Madras Agric. J., 105 (7-9): 313-316, September 2018

Comparison of Prediction Accuracy of Multiple Linear Regression, ARIMA and ARIMAX Model for Pest Incidence of Cotton with Weather Factors

V.S. Aswathi* **and M.R. Duraisamy**
Department of Physical Science and Information Technology,
Tamil Nadu Agricultural University, Coimbatore - 641 003.

Identifying suitable statistical model for predicting pest incidence have important role in pest management programmes. For this study weekly data of aphid, thrips, jassid and whitefly incidence of cotton at the TNAU region, Coimbatore and the weather factors influencing these pests incidence were used for model development. Rainfall, maximum temperature, minimum temperature, morning humidity, evening humidity were used as the independent variables and MLR, ARIMA, ARIMAX models built for each pests. Comparison of these three models was done and checked the model accuracy using root mean square error value. It was found that for all pests ARIMAX model posses lowest RMSE value compared to ARIMA and MLR. So ARIMAX model was selected as best fit model.

Key words*:* Cotton pests, Multiple linear regression, ARIMA, ARIMAX, Weather factors

Cotton (*Gossypium* spp.) is one among the most important cash crops playing a key role in Indian economy and is designated as “king of fibre crops”. India is the largest producer of cotton in the world accounting for about twenty six per cent of global cotton production. Main factors affecting cotton yield are climatic conditions, pests and diseases. Cotton is ravaged by many insect pests starting from sowing to harvesting stage. The pests which are more problematic are sucking pests and bollworms. Understanding the factors affecting population abundance of the pest during the crop as well as off seasons would guide in formulating strategies for their management. Pest incidence is highly influenced by weather factors. Hence an attempt has been made to develop pest forewarning models using weather factors to reduce yield loss in advance. The development of models can depict the interaction of the pests and environment over time. In the present study, identified the suitable model for prediction of pest incidence and these findings may give reliable methods to identify environmental condition that are conducive for a particular pest development in cotton.

Material and Methods

Standard weekly data for pest incidence (pests per three leaves) were collected for cotton variety DCH-32 from Cotton Department, TNAU, Coimbatore. The data corresponding to crop periods from September to January for four major cotton pests such as aphid, thrips, jassid and whitefly were selected for the study. The corresponding pest incidence data for aphids and thrips were recorded for the period of 2008-2012 while for Jassids and Whitefly were recorded for the period of 2008-2017 and 2008-2015 respectively.

*Corresponding author’s email: aswathivs3993@gmail.com

Weekly weather data for Coimbatore (TNAU region) were collected from Agro Climatic Research Centre, TNAU, Coimbatore. The weather factors selected for the study were maximum temperature(°C), minimum temperature (°C), morning humidity(%), evening humidity (%) ,rainfall(mm). Here cotton pest data did not follow normality, so *square root transformation* ($\sqrt{X + 0.5}$) was performed.

Multiple linear regression (MLR)

The multiple regression procedure was utilized over the training data set to estimate the significant regression coefficients of the linear equation

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots + \beta_i X_i + \varepsilon$$

Where: β_0 = Intercept,

β_i = regression coefficient of i^{th} independent parameters, (i = 1,2,…, n),

X_i = i^{th} weather parameter (independent variable) and ε = error term.

$$Y = X \beta + \varepsilon.$$

Using least square criteria, regression coefficients were estimated. $\hat{\beta} = (X' X)^{-1} X' Y\hat{Y}$

Autoregressive integrated moving average(ARIMA) model

An autoregressive model of order p is conventionally classified as AR (p) and a moving average model with q terms is known as MA (q). A combined model that contains p autoregressive terms and q moving average terms is called ARMA (p,q). If the object series is differenced d times to achieve stationary, the model is classified as ARIMA (p, d, q). Thus, an ARIMA model is a combination of an autoregressive (AR) process and a moving average

(MA) process applied to a non-stationary data series. Since weekly data on pest has been used in this study, ARIMA models mentioned as Seasonal ARIMA models (SARIMA).

Seasonal ARIMA models are an adaptation of autoregressive integrated moving average (ARIMA) models to specifically fit seasonal time series. Their construction includes the seasonal part in model fitting. A combination of non-seasonal and seasonal process yields the multiplicative Seasonal ARIMA, $(p,d,q) \times (P,D,Q)_m$, where m is the periods per season. The general forms as:

$$\Phi_P(B^s)\, \phi_p(B)\, (1-B)^d\, (1-B^s)^D Y_t = \Theta_Q(B)\, \varepsilon_t$$

Where: $\phi_p(B)$, $\theta q(B)$ and $(1-B)^d$ are non-seasonal autoregressive operator of order p, non-seasonal moving average operator of order q and non-seasonal differencing operator of order d, respectively.

Where $\Phi_P(B^s)$, $(1-B^s)^D$ and $\Theta_Q(B^s)$ are seasonal autoregressive operator, moving average operator and differencing operator their orders are P,D and Q respectively, and are represented by:

$$\Phi_P(B^s) = 1- \Phi_1 B^s- \Phi_2 B^{2s} - \text{.....} - \Phi_P B^{Ps}$$

$$\Theta_Q(B^s) = 1- \Theta_1 B^s- \Theta_2 B^{2s} - \text{.....} - \Theta_Q B^{Qs}$$

Where: s indicates the length of seasonality.

ARIMAX or regression with ARIMA errors

ARIMAX method is an extension method of ARIMA model. In forecasting, this method involves independent variables also.The independent variables used in this study were weather factors that affect the pest incidence. Multiple regression models of the form $Y = \beta_0 + \beta_1 X_1 + \text{....} + \beta_i X_i + \varepsilon$. Where Y is a dependent variable of the X_i predictor variables and ε usually assumed to be an uncorrelated error term (i.e., it is white noise). We considered tests such as the Durbin-Watson test for assessing whether ε was significantly correlated. We will replace ε by n_t in the equation. The error series n_t is assumed to follow an ARIMA model. For example, if n_t follows an ARIMA (1,1,1) model, we can write

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{.....} + \beta_i X_i + n_t$$

$$(1- \phi_1 B)(1-B)\, n_t = (1+\theta_1 B)\varepsilon_t$$

Where ε_t, is a white noise series. ARIMAX model have two error terms here the error from the regression model which we denote by n_t and the error from the ARIMA model which we denote by ε_t. Only the ARIMA model errors are assumed to be white noise.

Root mean square error

The root mean square error (RMSE) is a frequently used measure of the difference between values predicted by a model and the actually observed values. Less RMSE needed for the best fit model. RMSE can be calculated by using the equation

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

Where y_i =actual value and $\hat{y}_i$=estimated value.

Data analysis was done by using Excel,SPSS, R software.

Akaike's information criterion (AIC)

The AIC computation is based on the mathematical formula AIC = −2log L + 2m, where m = p + q + P + Q is the number of parameters in the model and *L* is the likelihood function. Since -2 log L + 2m is approximately equal to $\{n (1+\log 2\pi) + n \log \sigma^2\}$ where σ^2 is the model MSE. The best model is the one with the lowest AIC value.

Results and Discussion

Multiple linear regression of pests with weather factors

Multiple linear regression equation of aphid, thrips, jassid, whitefly were given in the Table 1. MLR for aphid was found to be significant with R^2value of 0.232 and RMSE 0.87699. MLR for thrips was found to be significant with R^2 value of 0.098 and RMSE

Table 1. Multiple linear regression of pests with weather factors

Pests	Equation	R^2	RMSE
Aphid	Y= 7.027-0.006(rain)-0.286 (maximum temperature) +0.178(minimum temperature)+ 0.046(morning humidity)-0.066(evening humidity)	0.232**	0.87699
Thrips	Y= 2.564+0.002(rain)+0.043 (maximum temperature) -0.034(minimum temperature)-0.029(morning humidity)+0.008(evening humidity)	0.098*	0.32686
Jassid	Y= 5.101-0.002(rain)+0.047 (maximum temperature)-0.049(minimum temperature)-0.033(morning humidity)-0.006 (evening humidity)	0.141**	0.71047
Whitefly	Y= 4.766-0.000032(rain)-0.084 (maximum temperature) + 0.008(minimum temperature)- 0.001(morning humidity)-0.015(evening humidity)	0.062	0.56966

Note: ** significant at 1%, *significant at 5%.

0.32686. MLR for jassid was found to be significant with R^2 value of 0.141 and RMSE 0.71047. MLR for whitefly was not significant with R^2 value of 0.062 and RMSE 0.5696. Akram *et al* (2013) also used same method to study impact of weather on whitefly and thrips incidence of cotton. Raghavendra *et al.* (2014) also used MLR for predicting pest incidence of cotton.

Table 2. Fitted ARIMA and ARIMAX models of pests

Pests	Models		RMSE	AIC
Aphid	SARIMA	$(0,0,3)\ (1,0,1)_{20}$	0.635	210.4
	SARIMAX	$(1,0,0)\ (1,0,0)_{20}$	0.616	204.9
Thrips	SARIMA	$(1,0,0)_{20}$	0.2872	40.63
	SARIMAX	$(1,0,0)_{20}$	0.2730	40.53
Jassid	SARIMA	$(0,1,0)\ (2,0,1)_{20}$	0.4188	207
	SARIMAX	$(0,1,0)\ (2,0,1)_{20}$	0.3980	197.7
Whitefly	SARIMA	$(1,0,0)\ (2,0,0)_{20}$	0.3308	113.8
	SARIMAX	$(1,0,0)\ (2,0,1)_{20}$	0.2932	100.3

ARIMA and ARIMAX models for pests

Weekly data of pest incidence were used to fit ARIMA and Autoregressive Integrated Moving Average with regression (ARIMAX) model. So ARIMA and ARIMAX model mentioned as seasonal ARIMA(SARIMA) and SARIMAX respectively. Best fitted models are presented in the Table 2,and RMSE and AIC values were calculated.

For aphid SARIMA $(0,0,3)\ (1,0,1)_{20}$ and SARIMAX$(1,0,0)\ (1,0,0)_{20}$ with 0.63 and 0.61 RMSE respectively. AIC values are 210.4 and 204.9 respectively. For Thrips SARIMA $(1,0,0)_{20}$ and SARIMAX $(1,0,0)_{20}$ with 0.2872 and 0.2730 RMSE respectively. AIC values are 40.63 and 40.53 respectively. In case of Jassid SARIMA (0,1,0) $(2,0,1)_{20}$ and SARIMAX $(0,1,0)\ (2,0,1)_{20}$ with 0.4188 and 0.3980 RMSE values respectively. AIC values are 207 and 197.7 respectively. For Whitefly SARIMA $(1,0,0)\ (2,0,0)_{20}$ and SARIMAX $(1,0,0)\ (2,0,1)_{20}$ with 0.3308 and 0.2932 RMSE respectively. AIC values are 113.8 and 100.3 respectively.

Table 3. RMSE value of MLR,ARIMA, ARIMAX for different pests

Pests	Models	RMSE
Aphid	MLR	0.876
	ARIMA	0.635
	ARIMAX	0.616
Thrips	MLR	0.326
	ARIMA	0.287
	ARIMAX	0.273
Jassid	MLR	0.710
	ARIMA	0.418
	ARIMAX	0.398
Whitefly	MLR	0.569
	ARIMA	0.330
	ARIMAX	0.293

ARIMAX showed lowest RMSE and AIC values. From the best fitted model, the result concluded that ARIMAX showed better result compared to ARIMA. Arya *et al*. (2015) predicted pest population using ARIMAX model. Anggraeni *et al.*(2017) also used ARIMAX method to predict the number of dengue fever patients in Indonesia by using trend data and the results showed that ARIMAX was suitable for irregular patterned data. Kongcharoen and Kruangpradit (2013) also used ARIMAX for forecasting export to major trade partners.

Comparison between MLR, ARIMA and ARIMAX for different pests

Accuracy of models was checked by using RMSE value and it was given in the Table 3. For aphid lowest RMSE value obtained for ARIMAX model (0.616) followed by ARIMA model(0.635) and then MLR model (0.876). For thrips lowest RMSE value for ARIMAX model (0.273) followed by ARIMA (0.287) and then MLR model (0.326). For jassid lowest RMSE value for ARIMAX model(0.398) followed by ARIMA model (0.418) and then MLR model(0.710).For whitefly lowest RMSE value for ARIMAX model(0.293) followed by ARIMA(0.330) and then MLR model(0.569).Based on the lowest RMSE value, ARIMAX model recorded highest accuracy of prediction compared to ARIMA and MLR models.

Conclusion

ARIMAX model recorded the lowest RMSE values for all the pests compared to MLR and ARIMA. So ARIMAX selected as best model for predicting pest incidence of aphid, thrips, jassid and whitefly of cotton using weather factors in the study area.

Acknowledgement

Our sincere gratitude to Dr. Patil Santosh Ganapati, who has helped us throughout the research programme.

References

Akram, M, Hafeez, F, Farooq, M, Arshad, M, Hussain, M, Ahmed, S. and Khan, H.A. 2013. A case to study population dynamics of *Bemisia tabaci and Thrips tabaci* on Bt and non-Bt cotton genotypes. *Pakistan Journal of Agricultural Sciences,* **50(4)**, 617-623.

Anggraeni,W,Pusparinda,N, Riksakomara, E, Samopa, F.and Pujiadi. 2017. The Performance of ARIMAX Method in Forecasting Number of Tuberculosis Patients in Malang Regency, Indonesia. *International J. Applied Engineering Research,* **12(17)**,189-196.

Arya, P, Paul, R. K, Kumar, A, Singh, K. N.and Sivaramne, N. 2015. Predicting pest population using weather variables : An ARIMAX time series framework. *International Journal of Agricultural and Statistical Sciences,* **11(2)**, 381-386.

Boopathi,T, Singh,S.B, Manju,T, Ramakrishna,Y, Akoijam,R.S, Samik Chowdhury and Ngachan,S.V.2015. Development of temporal modeling for forecasting and prediction of the incidence of lychee, *Tessaratoma papillosa* (Hemiptera: Tessaratomidae), using time series (ARIMA) analysis. *Journal of Insect Science,***15(1)**,55-59.

Campos, A. C, Campos, M. S, Bustillo, C. W. G, Villafranca, M. H.and Johanssonb, T. 2012. Non-parametric statistical methods and data transformations in agricultural pest population studies. *Chilean Journal of Agricultural Research*, **72(3)**,440-443.

Dhar, T, Ghosh, A, Senapati, S. K.and Bhattacharya, S. 2014. Identification of prediction model on population build up of *Singhiella pallida* on *Piper betle* L. for timely intervention. *International Journal of Agriculture, Environment and Biotechnology,* **7(4)**, 883.

Hameed, A, Shahzad, M.S, bidmehmood, Ahmad, S.and Noor-UlIslam.2014. Forecasting and modeling of sucking insect complex of cotton under agro-ecosystem of Multan- Punjab, Pakistan. *Pakistan Journal of Agricultural Sciences,* **51(4)**, 997-1003.

Janu, A. and Dahiya, K. 2017. Influence of weather parameters on population of whitefly, *Bemisia tabaci* in American cotton (*Gossypium hirsutum*). *Journal of Entomology and Zoology Studies,* **5(4)**, 649-654.

Kadam, D. B, Kadam, D. R. and Umate, S. M. 2015. Effects of weather parameters on incidence sucking pests on Bt cotton. *International Journal of Plant Protection,* **8(1)**, 211-213.

Kongcharoen, C, and Kruangpradit, T. 2013. Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export. Paper presented at the 33rd International Symposium on Forecasting,South Korea, p.1-8, www.researchgate.net/publication/255731345.

Panwar, T. S, Singh, S. B. and Garg, V. K. 2015. Influence of meteorological parameters on population dynamics of thrips and aphid in Bt and non Bt cotton at Malwa region of Madhya Pradesh. *Journal of Agrometeorology,* **17(1)**, 136-138.

Raghavendra, K. V, Naik, D. S. B, Mieee, S.Venkatrama phanikumar.and Mieee. 2014. Weather Based Prediction of Pests in Cotton. Paper presented at the sixth International Conference on Computational Intelligence and Communication Networks, p.570-574,www.researchgate.net/publication/283649380.

Received : February 20, 2018; Revised : March 10, 2018; Accepted : March 25, 2018