang. Biol.*, **8**, 1–16.

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

Cammell, M. and J. Knight (1992). Effects of climatic change on the population dynamics of crop pests. *Adv. Ecol. Res.*, **22**, 117–162.

Cannon, R. (1998). The implications of predicted climate change for insect pests in the UK, with emphasis on non-indigenous species. *Glob. Chang. Biol.*, **4**, 785–796.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for the autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427-31.

Hagstrum, D. and G. Milliken (1988). Quantitative analysis of temperature, moisture and diet factors affecting insect development. *Ann. Entomol. Soc. Am.*, **81**, 539–546.

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

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

Meehl, G., T. Stocker, W. Collins, P. Friedlingstein, A. Gaye, S. Solomon, D. Qin, M. Manning, Z. Chen and M. Marquis (2007). *Climate Change, 2007: the Physical Science Basis.* Contribution of Working group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Merrill, R., D. Gutie'rrez, O. Lewis, J. Gutie'rrez, S. Diez and R. Wilson (2008). Combined effects of climate and biotic interactions on the elevational range of a phytophagous insect. *J. Anim. Ecol.*, **77**, 145–155.

Parmesan, C. (2007). Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Glob. Chang. Biol.*, **13**, 1860–1872.

Paul, R. K., Prajneshu and H. Ghosh (2013). Modelling and forecasting of wheat yield data based on weather variables. *Indian Journal of Agricultural Science*, **83(2)**, 180-183.

Paul, R. K. and M. K. Das (2010). Statistical modelling of inland fish production in India. *Journal of the Inland*

Fisheries Society of India, **42(2)**, 1-7.

Paul, R. K., W. Alam and A. K. Paul (2014). Prospects of livestock and dairy production in India under time series framework. *Indian Journal of Animal Sciences*, **84(4)**, 130-134.

Porter, J. H., M. L. Parry and T. R. Carter (1991). The potential effects of climatic change on agricultural insect pest. *Agric. For. Meteorol.*, **57**, 221–240.

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

Samways, M. (2005). *Insect Diversity Conservation*. Cambridge University Press, Cambridge, 342 p.

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

Woiwod, I. (1997). Detecting the effects of climate change on Lepidoptera. *J. Insect Conserv.*, **1**, 149–158.

Yamamura, K. (1999). Transformation using (x+0.5) to stabilize the variance of populations. *Res. Popul. Ecol.*, **41**, 229–234.