Journal of Entomology and Zoology Studies
Available online at www.entomoljournal.com
JEZS
Journal of Entomology and Zoology Studies

E-ISSN: 2320-7078
P-ISSN: 2349-6800
www.entomoljournal.com
JEZS 2022; 10(1): 388-394
© 2022 JEZS
Received: 10-11-2021
Accepted: 12-12-2021

Kari Iamba
Department of Agriculture, PNG University of Natural Resources and Environment, Private Mail Bag, Vudal, East New Britain, Papua New Guinea

Corresponding Author:
Kari Iamba
Department of Agriculture, PNG University of Natural Resources and Environment, Private Mail Bag, Vudal, East New Britain, Papua New Guinea

Data analysis of flea beetle (*Psylliodes chrysocephala* L.): comparing three (3) distribution families of Generalized Linear Model

Kari Iamba

DOI: https://doi.org/10.22271/j.ento.2022.v10.i1e.8958

Abstract

Generalized linear models (GLMs) are used when the variance is not constant, and when the errors are not normally distributed. Some ecological and entomological response variables invariably suffer from these two standard assumptions, and GLMs are excellent at dealing with them. Three distribution families of GLM: (1) Linear, (2) Poisson and (3) Gamma, were fitted to the null, reduced and full models with the log link function. The data used was derived from a study on the cabbage flea beetle (*Psylliodes chrysocephala* L.). According to the residual deviance (Goodness of Fit) and Akaike information criterion (AIC) as an estimator of model quality, it was confirmed that Gamma GLM is the best fit for the data set. Both the AIC and deviance were low in the Gamma model, while high values were noted for Poisson and Linear GLMs. Our study confirms that severe skewness often exists in data sets pertaining to parasitology and entomology. The Gamma distribution provided a better and more robust alternative estimator than Poisson and Linear models. Poisson distribution is mostly used to model the count of events occurring within a given time interval. Poisson and linear GLMs did not fit well with the data set, which was evident by their high scaled deviance (G^2).

Keywords: GLM, residuals, null model, reduced model, full model

Introduction

Non-normal data that deviates from the normal distribution is frequently observed by field entomologists, biologists and ecologists. Although Analysis of Variance (ANOVA) have been widely used in data analysis, the abundance and incidence data often violate the assumptions of ANOVA [1]. Most data pertaining to insect or weed abundance do not meet the assumptions of normality and homogeneity of variance [1-4]. Therefore appropriate analytical tools are at this moment needed to analyse data that is not normally distributed and non-linear. The generalized Linear Model (GLM) thus offers an alternative to address such skewness since it provides a unified application to other common statistical procedures [5]. The traditional linear model assumes that errors have normal distributions [6]. GLMs as a class of statistical models provide an abstract and simplified representations of the real data. They are called GLM since they generalize the classical linear models based on the normal distributions [5]. In addition to the linear regression component, GLMs include a special exponential family which transforms the mean via a “link function” and links the regression part to the mean of one of these distributions [6, 7].

GLM is a combination of systematic and random components of a linear model which has three (3) characteristics: (i) a dependent variable z whose distribution with parameter θ as one of the class, (ii) a set of independent variables $x_1, \ldots, x_m$ and predicted systematic component $Y=\sum=\beta(_iX_i)$, and (iii) the linking function $\theta=f(Y)$ connecting the parameter θ of the distribution of z with the Y*’s* of the linear model [6]. Different statistical criteria are used for assessing a model quality and best fit. The Analysis of Deviance involving the residual deviance is a good test to determine the Goodness of Fit of a model [6, 8, 9]. On the other hand, the Akaike information criterion (AIC) estimates the quality of models in the class of both linear and Generalized Linear Model (GLM) [10]. Both AIC or the Bayesian information criterion (BIC) are used for model comparisons [11]. Burnham and Anderson [12] emphasised the use of AIC and BIC for comparing statistical models and transformations. Sileshi [1] and Yabeja [13] used the Poisson regression model to analyse the population of psyllid and whitefly population.

For this study, we made comparisons among three distribution families of GLM using AIC, BIC and residual deviance. Linear (gaussian), Poisson, and Gamma distributions used the three exponential families. From the AIC and deviance analysis, the model with the best fit and good quality was considered the best fit for the pest data. These three distribution families were used because they deal with real numbers (> 0). All data collected from the pest study were real numbers collected through measurements of pest counts, leaf area index and defoliation values.

Materials and Methods

(a) Study site

The data was derived from a study done at the University of Natural Resources and Environment (UNRE) in East New Britain, Papua New Guinea (PNG). PNG UNRE is an Agricultural and Environmental institution that focuses its academic programs on natural resource management. The campus is situated relatively at 4°21'01.90" S and 152° 54 00'33.44" E with an elevation of 51 meters above sea level at [14]. The soil has been categorized as more calcareous in nature and relatively sandy loamy with high alkalinity [15]. It has a tropical climate that experiences a great deal of rainfall all year round, even in the driest month.

(b) Nursery

Before the study, a nursery was established via seed sowing in a nursery house. We used the cabbage variety, K-K cross (*Oleracea* var. *capitata*), as the host plant since it is popular in the tropical regions, proven to be heat tolerant and has quick maturity (58 days after transplant). The seedling phase lasted for three (3) weeks in the nursery house. To ensure viability and robust growth, seedlings were kept for 3 weeks until plantlets had acclimatized to the local environmental conditions before transplanting.

(c) Sampling and data collection

Three response variables were measured to calculate the impact of cabbage flea beetle under three treatments; T1=lemon grass + cabbage, T2= marigold plant + cabbage, and T3= monocrop cabbage (control), with each treatment, replicated three times. The abundance of flea beetles was counted during each sampling time. Due to the mobility of the beetles, less disturbance was done to the foliar. Defoliation (%) as a response variable and leaf area index (LAI) were calculated using BioLeaf Foliar Analysis in an android phone where a sample of the leaf was placed against a white paper, a clear photo taken using the phone camera (13-megapixel), and then the app automated all scanning and the defoliation (%) and LAI were calculated. Sampling was done randomly at all growth stages of cabbage (i.e. seedling, crowning, harvesting) on 5 plants per treatment per week. Data was collected once per week for a total period of 5-weeks. So, in total, there were 15 data (5 plants x 3 measured variables) collected per treatment per replicate. Multiplying that by 3 treatments and 3 replicates each (15x3x3) produced 135 data points per sampling per week. Therefore the grand total of collected data for 5 weeks was 675 (5x135) data points.

(d) Data analysis

The first objective of this study was to match the three distribution families to fit the data. We used the full model, $y{\sim}x_1{+}x_2{+}x_3$, where the response variable (y) is a function of sets of independent variables (x). Here, y represents defoliation (%), and it depends on abundance (x_1), LAI (x_2) and treatments (x_3). In GLM, we can model the relationship between predictor X and response Y as $g(\mu) = \beta_0 + \beta_1^T \mathrm{X}$, where $\mu = E(Y;\ \theta;\ \varphi) = b'(\theta)$ and g is referred to as the link function [16]. The three important properties of a generalized linear model are the error structure, the linear predictor and the link function. The Gaussian, Poisson and Gamma distributions were investigated for these properties and fitted to the pest data. Although all computations were done in RStudio (version 4.0.3), only the simplified formulas are shown as follows for the sake of clarity.

(i). Linear model: $y = \beta_0 + \beta_i X_i + \varepsilon_i$, where y is the dependent variable, β_0 is the y-intercept, β_i is the slope coefficient, X_i is the independent variable and ε_i is the random error term [17].

(ii). Poisson model:
$log(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$, where y is the response variable, α and β are numeric coefficients, α being the intercept, x is the predictor/explanatory variable, β_p is the i-th coefficient and x_p is the i-th predictor variable [18].

(iii). Gamma model: $g(\mu_i) = n_i = x_i\beta$, where g is the link function, $\beta = (\beta_0 \,..,\beta_p)$ is the vector of mean regression parameters, x_i is the i-th vector value of the explanatory variables, and η_i is a linear predictor [19].

In R, the error structure is defined by means of the family directive, used as part of the model formula. Examples are glm(y ~ z, family = Gaussian) which means that the response variable y has normal errors, glm(y ~ z, family = Poisson) means that the response variable y has Poisson errors, and glm(y ~ z, family = Gamma) means that the response has Gamma errors. The explanatory variable z can be continuous, leading to a regression analysis or categorical, leading to an ANOVA-like procedure called analysis of deviance [20].

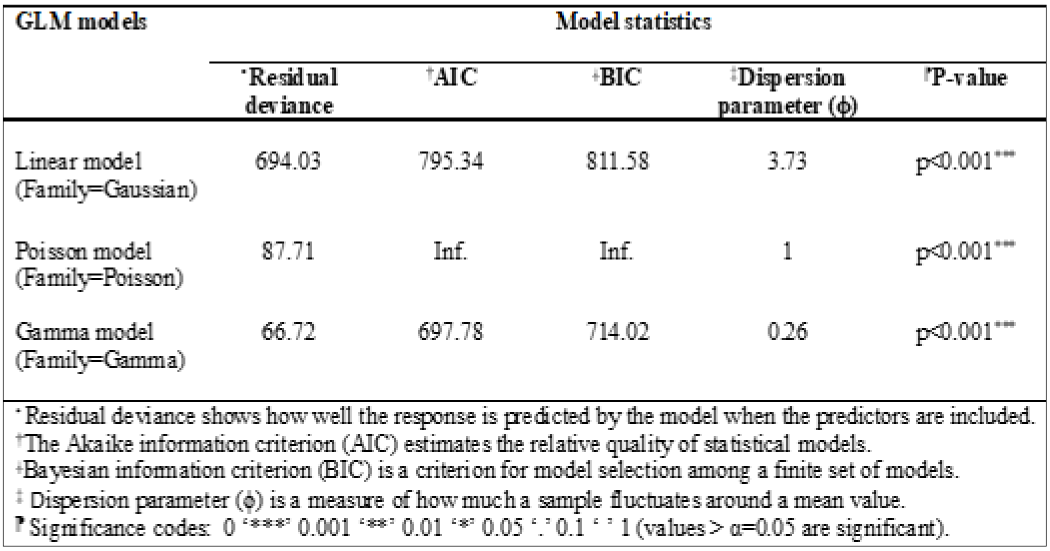
For measuring the Goodness of Fit of a GLM, an Analysis of Deviance involving the residual deviance was used. The deviance formula for normal linear is $\sum(y - \bar{y})^2$, for Poisson it is $2\sum y\ln(y/\mu)$-$(y - \mu)$ and for Gamma it is $2\sum(y - \mu)/y - \ln(\gamma/\mu)$ where y is observed data, $\bar{y}$ the mean value of y, and μ are the fitted values of y from the maximum likelihood model [20]. The Akaike information criterion (AIC) was used to estimate the quality of the three GLM models. The general formula for AIC is AIC = 2*ln(*likelihood*) + 2**k* where ln is the natural logarithm, *k* is the number of parameters in the statistical model and RSS is the residual sums of squares [21]. Bayesian information criterion (BIC) also compares different GLMs and is often used alongside AIC. The BIC statistic calculated for logistic regression is, BIC = -2 * *LL* + *log(N)* * *k*, where log() has the base-e called the natural logarithm, *LL* is the log-likelihood of the model, *N* is the number of examples in the training dataset, and *k* is the number of parameters in the model [12].

Results

The model comparisons were based on the dataset of cabbage flea beetle (*Psylliodes chrysocephala* L.). Three response variables measured were defoliation (%), abundance and leaf area index (LAI). These variables were tested under three treatments: T1=lemon grass + cabbage, T2= marigold plant + cabbage, and T3= monocrop cabbage (control) with each treatment replicated three times. A total of 675 plants were sampled during the study, with 654 data points for *P. chrysocephala*. abundance, 1238.09 for defoliation (%) and 749.51 for leaf area index (LAI).

The data pertaining to the three variables were not normally distributed according to the Shapiro-Wilk test ($p < 0.05$). To correct the skewness in data distribution, we used Generalized Linear Model (GLM). The data were log transformed and analysed using the exponential family link function in RStudio (version 4.0.3). The aim was to find out the relationship between the values of the response variable (as measured in the data and predicted by the model in fitted values) and the linear predictor. The link function relates the mean value of *y* to its linear predictor. We used link functions (e.g. log link) and transformed the response variables (i.e. log (*y*) as the response variable rather than *y*). The following default canonical link functions were used for each GLM Family: Linear (link=identity), Poisson model (link=log) and Gamma (link=reciprocal) (table 1). The fitted values produced by each family were bit able to match the values of the data perfectly.

Table 1: The statistics of the three generalized linear models (GLM) were computed using the residual deviance, AIC, BIC, dispersion parameter (ϕ) and p-value.



GLM models	**Model statistics**				
	[*]**Residual deviance**	[†]**AIC**	[ǂ]**BIC**	[‡]**Dispersion parameter (ϕ)**	[¶]**P-value**
Linear model (Family=Gaussian)	694.03	795.34	811.58	3.73	p<0.001***
Poisson model (Family=Poisson)	87.71	Inf.	Inf.	1	p<0.001***
Gamma model (Family=Gamma)	66.72	697.78	714.02	0.26	p<0.001***

[*] Residual deviance shows how well the response is predicted by the model when the predictors are included.
[†] The Akaike information criterion (AIC) estimates the relative quality of statistical models.
[ǂ] Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models.
[‡] Dispersion parameter (ϕ) is a measure of how much a sample fluctuates around a mean value.
[¶] Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (values > α=0.05 are significant).

Deviance (G^2) was used as Goodness of Fit (GOF) and as a measure of discrepancy to assess the GOF in the GLM families. Akaike Information Criterion (AIC) was used to compare the three GLM models and determine which one was the best fit for the data. In addition, the Bayesian Information Criterion (BIC) was also used as it is closely related to AIC however the penalty term is larger in BIC than in AIC. Gamma GLM had lower AIC (697.78) and BIC (714.02) when compared to Poisson GLM and Gaussian GLM (Table 1). Poisson model had infinite values for both AIC and BIC, which were considered higher than Gamma. The dispersion parameter (ϕ=0.26) of Gamma was much lower than the other two GLM models. Gamma also had the lowest residual deviance (66.72) which was deemed as a good fit for the data. The linear model had the highest residual deviance (694.03) followed by the Poisson model (87.71) therefore both were not considered as a good fit. The p-values were deduced by comparing the GLM models with a null model.

According to the Analysis of Deviance table (Table 2), three sub-models from the three GLM models were analysed. Each of the three GLM models was reduced by omitting a *x*-variable and then comparing it with a full and null model. The reduced and null models were derived from the full GLM formular, $y \sim x_1 + x_2 + x_3$, where x_1=Abundance, x_2=LAI, x_3=Treatment and *y*=Defoliation. The reduced model had only two x-variables; abundance and LAI, while treatment was omitted. The null model did not have any *x*-variable but only the response variable (*y*) (Table 2). The null Linear (gaussian) model $(y \sim 1)$ had the highest residual deviance (RD) (1610.18) therefore it was not a good fit. The reduced linear model, $y \sim x_1 + x_2$, had a residual deviance of 712.55 ($p < 0.05$*). The full linear model, $y \sim x_1 + x_2 + x_3$, had the lowest deviance (694.03, $p <$ 0.001***) therefore we considered it as a better linear model.

The null Poisson model $(y \sim 1)$ had the highest residual deviance (389.77) therefore, it was not a good fit. The reduced Poisson model, $y \sim x_1 + x_2$, had a residual deviance of 90.09 ($p >$ 0.05). The full Poisson model, $y \sim x_1 + x_2 + x_3$, had the lowest deviance (87.71, $p <$ 0.001***) therefore we considered it as a better Poisson model. The null Gamma model (y~1) had the highest residual deviance (142.25) therefore it was not a Goodness of Fit (GOF). The reduced Gamma model, $y \sim x_1 + x_2$, had a residual deviance of 66.73 ($p >$ 0.05). The full Gamma model, $y \sim x_1 + x_2 + x_3$, was slightly lower than the reduced model (66.72, $p <$ 0.001***) therefore, it was considered as a better model.

The relationship between *x* (predicted values) and *y* (residuals) of the Linear GLM is non-linear (Fig 1). The fitted red line is not close to the dash line and residuals have many outliers. The red line is not linear but quadratic in nature therefore it violates the linear distribution pattern. The Normal Q-Q plot of the linear model (lm) shows that the dependent variable is not normally distributed (Fig. 1). A few outlier data points violated a normal distribution trend (lower and upper tail). Therefore we cannot accept Linear GLM as a good model. For the Poisson model, the red line is close to the dashed zero line except in the lower tail (Fig. 2). The red fitted line does not perfectly fits the residual data points as there are still distinct outliers. The similar trend is also visible in the Q-Q plot (Fig. 2) where both the tail and head have deviated from the theoretical quantile distribution line. Therefore we cannot conclude that Poisson GLM is a good model fit. The Gamma model has a better fit than the Linear and Poisson GLMs. The red line is much closer to the dashed line and fits well with the residuals (Fig. 1). The fitted red line is not linear or quadratic but a polynomial regression that often fits a nonlinear relationship between the values of *x* and the corresponding mean of *y*, denoted $E(y \mid x)$. The tail of the Q-Q plot lies approximately on the theoretical line with less head deviation. Therefore Gamma model is the better model to use than the other two GLMs.

Table 2. This Analysis of Deviance table compares different GLM models. Each of the three generalized linear models (Gwas reduced by omitting a x-variable and hen comparing it with a full and null model. Model comparisons were computed using the residual df, residual deviance, DF, Deviance and p-value.

Null, Reduced and Full GLM models	**Model statistics**				
	[†]Residual df	**[•]Residual deviance**	**[ǂ]DF**	**[‡]Deviance**	**[⁋]P-value**
Linear model					
Model 1: $\gamma \sim 1$	189	1610.18			
Model 2: $\gamma \sim x_1 + x_2$	187	712.55	-1	-18.52	$p<0.05^{*}$
Model 3: $\gamma \sim x_1 + x_2 + x_3$	186	694.03	3	916.15	$p<0.001^{***}$
Poisson model					
Model 1: $\gamma \sim 1$	189	389.77			
Model 2: $\gamma \sim x_1 + x_2$	187	90.09	-1	-2.38	$p > 0.05$
Model 3: $\gamma \sim x_1 + x_2 + x_3$	186	87.71	3	302.07	$p<0.001^{***}$
Gamma model					
Model 1: $\gamma \sim 1$	189	142.25			
Model 2: $\gamma \sim x_1 + x_2$	187	66.73	-1	-0.003	$p > 0.05$
Model 3: $\gamma \sim x_1 + x_2 + x_3$	186	66.72	3	75.53	$p<0.001^{***}$

[•] Residual deviance shows how well the response is predicted by the model when the predictors are included.
[†] Degrees of freedom (df) for residuals and [ǂ] model (DF).
[‡] Deviance measures the goodness-of-fit of a statistical models (i.e. hypothesis testing).
[⁋] Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 (values > α=0.05 are significant).

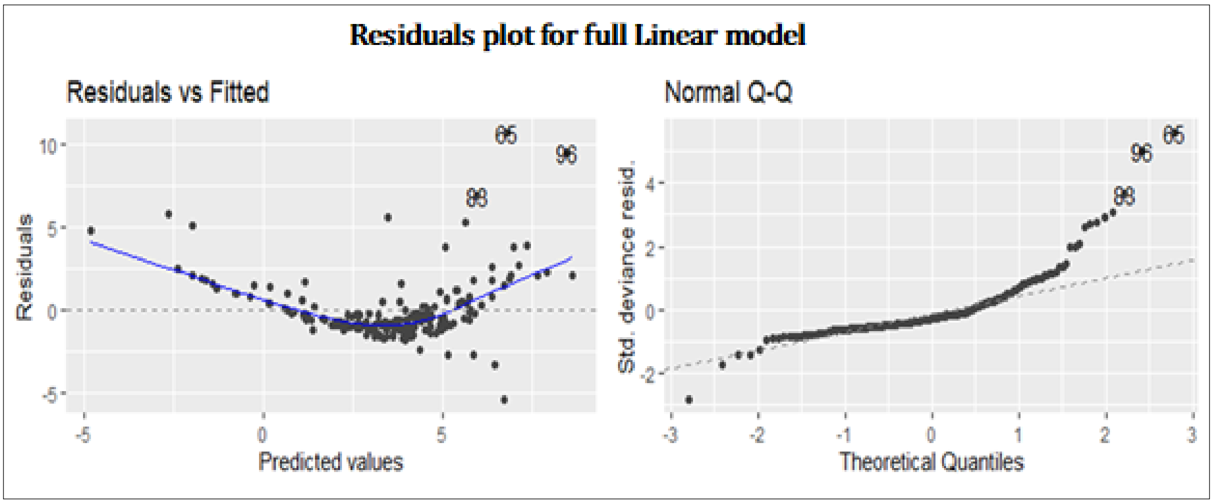


Fig 1: The residuals vs fitted plot (left) represents the full Linear model (formula = Defoliation ~ Abundance + LAI + Treatment). The relationship between *x* (predicted values) and *y* (residuals) is non-linear with distinct outliers. The normal Q-Q plot (right) shows that points fall along the normality line in the middle of the graph, but curve off in the extremities. This usually mean the data have more extreme values than would be expected in a normal distribution.

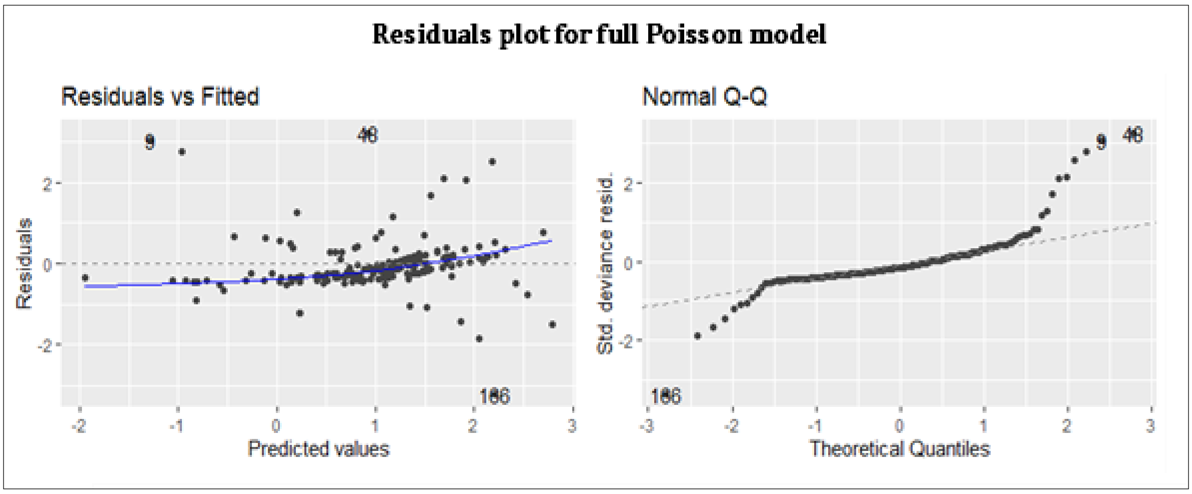


Fig 2: The residuals vs fitted plot (left) represents the full Poisson model (formula = Defoliation ~ Abundance + LAI + Treatment). The relationship between *x* (predicted values) and *y* (residuals) is non-linear with visible outliers. The normal Q-Q plot (right) shows that points fall along the normality line in the middle of the graph, but curve off in the extreme head. The presence of extreme values violates a possibility of normal distribution.

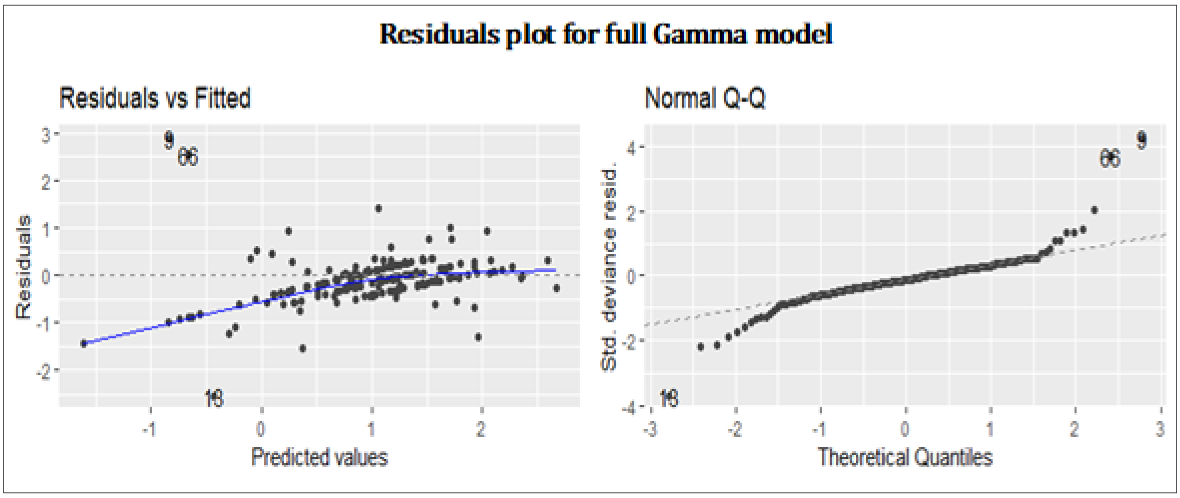


Fig 3: The residuals vs fitted plot **(left)** represents the full Gamma model (formula = Defoliation ~ Abundance + LAI + Treatment). The relationship between *x* (predicted values) and *y* (residuals) is non-linear, with some outliers. The normal Q-Q plot **(right)** shows that points fall along the normality line in the middle of the graph and curve off in the extreme tail. However the data lies close to the normality line at the extreme head making Gamma as a possible good fit.

Discussion

According to the residual deviance (Goodness of Fit) and AIC as an estimator of model quality, it was confirmed that Gamma GLM is the best fit for the data set on cabbage flea beetle (*Psylliodes chrysocephala* L.). Both the AIC and deviance were low in the Gamma model, while high values were noted for Poisson and Linear GLMs. The use of deviance as a measurement of Goodness of Fit for GLM models has been described extensively by other researchers [6, 24, 42]. Akaike Information Criterion (AIC) is a common statistical criteria used in selecting a model of best fit and it has had a fundamental impact in evaluating problems relating to statistical models [25-27]. The AIC was also computed in conjunction with the Generalized Linear Model in a study by Iamba and Waiviro [15] to correct the skewness in data distribution. Both AIC and BIC are helpful in comparing different statistical models [12, 28]. Since the response variables were not normally distributed (Shapiro-Wilk test: $p< 0.05$), GLM was fitted to analyse the non-linearity of data distribution [14].

The GLM differs from ordinary least squares (OLS), even with a normally distributed dependent variable and OLS with “manually transformed” data would lead to wrong parameter estimates [29]. The gamma is a two-parameter continuous distribution family over positive values [30]. The Gamma GLM is typically applied to models having right-skewed data [41]. According to Evans, Hastings [32] if Y is a random variable with shape parameter κ and scale parameter θ, then $E(Y) = \kappa\theta \equiv \mu$ and $V(\mu) = \kappa\theta 2 = \mu 2/\kappa$. Our study confirms the findings of Cundill and Alexander [30] where the negative binomial and gamma distributions captured severe skewness in the data sets relating to parasitology and entomology. Gamma distribution provided a better and more robust alternative estimator than other standard alternatives [41]. On the other hand, the Poisson distribution is used to model the number or count of events occurring within a given time interval [30]. The Poisson model did not fit well with the data set, which is evident by its high-scaled deviance (G^2). The Poisson model fits count data well and has been used to analyze abundance data of insects, weeds, and diseases [1].

The study found that flea beetle (*P. chrysocephala*) abundance was significantly lower in lemon grass-cabbage intercrop plots hence the leaf damathe ge while increase in leaf area was evident [14]. By nature, abundance is not normally distributed [33, 34] and is quantified by discrete variables therefore can be best described by the Poisson or negative binomial distributions [1]. It is noticeable that the choice of link function (transformation) can influence the statistical significance the and is the power of test. It also important to test whether the transformation has corrected the problem [35]. Nonparametric methods can be used as alternatives to parametric tests for analyses of abundance and incidence when the assumptions of ANOVA are violated [2, 3]. On the contrary, nonparametric approaches are less powerful than parametric methods since they are unable to analyse data from factorial designs and repeated measures [36-38]. GLMs provide a more robust analysis than parametric ANOVA and nonparameteric tests since they enable appropriate analyses of skewed frequency or binary data. Moreover, GLMs considers the properties of data from discrete distributions such as Poisson and negative binomial distribution (counts) and binomial distribution (proportions) [39, 40].

Conclusion

The use of residual deviance as Goodness of Fit and AIC as an estimator of model quality are two important criterion of Generalized Linear Model (GLM). Most ecological as well as entomological and parasitological data are non-normal and thereby requires GLM with appropriate families and canonical links to analyse the non-linearity. From this study, Gamma model was found to be a better model to analyse the data on cabbage flea beetle (*Psylliodes chrysocephala* L.). We thereby recommend future studies to take into the consideration the direction of skewness [41].

Acknowledgments

A special thankyou to Tonavu Yaubi for his assistance in collecting the data on flea beetle. And grateful for the technical and logistical support from technical officers of PNG University of Natural Resources and Environment (PNG UNRE). Acknowledgement also extends to Dr Mauricio González-Chang of Universidad Austral de Chile for his suggestions to explore insect-plant interactions using a generalized model.

References

1. Sileshi G. Evaluation of statistical models for analysis of insect, disease and weed abundance and incidence data. East African Journal of Sciences. 2007;1(1):1-9.
2. Sileshi G, Mafongoya, Incidence of Mesoplatys ochroptera Stål (Coleoptera: Chrysomelidae) on Sesbania sesban in pure and mixed species fallows in eastern Zambia. Agroforestry systems. 2002;56(3):225-231.
3. Sileshi G, Mafongoya P. Effect of rotational fallows on abundance of soil insects and weeds in maize crops in eastern Zambia. Applied Soil Ecology. 2003;23(3):211-222.
4. Sileshi G, *et al*. Spatial distribution and sampling plans for mesoplatys ochroptera (Coleoptera: Chrysomelidae) on Sesbania. Journal of economic entomology. 2002;95(2):499-506.
5. Lindsey JK. Applying generalized linear models. Springer Science & Business Media 2000.
6. Nelder JA, Wedderburn RW. Generalized linear models. Journal of the Royal Statistical Society: Series A (General). 1972;135(3):370-384.
7. Myers RH, *et al*. Generalized linear models: with applications in engineering and the sciences. Vol. 791. John Wiley & Sons, 2012.
8. Pierce DA, Schafer DW. Residuals in generalized linear models. Journal of the American Statistical Association. 1986;81(396):977-986.
9. Cameron AC, Windmeijer FA. An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of econometrics. 1997;77(2):329-342.
10. Osuji G, *et al*. Effect of akaike information criterion on model selection in analyzing auto-crash variables. Int J Sci Basic Appl Res. 2016;26(1): 98-109.
11. Li W, Nyholt DR. Marker selection by Akaike information criterion and Bayesian information criterion. Genetic Epidemiology. 2001;21(S1):S272-S277.
12. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research. 2004;33(2):261-304.
13. Yabeja JW. Statistical models for yield, pest abundance and disease incidence of tomato. Jomo Kenyatta University of Agriculture and Technology, 2013.
14. Iamba K, Yaubi T. Incorporating lemon grass (Cymbopogon citratus L.) and marigold (Tagetes erecta L.) as non-host barrier plants to reduce impact of flea beetle (Chaetocnema confinis C.) in cabbage (Brassica oleracea var. capitata L.). Acta Entomology and Zoology. 2021;2(1):95-101.
15. Iamba K, Waiviro M. Concentration effects of chili and ginger extracts against diamondback moth (Plutella xylostella L.) on round cabbage. Acta Entomology and Zoology. 2021;2(1):115-122.
16. James GM. Generalized linear models with functional predictors. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64(3):411-432.
17. Andrejiová M, Marasová D. Using the classical linear regression model in analysis of the dependences of conveyor belt life. Acta Montanistica Slovaca. 2013;18(2).
18. Hinde J. Compound Poisson regression models. in GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Springer, 1982.
19. Cepeda-Cuervo E, Corrales-Bosio M. Gamma regression models with the Gammareg R package. Departamento de Estadística, 2015.
20. Crawley MJ. The R book. John Wiley & Sons, 2012.
21. Panchal G, *et al*. Searching most efficient neural network architecture using Akaike's information criterion (AIC). International Journal of Computer Applications. 2010;1(5):41-44.
22. Sahoo RK, Sabat AK, Nayak RK, Sahoo LN. On a method of estimating variance of the product estimator. International Journal of Statistics and Applied Mathematics. 2021;6(6):16-23.
23. Myers RH, Montgomery DC. A tutorial on generalized linear models. Journal of Quality Technology. 1997;29(3):274-291.
24. Pregibon D. Goodness of link tests for generalized linear models. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1980;29(1):15-24.
25. Bozdogan H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika. 1987;52(3):345-370.
26. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike

information criterion (AIC) and the Bayesian information criterion (BIC). Psychological methods. 2012;17(2):228.

27. Yamaoka K, Nakagawa T, Uno T. Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. Journal of pharmacokinetics and biopharmaceutics. 1978;6(2):165-175.
28. Burnham KP, Anderson DR. A practical information-theoretic approach. Model selection and multimodel inference, 2002, 2.
29. Johnson PE. GLM with a Gamma-distributed dependent variable, 2016.
30. Cundill B, Alexander ND. Sample size calculations for skewed distributions. BMC medical research methodology. 2015;15(1):1-9.
31. Kikechi CB. On local polynomial regression estimators in finite populations. Int. J. Stats. Appl. Math. 2020;5:58-63.
32. Evans M, Hastings N, Peacock B. Statistical distributions. IOP Publishing, 2001.
33. Madden L, Hughes G. Plant disease incidence: distributions, heterogeneity, and temporal analysis. Annual Review of Phytopathology. 1995;33(1):529-564.
34. Garrett K, *et al*. New applications of statistical tools in plant pathology. Phytopathology. 2004;94(9):999-1003.
35. McArdle BH, Anderson MJ. Variance heterogeneity, transformations, and models of species abundance: a cautionary tale. Canadian Journal of Fisheries and Aquatic Sciences. 2004;61(7):1294-1302.
36. Shah D, Madden L. Nonparametric analysis of ordinal data in designed factorial experiments. Phytopathology. 2004;94(1):33-43.
37. Brunner E, Puri ML. Nonparametric methods in factorial designs. Statistical papers. 2001;42(1):1-52.
38. Turechek W. Nonparametric tests in plant disease epidemiology: Characterizing disease associations. Phytopathology. 2004;94(9):1018-1021.
39. Hughes G, Madden L. Some methods allowing for aggregated patterns of disease incidence in the analysis of data from designed experiments. Plant Pathology. 1995;44(6):927-943.
40. Collett D. Modelling binary data. CRC press, 2002.
41. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. Journal of health economics. 2005;24(3):465-488.
42. McCullagh P, Nelder JA. Generalized linear models. Routledge, 2019.