Environment and Ecology 38 (2) : 277—281, April—June 2020
ISSN 0970-0420

Study of Predictive Ability of Generalized Linear Models (GLM's) for Population Dynamics of Brinjal Mites

Yogeesh K. J., Abhishek Singh, Raghuraman M., Nema A. K.

Received 25 January 2020 ; Accepted 13 March 2020 ; Published on 3 April 2020

ABSTRACT

One of the important limitation of all time in agricultural is pest damage our research attempted to find out the best model for brinjal mites prediction efficiently using the Generalized model because the nature of insect counts are non-continues or non-negative and the test like the Kolmogorov–Smirnov test (0.21** for 168 df) and Shapiro–Wilk's test (0.84** for 168 df) and the high variance mean ratio (12) provide the evidence that data nature are over dispersed and positively skewed (skewness 1.6) and direction to go for GLM's. In GLM's the models like Poisson model, Quassi-Poisson model, Negative binomial model fitted and we were obtained good predictive ability measures for Negative binomial model with AIC and BIC is about 473.28 and 512.67 respectively with Pseudo R^2 of 0.72 and 1.495, 0.703 of RMSE and MAPE respectively. The weather parameters like Minimum temperature, Maximum Relative humidity, time lag plays a important role in population dynamics of mites.

Keywords Pest, Poisson, Quassi-Poisson, Negative binomial, Prediction.

Yogeesh K. J.[1]*, Abhishek Singh[1], Raghuraman M.[2], Nema A. K.[1]
[1]Department of Farm Engineering, [2]Department of Entomology and Agriculture Zoology, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh 221005, India
Email: kjyogeesh.agri@gmail.com,asbhu2006@gmail.com
*Corresponding author

INTRODUCTION

Brinjal (*Solanum melongena* L.) is an important solanacious crop of sub-tropics and tropics. Among all summer grown vegetables with semi-perennial nature brinjal is almost available throughout the year and consumed in various forms. It is a versatile crop adapted to different agro-climatic regions and can be grown throughout the year. It is a perennial but grown commercially as an annual crop. It is a good source of minerals and vitamins and is rich in total water soluble sugars, free reducing sugars, amide proteins among other nutrients (Mondal et al. 2017, Duxbury et al. 2003, Bisseleua et al. 2011). It is amazing to record that eggplant possesses the highest nutritive value, providing energy of 24 Calories.

Insect pests infestation is one of the most limiting factors for accelerating yield potential of brinjal. The crop is prone to damage by various insects, although there is wide variability in their degree of infestation since its growing as a semi-perennial and available throughout the year hence crop get affected by very diversified insect pest in different location and season. Mites is also a important seriously pest cause damage to the important vegetables like brinjal, tomato. They are minute in size and vary in color likes green, greenish yellow, brown, or orange red with two dark spots on the body. Spider mites usually extract the cell contents from the leaves using their long, needle - like mouthparts. This results in reduced chlorophyll content in the leaves, leading to the formation of white or yellow speckles on the leaves. In severe infestations, leaves will completely desiccate and drop off.

Most of the model based on insect pest in agriculture crops forecasting techniques mostly come across with count time series data than any continues or non-negative data points (Prabhakar Mathyam 2012). Our emphasis is more given to insect count time series data. Count data are the data in which observations can take only the non-negative integer value. Among the characteristics of such data are those including rare events, skewed, discrete and often exhibit over-dispersion and excess zeroes. Generalized linear models (GLM's) such as Poisson, Quasi-Poisson and Negative binomial models have been recommended and frequently used to deal with the count time series data depending on the nature and pattern of the data we recorded in pest forecasting, several intrinsic attributes of the insects and the determining environmental and host factors need to be considered. The collection and analysis of weather data from pest-affected areas are an essential input for models (Cohen et al. 2002, Gent et al. 2012, Maindonald and Braun 2007). The practical application of model outputs is aided by decision support systems. Mainly these are as follows accurate forecasting of pest attacks before they actually take place is desired in pest control programs, so that control measures can be planned with maximum efficiency so to get the efficient model with good predictive ability so that our plan and control measures planned accordingly.

MATERIALS AND METHODS

This study was undertaken to study the predictive ability of GLM's for mites infestation in brinjal at Varanasi region. The week wise data collected for the brinjal mites. Numbers / square cm from the period 9th to 36th standard meteorological week that is from February 27th to September 9th from 2011 to 2016 and respective period collected weather parameters like Rainfall (mm), Maximum temperature (°C), Minimum temperature (°C), Maximum Relative humidity (%), Minimum Relative humidity (%) Sunshine hours (h), Lagged values (Yt–1) were considered in addition recorded the different growing stages of the crop during the study period.

To summarize data sets we used exploratory data analyses and the descriptive statistical measures and to study the relationship between mites infestation at different crop growth stages with weather parameters can be obtained through Pearson correlation coefficient.

To study the pattern of an event count non-negative integer-valued random variable data distribution we employed, Variance - Mean ratio ($D = \frac{\sigma 2}{\mu}$) test, Kolmogorov–Smirnov test and Shapiro–Wilk's test (Raghu et al. 2014, Rodrigues–Motta et al. 2013). For modeling the count time series data or discrete distributions data we used Generalized linear model (GLM's) like Poisson regression model, Quassi-poisson regression model, Negative binomial regression model is used.

Poisson distribution is a class of Generalized linear model which follows Poisson distribution. Let us consider a random variable y follows a Poisson distribution with parameter μ, if it takes integer values $y = 0, 1, 2, \ldots$ with probability distribution.

$$P(Y = y) = \frac{e^{-\mu} \quad \mu^y}{y!} \quad \mu > 0$$

The parameters of Poisson regression model.

$$yi = \exp(X_i' \beta) + \varepsilon_i$$

Can be estimated using maximum likelihood method.

Similarly, the estimating equation used for the quassi-likelihood approach for count data are :

$$\sum_{i=1}^{n} \frac{Y_i - \mu_i}{\varphi \mu_i} = 0$$

Let Y is random variable which follows the negative binomial distribution with parameters (r, θ), where $\theta \in (0, 1)$ and r an integer, then its probability mass function is given by :

$$P(Y=y) = \frac{y+r-1}{y} \theta^{y}(1-\theta)^{r}, \quad y=0,1,2, ...$$

The in negative binomial regression, the mean of y is determined by the exposure time t and a set of k regressor variables (the $x^{'s}$). The expression relating these quantities is :

$$u_i = \exp(\ln(t_i) + b_1 x_{1i} + b_2 x_{2i} + ... + b_k x_{ki})$$

$y \sim NegBin(r, \theta)$ Therefore $E(Y) = \frac{r\theta}{(1-\theta)}$ and $Var(Y) = \frac{r\theta}{(1-\theta)^2}$

RESULTS AND DISCUSSION

The summarization of the insect count data were done by descriptive and exploratory data techniques mean count of mites/cm^2 was about 13.3 with standard deviation 12.4 the maximum counts of mites was recorded to be 55/cm^2 and minimum counts recorded was 0. The skewness is found to be more than 1 i. e., 1.6 this value characterizes the positively skewed distribution. The frequency of insect counts was observed on week interval basis from the 9th to 36th standard meteorological week that is February 27th to September 9th from 2011 to 2016 and the pattern of frequency is depicted in Fig. 1 clearly shows that the frequency not in a normal pattern and likely to be skewed distribution (positive skew with value 1.6).

The correlation between population dynamics of brinjal mites with weather parameters were computed through Pearson product moment method and the result shows that the parameters like Maximum temperature (0.21), Minimum temperature (0.37*), Maximum Relative humidity (0.32*), Minimum Relative humidity (0.13) and Sunshine hours (0.06), Lagged value (Yt-1) (0.63**) found to be positively significant with insect occurrences similarly rainfall shows negative correlation (–0.36*) with weather parameter.

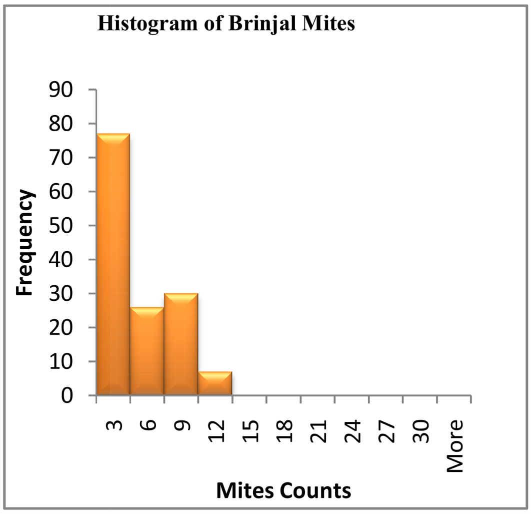

Fig. 1. Frequency distribution of brinjal mites over a period of time.

The test like the Kolmogorov–Smirnov test (0.21** for 168 df) and Shapiro–Wilk's test (0.84** for 168 df) and the high variance mean ratio (12) concludes the data nature and pattern that non-normality of insect counts and there will be more degree of dispersion or variations in mites population dynamics and it gives the direction for choosing the models which is more appropriate for the given data set. Discrete, non-negative, more dispersed data can be modeled by Generalized linear model for best forecasting of the insect appearances.

The Generalized linear models considered to fit the data sets are Poisson regression, Quassi-Poisson regression and Negative binomial regression models. The goodness of fit for different GLM models depicted in the Table 1. The result shows that the

Table 1. Descriptive statistics of brinjal mites in the study area.

Measures	Mean	Std dev	cv (%)	Minimum	Maximum	Skewness	Kurtosis
Mites counts	13.3	12.4	93	0	55	1.6	1.9

Table 2. Chi-square goodness of fit for GLM's.

Test statistics	Df	P	Model
21.131	61	0.042	Poisson
21.131	61	0.042	Quassi-Poisson
8.271	61	0.748	Negative Binomial

chi-square test statistics for goodness of fit is found to be significant for Poisson (21.13, 0.042* for 61 df) and Quassi-Poisson model (21.13, 0.042* for 61 df) so we should reject Null hypothesis that there is no significant difference between observed values and the expected value but for Negative binomial model was found to non-significant value (8.27, 0.748 for 61 df) so we conclude that the mite population is better fitted to Negative binomial model compared to Poisson and Quassi-Poisson. The model coefficient for Negative binomial regression is given by ln (y_t) = 1.12–0.04 (Rainfall) + 1.02 (Min temp) + 0.01 (Max tem) + 2.27 (Y_{t-1}).

The model selection criteria for the best prediction and forecasting of the brinjal mites pest incidences is referred to the Table 2 it explains that in Generalized linear models like Poisson model, Quassi-Poisson model, Negative binomial model we were obtained good predictive ability measures for Negative binomial model with AIC and BIC is about 473.28 and 512.67 respectively with Pseudo R^2 of 0.72 and 1.495, 0.703 of RMSE and MAPE respectively. So we proceeded to for forecasting of pest incidences of brinjal mites using Negative binomial model and the result of actual observation and predicted observation for the year of 2016 for brinjal mites / cm^2 was depicted in Tables 3 and represent

Table 3. Selection of best model for prediction of brinjal mites per plot based on key performance indicators (KPI).

Models	AIC	BIC	Pseudo R^2	RMSE	MAPE
Poisson	498.28	585.11	0.63	2.315	1.013
Quassi-Poisson	_	_	0.63	2.315	1.013
Negative Binomial	473.28	512.67	0.72	1.495	0.703

Table 4. Predicted brinjal mite per cm square using Negative Binomial Regression Model.

Weeks No.	Dates	Brinjal growth stages	Actual	Predicted
9	26 Feb—04 Mar		1	1
10	05 Mar—11 Mar	Plant establishment and seedling	2	3
11	12 Mar—18 Mar		2	2
12	19 Mar—25 Mar		6	5
13	26 Mar—01 Apr		11	13
14	02 Apr—08 Apr	Vegetative stage	9	6
15	09 Apr—15 Apr		13	11
16	16 Apr—22 Apr		15	13
17	23 Apr—29 Apr	Flowering	37	32
18	30 Apr—06 May		35	33
19	07 May—13 May		38	41
20	14 May—20 May		33	32
21	21 May—27 May	Fruiting and first harvesting	22	20
22	28 May—03 Jun		16	14
23	04 Jun—10 Jun		16	18
24	11 Jun—17 Jun	Fruiting and subsequent harvesting	11	15
25	18 Jun—24 Jun		9	6
26	25 Jun—01 Jul		4	5
27	02 Jul—08 Jul		2	4
28	09 Jul—15 Jul		4	3
29	16 Jul—22 Jul		2	3
30	23 Jul—29 Jul		3	1
31	30 Jul—05 Aug		3	1

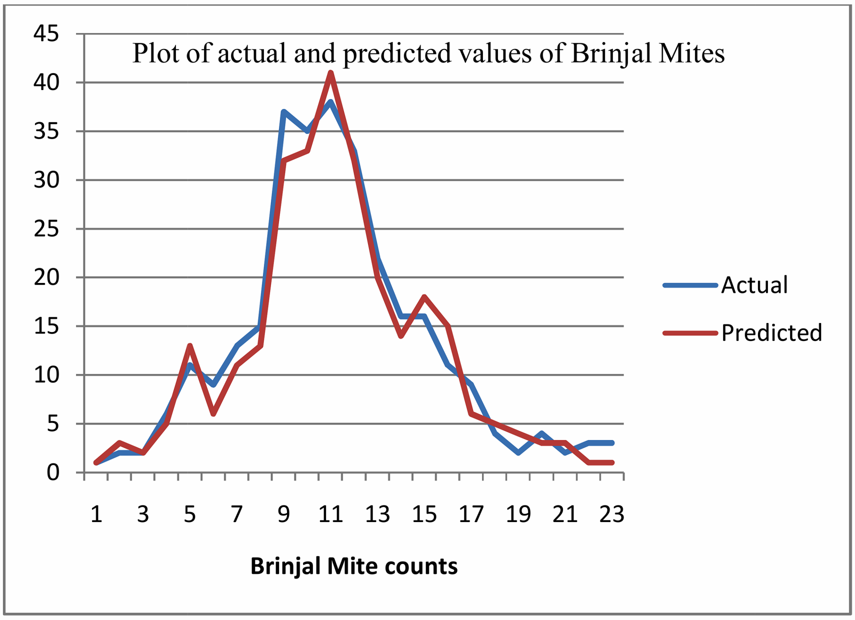

Fig. 2. Graph showing the line of actual and predicted model.

graphically it in the Fig. 2.

So for the given study area the brinjal mites population dynamics is explained by NB model compared with other models by giving lesser value of key performance indicator and good Pseudo R^2 value. The weather parameters like Minimum temperature, Maximum Relative humidity, time lag plays a important role in population dynamics of mites.

REFERENCES

Bisseleua D, Yede HB, Stefan Vidal S (2011) Dispersion models and sampling of cacao mirid bug *Sahlbergella singularis* (Hemiptera : Miridae) on Theobroma cacao in Southern Cameroon. Environm Entomol 40 (1) : 111—119.

Cohen J, Cohen P, West SG, Aiken LS (2002) Applied multiple regression / correlation analysis for the behavioral sciences. 3rd edn. Mahwah, New Jersey : Lawrence Erlbaum Associates.

Duxbury JM, Mayer AB, Lauren JG, Hassan N (2003) Food chain aspects of arsenic contamination in Bangladesh : Effects on quality and productivity of rice. J Environm Sci and Hlth, Part A : Toxic / Hazardous Substances and Environm Engg 38 (1) : 61—69.

Gent DH, Farnsworth JL, Johnson DA (2012) Spatial analysis and incidence-density relationships for downy mildew on hop. Pl Pathol 61 : 37—47.

Maindonald J, Braun J (2007) Data Analyses and Graphics using R - An Example - Based Approach. 2nd edn. Cambridge University Press, UK.

Mondal D, Ghosh A, Roy D, Kumar A, Shamurailatpam D, Bera S, Ghosh R, Bandopadhyay P, Majumder A (2017) Yield loss assessment of rice (*Oryza sativa* L.) due to different biotic stresses under system of rice intensification (SRI). J Entomol and Zool Stud 5 (4) : 1974—1980.

Prabhakar Mathyam (2012) Integrated Pest Management, Chapter : Pest Monitoring and Forecasting, Publisher : CABI, Abrol DP, Shankar U (eds), pp 41—57.

Raghu PT, Varghese Manaloor V, Nambi VA (2014) Factors influencing adoption of farm management practices in three agro biodiversity hotspots in India: An analysis using the count data model. J Natural Resour and Develop 4 : 46—53.

Rodrigues - Motta M, Pinheiro HP, Martins EG, Araujo MS, dos Reis SF (2013) Multivariate models for correlated count data. J Appl Statistics 40 (7) : 1586—1596.