Indian Journal of Ecology (2022) 49(3): 793-800
DOI: https://doi.org/10.55362/IJE/2022/3597

Manuscript Number: 3597
NAAS Rating: 5.79

Count Data Regression Models: An Application to the Major Rice Insect Pest Counts in the Terai Region of West Bengal

Pramit Pandit, K.N. Krishnamurthy and Bishvajit Bakshi*

Department of Agricultural Statistics, Applied Mathematics and Computer Science
University of Agricultural Sciences, Bengaluru-560 065, India
**E-mail: bishvajitb93@gmail.com*

Abstract: In spite of India's phenomenal rice production of 118.43 million tonnes in 2019-20, the potential yield and the yield realised at the farmers' fields are vastly different. Among the factors contributing towards this yield gap, the infestation of insect pests causes significant economic damage. Since these biotic menaces are largely weather-dependent, weather-based predictions of insect pests are often utilised to make economic decisions about insect pest management. Hence, an effort has been made in this study to comparatively assess the suitability of different count data regression models for weather-based prediction of three major rice insect pests (viz., gundhi bug, brown planthopper and green leafhopper) in the Terai region of West Bengal. As the input weather variables are related in a linear fashion, principal components have been obtained to be utilised subsequently in the regression analysis. Among the regression models considered, the recently developed modified Poisson quasi-Lindley regression model has empirically outperformed all of its counterparts in handling over-dispersion. However, the Poisson regression model has provided better result when no over-dispersion is evident. Outcomes emanated from the investigation have also revealed that the over-dispersion test plays a fairly good role in providing reliable guidance on the presence of over-dispersion. Hence, it is suggested that before adopting any weather-based count data regression model for predicting insect pest counts, one needs to check whether the count response variable is indeed over-dispersed.

Keywords: Count data regression, Insect pest count, Over-dispersion, Rice, Weather influence, Weather-based modelling

Rice (*Oryza sativa* L.) feeds more than half of the global population (Doliente and Samsatli 2021) and is the sole cereal crop grown under waterlogged conditions in irrigated as well as in rain-fed fields in India (Mishra et al 2007). In spite of India's phenomenal rice production of 118.43 million tonnes in 2019-20 (Directorate of Economics and Statistics 2021), the potential yield and the yield realised at the farmers' fields are vastly different. Among the factors contributing towards this yield gap, the infestation of insect pests causes significant economic damage. In the absence of stable, desirable and diversified sources of resistance to the biotic menaces, pesticides remain the only effective means to manage them (Kumar et al 2016, Sindhu et al 2016). Precise knowledge about the insect pest population dynamics can substantially help to formulate the necessary pesticide schedule for the region against the specific menace anticipated. Since these biotic menaces are largely weather-dependent (Kumar 2016), weather-based predictions of insect pests are often utilised to make economic decisions about insect pest management. Singh et al (2012) have investigated the incidence of insect pest damage in rice crop in Punjab in connection to meteorological parameters and also suggested that these can be used as a tool for the preparation of weather-based agro-advisory. Kaur and Bala (2014) have developed crop-weather-pest calendars for need-based spraying of the pesticides to manage the rice insect pests in Punjab. Yadav et al (2010) have employed weather-based log-linear models for agro-ecological zoning of brown planthopper incidence on rice. However, if the response variable of interest is a count (i.e., non-negative integer values), as in this case insect pest count, the applications of count data regression models are now well-recognised. Due to its increasing application in divergent disciplines such as actuarial science, biostatistics, demography, economics and so on, upsurging research interest in count data modelling has been evident in the last decade. The Poisson regression model has served well as a starting point for count data modelling. Tobías et al (2001) have employed the Poisson regression model to assess the short-term impact of environmental noise on daily emergency admissions in Madrid. Crowther et al (2012) have carried out a meta-analysis of survival data using the Poisson regression model. Li et al (2013) have achieved remarkable success in modelling county-level crashes by using the geographically weighted Poisson regression model. Despite the immense popularity and sheer power of the Poisson regression model, it suffers from the limitation of equi-dispersion (Osgood 2000). That is, the mean and variance of the count response variable are assumed to be equal in this model. In reality, many count data sets often violate this assumption because of their over-dispersed (variance>mean) or under-dispersed (variance<mean) nature (Lee et al 2021). To overcome this lacuna, several researchers have put great efforts to introduce and improve different regression models for the over-dispersed or under-dispersed count data (Cameron and Trivedi 2013, Wongrin and Bodhisuwan 2017). Among the alternatives available for accommodating over-dispersion, negative binomial regression is the most popular choice (Osgood 2000). It includes a parameter to inflate the Poisson dispersion as needed (Berk and MacDonald 2008). Another alternative model, which can capture over-dispersion or under-dispersion or no dispersion at all, is the generalised Poisson regression model given by Consul and Famoye (1992). Recently, Altun (2019) has also proposed a new regression model for over-dispersed count data via re-parametrisation of Poisson quasi-Lindley distribution.

Count data regression models have been compared in several studies, especially in presence of over-dispersion. Ismail and Jemain (2007) have compared the ability of negative binomial and generalised Poisson regression models in handling over-dispersion on three different sets of claim frequency data and obtained comparable performance. Gent et al (2012) have carried out a spatial analysis to derive an appropriate incidence-density relationship for downy mildew. Outcomes emanated from their study have revealed the advantages of employing a negative binomial regression model over a Poisson regression model. Melliana et al (2013) have compared negative binomial and generalised Poisson regression models in determining the factors responsible for cervical cancer cases in East Java and found that the former one has handled over-dispersion in a much better way. Yusuf and Ugalahi (2015) have observed the superiority generalised Poisson regression model in identifying the factors associated with the number of antenatal care visits in Nigeria. However, most of these studies are devoid of any pretesting of over-dispersion. This will enable us to assess whether the over-dispersion test (Cameron and Trivedi 1990) provides any indication to the suitability of count data regression models. Besides, the literature also suggests to consider various information criteria along with the significance of regression coefficients while evaluating the model performance. This is due to the fact that Poisson regression has the property to underestimate the standard error and consequently, to exaggerate the significance of the model coefficients in presence of over-dispersion (Sileshi 2006, Ismail and Jemain 2007). The above facts explicitly indicate that there is a lack of systematic investigation on count data regression modelling, especially in the field of weather-based prediction of insect pest counts. Since improved models have continuously been added to the model builders' arsenal, there is a necessity to investigate afresh the insect pest population dynamics concerning the weather parameters. With this backdrop, an effort has been made in this paper to comparatively assess the suitability of different count data regression models for weather-based prediction of three major rice insect pests (viz., gundhi bug, brown planthopper and green leafhopper) in the Terai region of West Bengal.

MATERIAL AND METHODS

Collection of data: For the present investigation on count data regression modelling, count data of three major rice insect pests, viz. gundhi bug (GB), brown planthopper (BPH) and green leafhopper (GLH) have been obtained from the Department of Agricultural Entomology, Uttar Banga Krishi Viswavidyalaya, Pundibari. The collected data consists of insect pests counts of the seven consecutive boro seasons (from 2011 to 2017). It is noteworthy to mention that the data of a particular field, which was devoid of any insecticide application, have been considered in this study with a view to analyse the natural insect pest population build-up in relation to weather parameters. Rice variety 'Annada' has been utilised for the experimental purpose. The sampling method employed to obtain the counts from the experimental field is adapted from Daorai et al (2005) and Arbab (2014). The experimental field has been divided into four strata and from each stratum, one square metre area has been selected randomly. The total count of an insect pest from those four selected areas is noted as the count of that insect pest for that particular day. The same sampling procedure is repeated at three days' intervals during the period under observation. Table 1 briefs the basic characteristics of the insect pest counts. Daily data pertaining to five weather variables (viz. T_{max}, T_{min}, RH_{max}, RH_{min} and Rainfall) have been obtained from Gramin Krishi Mausam Sewa (GKMS), Uttar Banga Krishi Viswavidyalaya, Pundibari for the same period (2011 to 2017).

Count Data Regression Models

Poisson regression model: The Poisson regression model, which allows the intensity parameter μ to depend on the explanatory variables (regressors), is often considered as the benchmark model for modelling count data. Even though this model is constrained to the restrictive assumption of equi-dispersion, it still dominates the sphere of count data regression modelling due to its simpler form and flexibility of re-parameterisation into other forms of distributional functions. In this model, the count response variable Y follows a Poisson distribution with the probability mass function (pmf)

$$P\ (Y=y) = \frac{e^{-\mu}\mu^{y}}{y!}, y = 0, 1, 2, \ldots$$
$$= 0, \text{otherwise}$$

where E(Y) = V(Y) = μ. In the log-linear version of the model, the mean parameter is parameterised as $\mu = \exp(\beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_kX_k)$ to ensure μ > 0.

Negative binomial regression model: The negative binomial regression model introduces a dispersion parameter to accommodate for the unobserved heterogeneity present in the count data. In literature, there exist different parameterisations leading to the generation of different types of negative binomial models. The most popular one among these can be mathematically expressed as

$$P\ (Y=y) = \frac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}}\left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{y},$$
$$y = 0, 1, 2, \ldots, \alpha \geq 0$$
$$= 0, \text{otherwise}$$

where E(Y) = μ, V(Y) = μ (1+αμ) and α denotes the dispersion parameter. For α = 0, the mean and variance will be equal, i.e. E(Y) = V(Y) = μ, resulting in the distribution to be a Poisson. For α > 0, the variance will exceed the mean, i.e. E(Y) < V(Y), leading to over-dispersion.

Generalised Poisson regression model: The generalised Poisson regression model is a natural extension of the standard Poisson regression model. It includes a dispersion parameter, which adjusts for both under-dispersion and over-dispersion. The count response variable Y follows a generalised Poisson distribution with pmf specified as (Famoye et al 2004)

$$P\ (Y=y) = \left(\frac{\mu}{1+\alpha\mu}\right)^{y}\frac{(1+\alpha y)^{y-1}}{y!}\exp\left[\frac{\mu\,(1+\alpha y)}{(1+\alpha\mu)}\right], \quad y = 0, 1, 2, \ldots$$
$$= 0, \text{otherwise}$$

where E(Y) = μ, $V(Y) = \mu\,(1+\alpha\mu)^2$ and α denotes the dispersion parameter. Similar to the negative binomial regression model, α = 0 reduces the distribution into Poisson. For α > 0, and α < 0, it will adjust for over-dispersion and under-dispersion, respectively. Concerning the mean-variance structure, this model possesses substantial similarity with the generalised event count (GEC_k) model proposed by Winkelmann and Zimmermann (1994).

Modified Poisson quasi-Lindley regression model: Grine and Zeghdoudi (2017) have first introduced a mixed Poisson distribution, namely Poisson quasi-Lindley distribution, by compounding the Poisson distribution with the quasi-Lindley distribution. Being motivated by the approach of the generalised linear model, Altun (2019) has further proposed a re-parametrisation of the already developed Poisson quasi-Lindley distribution by putting $\theta = \frac{(2+\alpha)}{(1+\alpha)\mu}$ in its pmf. The pmf of the modified Poisson quasi-Lindley distribution is given by

$$P\ (Y=y) = \frac{(2+\alpha)}{(1+\alpha)^2\mu}\left\{\frac{\alpha + (2+\alpha)\,(\mu+\alpha\mu)^{-1}\ (y+\alpha+1)}{[1+(2+\alpha)\,(\mu+\alpha\mu)^{-1}]^{y+2}}\right\}$$
$$, y = 0, 1, 2, \ldots, \alpha > 0, \mu > 0$$
$$= 0, \text{otherwise}$$

where E(Y) = μ and $V(Y) = \mu + \frac{\mu^2}{(2+\alpha)^2}(2 + 4\alpha + \alpha^2)$. As the variance of this distribution is always greater than its mean, it can be a great pick for modelling the over-dispersed data sets. However, the dispersion of this distribution, as expressed in the variance function, does not depend only on α, but also on μ.

The parameters of all the count data regression models employed in this study have been estimated by the method of maximum likelihood (Famoye et al 2004, Cameron and Trivedi 2013).

Comparative assessment: The performance of the count data regression models is assessed in terms of two common information criteria (Bozdogan 2000, Altun 2019), viz., the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). In estimating the amount of information loss, AIC deals with the trade-off between the goodness of fit and the simplicity of the model. AIC is mathematically expressed as

$$AIC = -2\ln L(\hat{\theta}) + 2k$$

where $L(\hat{\theta})$ is the log-likelihood value and k is the number of model parameters. Similar to AIC, BIC is also a penalised-likelihood criterion and is defined as

$$BIC = -2\ln L(\hat{\theta}) + k\ln(n)$$

where n is the number of observations. $L(\hat{\theta})$ and k have the same meanings as above. Models with the lowest AIC and BIC values are considered best.

RESULTS AND DISCUSSION

To test for normality, the Shapiro-Wilk test (Yap and Sim 2011) has been employed. Table 2 reflects the strong rejection of normality for all three cases despite the large sample size (>30). This significant result clearly indicates the possible effective implementations of the count data regression models to these data sets.

Since the pest count of the current day depends on the weather of preceding weeks, weather variables are considered accordingly. However, these variables are usually related in a linear fashion. Hence, principal component analysis (PCA), which is basically a dimensionality reduction technique, has been employed to handle the problem of multicollinearity. As the emergence pattern of the hoppers (BPH and GLH) varies from GB, different sets of weather observations have served as inputs. Bartlett's sphericity test (Bucci et al 2018), which compares the sample correlation matrix to the identity matrix, has been applied before proceeding to PCA in order to check whether there is a redundancy between the weather variables that can be summarised with fewer factors. The results (Table 3) confirm the appropriateness of PCA application to the present data.

PCA has been carried out using the ten weather variables (five weather variables of each 31 DAT – 37 DAT and 38 DAT – 44 DAT for GB, and of each 1 DAT – 7 DAT and 8 DAT – 14 DAT for BPH and GLH) for feature extraction of the weather data. The sensitivity of weather variables is evaluated in terms of principal component loadings, which are nothing but the correlations among the PC scores and the attributing weather variables. For a better interpretation of factor loadings, principal components (PCs) are obtained by varimax rotation. As the retention criterion, PCs with Eigen values higher than one are further considered for the regression analysis (Jolliffe and Cadima 2016). Table 4 provides the loadings of the retained PCs. All the ten weather variables are included in the three selected PCs in both

Table 2. Results of the Shapiro-Wilk test

Insect pest	Statistic	p-value
GB	0.947	0.002
BPH	0.928	<0.001
GLH	0.936	<0.001

Table 3. Results of the Bartlett's test of sphericity

Insect pest	Statistic	p-value
GB	361.044	<0.001
BPH and GLH	620.092	<0.001

Table 1. Characteristics of the insect pest count

Insect pest	Total no. of observations	Period under observation in each year	Mean	Variance
GB	84	45 DAT – 78 DAT	7.31	9.42
BPH	119	15 DAT – 63 DAT	114.12	757.79
GLH	119	15 DAT – 63 DAT	228.57	1861.84

Start year: 2011; End year: 2017; DAT: Days after transplanting

Table 4. Loadings of the selected principal components

Insect pest	Period considered in each year	Variable	PC_1	PC_2	PC_3
GB	38 DAT – 44 DAT	T_{max}	-0.722	0.173	0.450
		T_{min}	-0.028	0.126	0.777
		RH_{max}	0.887	0.053	0.162
		RH_{min}	0.881	-0.017	0.091
		Rainfall	0.731	-0.005	0.189
	31 DAT – 37 DAT	T_{max}	0.076	-0.745	0.386
		T_{min}	0.347	0.029	0.760
		RH_{max}	0.137	0.834	0.316
		RH_{min}	0.040	0.827	0.297
		Rainfall	-0.105	0.589	0.001
% of variance explained			35.76	31.31	25.92
BPH and GLH	8 DAT – 14 DAT	T_{max}	0.505	-0.662	0.323
		T_{min}	0.680	0.305	0.418
		RH_{max}	0.265	0.852	0.077
		RH_{min}	0.295	0.826	0.159
		Rainfall	0.100	0.727	0.066
	1 DAT – 7 DAT	T_{max}	-0.233	0.067	0.880
		T_{min}	0.636	0.458	0.397
		RH_{max}	0.837	0.259	-0.155
		RH_{min}	0.853	0.237	-0.107
		Rainfall	0.635	-0.066	-0.147
% of variance explained			36.74	33.09	18.04

cases. Only a few variables have displayed the high loading within each PC. In the case of GB, PC_1 has accounted for 35.76 per cent of variations in the input data and has loadings of more than 0.7 with a combination of T_{max}, RH_{max}, RH_{min} and rainfall of 38 DAT – 44 DAT. PC_2 on the other hand, has explained 31.31 per cent of the total variation by extracting the information from the same weather variables (T_{max}, RH_{max}, RH_{min} and rainfall) but of 31 DAT – 37 DAT. PC_3 depends on T_{min} of both 31 DAT – 37 DAT and 38 DAT – 44 DAT to explain 25.92 per cent of the variation. However, in the case of BPH and GLH, PC_1 has accounted for 36.74 per cent of variations by exhibiting higher loadings with T_{min}, RH_{max}, RH_{min} and rainfall of 1 DAT – 7 DAT and T_{min} of 8 DAT – 14 DAT. PC_2 has explained 33.09 per cent of variations and has higher loadings with a combination of T_{max}, RH_{max}, RH_{min} and rainfall of 8 DAT – 14 DAT. PC_3 relies on T_{max} of 1 DAT – 7 DAT to account for 18.04 per cent of variations in the input data. In both cases, the retained PCs cumulatively explain around 90 per cent of the input data variations.

In the next step, to find out the best model under each count data regression set up, the stepdown method has been employed. In our study, the stepdown regression procedure starts with considering all the retained PCs in the model, i.e. with a full model and then, variable selection and model building are carried out simultaneously. For GB and GLH count data, the final model consists of PC_2 and PC_3 as explanatory variables in all the regression set-ups under investigation. However, in the case of BPH count data, PC_1 and PC_2 comprise the final models. To examine whether the over-dispersion test provides any reliable guidance for model selection, we have applied the over-dispersion test to all the count data sets under study. Table 5 deciphers the presence of overdispersion in BPH and GLH count data sets indicating that in both these cases, the models with the ability of accommodating over-dispersion (Negative binomial regression model, Generalised Poisson regression model and Modified Poisson quasi-Lindley regression model) may perform better than the model assuming equi-dispersion (Poisson regression model).

The parameter estimates of the Poisson regression, negative binomial regression, generalised Poisson regression and modified Poisson quasi-Lindley regression models are obtained in the next step and provided in Table 6-

Table 5. Results of the over-dispersion test

Insect pest	Statistic	p-value
GB	1.272	0.207
BPH	11.603	<0.001
GLH	10.153	<0.001

Table 6. Parameter estimates of the poisson regression models

Insect pest	Parameter	Estimate	Standard error	p-value
GB	β_0	3.034	0.024	<0.001
	β_2	-0.037	0.004	<0.001
	β_3	0.036	0.002	<0.001
BPH	β_0	4.737	0.008	<0.001
	β_1	-0.028	0.001	<0.001
	β_2	0.013	0.002	<0.001
GLH	β_0	5.431	0.006	<0.001
	β_2	0.028	0.001	<0.001
	β_3	0.009	0.001	<0.001

Table 7. Parameter estimates of the negative binomial regression models

Insect pest	Parameter	Estimate	Standard error	p-value
GB	β_0	3.032	0.032	<0.001
	β_2	-0.038	0.016	0.017
	β_3	0.037	0.017	0.029
	Dispersion	0.039	0.031	0.208
BPH	β_0	4.734	0.023	<0.001
	β_1	-0.025	0.007	<0.001
	β_2	0.012	0.005	0.016
	Dispersion	0.049	0.022	0.026
GLH	β_0	5.425	0.017	<0.001
	β_2	0.031	0.007	<0.001
	β_3	0.012	0.006	0.045
	Dispersion	0.031	0.014	0.031

Table 8. Parameter estimates of the generalised Poisson regression models

Insect pest	Parameter	Estimate	Standard error	p-value
GB	β_0	3.029	0.026	<0.001
	β_2	-0.036	0.012	0.003
	β_3	0.040	0.014	0.004
	Dispersion	0.018	0.011	0.102
BPH	β_0	4.729	0.020	<0.001
	β_1	-0.023	0.005	<0.001
	β_2	0.015	0.003	<0.001
	Dispersion	0.014	0.006	0.019
GLH	β_0	5.421	0.011	<0.001
	β_2	0.034	0.005	<0.001
	β_3	0.011	0.004	0.006
	Dispersion	0.008	0.004	0.026

9 respectively, where β_0 denote the intercept term and β_i (i= 1, 2, 3) represents the coefficient of the i^{th} PC.

The comparative assessment (Table 10) reflects that the modified Poisson quasi-Lindley regression model has provided better results than all other count data regression models in presence of over-dispersion. The Poisson regression model has clearly failed to account for over-dispersion in the case of BPH and GLH data sets (Table 6). The significance of regression parameters has exhibited an upward bias due to the under-estimation of standard errors (Berk and MacDonald 2008). However, as no over-dispersion is evident in the GB data set, the Poisson regression model has outperformed all of its counterparts. As

Table 9. Parameter estimates of the modified Poisson quasi-Lindley regression models

Insect pest	Parameter	Estimate	Standard error	p-value
GB	β_0	3.033	0.035	<0.001
	β_2	-0.035	0.015	0.019
	β_3	0.034	0.016	0.034
	Dispersion	0.446	0.378	0.238
BPH	β_0	4.736	0.024	<0.001
	β_1	-0.023	0.006	<0.001
	β_2	0.011	0.004	0.006
	Dispersion	0.028	0.014	0.046
GLH	β_0	5.428	0.019	<0.001
	β_2	0.026	0.006	<0.001
	β_3	0.013	0.005	0.009
	Dispersion	0.014	0.007	0.049

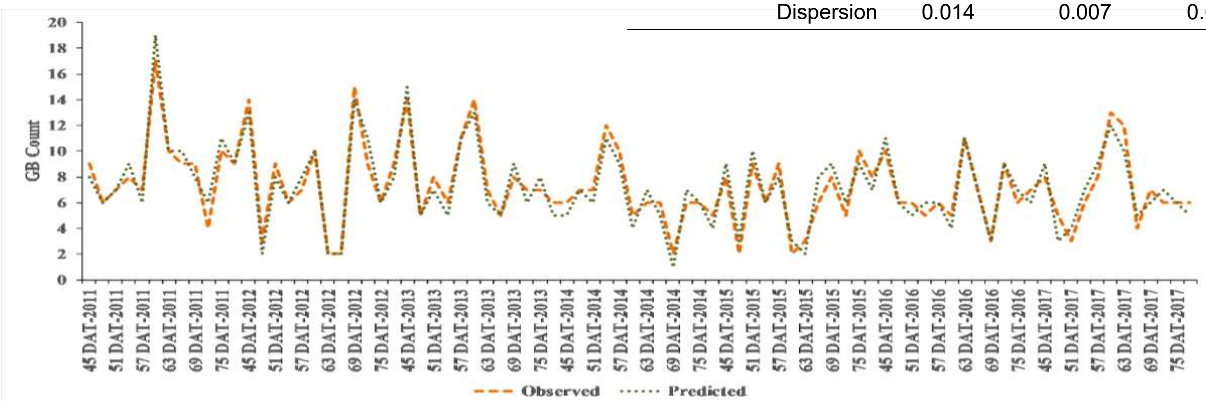

Fig. 1. Observed and the Poisson regression model predicted GB counts for the years under study

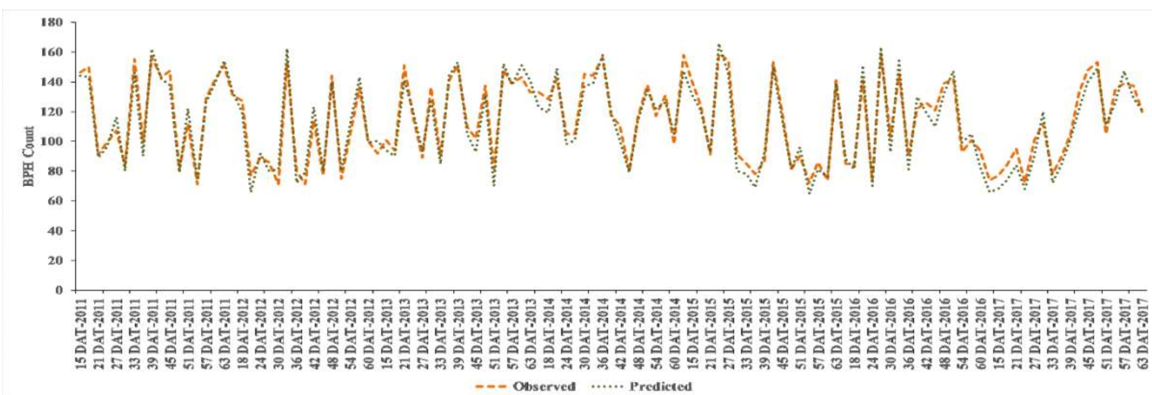

Fig. 2. Observed and the modified Poisson quasi-Lindley regression model predicted BPH counts for the years under study

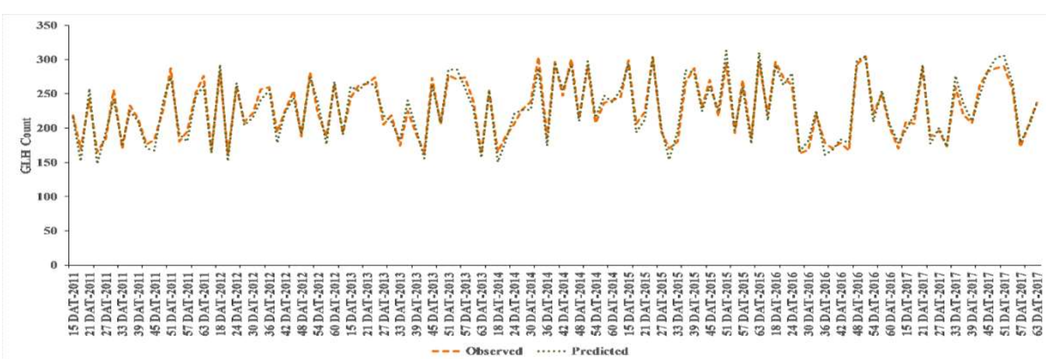

Fig. 3. Observed and the modified Poisson quasi-Lindley regression model predicted GLH counts for the years under study

Table 10. Comparative assessment of the count data regression models under study

Model	GB		BPH		GLH	
	AIC	BIC	AIC	BIC	AIC	BIC
Poisson regression model	1.90	9.19	866.40	874.74	1064.97	1073.31
Negative binomial regression model	7.01	16.73	492.50	503.62	575.80	586.92
Generalised Poisson regression model	5.73	15.45	501.05	512.17	580.94	592.06
Modified Poisson quasi-Lindley regression model	9.96	19.68	463.39	474.51	562.32	573.44

the estimation procedure of these models is directly linked to the existence of over-dispersion of the count response variable conditional to the explanatory variables (Cameron and Trivedi 2013), the detection of over-dispersion is of prime importance to ensure that the inferences drawn from the employed count data regression model are appropriate. The mean and variance values of GB count are somewhat comparable, whereas in case of BPH and GLH counts, there exist substantial differences indicating possible over-dispersion, which has further been confirmed from the results of the over-dispersion test. Year-wise observed and fitted counts for the three rice insect pests under study by the respective best-fitted model are graphically represented in Figure 1-3, respectively.

CONCLUSIONS

Among the count data regression models under investigation, the recently developed modified Poisson quasi-Lindley regression model has empirically outperformed all of its counterparts in handling over-dispersion. However, the Poisson regression model has provided better result when no over-dispersion is evident. We also find that the over-dispersion test plays a fairly good role in providing reliable guidance on the presence of over-dispersion. Even though the weather-based modified Poisson quasi-Lindley regression model has satisfactorily accommodated for over-dispersion, the scope still remains for further modification and exploration to predict the over-dispersed count response variable more accurately.

ACKNOWLEDGEMENT

The authors deeply acknowledge the resources and financial aid received from the Indian Council of Agricultural Research (ICAR), New Delhi, India, and the University of Agricultural Sciences, Bengaluru, India. The authors also sincerely acknowledge Uttar Banga Krishi Viswavidyalaya, Pundibari for providing the data sets necessary for the investigation.

REFERENCES

Altun E 2019. A new model for over-dispersed count data: Poisson quasi-Lindley regression model. *Mathematical Sciences* **13**(3): 241-247.

Arbab A 2014. Spatial distribution and minimum sample size for overwintering larvae of the rice stem borer *Chilo suppressalis* (Walker) in paddy fields. *Neotropical Entomology* **43**(5): 415-420.

Berk R and MacDonald JM 2008. Overdispersion and Poisson regression. *Journal of Quantitative Criminology* **24**(3): 269-284.

Bozdogan H 2000. Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology* **44**(1): 62-91.

Bucci N, Luna M, Viloria A, García JH, Parody A, Varela N and López LAB 2018. Factor analysis of the psychosocial risk assessment instrument, pp. 149-158. In: Ying Tan, Yuhui Shi, Qirong Tang (eds), *Proceedings of the international conference on data mining and big data*, June 17-22, 2018, Shanghai, China.

Cameron AC and Trivedi PK 1990. Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics* **46**(3): 347-364.

Cameron AC and Trivedi PK 2013. *Regression analysis of count data*, Cambridge University Press, Cambridge.

Consul P and Famoye F 1992. Generalized Poisson regression model. *Communications in Statistics-Theory and Methods* **21**(1): 89-109.

Crowther MJ, Riley RD, Staessen JA, Wang J, Gueyffier F and Lambert PC 2012. Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Medical Research Methodology* **12**(1): 1-14.

Daorai A, Jamornmarn S and Thongdeethae S 2005. Sampling technique in rice pest management: A tool for the rice growing farmers at Buang Thong Lang village in Lam Luk Ka sub-district, Pathum Thani province, Thailand. *Agriculture and Natural Resources* **39**(3): 424-430.

Directorate of Economics and Statistics 2021. *Agricultural Statistics at a Glance 2020*. Department of Agriculture, Cooperation & Farmers Welfare, Ministry of Agriculture & Farmers Welfare, Government of India.

Doliente S and Samsatli S 2021. Integrated production of food, energy, fuels and chemicals from rice crops: Multi-objective optimisation for efficient and sustainable value chains. *Journal of Cleaner Production* **285**: 1-45.

Famoye F, Wulu JT and Singh KP 2004. On the generalized Poisson regression model with an application to accident data. *Journal of Data Science* **2**(3): 287-295.

Gent DH, Farnsworth JL and Johnson DA 2012. Spatial analysis and incidence–density relationships for downy mildew on hop. *Plant Pathology* **61**(1): 37-47.

Grine R and Zeghdoudi H 2017. On Poisson quasi-Lindley distribution and its applications. *Journal of Modern Applied Statistical Methods* **16**(2): 403-417.

Ismail N and Jemain AA 2007. Handling overdispersion with negative binomial and generalized Poisson regression models. *Casualty Actuarial Society Forum* **7**: 103-158.

Jolliffe IT and Cadima J 2016. Principal component analysis: A review and recent developments. *Philosophical Transactions of* *the Royal Society A: Mathematical, Physical and Engineering Sciences* **374**(2065): 1-16.

Kaur P and Bala A 2014. Development of crop-weather-pest calendars for rice crop in Punjab. *Journal of Research* **51**(1): 18-27.

Kumar S 2016. Impact of climate variability and crop phenology on abundance of mango thrips complex. *Indian Journal of Ecology* **43**(2): 637-644.

Kumar A, Jain A, Bhattacharya B, Kumar V, Mishra AK and Chattopadhyay C 2016. Epidemiology and forecasting of insect-pests and diseases for value-added agro-advisory. *Mausam* **67**(1): 267-276.

Lee C, Famoye F and Akinsete A 2021. Generalized count data regression models and their applications to health care data. *Annals of Data Science* **8**(2): 367-386.

Li Z, Wang W, Liu P, Bigham JM and Ragland DR 2013. Using geographically weighted Poisson regression for county-level crash modelling in California. *Safety Science* **58**: 89-97.

Melliana A, Setyorini Y, Eko H, Rosi S and Purhadi S 2013. The comparison of generalized Poisson regression and negative binomial regression methods in overcoming overdispersion. *International Journal of Engineering, Science and Technology* **8**(2): 255-258.

Mishra M, Sahu RK and Padhy RN 2007. Growth, yield and elemental status of rice (*Oryza sativa*) grown in fly ash amended soils. *Ecotoxicology* **16**(2): 271-278.

Osgood DW 2000. Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology* **16**(1): 21-43.

Sileshi G 2006. Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bulletin of entomological research* **96**(5): 479-488.

Sindhu N, Malik JS and Sheharawat PS 2016. Farmers' perception in context of impact of agrochemicals on environment. *Indian Journal of Ecology* **43**(2): 806-810.

Singh S, Kaur P, Kumar V and Singh H 2012. Incidence of insect pest damage in rice crop in relation to meteorological parameters in Punjab: A plant clinic data based case study. *Journal of Agrometeorology* **14**(1): 50-53.

Tobías A, Díaz J, Saez M and Alberdi JC 2001. Use of Poisson regression and Box–Jenkins models to evaluate the short-term effects of environmental noise levels on daily emergency admissions in Madrid, Spain. *European Journal of Epidemiology* **17**(8): 765-771.

Winkelmann R and Zimmermann KF 1994. Count data models for demographic data. *Mathematical Population Studies* **4**(3): 205-221.

Wongrin W and Bodhisuwan W 2017. Generalized Poisson–Lindley linear model for count data. *Journal of Applied Statistics* **44**(15): 2659-2671.

Yadav DS, Chander S and Selvaraj K 2010. Agro-ecological zoning of brown planthopper [*Nilaparvata lugens* (Stal)] incidence on rice (*Oryza sativa* L.). *Journal of Scientific and Industrial Research* **69**: 818-822.

Yap BW and Sim CH 2011. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation* **81**(12): 2141-2155.

Yusuf OB and Ugalahi LO 2015. On the performance of the Poisson, negative binomial and generalized Poisson regression models in the prediction of antenatal care visits in Nigeria. *American Journal of Mathematics and Statistics* **5**(3): 128-136.

Received 08 January, 2022; Accepted 03 June, 2022