



# Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)  
Vol. No. 24 (4) : 409-416 (December- 2022)

<https://journal.agrimetassociation.org/index.php/jam>



## Research Paper

### Modelling of tea mosquito bug (*Helopeltis theivora*) incidence on neem tree: A zero inflated count data analysis

S. VISHNU SHANKAR<sup>1\*</sup>, R. AJAYKUMAR<sup>2</sup>, P. PRABHAKARAN<sup>3</sup>, R. KUMARAPERUMAL<sup>4</sup> and M. GUNA<sup>5</sup>

<sup>1</sup>Department of Basic Sciences, Dr. Y. S. Parmar University of Horticulture & Forestry, Solan-173230, Himachal Pradesh, India

<sup>2</sup>Department of Agronomy, Vanavarayar Institute of Agriculture, Pollachi-642103, Tamil Nadu, India

<sup>3</sup>Department of Forestry, Vanavarayar Institute of Agriculture, Pollachi-642103, Tamil Nadu, India

<sup>4</sup>Department of Remote Sensing and Geographic Information System, Tamil Nadu Agricultural University, Coimbatore- 641003, Tamil Nadu, India

<sup>5</sup>Agro Climate Research Centre, Tamil Nadu Agricultural University, Coimbatore- 641003, Tamil Nadu, India

\*Corresponding author email: [s.vishnushankar55@gmail.com](mailto:s.vishnushankar55@gmail.com)

#### ABSTRACT

Neem (*Azadirachta indica*) is an evergreen tree belonging to the Meliaceae family and is highly infected by the seasonal pest called *Helopeltis theivora*, the tea mosquito bug. The study monitors the pest infection between May 2019 and April 2021 by the direct counting method. Weekly counts of insect pest population were found to be correlated with weather parameters. Zero inflated count data techniques were opted for modelling the pest dynamics of tea mosquito bug as the data was featured by excess zeroes and heteroscedasticity nature. Poisson, Negative Binomial (NB), zero-inflated Negative Binomial (ZINB) and Negative Binomial Hurdle (NBH) models were fitted for the collected data and compared. The results of different count data models show that the negative binomial hurdle model is a good fit for given data, followed by the zero-inflated negative model. The fitted models show the weather covariates, which highly influencing the pest infestation on neem tree.

**Keywords:** Count data models, *Helopeltis theivora*, statistical models, weather parameters

Neem (*Azadirachta indica*) is one of the fast-growing tree species of the Indian subcontinent. It belongs to the family of Meliaceae and is extensively adapted to different climate and soil types. Different species of neems are found worldwide, covering 80 countries with a global population of 120 million trees (Dhillon *et al.*, 2008). This species is found in different parts of India, where Uttar Pradesh (55.7%), Tamil Nadu (17.8%) and Karnataka (5.5%) occupy the first three places. Over 53,800 tonnes of neem cake and 88,400 tonnes of neem oil were produced using the 4,42,300 tonnes of seeds harvested annually from the tree. They eventually generate million-dollar business for the Indian economy (Girish and Shankara, 2008). Neem is regarded as a sacred tree in India due to its multifaceted qualities, and the Vedic period is when it was first made aware of its commercial significance. Neems are highly known for their insecticidal property because of their non-toxicity, biodegradables, and low cost. Moreover, they believed in resisting 250 species of insects with eco-friendly nature. Its medicinal properties are highly notable for their quality standards. Nearly 150 nature compounds are produced from various tree parts (Nagendra

and Gopal, 2010). The trend towards organic farming expands the market for products made from neem. Neem is recognized as a “golden tree” that has gained prominence due to its numerous benefits. In addition to agroforestry, it is employed in energy generation, plant and animal nutrition, cosmetics, pharmaceuticals, and pest control.

According to reports, *Helopeltis theivora* is a major sucking pest, affecting 80% of tea plantations (*Camellia sinensis*). It results in a sizable (10–50%) production loss in Northeast India. They are not only the major pests of neem but also the serious pest of cashew (*Anacardium occidentale* L.), cocoa (*Theobroma cocoa* L.), cinchona (*Cinchona spp.*), black pepper (*Piper nigrum* L.), guava (*Psidium guajava* L.) and tea (*Camellia sinensis* L.). The pest was first reported in Coimbatore district (Rao, 1915). However, due to environmental changes, its destruction has recently been more noticeable in North regions of India (Handa *et al.*, 2022). Changes in the climate cause unexpected meteorological phenomena, such as unusually heavy rain and extremely high or low temperatures,

**Article info - DOI:** <https://doi.org/10.54386/jam.v24i4.1891>

Received: 14 October 2022; Accepted: 19 November 2022; Published online : 1 December 2022

This work is licenced under a Creative Common Attribution 4.0 International licence @ Author(s), Publishing right @ Association of Agrometeorologists

which increase pest and disease outbreaks. Only the young shoots that produce the actual harvest of tea are attacked (Ahmed *et al.*, 2013). Both nymphs and adults consume plant components by swallowing them down by secreting polyphenoloxidase from their salivary glands (Saroj *et al.*, 2016). Sap-feeding ultimately affects the yield, reduces the plant's growth, and causes the infested buds or shoots to become curled, dried and black (Prasada Rao *et al.*, 2002). In addition, insects consume food, grow, breed, and spread when the warm climate, but they also live less time (Reynolds *et al.*, 2017). Because of the accessibility of food sources and the macro- and micro-climatic conditions, the number of insects in a specific crop change throughout time. Therefore, understanding the weather factors as a prerequisite for insect pest population dynamics and developing effective control methods require understanding the weather-based forecasting system (Patel *et al.*, 2021). Plant protectionists were driven to hunt for safer substitutes and use every available tool to lower the load of pesticides due to increased consumer and planter awareness of synthetic insecticide residue in tea. The main obstacle for nations that export tea, like India, is the Maximum Residue Limit (MRL), which has been set at 0.1 mg kg<sup>-1</sup> and below for most contaminants in the European Union. The information on pest dynamics helps the farmers manage the pest controlling practices (Elango *et al.*, 2021). Considering the above facts, the present study tries to understand the pest dynamics with the help of statistical analysis.

## MATERIALS AND METHODS

### Data

Seasonal abundance of major insect pest, tea mosquito bugs, was studied in 5-year-old neem (*A. indica*) trees at Vanavarayar Institute of Agriculture, Pollachi, Western zone of Tamil Nadu from May 2019 and April 2021. The pest incidence was noted from September to April months and absent during rest of year. Observations were collected every week regarding the prevalence of insect pests on the ten randomly chosen neem trees. The three randomly chosen branches in each tree (15 cm terminal shoots each), the number of nymphs and adults of tea mosquito bugs was counted and expressed as a number per tree. The collection of count data is characterized by numerous zero, as the study is on seasonal pest. Daily weather data such as maximum temperature (T<sub>max</sub>), minimum temperature (T<sub>min</sub>), relative humidity [morning (07.22hrs) and afternoon (14.22hrs) (RH)], rainfall (mm/day) and wind speed (km/h) was obtained from Vanavarayar Institute of Agriculture, Pollachi. A total of 100 observations collected from the study in which pest count was considered the dependent variable and weather parameters as the independent variable. No transformation was done before modelling the data, as the process of transformation would affect the nature of data i.e., zero count data.

### Count data models

Count data are the type of statistical data where the dependent/response variables are counts, discrete, non-negative, analyzed with a set of independent/covariate variables. Count data are collected by counting the occurrence of an event during a fixed period interval. Count data are related to the properties of non-linearity and discreteness. These data types usually occur more in

entomology, pathology, epidemiology, sociology, demography, and insurance studies. Modelling such count data is quite different from modelling the usual data.

The selection of suitable count data models depends on the nature of dispersion present in the data (Cheung, 2002). When the mean is greater than the variance, it is called under dispersion and if it is less, the condition is called overdispersion. Equal dispersion is the condition where the mean and variance are equal. Generalized Poisson and Conway-Maxwell-Poisson models were the commonly used count data models when the data is under dispersion. Traditional Poisson regression can be used to fit the equal dispersion data. Data containing overdispersion can be modelled by negative binomial regression. But the most common problem in count data was excess zero. Some studies contain data where the observation has more zeros than values. These data types were referred to as zero-inflation (Cui and Yang, 2009). Ordinary least square methods fail to produce a better estimate for zero inflation data. This indicates the absence of normality in the data. Poisson and negative binomial distributions are not appropriate methods for zero inflation data. Modelling the data by removing the excess zeros also makes no sense. Such data types can be handled only with zero inflated or hurdle models.

The zero-inflated models can be extended to Poisson and negative binomial models. When equal dispersion data contains excess zero, the zero-inflated – Poisson model can be used, and zero-inflated negative binomial distribution is for overdispersion data with excess zero. Another popular truncated model for handling excess zeros was the hurdle model. The hurdle method can also be used for equal and overdispersion data, i.e., the Hurdle – Poisson model and Hurdle – Negative binomial model. Hurdle and zero-inflated models differ in the way of assuming the zeros in the data. The zero-inflation model assumes zero in two ways, whereas the hurdle model assumes only one way. First, both models assume the method of the binary process (On-Off part). The counting method is assumed only by the zero-inflation model. The natural log link function is a highly used link function in the count data model, resulting in positive predictive values ( $m > 0$ ).

### Poisson regression

Poisson regression has its application on social data to model the number of events ( $y$ ) or rate ( $r$ ). It is highly used for rare events data. The probability of an occurrence is frequently assumed to follow a Poisson distribution with a conditional mean ( $\mu$ ) that depends on a set of regressors ( $x$ ) and corresponding parameters ( $\beta$ ) for a participant's linear predictor. Log link function can be used to express the expected number of events for participant  $i$  at variable  $j$   $\mu_{ij} = E(y_{ij} | x_{ij}) = e^{\beta'x_{ij}}$ . The Poisson probability distribution of  $y_{ij}$  given  $x_{ij}$  can be expressed as

$$P(Y_{ij} = y_{ij}) = \frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{y_{ij}!} \quad (1a)$$

Where  $y$  is a non-negative integer. The contribution of the  $i^{\text{th}}$  variable at  $j^{\text{th}}$  variable to the log-likelihood for the Poisson model can be expressed as

$$LL(\beta) = y_{ij}(\beta x_{ij}) - e^{\beta x_{ij}} - \ln(y_{ij}!) \quad (1b)$$

**Table 1:** Summary of the *Helopeltis theivora* incidence and weather parameters (100 observations) and their correlation with insect counts.

Particulars	Maximum	Minimum	Mean	Variance	Correlation Coeff.
Insects	17.0	0.0	3.79	20.13	
T <sub>max</sub>	37.8	24.8	32.17	7.41	-0.63
T <sub>min</sub>	26.5	10.3	22.64	3.87	-0.47
Rainfall	50.4	0.0	4.98	51.32	0.27
Wind speed	15.2	3.2	7.15	8.21	-0.38
Relative humidity	90.0	75.0	84.93	9.44	0.27

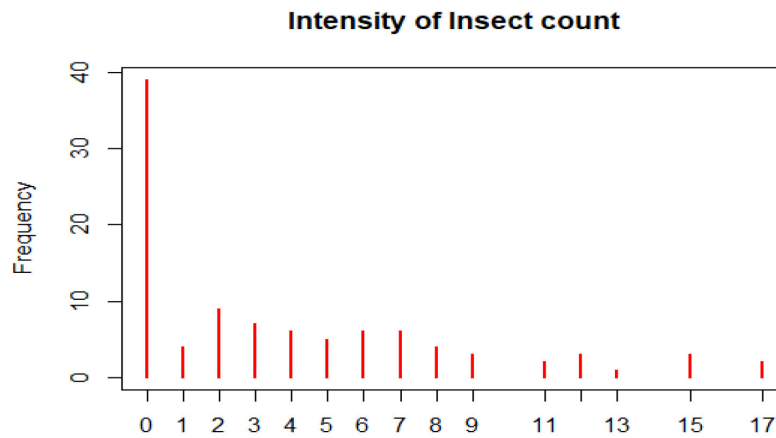
**Table 2:** Poisson model results

	Estimate	Std. Error	z value	Pr (> z )
Intercept	12.843	2.079	6.18	6.47 e-10***
T <sub>max</sub>	-0.211	0.021	-10.23	<2e-16***
T <sub>min</sub>	-0.122	0.019	-6.42	1.40e-10***
Rainfall	0.008	0.007	1.29	0.1970
Wind speed	-0.150	0.027	-5.45	4.96e-08***
Relative humidity	-0.015	0.022	-0.714	0.4750

*Number of Fisher Scoring iterations: 6*

**Over dispersion test:** z = 2.98, p-value = 0.002, Dispersion value = 2.750

Significant ratios: 0 (\*\*\*), 0.001 (\*\*), 0.01 (\*), 0.05 (.)



**Fig.1:** Distribution of intensity of *Helopeltis theivora* incidence in the field

The log-likelihood of all other considered models can be expressed using their respective distribution. The main assumption of Poisson regression, which is often violated, is the variance that equals the mean, i.e.,  $Var(y) = \sigma^2 = \mu$ . If overdispersion is a problem (i.e., variance exceeds the mean) the derived parameters based on Poisson regression will be inefficient (Cameron and Trivedi, 1998).

**Negative binomial regression**

The Negative binomial is an extension of the Poisson model, which relaxes variance assumptions. They can handle the data under overdispersion conditions. The overdispersion is

accounted for NB model by the addition of an error term,  $e$  to the conditional mean of the passion model (Sheu *et al.*, 2004) i.e.,  $\mu_{ij} = \dots$ . The  $\exp(c_{ij})$  from the NB model has a gamma distribution with mean 1 and variance  $a$ . so the conditional mean of  $y_{ij}$  is still  $\mu_{ij}$  where the conditional variance of  $y_{ij}$  becomes  $\mu_{ij}(1+a\mu_{ij})$ . The model becomes more dispersed as  $a$  increases and becomes Poisson distribution as it approaches zero. The NB probability distribution for  $i$  at variable  $j$  is given by:

$$P(Y_{ij} = y_{ij}) = \frac{\Gamma(y_{ij}+1/a)}{\Gamma(y_{ij}+1)\Gamma(1/a)} \frac{(a\mu_{ij})^{y_{ij}}}{(1+a\mu_{ij})^{y_{ij}+1/a}} \tag{2}$$

**Table 3:** Negative binomial model results

	Estimate	Std. Error	z value	Pr (> z )
Intercept	17.717	4.092	4.329	1.50e-5***
T <sub>max</sub>	-0.244	0.044	-5.571	2.53e-08***
T <sub>min</sub>	-0.243	0.049	-5.016	5.27e-07***
Rainfall	0.000	0.014	-0.015	0.9880
Wind speed	-0.213	0.049	-4.376	1.21E-05***
Relative humidity	-0.024	0.040	-0.599	0.5500

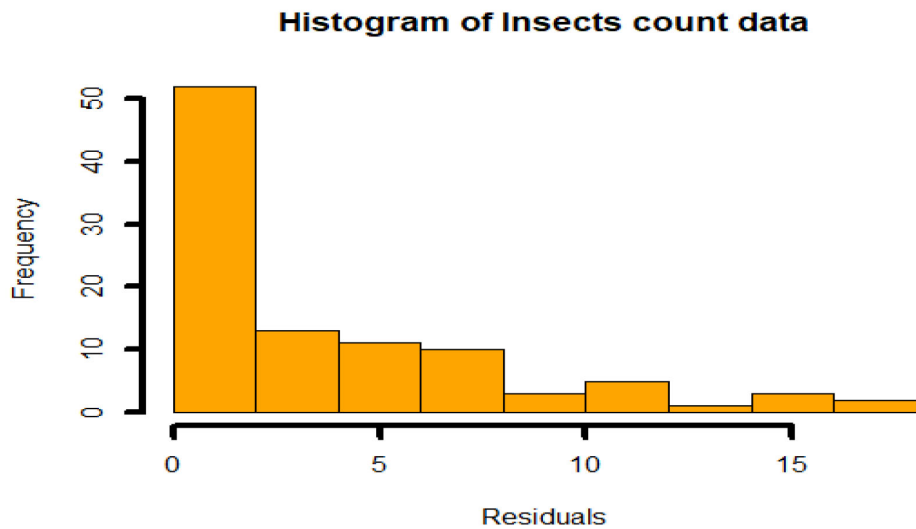
Number of Fisher Scoring iterations: 1  
 Theta: 1.900

Significant ratios: 0 (\*\*\*), 0.001 (\*\*), 0.01 (\*), 0.05 (.)

**Table 4:** Negative binomial hurdle model results

	Estimate	Std. Error	z value	Pr (> z )
Intercept	10.567	3.086917	3.42	0.001***
T <sub>max</sub>	-0.159	0.032	-4.960	7.0e-07***
T <sub>min</sub>	-0.091	0.037	-2.471	0.013*
Rainfall	0.011	0.009	1.191	0.234
Wind speed	-0.021	0.044	-0.485	0.627
Relative humidity	-0.022	0.030	-0.712	0.476
Log(theta)	2.174	0.494	4.402	1.07e-05***

Number of iterations in BFGS optimization: 30 Theta = 8.7942 Link function = logit



**Fig.2:** Histogram of intensity of *Helopeltis theivora* incidence in the field

Where  $\mu_{ij}$ ,  $a$ , and  $\Gamma(\bullet)$  refer to the mean of the count distribution, the NB dispersion parameter, and the gamma function. The NB model can handle the overdispersion data resulting from unobserved heterogeneity or temporal dependency. But the model may fail to take the overdispersion, which results from excess

zeroes.

**Zero-inflated negative binomial regression**

The zero-inflated negative binomial (ZINB) model has

**Table 5:** Zero inflated negative binomial models

	Estimate	Std. Error	z value	Pr (> z )
Intercept	10.228	3.060	3.342	0.000***
T <sub>max</sub>	-0.155	0.031	-5.015	5.32e-07***
T <sub>min</sub>	-0.090	0.036	-2.491	0.012*
Rainfall	0.010	0.009	1.139	0.255
Wind speed	-0.024	0.043	-0.546	0.585
Relative humidity	-0.019	0.031	-0.623	0.533
Log(theta)	2.239	0.493	4.544	5.52e-06***
<i>Number of iterations in BFGS optimization: 25</i>				
<i>Theta = 9.3803</i>				
<i>Link function = logit</i>				

**Table 6:** Model comparison

Models	AIC	BIC	Log likelihood	RMSE
Poisson	481.05	496.68	-234.52	3.26
Negative binomial	421.80	440.03	-407.795	8.05
Negative binomial hurdle	388.80	422.67	-180.46	2.73
Zero inflated negative binomial	389.82	423.68	-181.91	2.82

**Table 7:** Vuong’s test

<b>Vuong Non-Nested Hypothesis Test-Statistic</b> <i>(Test-statistic is asymptotically distributed N (0,1) under the null that the models are indistinguishable)</i>		
Model combination	Results	P - Value
Poisson vs Negative binomial	Model2 > Model1	0.000***
Negative binomial Vs Hurdle	Model2 > Model1	0.000***
Negative binomial Vs Zero inflated	Model2 > Model1	0.000***
Hurdle vs Zero inflated	Model1 > Model2	0.000***

advantages for excess zero condition because of its flexibility in the variance (Hall, 2000). Using Eq. (2), the ZINB model for i at variable j can be expressed as

$$P(Y_{ij} = y_{ij}) = \begin{cases} p_{ij} + (1 - p_{ij}) \frac{1}{(1 + a\mu_{ij})^{\frac{1}{\alpha}}} & y_{ij} = 0 \\ (1 - p_{ij}) \frac{r(y_{ij} + 1/a)}{r(y_{ij} + 1/a) \Gamma(1/a)} \frac{(a\mu_{ij})^{y_{ij}}}{(1 + a\mu_{ij})^{y_{ij} + 1/a}} & y_{ij} > 0 \end{cases} \quad (3)$$

Where all terms are same as previously defined. The mean is the same as for the ZIP model. The variance depends on the value of p and the dispersion parameter, and it is given by  $\sigma^2 = \mu(1-p)[1 + \mu(p + \alpha)]$ . Thus, the ZINB model adds more flexibility compared to the ZIP model. The ZINB model takes overdispersion from excess zeroes and heterogeneity, whereas the ZIP model only takes overdispersion from excess zeroes. Interpretation of the ZINB model is the same as for the ZIP model.

**Hurdle models: Poisson and negative binomial**

The hurdle models can be interpreted as two-part in contrast to zero-inflated models. Typically, the first component is a binary response model, and the second component is a truncated-at-zero count model. (Cameron and Trivedi, 1998). Hence, the hurdle model is a modified count model which produces positive and

zero counts that are not identical. This enables us to interpret the favourable results (>0) that follow overcoming the barrier of zero (threshold). The two-part model’s hurdle component calculates the likelihood that the threshold will be crossed. Although the threshold might theoretically have any value, it is typically set to zero because it is most frequently used to connect the study objectives. Mullahy (1986) laid out the basic form of hurdle count models. Assume that  $f_1$  and  $f_2$  are any probability density functions for nonnegative integers. A hurdle model can be expressed as

$$P[y=0] = f_1(0) = p \quad (4a)$$

$$P[y=i] = (1-p) \frac{f_2(i)}{1 - f_2(0)} = (1-p)f_2^*(i) \quad i > 0 \quad (4b)$$

Note that  $f_1(\bullet)$  governs the hurdle part and  $f_2(\bullet)$  oversees the counting process once the hurdle has been crossed. Furthermore,  $f_1(0)$  is the probability of crossing the hurdle and  $f_2(\bullet)$  is the truncated normalization of  $f_2(\bullet)$ . If  $f_1(\bullet) = f_2(\bullet)$ , the hurdle model breakdowns to the standard count model. Hurdle models can be given in various ways by choosing different distributions for  $f_2(\bullet)$  and  $f_2(\bullet)$ . As for the zero-inflated models, we use logistic regression to model p. Here we define two hurdle models by specifying  $f_2(\bullet)$  as the Poisson and NB distributions. For example, the substitution of Eq. (1) into Eq.

(4a) results in the Poisson hurdle model for participant  $i$  at variable  $j$  being defined as

$$P(Y_{ij} = y_{ij}) = \begin{cases} p_{ij} & y_{ij} = 0 \\ (1 - p_{ij}) \frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{(1 - e^{-\mu_{ij}})^{y_{ij}!}} & y_{ij} > 0 \end{cases} \quad (5)$$

All terms are defined as before, and Eq (5) can be used to get the log-likelihood definition. The expected value for the Poisson hurdle (PH) model is given by

$$E(y) = \frac{(1-p)\mu}{1-f_2(0)} \quad (6)$$

Substitution of Eq. (2) into Eq. (4b) for  $f_2(\bullet)$  results in the Negative Binomial hurdle (NBH) model. Computing the expected value for the NBH model is as for the PH model.

### Model validation

Akaike Information Criteria (AIC) and Bayes Information Criterion (BIC) were the selection criteria for picking the best-fitted models. BIC impose strict penalty compared to the AIC. Log-Likelihood and Root mean square error can also be used to select the best-fitted model. The best model is determined based on the least value of AIC and BIC. Selection of best-fitted models from  $n$  observation containing  $k$  model parameters to be estimated where  $L$  is log-likelihood using AIC and BIC is given by

$$AIC = -2 \log(L) + 2k \quad (7)$$

$$BIC = -2 \log(L) + k \log(n) \quad (8)$$

### Vuong test

Vuong test is a likelihood ratio test which compares the closeness of fitted models based on the kullback - Leibler information criteria. Probability statements on two fitted models are considered where the null hypothesis assumes no difference between two fitted models. Vuong test also indicates the best-fitted model. Software packages like R, SAS, and STATA can be used to fit the count data models. In this study, R software was used to fit the models.

## RESULTS AND DISCUSSION

The information statistics of the insect count data and weather parameters have been presented in Table. 1. It shows the mean, variance, maximum value, and minimum value of variables taken for the study. The mean of the dependent variable (tea mosquito bug) is found to be smaller than its variance, which indicates the overdispersion. Fig.1 shows the distribution of intensity of pest incidence in the field. The graph revealed that data has excess zero values indicating the incidence of pests only during a particular year's season.

### Correlation analysis

The correlation between the insect count and weather parameters is shown in Table 1. It is found that rainfall and relative humidity positively influence the incidence of pests (Singh 2021), whereas maximum temperature, minimum temperature, and wind speed negatively influence pest incidence (Roy *et al.*, 2015).

The maximum and minimum temperature is found have high significance on pest incidence compared to other parameters. This eventually shows that pest incidence high during rainy season and low during summer season (Pillai *et al.*, 1979). In south India, the pest incidence on Neem is noticed during September to march where pest is peak during December and January (Swaine *et al.*, 1959). The graph (Fig.2) indicates that *Helopeltis theivora* incidence is not normally distributed, which is a sign of zero count data (Kalloor *et al.*, 2020). Though the data is characterised by zero count nature, Poisson and Negative Binomial models were fitted in this study to show its discrepancy in handling such data. Usage of other statistical techniques like ARIMA, multiple linear regression etc., for modelling the data would be inappropriate in this case.

Poisson regression analysis (Table. 2) shows that variable maximum temperature, Minimum temperature, and wind speed are significant. Even though the variance is greater than the mean, the Poisson model is fitted to check the values. The results of overdispersion test (Table. 2) show that values are greater than zero, recommending the need of zero-inflated models to fit the data. Negative Binomial regression analysis (Table. 3) shows that variable maximum temperature, Minimum temperature, and wind speed are significant. It takes one fisher scoring iteration with the value 1.900. Both these models were fitted only for comparison purposes.

The hurdle model on negative binomial (Table. 4) shows that maximum temperature and Minimum temperature are significant (Sanjay *et al.*, 2022). Maximum temperature, minimum temperature, relative humidity and wind speed are found to have negative coefficient for pest incidence. Rainfall was found to have positive coefficient for pest incidence (Dutta *et al.*, 2013). The model used the logit link function and had theta value of 8.7942 with 30 iterations.

Zero-inflated negative binomial models (Table 5) also shows the same results as shown by negative binomial hurdle model. This too shows that pest incidence is influenced by temperature and rainfall (Kumar and Naik, 2002). This model also used the logit link function and had theta value of 9.3802 with 25 iterations.

### Model selection

Qualitative information and quantitative results are considered for selecting the best model. Model selection is carried out based on Table. 6's findings. Based on the model selection criteria like AIC and BIC, the negative binomial hurdle model is known to have low score values. It is followed the zero-inflated negative binomial model. Both these models have only minute differences in their scores (Naffees Gowsar *et al.*, 2019). Log-likelihood value is also high for the hurdle model, followed by zero-inflated models.

Error measures like Root mean score error is low for the hurdle model. All the results confirm that the negative binomial hurdle model is the best-fitted model for given data, followed by zero-inflated negative binomial models. From Table. 7, Vuong test shows that the hurdle model is the best fit for given data, followed by the zero-inflated model, the negative binomial model, and the Poisson model.

## CONCLUSION

Despite several statistical techniques available for modelling the insect incidence, using the appropriate is important which is always decided based on the nature of data. Since the study is on seasonal pest, the data contain numerous zeros on non-seasonal periods and are characterized by count rather than whole values. Based on the results of different count data models, it is found that the negative binomial hurdle model is a better fit for given data. It is followed by zero-inflated negative binomial model. The obtained results also support the zero counted model by listing hurdle and zero inflated models as best than other models. Moreover, these models give information regarding the weather covariates, which highly influence tea mosquito bug (*Helopeltis theivora*) incidence in neem tea. Therefore, selecting the best model for forecasting the incidence of pest dynamics using the above procedure will be appropriate.

**Conflict of Interest Statement:** The author(s) declare(s) that there is no conflict of interest.

**Disclaimer:** The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

**Publisher's Note:** The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

- Ahmed, P., Bhagawati, P., Das, S. K., Kalita, M. C., and Das, S. (2013). Hypersensitive reaction and anatomical changes of young tea leaf (*Camellia sinensis*, clone TV1) during feeding by tea mosquito bug (*Helopeltis theivora* Waterhouse: Hemiptera: Miridae). *Intern. J. Current Microbiol. Applied Sci.*, 2: 187-195.
- Cameron, A. C., and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- Cheung, Y. B. (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. *Stat. Med.*, 21(10): 1461-1469. DOI: [10.1002/sim.1088](https://doi.org/10.1002/sim.1088)
- Cui, Y. and Yang, W. (2009). Zero-inflated generalized Poisson regression mixture model for mapping quantitative trait loci underlying count trait with many zeros. *J. Theor. Biol.*, 256: 276-285.
- Dhillon, R. S., Ahlawat, K. S., and Pundeer, J. S. (2008). Botany of neem. *Neem-A Treatise*, 30-43.
- Dutta, P., Reddy, S. G. E., and Borthakur, B. K. (2013). Effect of neem kernal aqueous extract (NKAE) in tea mosquito bug, *Helopeltis theivora* (Waterhouse, 1886) (Heteroptera: Miridae). *Munis Entomol Zool.*, 8(1): 213-218.
- Elango, K., Nelson, S. J., and Dineshkumar, P. (2021). Incidence forecasting of new invasive pest of coconut rugose spiraling whitefly (*Aleurodicus rugioperculatus*) in India using ARIMAX analysis. *J. Agrometeorol.*, 23(2):194-199. <https://doi.org/10.54386/jam.v23i2.67>
- Girish, K., and Shankara, B. S. (2008). Neem—a green treasure. *Elect. J. Biol.*, 4(3): 102-111.
- Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56:1030–39.
- Handa, A. K., Sirohi, C., Arunachalam, A., Ramanan, S. S., Rajarajan, K., Krishna, A., and Kolse, R. H. (2022). Surge in neem tea mosquito bug incidence in India. *Current Sci.*, 122(6): 651-651.
- Kalloor, B. J., Suganthy, M., Balasubramanian, A., and Kumar, P. R. D. M. S. (2020). Seasonal abundance of tea mosquito bug, *Helopeltis antoni signoret* infesting neem. *J. Entomol. Zool. Studies*, 8(6): 2006-2009. <https://doi.org/10.22271/j.ento.2020.v8.i6aa.8118>
- Kumar, S., and Naik, L. K. (2002). Seasonal Incidence of *Helopeltis antonii* Signoret (Heteroptera: Miridae) on Guava, Karnataka. *J. Agric. Sci.*, 15(3), 530-533.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *J. Econometrics*, 33:341–365
- Nagendra, H., and Gopal, D. (2010). Street trees in Bangalore: Density, diversity, composition and distribution. *Urban Forest. Urban Green.*, 9(2): 129-137. DOI: [10.1016/j.ufug.2009.12.005](https://doi.org/10.1016/j.ufug.2009.12.005)
- Naffees Gowsar, S.R Radha, M and M Nirmala Devi. (2019). A Comparison of Generalized Linear Models for Insect Count Data. *Intern. J. Stat. Analysis*, 9 (1):1-9.
- Patel, R. M., Sharma, A. N., and Sharma, P. (2021). GIS based pest-weather model to predict the incidence of Girdle beetle (*Oberiopsis brevis*) in Soybean crop. *J. Agrometeorol.*, 23(2): 183-188. <https://doi.org/10.54386/jam.v23i2.65>
- Pillai, G. B., SINGH, V., Dubey, O. P., and Abraham, V. A. (1979, March). Seasonal abundance of tea mosquito (*Helopeltis antonii* Sign.) on cashew in relation to meteorological factors. In *Intern. Cashew Symp.*, 108 (pp. 101-108).
- Prasada Rao, G.S.L.H.V., G. T. Thirumala Raju, Puttaswamy, K. N. Krishnakumar, and Xavier Tony. (2002). Forecasting tea mosquito bug (*Helopeltis antonii*) of cashew (*Anacardium occidentale* L.). *J. Agrometeorol.*, 4(1): 45–52. <https://doi.org/10.54386/jam.v4i1.424>
- Rao, Y. R. (1915). *Helopeltis antonii* as a pest of neem trees. *Agric. J. India*, 10:412-6.
- Reynolds, D. R., Chapman, J. W., and Drake, V. A. (2017). Riders on the wind: the aerocology of insect migrants. In *Aeroecol.*, (pp. 145-178). Springer, Cham.

- Roy, S., Muraleedharan, N., Mukhopadhyay, A., and Handique, G. (2015). The tea mosquito bug, *Helopeltis theivora* Waterhouse (Heteroptera: Miridae): its status, biology, ecology and management in tea plantations. *Intern. J. Pest Manag.*, 61(3): 179-197.
- Sanjay, H. N., Manjunatha, R., Sumithramma, N., Mulimani, V., and Reddy, G. N. (2022). Seasonal incidence, damage and impact of abiotic factors on tea mosquito bug, *Helopeltis antonii* Signoret (Hemiptera: Miridae) infesting tamarind. *The Pharma Innovation*, 11(8): 1340-1343
- Saroj, P. L., Bhat, P. S., and Srikumar, K. K. (2016). Tea mosquito bug (*Helopeltis* spp.)—A devastating pest of cashew plantations in India: A review. *Indian J. Agric. Sci.*, 86(2): 151-162.
- Sheu, M. L., Hu, T. W., Keeler, T. E., Ong, M., and Sung, H. Y. (2004). The effect of a maor cigarette price change on smoking behavior in California: a zero-inflated negative binomial model. *Health Econom.*, 13(8):781-91. DOI: [10.1002/hec.849](https://doi.org/10.1002/hec.849)
- Singh, J. (2021). Formulation of weekly and monthly thumb rule models for prediction of potential productivity of sunflower in Punjab. *J. Agrometeorol.*, 23(2): 176-182. <https://doi.org/10.54386/jam.v23i2.64>
- Swaine, G. (1959). A preliminary note on *Helopeltis* spp. damaging cashew in Tanganyika territory. *Bull. Entomol. Res.*, 50(1):171-181.