J. Cotton Res. Dev. **37**(2), 233-240 (July, 2023) **ISSN No. 0972-8619**

Application of count data model for prediction of whitefly count of cotton crop

MANOJ KUMAR*, ANIL JAKHAR, ANURAG, HEMANT POONIA AND POOJA RAWAT
Department of Mathematics and Statistics, CCS Haryana Agriculture University, Hisar-125004
**Email : m25424553@gmail.com*

Abstract : An attempt has been made to predict the whitefly count data of cotton crop using weather variables through count data modeling and comparison of the ordinary regression techniques with count data modeling. For the purpose average weekly data of whitefly count (Adults/3 leaves) on three cotton varieties (Ganganagar Ageti, HS 06 and RS 2013) and one hybrid (RCH 650 BGII) grown in Cotton Section, CCS Haryana Agriculture University, Hisar was recorded from 23 to 43 standard meteorological weeks for the year 2021-2022 and pooled data of three year from 2019 to 2022 has also been made. The average weekly meteorological data on rainfall, mean temperature, mean relative humidity, sunshine hours, wind speeds *etc.* for the same period were considered for both the year 2021-2022 and pooled of three years. Data was tested for normality using Q-Q plot graph which shows that there is no normality in the data and data is over dispersed so ordinary normal regression cannot be applied Hence, negative binomial model was found to be best with low AIC and BIC value and high R2 for all the studied cotton varieties.

Keyword: Binomial model, Negative, normal model, poisson model, whitefly count

Agriculture plays a predominant role in India's economy as it contributes about seventeen percent of total gross domestic product (GDP) and about 60 per cent of population depends on agriculture sector directly or indirectly.

Agriculture being highly cost intensive and full of uncertainties and timely measures can minimize the farmer's risk. Threats from pest attacks are often localized but underlines the multitude of risks apart from those related to monsoon failure or a crash in crop prices. These incidences of pest and diseases in crops have made agriculture very risky venture and due to high seed cost and cost of cultivation farmers are very apprehensive in adopting new technologies. About one fourth of total crops yield is lost each year due to pest attack. To mitigate these problems, reliable and timely forecast provides an important and extremely useful input in formulation of policies. Cotton is a major commercial crop for sustainable economy of India and livelihood of the Indian farming community. It is cultivated in 11 million hectares in the country. India accounts for about 32 per cent of the global cotton area and contributes to 21 per cent of the global cotton produce, currently ranked second after China, but the productivity is found to be very low because it is prone to pest attacks and damage largely by many pests. The main pests of cotton crops are American bollworm (*Helicoverpa armigera*), whitefly (*Bemisia tabaci*), jassids (*Amrasca bigutella bigutella*), and pink bollworm (*Pectinophora gossypiella*) *etc.* The cotton crop requires an intensive use of pesticides to overcome the incidence of damage from these pests. The major cotton producing states include Maharashtra, Gujarat, Andhra Pradesh, Punjab, Karnataka, and Madhya Pradesh. But to enhance the productivity and income to the farmers, forewarning of pest and disease is crucial. In agriculture, disease and pest management is very much important. Because these management practices potentially increase the yield of different crops (Report, 2020).

Table 1.1: Summary of descriptive statistics of whitefly counts for four cotton varieties for the year 2021-2022 and pooled data

	Ganganagar Ageti		HS 6		RS 2013		RCH 650	
	2021-2022	**Pooled data**	2021-2022	**Pooled data**	2021-2022	**Pooled data**	2021-2022	**Pooled data**
Min.	0.20	**1.83**	0.27	**1.33**	0.47	**0.93**	0.27	**0.80**
Mean	**8.42**	**129.98**	**9.63**	**107.18**	**9.02**	**98.91**	**8.127**	**116.33**
S.D.	7.90	**47.75**	8.56	**48.18**	7.56	**116.33**	6.75	**46.05**
Max.	29.93	**37.94**	29.4	**35.65**	25.53	**41.73**	24.13	**35.83**
Skewness	1.27	**0.545**	1.03	**0.218**	0.853	**0.308**	0.939	**0.379**
Kurtosis	1.37	**-0.832**	0.383	**-1.399**	0.939	**-1.334**	0.424	**-1.154**

Table 1.2: Correlation coefficient between yield and weather parameter of whitefly counts on four cotton varieties for the year 2021-2022 and pooled data

	Ganganagar Ageti		HS06		RS2013		RCH 650	
	2021-2022	**Pooled data**	2021-2022	**Pooled data**	2021-2022	**Pooled data**	2021-2022	**Pooled data**
Tmax	-0.17	**-0.35**	-0.06	**-0.36**	-0.05	**-0.36**	-0.09	**-0.37**
Tmin	0.37	**0.34**	0.38	**0.37**	0.42	**0.39**	0.32	**0.33**
Rhm	0.21	**0.72****	0.27	**0.75****	0.30	**0.74****	0.326	**0.74****
Rhe	0.42**	**0.76****	0.49**	**0.79****	0.52**	**0.80****	0.53**	**0.74****
Aws	0.20	**-0.09**	0.20	**-0.04**	0.20	**-0.03**	0.15	**-0.12**
SS	-0.13	**-0.47***	-0.18	**-0.52***	-0.25	**-0.56****	0.20	**-0.47***
RF	-0.06	**0.38**	0.04	**0.43**	0.06	**0.44**	0.12	**0.37**

*'Indicates significance at 5% level of significance, '**'indicates significance at 1% level of significance

The predictors and response variables which follow non normal distributions are linearly modeled, it suffers from methodological limitations and statistical properties. Hence, generalized linear model is used. Generalized linear model is the extension of general linear model. The advanced models like poisson regressin moel and negative binomial models long with weather parameters may address appropriate solutions for early warning of pest/disease infestation for investigating and predicting pest/disease status. The objective of this paper is to determine appropriate generalized linear models (GLM) that are suitable for count data and investigate the presence of over dispersion in the model parameter.

MATERIALS AND METHODS

Weekly data of whitefly count (Adults/3 leaves) on three cotton varieties (Ganganagar Ageti, HS 06 and RS 2013) and one hybrid (RCH 650 BGII) were obtained from the Cotton Section, CCS HAU, Hisar during 23 to 43 standard meteorological weeks of *kharif* season for the year 2021-2022 and pooled data of three years from 2019 to 2022. Three models namely Ordinary Least Square, Poisson and Negative Binomial Regression were applied on three varieties and one hybrid of cotton for modeling and prediction of whitefly population count. The weekly meteorological data on rainfall, mean temperature, mean relative humidity, sunshine hours, wind speeds etc. for the same period were considered in the present study. R-code were compiled from different sources for the analysis purpose at Department of Mathematics and Statistics, CCS HAU, Hisar. Data analysis and programming codes for proposed methodologies were developed using different R packages like tscount, forecast, lmtest, tseries etc. Generalized Linear Models (GLMs) represent a class of regression models that allow us to generalize the linear regression approach to accommodate many types of response variables including count, binary, proportions and positive valued continuous distributions. Because of its flexibility in addressing a variety of statistical

Table 1.3: Parameter coefficients and model selection of fitted model for Gangnagar Ageti for the year 2021-2022 and pooled data

	Normal Model		Poisson Model		Negative Binomial Model	
	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**
Intercept	0.01(6.42)	**1.02**(98.20)**	1.46(0.92)	**0.85**(-9.74)**	1.01(0.80)	**0.01**(1.74)**
Rhm	1.25*(0.11)	**9.26* (0.89)**	1.03**(0.01)	**0.01(10.64)**	1.04**(0.01)	**1.09**(0.01)**
Rhe	-	**7.62**(0.62)**		**0.01**(8.73)**		**1.04**(0.01)**
SS	-	**-104401.6(6.47)**		**0.04(5.61)**		**1.23(0.10)**
AIC	154.00	**204.75**			140.74	**184.63**
RMSE	6.99	**20.22**	7.16	**17.77**	7.28	**20.12**
MASE	0.90	**0.48**	0.92	**0.33**	0.95	**0.40**
R2	0.81	**0.70**	0.82	**0.83**	0.89	**0.92**

Table 1.4: Parameter coefficients and model selection of fitted model for HS-06 for the year 2021-2022 and pooled data

	Normal Model		Poisson Model		Negative Binomial Model	
	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**
Intercept	0.01(6.7)	**0.01**(92.71)**	1.35(0.35)	**0.01**(0.81)**	0.94(0.74)	**1.86(0.00))**
Rhm	1.33*(0.1)	**8.78*(0.83)**	1.03**(0.00)	**1.08**(0.00)**	1.04**(0.01)	**1.08(0.02)**
Rhe		**5.73**(0.59)**		**1.03**(0.00)**		**1.05(0.01)**
SS		**-44077.29(6.22)**		**-1.23**(0.04)**		**1.28(0.11)**
Rainfall		**1.27(0.28)**		**1.00**(0.00)**		**1.00(0.00)**
AIC	155.71	**200.89**			144.01	**188.92**
RMSE	7.27	**17.70**	7.47	**17.44**	7.60	**19.19**
MASE	0.79	**0.44**	0.80	**0.37**	0.83	**0.40**
R2	0.42	**0.74**	0.62	**0.80**	0.82	**0.81**

problems and the availability of software to fit the models, it is considered a valuable statistical tool and is widely used. In fact, the generalized linear model has been referred to as the most significant advance in regression analysis in the past twenty years. A generalized linear model (GLM) consists of three components *viz.* random Component, Linear Predictor, Link Function. Muliple linear regression modle, Poisson regression model and negative binomial regress in model were applied to predtc the whitefly diseas count. The best fit model was found out by the Akaike Information Criterions (AIC) and Coefficient of determination (R2). The smaller the AIC the better fitted models of the parameter estimate. The model which has high R2 considered best for forecasting purpose.

RESULTS AND DISCUSSION

It is clear from the Table 1.1 that variance is greater than mean for all four varieties for the year 2021-2022 and pooled data. So, assumptions of ordinarily normal regression are not satisfied. Hence, this regression cannot be applied for prediction purpose. It is also observed that whitefly counts for all the varieties were positively skewed. The value of Skewness is 1.27, 1.03, 0.853, 0.939 for the year 2021-2022 whereas these values were found to be 0.545, 0.218, 0.308, and 0.379 for the pooled data. Histograms of whitefly counts over frequencies are also obtained. Histograms are presented in Figures 1.1(a) to 1.1(d) for year 2021-2022 and Fig. from 3.2 (a) to 3.2 (d) for the pooled data. These graphs also show no systematic pattern in occurrence of whitefly count. In most of the cases, occurrence of whitefly scatters around its mean occurrence. From these graphs, it is evident that occurrence of whitefly is very erratic. There is no linear relationship with time.

Normal Q-Q plot for the year 2021-2022 and Pooled Data:

Firstly, the data was tested for normality

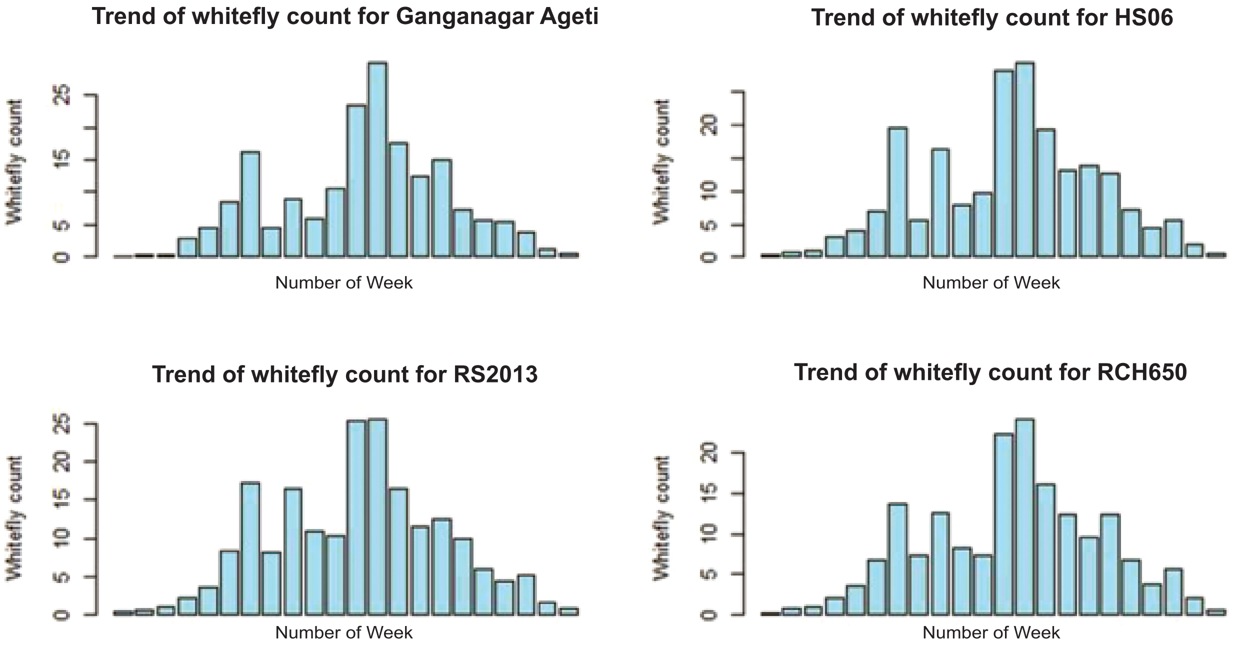
Table 1.5: Parameter coefficients and model selection of fitted model for RCH650 for the year 2021-2022 and pooled data

	Normal Model		Poisson Model		Negative Binomial Model	
	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**
Intercept	0.00 (5.78)	**0.00* (86.15)**	1.27 (0.36)	**0.01** (0.85)**	0.95 (0.71)	**0.00** (1.75)**
Rhm	1.31* (0.10)	**5.82* (0.77)**	1.03** (0.01)	**1.08** (0.00)**	1.04** (0.01)	**1.07** (0.01)**
Rhe		**5.12** (0.54)**		**1.03** (0.00)**		**1.05** (0.01)**
SS		**-1793.81 (5.78)**		**-1.17** (0.05)**		**1.25* (0.11)**
Rainfall		**1.19 (0.2632)**		**1.00* (0.00)**		**1.00(0.00)**
AIC	149.41	**197.66**	142.32	**195.3**	140.73	**182.94**
RMSE	6.30	**16.45**	6.46	**15.31**	6.56	**16.97**
MASE	0.82	**0.47**	0.83	**0.34**	0.85	**0.36**
R2	0.27	**0.74**	0.28	**0.82**	0.26	**0.81**

Table 1.6: Parameter coefficients and model selection of fitted model for RS 2013 for the year 2021-2022 and pooled data

	Normal Model		Poisson Model		Negative Binomial Model	
	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**	2021-2022	**Pooled Data**
Intercept	-5.96 (0.00)	**95.53** (-3.20)**	0.09 (1.10)	**0.00** (0.84)**	-0.15 (0.86)	**0.00** (1.97)**
Rhm	0.24* (1.28)	**0.87* (2.67)**	0.03** (1.03)	**1.10** (0.00)**	0.04** (1.04)	**1.09** (0.01)**
Rhe		**0.60* (2.81)**		**1.03** (0.00)**		**1.05** (0.01)**
SS		**6.29 (1.65)**		**1.21** (0.04)**		**1.32* (0.12)**
AIC	143.95	**203.54**			135.37	**188.6**
RMSE	5.56	**19.67**	5.71	**18.78**	5.77	**21.06**
MASE	0.79	**0.50**	0.79	**0.39**	0.82	**0.40**
R2	0.69	**0.68**	0.60	**0.78**	0.68	**0.86**

Histogram for the year 2021-2022:



Trend of whitefly count for Ganganagar Ageti

Trend of whitefly count for HS06

Trend of whitefly count for RS2013

Trend of whitefly count for RCH650

Fig. 1.1 (a) to 1.1(d): Trend of whitefly count for Gangnanagar Ageti, HS06, RS2013 and RCH 650 for the year 2021-2022

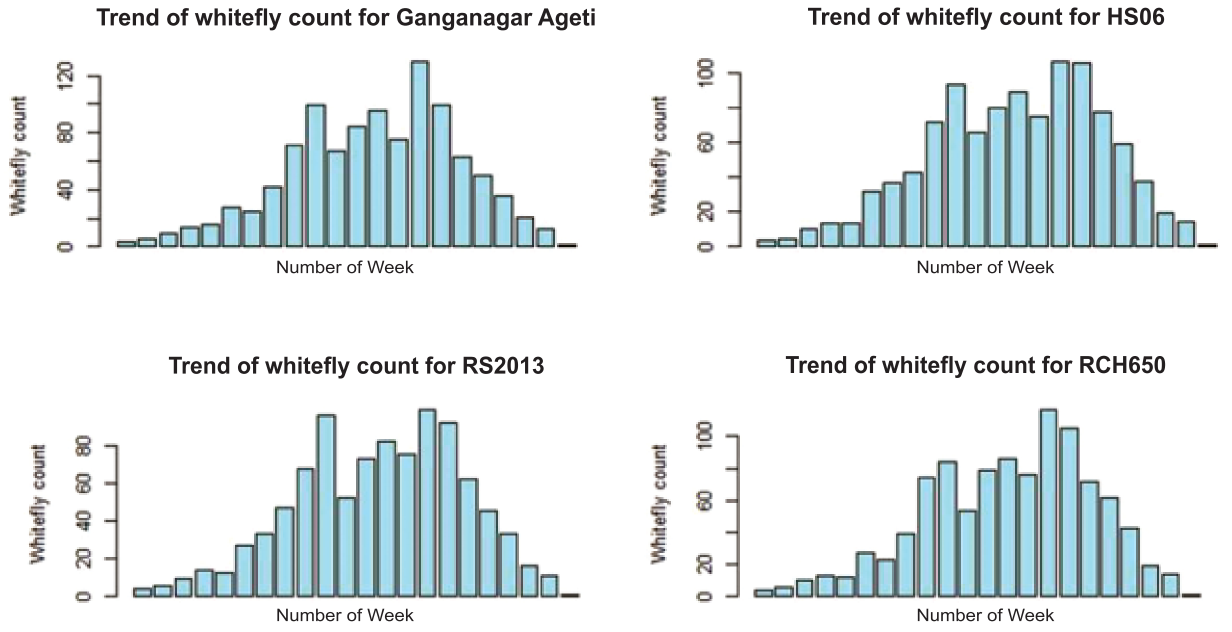


Fig. 1.2 (a) to 1.2 (d): Trend of whitefly count for Gangnanagar Ageti, HS 6,RS2013 and RCH 650 for the pooled data

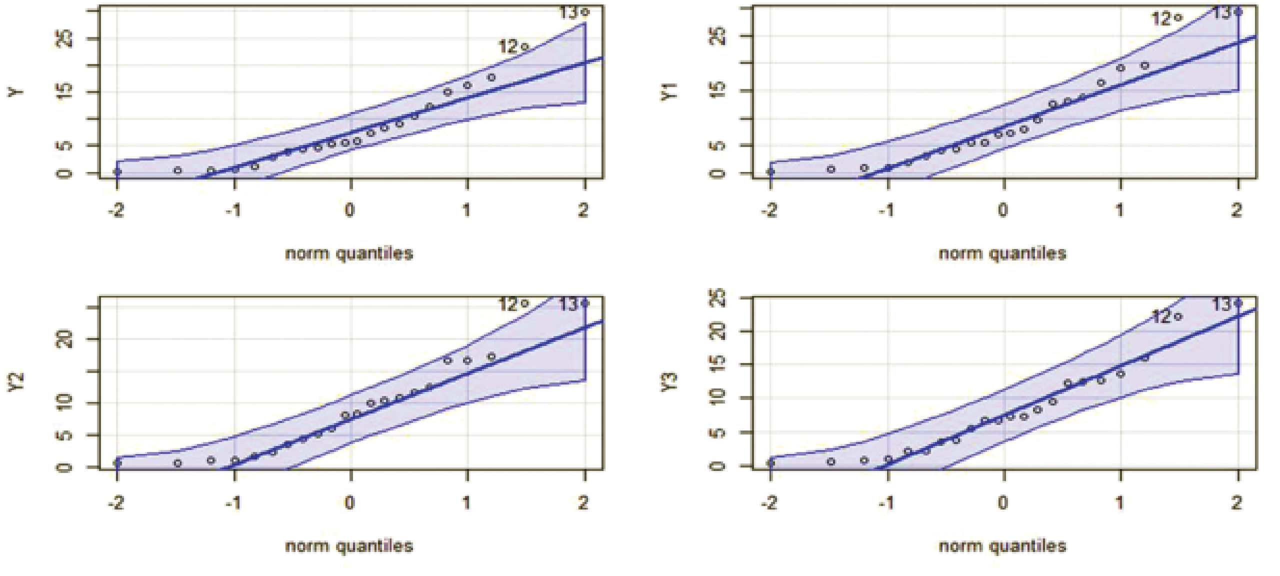


Fig. 1.3 (a) to 1.3 (d): Normal Q-Q plot for Ganganagar Ageti, HS 06, RS2013 and RCH650 for the year 20121-2022

using Q-Q plot which showed that there was no normality in the data. The data was found over dispersed and same were shown in the Fig. 1.3(a) to 1.3(d) for the year 2021-2022 and Fig. from 1.4(a) to 1.4(d) for the pooled data.

The correlation coefficients between weather variables and whitefly counts of cotton varieties are given in Table. It is observed from the Table that the relative humidity morning (Rhm) has positive and highly significant with whitefly counts for all cotton varieties for the year 2021-2022. Relative humidity morning, Relative

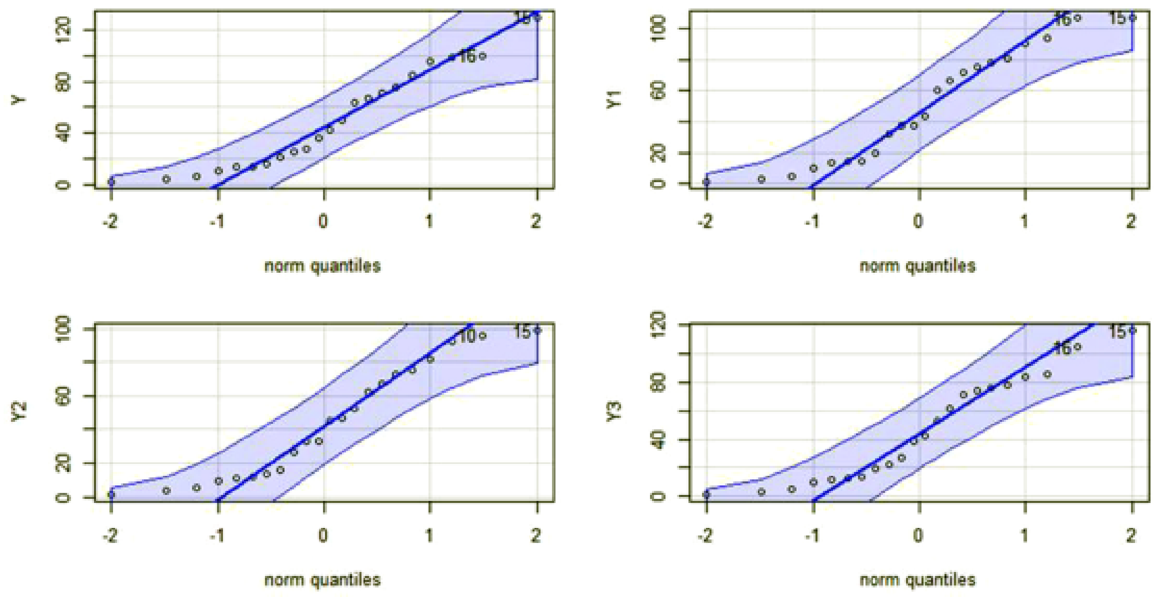


Fig. 1.4 (a) to 1.4 (d): Normal Q-Q plot for Ganganagar Ageti, HS 06, RS 2013 and RCH 650 for pooled data

humidity evening have very high positive and significant correlation for all varieties for the pooled data whereas sun shine has negative and significant correlation.

The parameter estimation for various model for four cotton varieties were shown in tables from 1.3 to 1.6 for the year 2021-2022 and pooled data. It is also observed that on the basis of various goodness of fit criteria negative binomial model was found to be best fit with low AIC value and BIC value and high R2 for all the studied cotton varieties.

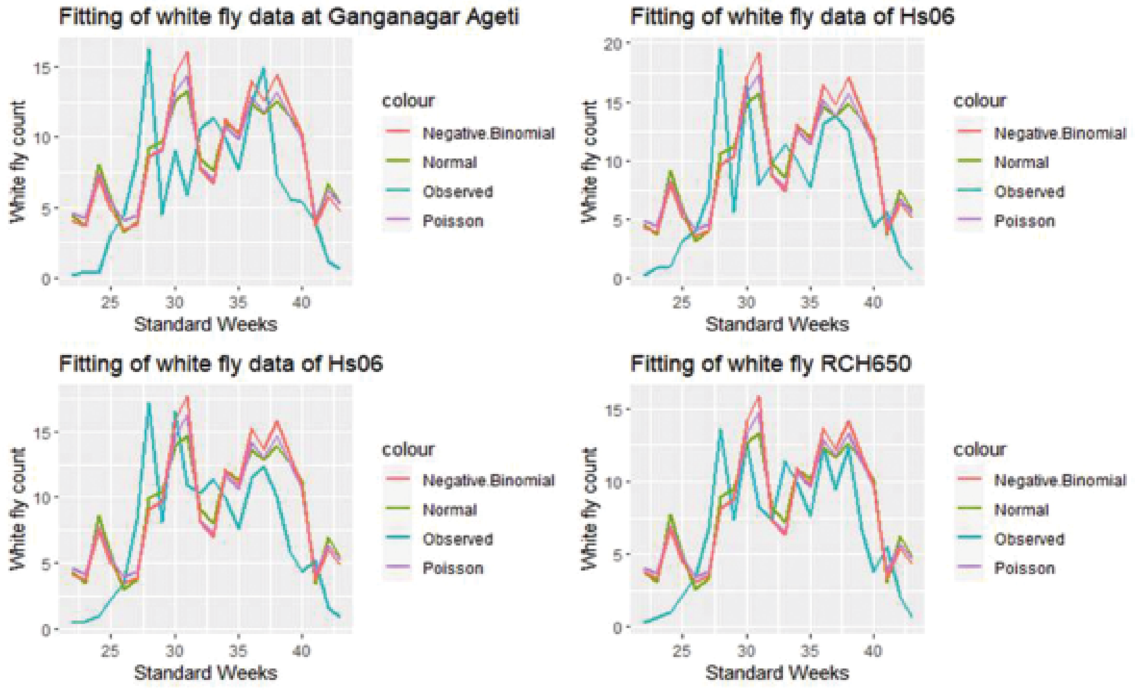


Fig. 1.5 (a) to 1.5 (d): Fitting of best fit model for Ganganagar Ageti, HS 6, RS2013 and RCH650 for the year 20121-2022

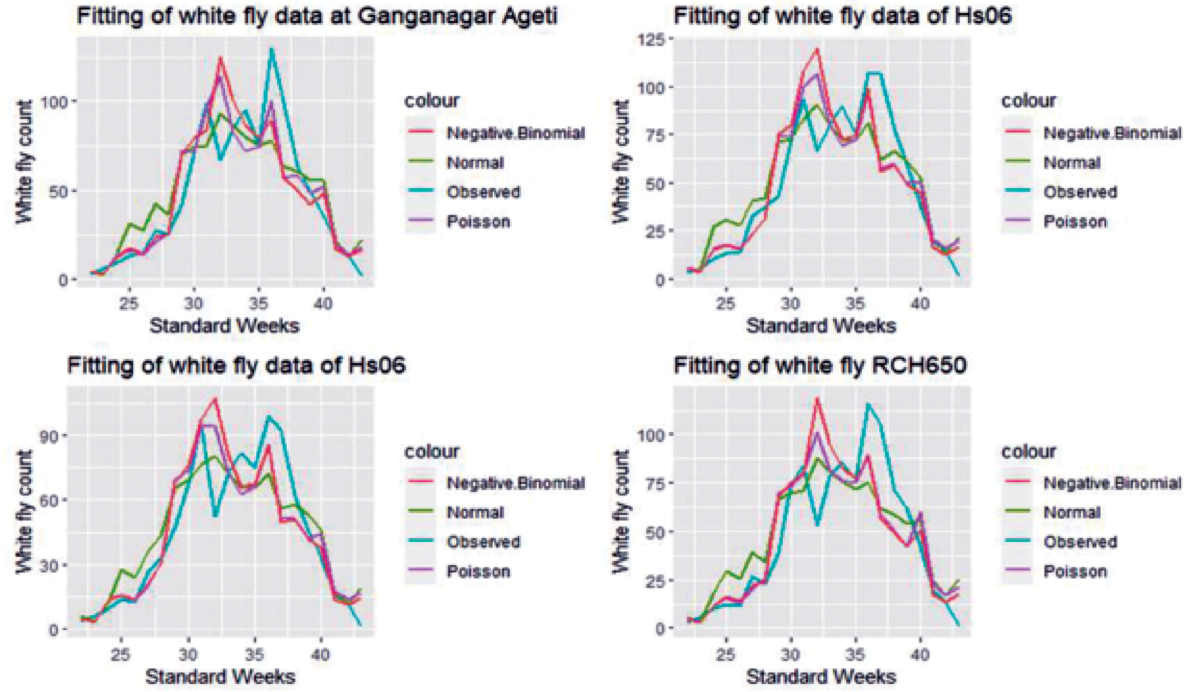


Fig. 1.6 (a) to 1.6 (d): Fitting of best fit model for Ganganagar Ageti, HS 6, RS2013 and RCH650 for the year 20121-2022

On the basis of fitted model, the predicted value was calculated and shown in Fig. from 1.5 (a) to 1.5 (d) for the year 2021-2022 and in Fig. 1.6 (a) to 1.6 (d) for the pooled data.

CONCLUSION

In this study the variable under study is whitefly disease count of cotton crop varieties were taken along with standard meteorological weekly (SMW) weather data were used. R code was developed compiled from various sources for fitting the count data models. Modeling of 2021-2022 data and pooled data concluded that Negative binomial regression was the best fit instead of Normal linear regression model. Therefore, it is recommended that Negative Binomial regression should be used for count data for early warning of disease infestation for investigating and predicting disease status.

REFERENCES

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. Journal of the Royal Statistical Society, Series A **135**: 370-84.

Hoffmann, J.P. 2004. Generalized linear models: An applied approach. Pearson: Boston.

McCullagh, P. and Nelder, J.A. 1989. Generalized Linear Models (2nd Edn). *London*: *Chapman* and *Hall.*

Araya, Prawin, Alam, MD. Wasi and Gurung, Bishal 2020. Development of count time series model s for predicting pest dynamics using weather variable. *Project Report IASRI*, 1-88

Agrawal, R. and Mehta, S.C. 2007. Weather based forecasting of crop yields, pest and diseases – IASRI Models. *J. Ind. Soc. Agril. Stat.* **61** : 255-63.

Bhardwaj, T. and Sharma, J.P. 2013. Impact of Pesticides Application in Agricultural Industry: An Indian Scenario. *Internat. Jour. Agric. Food Sci. Tech.* **4**: 817-22.

Kumar, A., Ranjana Agrawal, R. and Chattopadhyay, C. 2013. Weather based forecast models for diseases in mustard crop, *Mausam.* **64**: 663-70.

Kumari, Prity, Mishra, G.C. and Srivastava, C.P. 2014. Time series forecasting of losses due to pod borer, pod fly and productivity of pigeonpea (Cajanus cajan) for North West Plain Zone (NWPZ) by using artificial neural network (ANN). *Internat. Jour. Agric. Stati. Sci.* **10**:15-21.

Arya Prawin, Ranjit Kumar Paul, Anil Kumar, K. N. Singh, N. Sivaramne and Pradeep Chaudhary 2015. Predicting pest population using weather variables: An ARIMAX time series framework. *Int. J. Agricult. Stat. Sci.* **11** : 381-86.

Roy, H.S., Paul, R.K., Bhar, L.M. and Arya, P. 2016. Application of INAR model on the pest population dynamics in Agriculture. *Journal of Crop and Weed.* **12**:96-101.

Dobson, A.J. 2002. An Introduction to Generalized Linear Models. *Chapman and Hall, London*, UK.

Received for publication : March 2, 2023
Accepted for publication : May 5, 2023