ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2521-5051 (Print)
ISSN: 2521-506X (Online)
CODEN: ASMCCQ

Acta Scientifica Malaysia (ASM)

DOI: http://doi.org/10.26480/asm.01.2024.11.14

RESEARCH ARTICLE

MODELLING ZERO-INFLATED OVER DISPERSED DENGUE DATA VIA ZERO-INFLATED POISSON INVERSE GAUSSIAN REGRESSION MODEL: A CASE STUDY OF BANGLADESH

Sukanta Chakraborty, Soma Chowdhury Biswas

Department of Statistics, University of Chittagong
**Corresponding Author E-mail: chakraborty.sukanta@cu.ac.bd*

ARTICLE DETAILS

Article History:

Received 10 February 2024
Revised 15 March 2024
Accepted 23 April 2024
Available online 28 April 2024

ABSTRACT

Bangladesh has been noted for experiencing some of the most susceptible dengue outbreaks in Asia; the country's location, dense population, and changing environment all play major roles in this. Determining the correlation between meteorological conditions and case count is critical for predicting about the characteristics of the DENV outbreak. Certain widely used models, such as the Poisson regression model or the negative binomial regression model, are insufficient to adequately predict dengue fever since many of these datasets are of the over-dispersed, long-tail variety, and zero-inflated. In this study, the Zero-inflated negative binomial regression model is compared with the Zero-inflated Poisson inverse Gaussian regression model. Depending on AIC and BIC criteria Zero-inflated Poisson inverse-Gaussian regression model is proposed. Then Zero-inflated Poisson inverse Gaussian regression model is used to model the dataset containing confirmed positive cases of dengue fever and seven meteorological variables. The proposed model shows that all the meteorological variables are significantly associated with the confirmed positive cases of dengue fever. That's why modeling a dengue-fever dataset with a Zero-inflated Poisson-inverse Gaussian regression model is suggested in this study.

KEYWORDS

Dengue fever, Zero-inflated Poisson regression model, Zero-inflated negative binomial regression model, Zero-inflated Poisson inverse Gaussian regression model, Maximum likelihood estimation.

1. INTRODUCTION

According to WHO, the RNA Dengue virus (DENV), which causes dengue, has an RNA genome of about 11 kb and is composed of three structural proteins: the envelope (E), the pre-membrane or membrane (prM/M), and the capsid (C). DEN-1, DEN-2, DEN-3, and DEN-4 are the four serotypes of Aedes female mosquitoes that are known to transmit the acute febrile virus that causes dengue fever. Every year, more than 390 million people worldwide experience this sickness, and 50% of the global population is at risk. Viral titers can be found using the gold standard RT-PCR, which helps with patient diagnosis and treatment planning. Clinical signs and symptoms of Dengue infection are similar to those of the flu and include fever, headache, myalgia, reduced platelet counts, and leukopenia. In the event of severe diseases, dengue hemorrhagic fever (DHF), dengue shock syndrome (DSS), and life-threatening situations are occasionally taken into consideration.

It was discovered through a review of earlier data that the climate, particularly temperature, precipitation, and humidity, had changed suddenly in the previous few years. The disparity in climate also led to an increase in dengue outbreaks and positive cases. Because Aedes aegypti's growth and life cycle are influenced by rainfall, temperature, humidity, and wind, either directly or indirectly, meteorological conditions are associated to dengue sickness. There have been reports that increased precipitation and temperatures have contributed to the spike in dengue cases. Such settings are advantageous for breeding places that promote vector growth as a result of an increase in human and vector interaction that enhances viral transmission from an infected person to a new person. Understanding the correlations between climatic variables and the number of dengue cases reported is important since the dengue-transmitting mosquito's multiplication depends on temperature, water availability, and other climatic parameters to complete its life cycle. For such applications, the Poisson regression (PR) model has been applied numerous times. For instance, using a Poisson regression model as the main model, the meteorological parameters that influenced the spread of dengue in the city of Colombo, Sri Lanka between 2010 and 2018 were investigated (Chandrakantha, 2019).

Using a Poisson regression model, the number of cases of dengue hemorrhagic fever in Medan was estimated also (Sinaga et al., 2021). Population density, the number of medical personnel, the number of medical facilities, area height, and average waste output were all taken into account by the authors as explanatory variables. A Poisson regression model with temperature and cumulative rainfall as explanatory variables to forecast the number of dengue fever cases in Bandung, West Java, Indonesia, between 2001 and 2016 was suggested (Mukhaiyar et al., 2022). Using data from 2000 to 2023 and GLM models, we thoroughly examined the association between meteorological factors and Dengue cases in this study. This study can guide fellow researchers and policymakers in tackling the Dengue outbreak.

2. MATERIALS AND METHODS

2.1 Dengue Cases and Meteorological Factors

Patients from eight divisions across Bangladesh were able to obtain

Quick Response Code	Access this article online	
	Website: www.actascientificamalaysia.com	DOI: 10.26480/asm.01.2024.11.14

Cite The Article: Sukanta Chakraborty, Soma Chowdhury Biswas (2024). Modelling Zero-Inflated Over Dispersed Dengue Data Via Zero-Inflated Poisson Inverse Gaussian Regression Model: A Case Study Of Bangladesh. *Acta Scientifica Malaysia, 8(1): 11-14.*

confirmed positive cases from the website of the Directorate General of Health Services (DGHS). This study employed the daily DENV new cases from..the..DGHS..website..(https://old.dghs.gov.bd/index.php/bd/home/81-english-root/5200-daily-dengue-status-report) between January 1, 2000, and January 31, 2023. Additionally, we collected meteorological data from the NASA website (https://www.nasa.gov/). In this study, we took into account the following variables: wind velocity (m/s), surface pressure (kPa), precipitation (mm), relative humidity (%), daily temperatures (∘ C), and dew point (∘ C) at a height of 2 m above ground level.

2.2 Statistical Modelling

The descriptions of the ZIPR, ZINBR, and ZIPIGR models can be found in Chapters 328 & 329 of NCSS Statistical Software as well as in a conference paper (Purhadi et al., 2023). Let $y = (y_1, \dots\dots, y_n)$ be a vector of data composed of the number of dengue-fever cases recorded in n months in a country, state, or city. Assume that the recorded value y_t is a realization of the random variable Y_t for $Y_t \in$,1,2,3 $\dots\dots$ }. In addition, assume that measurements of p explanatory variables are available, denoted by $X_1, \dots\dots\dots X_p$ that can be associated with mosquito reproduction and dengue transmission, and consequently also associated with the number of recorded cases of dengue. Consider x to be an $n \times (p+1)$ matrix in which the first column contains only values 1 and columns 2 to p + 1 are composed of the recorded measurements for variables X_1 to X_p respectively. Denote the t[th] line of x by $x_t = (1, x_{t1}, x_{t2}, \dots\dots\dots, x_{tp})$ for $t = 1,2, \dots\dots, n$.

2.3 Zero-inflated Poisson regression model

Suppose that for each observation, there are two possible cases. Suppose that if case 1 occurs, the count is zero. However, if case 2 occurs, counts (including zeros) are generated according to a Poisson model. Suppose that case 1 occurs with probability π and case 2 occurs with probability 1 - π. Therefore, the probability distribution of the ZIP random variable y_i can be written

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1-\pi_i)\exp(-\mu_i) & if\ j = 0 \\ (1-\pi_i)\dfrac{{\mu_i}^{y_i}\ \exp(-\mu_i)}{y_i!} & if\ j > 0 \end{cases}$$

Where, the logistic link function is given by $\pi_i = \frac{\lambda_i}{1+\lambda_i}$ and $\mu_i = \exp(\ln(t_i) + y_1x_{1i} + y_2x_{2i} + \dots + y_mx_{ki}$ and $\lambda_i = \exp(\ln(t_i) + y_1z_{1i} + y_2z_{2i} + \dots + y_mz_{mi}$.

The logistic component includes an exposure time t and a set of m regressor variables (the z's). Note that the z's and the x's may or may not include terms in common.

The logarithm of the likelihood function is

$$L = L1 + L2 - L3 \qquad (1)$$

Where

$$L1 = \sum_{(i:y_i=0)} \ln[\lambda_i + \exp(-\mu_i)]$$

$$L2 = \sum_{(i:y_i>0)} \{y_i \ln(\mu_i) - \mu_i - \ln(y_i!)\}$$

$$L3 = \sum_{i=1}^{n} \ln(1+\lambda_i)$$

An essential assumption of the Zero-inflated Poisson regression model is that the mean of the response variable is equal to the variance, a property known as equidispersion. However, dengue-fever data, in general, do not have this property. Therefore, the Zero-inflated Poisson regression model is not suitable for modeling such data, because the standard errors may be underestimated, leading to misleading inferences from the regression. For overdispersed data, an alternative is to consider the zero-inflated negative binomial regression model.

2.4 Zero-inflated negative binomial regression model

Suppose that for each observation, there are two possible cases. Suppose that if case 1 occurs, the count is zero. However, if case 2 occurs, counts (including zeros) are generated according to the negative binomial model. Suppose that case 1 occurs with probability π and case 2 occurs with probability 1 - π. Therefore, the probability distribution of the ZINB random variable y_i can be written

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1-\pi_i)g(y_i = 0) & if\ j = 0 \\ (1-\pi_i)g(y_i) & if\ j > 0 \end{cases}$$

Where, the logistic link function is given by $\pi_i = \frac{\lambda_i}{1+\lambda_i}$ and $\mu_i = \exp(\ln(t_i) + y_1x_{1i} + y_2x_{2i} + \dots + y_mx_{ki}$ and $\lambda_i = \exp(\ln(t_i) + y_1z_{1i} + y_2z_{2i} + \dots + y_mz_{mi}$.

The logarithm of the likelihood function is

$$L = L1 + L2 + L3 - L4 \qquad (2)$$

Where

$$L1 = \sum_{(i:y_i=0)} \ln[\lambda_i + (1+\alpha\mu_i]^{-\alpha^{-1}}$$

$$L2 = \sum_{(i:y_i>0)} \sum_{j=0}^{y_i-1} \ln(j + \alpha^{-1})$$

$$L3 = \sum_{(i:y_i>0)} \{-\ln(y_i!) - (y_i + \alpha^{-1})\ln(1+\alpha\,\mu_i) + y_i \ln(\alpha) + y_i \ln(\mu_i)\}$$

$$L4 = \sum_{i=1}^{n} \ln(1+\lambda_i)$$

When dealing with overdispersed and zero-inflated data, that is, when the response variable's variance exceeds its average, a Zero-inflated negative binomial regression (ZINBR) model is typically taken into consideration in statistical analysis. This method assumes that the response variable values follow a negative binomial distribution in their generation. This distribution is a hybrid of the Gamma and Poisson distributions. Similar to the ZIPR model, this method also uses a log-linear connection to relate the average value of the response variable to a set of p explanatory variables. But long-tailed datasets—that is, datasets with a small percentage of extremely large integer values that are far from the majority—cannot be adequately modeled by the ZINBR model.

As a result, we suggest modeling a dengue-fever dataset using the Zero-inflated Poisson-inverse Gaussian regression (ZIPIGR) model in place of the ZINBR model. Response-variable values in this model are thought to be produced using a Zero-inflated Poisson-inverse Gaussian distribution. This distribution is a hybrid of the inverse Gaussian and Poisson distributions. This distribution's primary benefit is that, compared to a negative binomial distribution, its wider range of skewness allows it to accurately represent overdispersed long-tail data. I also use a log-linear relationship for this model to connect the answer variable's predicted value to a collection of p explanatory variables.

2.5 Zero-inflated Poisson inverse Gaussian regression

The PIG distribution consists of two parameters that are λ (average) and τ (dispersion parameter), and π (zero inflation) is the parameter of ZI distribution, so the ZIPIG distribution consists of three parameters that are λ, τ and π The zero value on the response variable can from two possible states, that are Zero state or Poisson state. π or $(1-\pi)$ is a probability zero value of the response variable from the zero states or Poisson state. The probability function of $y_i \sim ZIPIG(\lambda_i, \tau_i, \pi_i)$ $where$ $\lambda_i > 0, \tau_i > 0$ is as follows:

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1-\pi_i)P(y_i = 0) & if\ j = 0 \\ (1-\pi_i)P(y_i) & if\ j > 0 \end{cases}$$

The probability function of Y if $y_i = 0$ can be written as

$$\Pr(y_i = 0) = \pi_i + (1-\pi_i)e^{\frac{1}{\tau_i}}\left(\frac{2}{\pi\tau_i}\right)^{\frac{1}{2}}(2\lambda_i\tau_i + 1)^{\frac{1}{4}}K_{-\frac{1}{2}}\left(\frac{1}{\tau_i}\sqrt{1+2\lambda_i\tau_i}\right)$$

The probability function of Y if $y_i > 0$ can be written as

$$\Pr(y_i = j) = (1-\pi_i)\frac{{\lambda_i}^{y_i}e^{\frac{1}{\tau_i}}}{y_i!}\left(\frac{2}{\pi\tau_i}\right)^{\frac{1}{2}}(2\lambda_i\tau_i + 1)^{\frac{(y_i-\frac{1}{2})}{2}}K_{y_i-\frac{1}{2}}\left(\frac{1}{\tau_i}\sqrt{1+2\lambda_i\tau_i}\right)$$

Where K is the modified Bessel function of the third kind.

The likelihood function of the ZIPIGR model is formed by two models including a model for zero response variables ($y_i = 0$) and a model for response variables that are not equal to zero ($y_i > 0$) can be written as follows:

$$L(\beta;\tau;\gamma) = \left(\prod_{\substack{i=1 \\ y_i=0}}^{p_0} P(Y_i = y_i)\right)\left(\prod_{\substack{i=1 \\ y_i>0}}^{p} P(Y_i = y_i)\right) = L_1(\beta;\tau;\lambda)\, L_2(\beta;\tau;\lambda)$$

Where,

$$L_1(\beta;\tau;\lambda) = \prod_{\substack{i=1 \\ y_i=0}}^{p_0}\left(\exp(x_i'\lambda_i) + \exp\left(\frac{1}{\tau_i} - \frac{1}{\tau_i}\sqrt{1 + 2\tau_i(1-\pi_i)\exp(x_i'\beta)}\right)\right)\frac{1}{1+\exp(x_i'\lambda_i)}$$

and

$$L_2(\beta;\tau;\lambda) = \prod_{\substack{i=1 \\ y_i>0}}^{p}\left(\frac{1}{1+\exp(x_i'\lambda_i)}\right)\left(\frac{((1-\pi_i)\exp(x_i'\beta))^{y_i}\exp\left(\frac{1}{\tau_i}\right)}{y_i!}\right)\left(\frac{2}{\pi\tau_i}\right)^{\frac{1}{2}}((1+2-\pi_i)\,\tau_i\exp(x_i'\beta))^{-\frac{\left(y_i-\frac{1}{2}\right)}{2}} K_{s_i}(z_i)$$

The maximum-likelihood estimates are the solutions of equations in 1, 2 & 3 when $\frac{\Delta L}{\Delta \beta_i} = 0$, for $i = 0, \ldots\ldots\ldots, p$. However, these equations do not have explicit analytic solutions. Therefore, we apply numerical methods to solve these equations. We can obtain the maximum-likelihood estimates $\hat{\beta}$ of the parameters β using the R software and the function *gamlss* (). The overdispersion test may be performed in the R software using the check_overdispersion () function of the *performance* package.

2.6 Application

In this section, we apply the ZIPR, ZINBR, and ZIPIGR models to a real data set containing the number of dengue fever cases recorded in Bangladesh from January 2000 (t = 1) to September 2023 (t = 286).

Dengue was first recorded in the 1960s in Bangladesh (then known as East Pakistan) and was known as "Dacca fever". Since 2010 cases of dengue appear to coincide with the rainy season from May to September and higher temperatures. Bangladesh's climate conditions are becoming more favorable for the transmission of dengue and other vector-borne diseases including malaria and chikungunya virus due to excessive rainfall, waterlogging, flooding, rise in temperature, and the unusual shifts in the country's traditional seasons.

The local government engineering department (LGED) is leading vector control activities including the elimination of mosquito breeding sites and larvicidal and adult mosquito control using different insecticides such as Temephos and Deltamethrin. The hospital-based surveillance system is actively collecting regular information from 57 hospitals in Dhaka city as well as all Upazila and district-level hospitals. Daily reports are also disseminated through the Health Emergency Operation Center (HEOC).

3. Results

Consider $y = (y_1, \ldots\ldots\ldots, y_n)$ to be the number of dengue-fever cases recorded in Bangladesh from January 2000 (t = 1) to September 2023 (t = 286). These measures are freely available on the website https://old.Dghs.gov.bd/index.php/bd/home/5200-daily-dengue-status-report and also can be obtained by emailing the authors of the present article.

Let x be a matrix of dimension $n \times 7$ composed of the recorded measures of the variables

X_1: Month of the year, coded from 1 to 12;

X_2: average wind velocity (m/s) at a level of 2 m height above ground level;

X_3: average daily temperature in the month (°C);

X_4: daily dew point (°C)

X_5: average humidity in the month (%)

X_6: rainfall in the month(mm)

X_7: surface pressure (kPa)

The recorded measures for variables X_2 to X_7 are freely available at https://www.nasa.gov/. Denote this dataset by D = (y, x), which is a matrix of dimension n × 8. The first column contains the recorded number of dengue fever cases in each of the 144 months considered in the study. Columns 2 to 8 contain the recorded values of the explanatory variables X_1 to X_7.

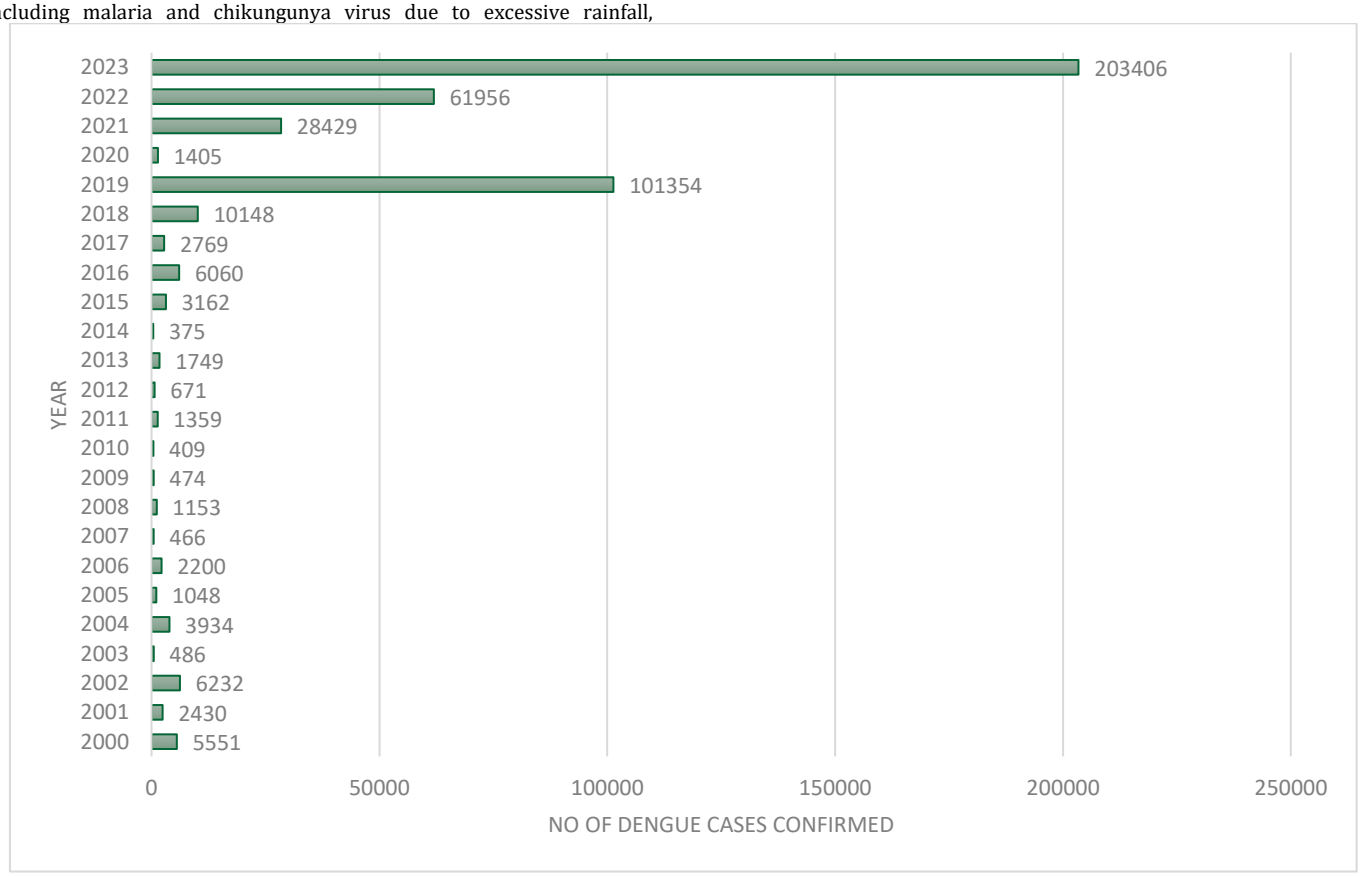

Figure 1: Number of recorded dengue-fever cases by year from 2000 to 2023.

Figure 1 shows the number of recorded dengue-fever cases from 2000 to 2023 which shows that the dengue outbreak in Bangladesh has taken a worrisome turn, with a worrying increase in cases and fatalities this year.

Table 1: Descriptive statistics of the recorded numbers of dengue fever cases.						
Minimum	1st Quartile	Median	Average	3rd Quartile	Maximum	Variance
0	0	44	1569	325	79598	61235680

Figure 2 shows the evolution of the number of dengue-fever cases by month for the period January 2022 to September 2023. Due mainly to the climate of the city, characterized by high heat and heavy rains from March to October, this period contains most of the recorded dengue-fever cases in the country. This fact shows the importance of having a model for projection for the number of dengue cases from environmental variables, to support actions to combat the proliferation of the mosquito and consequently the reduction of the number of cases.

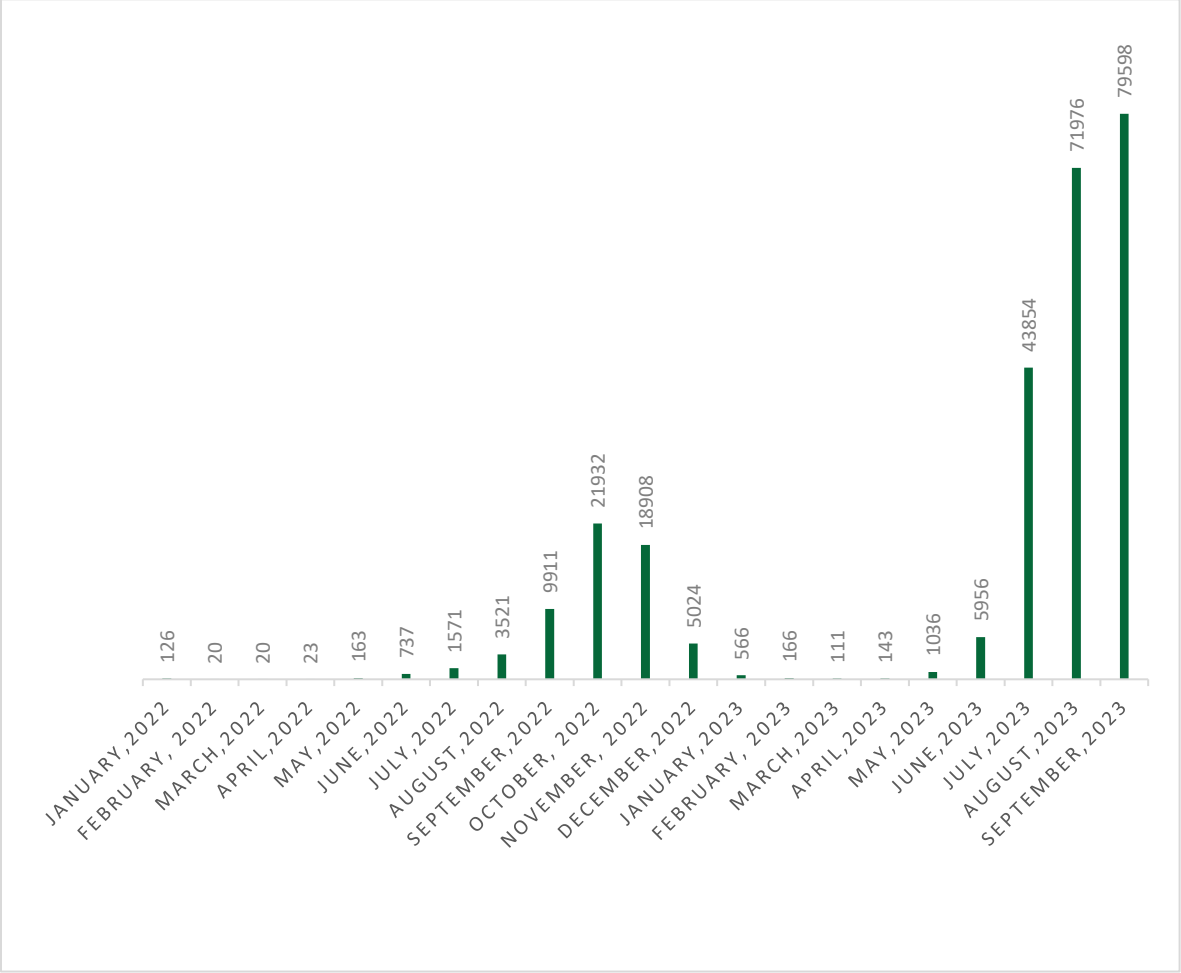

Figure 2: Evolution of the number of dengue-fever cases by month in the period considered (January 2022 to September 2023).

The descriptive statistics for the y values that were recorded between January 2000 and September 2023 are displayed in Table 1. In September 2023, the highest recorded number was 79598 instances. During the time under consideration, there were 1569 instances reported monthly on average. Table 2 Shows the correlations for each pair of variables.

Table 2: Correlations								
	Time	Wind velocity	Daily Temperatures	Dew Point	Relative Humidity	Rainfall	Surface Pressure	Dengue case
Time	1	-0.023	0.053	0.095	0.059	0.097	0.018	.415**
Wind velocity	-0.023	1	.769**	.650**	.363**	.688**	-.836**	-0.001
Daily Temperatures	0.053	.769**	1	.714**	.307**	.700**	-.839**	0.091
Dew Point	0.095	.650**	.714**	1	.870**	.856**	-.888**	.502**
Relative Humidity	0.059	.363**	.307**	.870**	1	.701**	-.637**	.629**
Rainfall	0.097	.688**	.700**	.856**	.701**	1	-.829**	.350**
Surface Pressure	0.018	-.836**	-.839**	-.888**	-.637**	-.829**	1	-.266**
Dengue case	.415**	-0.001	0.091	.502**	.629**	.350**	-.266**	1

Note: (*, **, *** Significant correlation)

3.1 Zero-inflation, over-dispersion checking and model comparison

For this data at first, we apply "*check_zeroinflation ()*" of "*Performance*" package in R for Poisson and Negative-binomial regression model. Figure 3 shows the output of the test in the R software. As the number of observed zeros is larger than the number of predicted zeros, the Poisson and Negative-binomial regression model are underfitting zeros, which indicates a zero inflation in the data. So, we can use Zero-inflated Poisson regression model, Zero-inflated negative binomial regression model & Zero-inflated Poisson inverse Gaussian regression.

```
> #Database
> D=data.frame(x1,x2,x3,x4,x5,x6,x7,y)
> # Poisson regression model
> PR.model=glm(y~1+x1+x2+x3+x4+x5+x6+x7,data=D,family=poisson())
> # Negative binomial regression model
> NBR.model=glm.nb(y~1+x1+x2+x3+x4+x5+x6+x7,data=D,init.theta = 1)
There were 27 warnings (use warnings() to see them)
> #checking zeroinflation
> check_zeroinflation(PR.model)
# Check for zero-inflation

   Observed zeros: 82
  Predicted zeros: 65
            Ratio: 0.79

Model is underfitting zeros (probable zero-inflation).
> check_zeroinflation(NBR.model)
# Check for zero-inflation

   Observed zeros: 82
  Predicted zeros: 72
            Ratio: 0.88

Model is underfitting zeros (probable zero-inflation).
```

Figure 3: Outputs of test for zero-inflation using the *check_zeroinflation ()* function

But the equidispersion property of Zero-inflated Poisson regression model is not satisfied here since dengue-fever data, in general, do not have this property. We fit ZIPR model using the *gamlss ()* function of the *gamlss* package of the R software. Moreover, we check the overdispersion of Zero-inflated Poisson regression model by using the *check_overdispersion ()* function of the R software. Figure 4 shows the output of the test in the R software.

```
> library(gamlss)
 #overdispersion test for Zero-inflated poisson
> ZIP.s=gamlss(y~pbc(x1)+pb(x2)+pb(x3)+pb(x4)+pb(x5)+pb(x6)+pb(x7),data=D,family=ZIP())
> check_overdispersion(ZIP.s)
# Overdispersion test

     dispersion ratio =   25503215.193
 Pearson's Chi-Squared = 7089893823.562
             p-value =        < 0.001

Overdispersion detected.
```

Figure 4: Outputs of the test for overdispersion using the *check_overdispersion ()* function.

The results described above indicate that the PR model is not appropriate for this dataset. Due to this, hereafter we fit the ZINBR and ZIPIGR models to the dataset and compare these two models according to the AIC and BIC model-selection criteria. The best model is the one that has the smallest AIC and BIC values.

We fit ZINBR and ZIPIGR models using the *gamlss ()* function of the *gamlss* package of the R software. Since the month variable has cyclical values, we fit both models by considering a cyclical P-spline term for this variable. For this, we use the *pbc ()* function inside the gamlss function. In addition, we fit both models by considering smooth terms for continuous variables. For this case, we use the *pb ()* function. We call the models fitted with *pb ()* function of ZINBR.S and ZIPIGR.S respectively

To significance level $\alpha = 0.01$, For all the fitted models, the variables are significant (p-values > α). Table 3 shows model-comparison criteria for the three fitted models. The smallest values are highlighted in boldface. Since the AIC and BIC values for the ZIPIGR are small, we opt to maintain the ZIPIGR as the best model because the smooth terms have not led to a significant improvement in the model.

Table 3: Model-comparison criteria.

Model	AIC	BIC
ZINBR	3883220.890	3883315.855
ZINBR.S	3042407.717	3042940.707
ZIPIGR	**3219.062**	**3304.254**
ZIPIGR.S	84619.867	85007.479

With the models fitted, it is important to perform a residuals analysis to identify the discrepancies between the models and the data, and to assess the overall model goodness-of-fit. In a normal linear regression scenario, the Pearson and deviance residuals are usually considered. However, these residuals are not suitable for problems in which the response variable is discrete because they are not normally distributed, and "have nearly parallel curves according to the distinct discrete response values, imposing great challenges for visual inspection" (Feng et al., 2020). To circumvent this issue, the use of randomized quantile residuals (RQR) was proposed (Dunn et al., 1996). According to the authors, this kind of residuals is particularly ideal for visualizing the goodness-of-fit of count regression models.

To calculate the RQR, we first need to obtain the cumulative distribution function, $F(y_t|\hat{\mu}_t, \hat{\tau})$ of the model considered, for $t = 1, \ldots \ldots, n$. For the continuous case, $F(.)$ values are uniformly distributed on interval (0, 1), and the RQR is defined as $r_t = \Phi^{-1}(F(y_t|\hat{\mu}_t, \hat{\tau}))$, where Φ $(\cdot)$ is the cumulative distribution function of the standard normal distribution. However, since the cumulative distribution function $F(.)$ for the models considered (ZINBR and ZIPIGR) is not strictly continuous, but a step function, a randomization is introduced to produce continuous normal residuals. Thus, to get the RQR, we use the above method. We obtained the RQR values for ZINBR and ZIPIGR models using the *residuals* function of the R software.

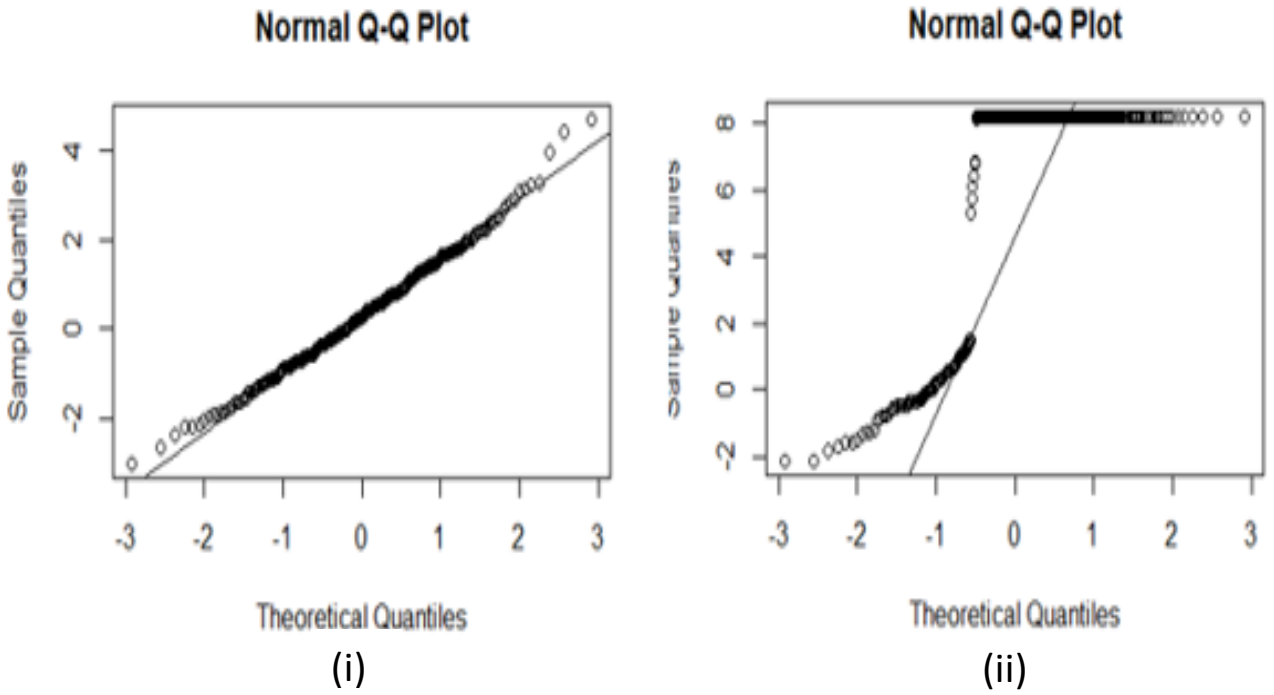

Figure 5 (a): Normal quantile-quantile plot for the residuals. (i) ZIPIGR model. (ii) ZINBR model.

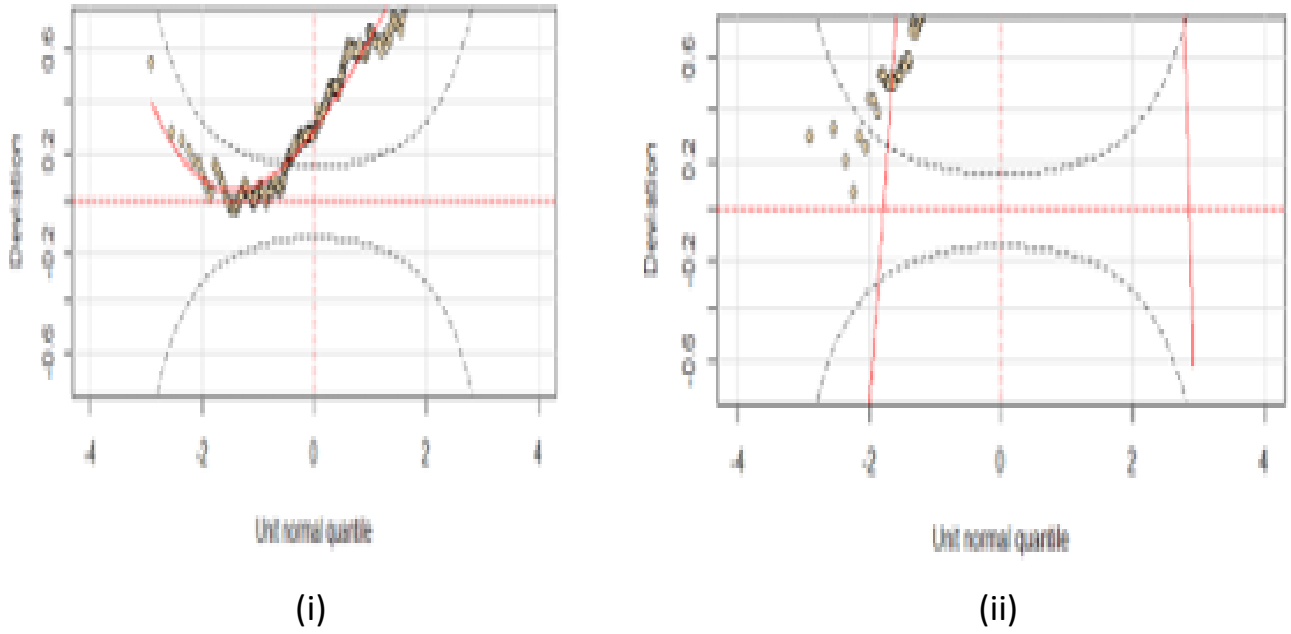

Figure 5 (b): Worm plot. (i) ZIPIGR model. (ii) ZINBR model.

Figure 5(a) shows the normal quantile-quantile plot (q-q plot) for the randomized quantile residuals of the ZINBR and ZIPIGR fitted models. Figure 5(b) shows the worm plot. This graph was proposed to identify regions (intervals) of the explanatory variable within which the model does not fit the data adequately (Buuren et al., 2001). In this graph, the upward line of the q-q plot is rotated to the horizontal to remove the trend and the Y axis contains the difference between its location in the theoretical and empirical distributions. If the residuals follow a normal distribution, then the Y values are near the horizontal line and consequently inside the confidence band. The R function *wp ()* provides the worm plot for a *gamlss* fitted model. As one can note, both figures indicate the ZIPIGR model performs better than the ZINBR model. In addition, the graphs of the residuals from the ZIPIGR model indicate that there is no reason to worry about the inadequacy of the fit.

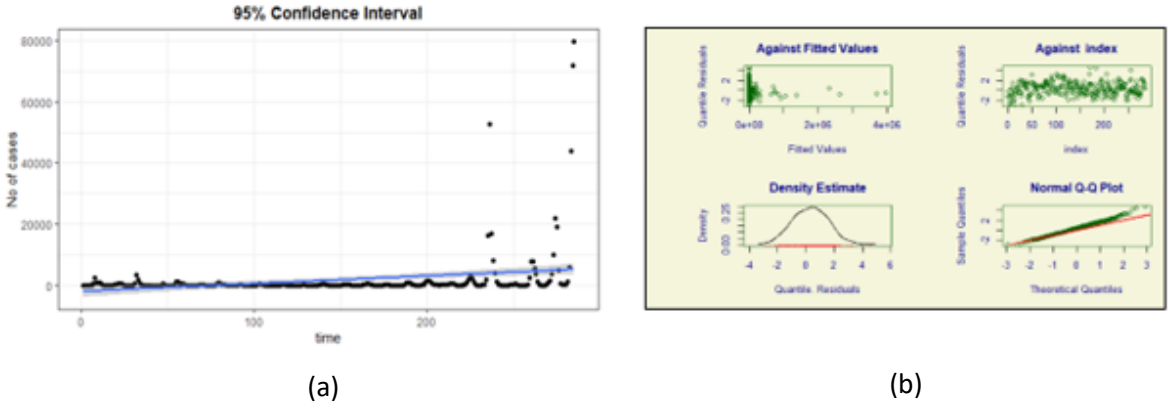

Figure 6: (a) Recorded values and confidence band (95%) generated from fitted model. (b) quantile residuals against fitted model

Figure 6 (a) shows the number of registered dengue cases (symbol •) and a confidence band of 95% generated from the fitted ZIPIGR model. As one can note, the fitted model indicates that every year a peak will occur (Saraiva et al., 2022). How high or low the recorded number of denguecases will be in relation to the expected peak (given by the fitted model) is controlled by action taken to combat the proliferation of the mosquito (Silva et al., 2008). Figure 6 (b) shows the quantile residuals against fitted model. Table 4 shows the estimates for the parameters of the ZIPIGR model.

Table 4: Estimates for parameters of ZIPIGR model

Parameters	Estimate	Std. Error	t value	Pr (> [t])
β_0	-170.85279	36. 51556	-4.679	4.63e - 06
β_2	-1.26211	0.36678	-3.441	0.000674
β_3	1.73122	0.67573	2.562	0.010969
β_4	-1.54311	0.72018	2.143	0.033064
β_5	0.75281	0.18546	4.059	6.51e - 05
β_6	-0.22219	0.03289	-6.756	9.18e - 11
β_7	1.08072	0.42203	2.561	0.011008
τ	4.4463	0.7682	5.788	2.04e-08
π	-3.9222	0.5868	-6.684	1.4e-10

4. Conclusion

Millions of people contract dengue fever every year, particularly in tropical countries, which has a significant negative influence on public health systems. As a result, there is interest in creating statistical models that can predict the number of instances of dengue fever and pinpoint the environmental factors that could be connected to the quantity of cases that are reported. This article presents statistical modeling for a dengue fever dataset that is zero inflated, long-tailed, and over dispersed. The underlying premise of the suggested modeling is that the number of dengue fever cases reported in a given month is distributed as a Zero-inflated Poisson inverse Gaussian distribution. his distribution may be used for modeling over dispersed, long-tailed datasets and presents a larger range of skewness than negative binomial distribution.

We use a log-linear function to explain how the expected number of dengue-fever cases is related to a collection of explanatory variables. A Zero inflated Poisson inverse Gaussian regression model (ZIPIGR) is the term used to describe this method. We use the maximum-likelihood approach to estimate the parameters of interest. We use the *gamlss ()* function from the *gamlss* package of the R software to acquire estimates numerically because the estimators lack known analytic answers.

We assess the differences between the standard methodology and the suggested modeling, which use a zero inflated negative-binomial regression (ZINBR) model. The AIC and BIC criteria were applied to compare the two models. We also present a comparison of the two models with respect to randomized quantile residuals.

For this application, the ZIPIGR model surpasses the ZINBR model according to two model-selection criteria. Additionally, the randomized quantile residuals show that ZIPIGR outperforms ZINBR. In other words, it has a worm plot with residuals close to the horizontal line and a quantile-quantile normal plot with residuals close to the line y = x. According to the fitted ZIPIGR model, there exists a relationship between all explanatory variables and the recorded number of cases of dengue fever. The number of dengue cases is strongly correlated with variables X_1, X_5, and X_7. That is, it is anticipated that a rise in the number of dengue cases reported will occur with increases in temperature, air humidity, and/or surface pressure.

This makes natural sense because the growth of the mosquito that spreads dengue disease is strongly correlated with these two characteristics. The female mosquito, when infected and exposed to temperatures of roughly 32 ◦ C, has 2.64 times more chance of finishing the incubation phase than when exposed to mild temperatures.

Ultimately, the fitted model predicts an annual peak in instances. Based on this finding, we hypothesize that human involvement is the only means of preventing the spread of the transmitting mosquito and avoiding the peak. The R program was used for all of the analyses, and the authors can be contacted via email to receive the full code.

We emphasize the following three aspects to wrap up the paper: While the suggested modeling has outperformed standard methods, it is based on the same fundamental assumptions as zero inflated Poisson and zero-inflated negative binomial regression models: log linearity in model parameters and individual observation independence; (ii) as previously mentioned, the primary benefit is the ability to model zero inflated & over dispersed long-tail datasets; (iii) On the other hand, because dengue fever data are longitudinally recorded, they might exhibit some type of temporal correlation. Therefore, additional work might be considered on developing a modeling strategy that takes into account the correlation among recorded values for the answer variable.

Declaration Of Conflicting Interests

The author declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Data Availability Statement

Data will be made available upon request.

References

Assessment of regression models performance [R package performance version 0.10.8] [Internet]. Comprehensive R Archive Network (CRAN); 2023 [cited 2023 Dec 30]. Available from: https://cran.r-project.org/web/packages/performance/index.html

Buuren, S.V., and Fredriks M., 2001. Worm plot: A simple diagnostic device for modeling growth reference curves. Stat Med., 20 (8), Pp. 1259–77. doi:10.1002/sim.746

Chandrakantha, L.,2019. Statistical analysis of climate factors influencing dengue incidences in Colombo, Sri Lanka: Poisson and negative binomial regression approach. Int. J. Sci. Res. Pub., 9 (2). doi:10.29322/ijsrp.9.02. 2019.p8616

Dengue - bangladesh [Internet]. World Health Organization; [cited 2023 Dec 30]. Available from: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON481

Dengue and severe dengue [Internet]. World Health Organization; [cited 2023 Dec 30]. Available from: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

Dunn, P.K., and Smyth, G.K., 1966. Randomized quantile residuals. J. Comput. Graph. Stat., 5 (3), Pp. 236–44. doi:10.1080/10618600.1996.10474708

Feng, C., Li, L., Sadeghpour, A., 2020. A comparison of residual diagnosis tools for diagnosing regression models for Count Data. BMC Med Res Methodol., 20 (1). doi:10.1186/s12874-020-01055-2

Instructions on how to use the GAMLSS package in R [Internet]. [cited 2023 Dec 30]. Available from:https://www.researchgate.net/publication/279233661_Instructions_on_how_ to_ use_ the_GAMLSS_package_in_R

Mukhaiyar, U., Huda, N.M., Andirasdini, I.G., 2022. Forecasting of dengue hemorrhages fever cases with autoregression distributed lag model using Poisson regression approach. Biostat. and Epidemiol., 7 (1). doi:10.1080/24709360.2022.2064636

CSS Statistical Software (https://www.ncss.com/wp-content/ themes /ncss/pdf/Procedures/NCSS/ZeroInflated_Negative_Binomial_Regression.pdfandhttps://www.ncss.com/wp-content/ themes/ncss/pdf/Procedures/NCSS/Zero-Inflated_Poisson_Regression.pdf)

Purhadi, E., and Rahayu, S.P., 2023. Parameter estimation and statistical test on zero inflated Poisson inverse gaussian regression model. The 8th International Conference And Workshop On Basic And Applied Science (Icowobas) 2021. https://doi.org/10.1063/5.0104166

R: A language and environment for statistical computing [Internet]. [cited 2023 Dec 30]. Available from: https://cran.r-hub.io/doc/manuals/fullrefman.pdf

Saraiva, E.F., Vigas, V.P., Flesch, M.V., 2022. Modeling overdispersed dengue data via Poisson inverse gaussian regression model: A case study in the city of Campo Grande, MS, Brazil. Entropy, 24 (9), Pp. 1256. doi:10.3390/e24091256

Silva, J., Mariano, Z., de, F., Scopel, I., 2008. The influence of urban climate on the proliferation of the Aedes aegypti mosquito in Jataí (GO), from the perspective of Medical Geography. Hygeia., 3 (5), Pp. 33–49. doi: https://doi.org/10.14393/Hygeia316883

Sinaga, J.P., Sinulingga, U., 2021. Poisson Regression Modeling Case Study Dengue Fever in Medan City in 2019. J Math Technol Educ., 1 (1), Pp. 94–102. doi: https:// doi.org/ 10.32734 /jomte. v1i1.7500

Srivastava, K., Purbey, S.K., Patel, R.K. and Nath, V., 2016. Managing Fruit-Borer for having Healthy Litchi. Indian Horticulture, 61 (3), Pp. 39-41.

Sultana, T., Uddin, M.M., Rahman, M.M. and Shahjahan, M., 2017. Host preference and eco-friendly management of cucurbit fruit fly under field condition of Bangladesh. Asian-Australasian Journal of Bioscience and Biotechnology, 2 (1), Pp. 55-59.

Taher, M.A., 2020. Development of an appropriate management strategy for litchi fruit borer, Conopomorpha sinensis bradley using non-chemical and chemical approaches. PhD thesis, Department of Entomology, Bangladesh Agricultural University, Bangladesh.

Taher, M.A., Rahman, M.M., Islam, K.S. and Uddin, M.M., 2023. Some significant threats in lychee production and their management options in Bangladesh. Arab Journal of Plant Protection, 41 (2), Pp. 114-118. https://doi.org/10.22268/AJPP-041.2.114118

Tuat, N.V., Son, N.H. and Liem, N.V., 2012. Research results on control of litchi fruit borer Conopomorpha sinensis. Retrievedfromhttp://www.Rauantoan.com/nd5/print/43.html.

Vercammen, J., and Schmitz, A., 2001. Chapter 20 Marketing and distribution: Theory and statistical measurement. Handbook of Agricultural Economics, 1, Pp. 1137-1181. https:// doi.org/10.1016/S1574-0072(01)10028-9.

Waite, G.K., and Hwang, J.S., 2002. Tropical fruit pests and pollinators: biology, economic importance, natural enemies and control: CABI Publishing, Wallingford, UK. Pp. 331-359.

Walter, J.F., 1999. Commercial experience with neem products: Methods in Biotechnology.5: Biopesticide (ed. by FR Hall, JJ Menn & N Totowa) Humana Press, Pp. 155-170.