International Journal of Agricultural Sciences
Volume 19 | Issue 1 | January, 2023 | 315-320 ■ ISSN : 0973-130X
DOI:10.15740/HAS/IJAS/19.1/315-320
Visit us : *www.researchjournal.co.in*

Research Paper

Forewarning the outbreak of Haemorrhagic Septicaemia disease by using suitable statistical models

H. T. Vinay[1], A. P. Jahnavi*, Nethravathi Ashok Patil[2] **and** A. Varalakshmi
Department of Agricultural Statistics, Applied Mathematics and Computer Application, University of Agricultural Sciences, GKVK, Bengalore (Karnataka) India (Email : jahnaviap0305@gmail.com)

Abstract : Haemorrhagic Septicaemia was considered as one of the lethal diseases for the bovines. Ingestion or inhalation were considered as the major sources of natural infection of the disease, but the mortality rate was high when casual agent introduced to non-endemic regions. This disease was observed as most deadly bacterial disease which was responsible for highest number of annual deaths of bovines. Three generalized linear models such as poisson regression, zero inflated poisson and zero inflated negative binomial models were fitted, among these models zero inflated models were found to be best fit and the risk factors *viz.,* land surface temperature, air temperature, potential evapotranspiration, rainfall and soil temperature were found to have significant effect in the outbreak of Haemorrhagic Septicaemia disease.

Key Words : Haemorrhagic Septicaemia, Poisson regression, Zero inflated poisson model, Zero inflated negative binomial model, Akaike information criterion, Akaike information criterion for correction

View Point Article : Vinay, H. T., Jahnavi, A.P., Patil, Nethravathi Ashok and Varalakshmi, A. (2023). Forewarning the outbreak of Haemorrhagic Septicaemia disease by using suitable statistical models. *Internat. J. agric. Sci.*, **19** (1) : 315-320, **DOI:10.15740/HAS/IJAS/19.1/315-320.**
Copyright@2023: Hind Agri-Horticultural Society.

Article History : Received : 28.09.2022; **Revised :** 24.11.2022; **Accepted :** 26.12.2022

Introduction

The animals do suffer from a wide range of infectious diseases. Amongst these, Haemorrhagic Septicaemia (HS), one of the lethal disease of cattle and buffaloes, caused by *Pasteurella multocida*type-I organism, is an important disease (Ataei *et al.,* 2009). It is a gram-negative coccoid, short rod or filamentous shaped organism. This disease is of great economic importance in South East Asia after the virtual eradication of Rinderpest from Asia (Bain *et al.,* 1982).The source of infective bacteria is thought to be the nasopharynx of bovine carriers. Natural infection occurs by ingestion or inhalation. The initial site of proliferation thought to be the tonsillar region. In susceptible animals, a septicemia develops rapidly, and death due to endotoxemia ensues within 8-24 hour after the first signs are seen depending on the severity of the disease. The mortality rate is high, when the agent is introduced to non-endemic regions. Losses vary widely in endemic areas due to exposure previously to bacterian resistance can be expected. Young growing cattle within the age group of six months to two years are most often affected. It is prevalent throughout the country and responsible for approximately 60 per

***Author for correspondence:**
[1]Department of Agricultural Statistics, Uttara Banga Krishi Viswavidyalaya, Pundibari, Coochbehar (W.B.) India
[2]Institute of Agribusiness Management. University of Agricultural Sciences, GKVK, Bengalore (Karnataka) India (Email: nethrauasgkvk@gmail.com)

cent bovine mortality Hemorrhagic Septicaemia is the most deadly bacterial disease which causes the highest number of annual deaths, resulting in huge economic losses (Singh *et al.,* 1996).

When the assumptions of linear regression models were violated, Generalized Linear Models were used to model the relationship between response and explanatory variables.A generalized linear model is an extension of a traditional or classical linear model. In Generalized Linear Models, the distribution of the response variable is any member of the exponential family. The non-linear link function is used to an established connection between the explanatory variables and the distribution of the response variable. Generally statistical models *viz.,* Poisson, Zero Inflated Poisson (ZIP) and Zero Inflated Negative Binomial Regression (ZINB) were used for the analysis of the data pertaining to forecasting the disease outbreak. Further, the models were evaluated using different methods like Akaike Information Criterion (AIC), Akaike Information Criterion for correction (AICC) and evaluation of the model parameters were done using Wald's Chi-square and t-statistic.Since the disease is endemic in certain pockets of the state, it is necessary to forecast the disease in these areas by employing appropriate statistical models.

Material and Methods

The data on the outbreak of Haemorrhagic Septicaemia (HS) in all districts of Karnataka for the period from 2006 to 2017 was obtained from the Department of Animal Husbandry and Veterinary Services, Government of Karnataka. For the analysis of the data, months were classified into three seasons pre-monsoon (March to May), Monsoon (June to November) and post- monsoon (December to February).

The risk factors for incidence of Hemorrhagic Septicaemia were classified into the following three categories:

Remote sensing variables :

Data on Land Surface Temperature and Normalized Difference Vegetative Index was collected using satellite sensors from Moderate Resolution Imaging Spectro-radiometer tools (MODIS tools) aboard the Terra satellite.

Meteorological variables:

Data on climatic variables like Rainfall, Air temperature and Relative Humidity were collected from Earth System Research Laboratory (ESRL), USA.

Soil parameters :

Data on Soil moisture was collected from the Earth System Research Laboratory (ESRL), USA. Information on soil pH of all the districts of Karnataka was collected from Soil fertility Atlas for Karnataka developed by Department of Agriculture, Government of Karnataka.

Poisson regression model :

A Poisson regression model is one of a generalized linear model, in which the response variable follows the Poisson distribution. The assumption made in the Poisson distribution model is equality of the conditional mean and the conditional variance *i.e.,* equi-dispersion of data.

The probability mass function of a Poisson distribution is :

$$P_{poi}\,(Y_i = y_i | x_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}\ ,$$

where,

y_i and x_i are dependent and independent variables

$i=0,1,2.....$

With mean and variance given by

$$E\,(y_i|x_i) = V(y_i|x_i) = \mu_i$$

Since the mean must be positive but the linear combination $\mu_i=\beta_0+\beta_1X_1+\beta_2X_2+...\beta_kX_k$ Can take on any value, we need to use a link Function for the parameter μ.

Natural Logarithm is the Standard link used for the Poisson distribution.

$\text{Log}\,(\mu) = \eta = \beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots \beta_kX_k$

So that $\mu = \exp(\eta)$ (Bret, 2007).

Zero Inflated Poisson model (ZIP) :

Zero Inflated Poisson model is used when the count data has excess of zeros and exhibits over dispersion. This model has two parts one is poisson count model and logit model for predicting excess zeros. If case 1 occurs of with probability of π then case 2 occurs with the probability of $1-\pi$. Response variable y takes the values 0, 1, 2 and so on. The probability mass function of Zero Inflated Poisson model is given by:

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

where,

Parameters $\pi_i = (\pi_1...\pi_n)$ and $\mu_i = (\mu_1, \mu_2 ... \mu_n)$

were assumed to satisfy

log (μ) = B β and log it (π_i) = log ($\pi_i i/(1-\pi_i i)$) = Gγ where B and G are matrices of explanatory variables.

Zero Inflated Negative Binomial model (ZINB) :

Zero Inflated Negative Binomial model was used when the count data has excess of zeros and exhibits over dispersion.The data follows combination of two distributions *i.e.,* Negative Binomial and logit distribution. Where Y takes non-negative integers: 0, 1, 2, 3, and so on. Suppose if each observation has two possible cases, if case 1 occurs with probability π and case two occurs with probability 1-π. The probability mass function for Zero inflated Negative Binomial model is given by:

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

where, π_i is the logistic link is function and g (γ_i) is the function of negative binomial given by:

$$g(y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}}\left(\left[\frac{\alpha\mu_i}{1+\alpha\mu_i}\right]\right)^{y_i}$$

where,

Parameters π_i = ($\pi_1 ... \pi_n$) and μ_i = ($\mu_1 1, \mu_1 2 ... \mu_1 n$) were assumed to satisfy log (μ) = Bβ and logit (π_i) = log ($\pi_i i/(1-\pi_i i)$)= Gγ where B and G are matrices of explanatory variables.

Goodness of fit for the fitted models :

The goodness of fit of models was evaluated using Pearson's Chi-Square test statistic, Akaike information criterion (AIC) and Akaike information criterion for correction of finite sample size (AICC). Lower values of AIC and AICC indicate a better fit of the models.

The Chi-Square Test Statistic (χ^2) for goodness of Fit is defined as : $\chi^2 = \sum \frac{[(O_i - E_i)]^2}{E_i}$

where,

O_i's are observed frequencies

E_i's are theoretical or expected frequencies for i = 1, 2… k.

The Chi Square Test Statistic (χ^2) follows Chi Square distribution with k-c degrees of freedom, where 'c' is the number of independent constraints.

Akaike Information Criterion (AIC) :

The Akaike Information Criterion (AIC) is model selection tool. It used to measure the relative quality of statistical model fitted to the observed set of data. Model with the lowest AIC is considered as "Best model" when compared to other models. AIC is given by the formula:

AIC = -2ln (L) + 2k

where, k is the number of fitted parameters, including intercepts and L is the maximum likelihood estimate for the model. Lowest AIC suggests the best fit model.

Akaike Information Criterion for correction (AICC):

For small sample sizes (approximated as n/k is less than 40 and k is the number of fitted parameters in the model), a modified version of AIC, AIC for correction (AICC) is suggested (Symonds *et al.*, 2011).

AICC is given by the formula :

$$AICC = AIC + \frac{2k(k+1)}{n-k-1}$$

where, n is the total number of observations used. Lowest AICC suggests the best fit model.

Estimated parameters for the fitted models were evaluated using Wald's 95 per cent confidence limits,Chi Square statistics, t statistic and z statistics.

Results and Discussion

The experimental findings obtained from the present study have been discussed in following heads :

Poisson Regression Model :

Poisson regression model was used to develop a model which can forewarn the outbreak of Haemorrhagic Septicaemia disease, since the outbreak was count.

The statistical packages such as python-3.6 and R studio were used to analyse the data and the outcome of the analysis were tabulated in Table 1 along with values of the parameter estimates along with their standard error and significance.

The fitted model for the disease outbreak was found to be y = 0.0246+1.1774 (Soil pH) +1.1543 (Air Temperature) + 1.0594 (LST) + 1.0091 (Relative humidity) + 0.9544 (Soil moisture) + 0.9130 (Rainfall) + 0.6350 (PET) + 0.5128 (NDVI). It can be inferred that the expected count of the disease outbreak increase by 1.1774, 1.1543, 1.0594, 1.0091, 0.9544, 0.9130, 0.6350 and 0.5128 units for every one unit increase in Soil pH, Air Temperature, Land Surface Temperature, Relative Humidity, Soil Moisture, Rainfall, PET and NDVI, respectively. The corresponding standard error and

Wald's 95 per cent Confidence Limits are provided in the Table 1.

The model indicated that the variables *viz.,* Soil pH, Air Temperature, Land Surface Temperature, Soil Moisture, Rainfall, PET and NDVI found to significant at 1 per cent level of significance, Relative humidity significant at 5 per cent level of significance and Wind Speed was found to be non-significant. From this it can be inferred that all the variables except wind speed contribute for the outbreak of disease. Further, the goodness of fit for the model developed was tested using the Pearson Chi-Square and the value was found to be 1.5050 which indicate over dispersion in the data. AIC and AICC were computed in order to compare this model with other similar models and values are presented in Table 2. Similar model were used by Suma (2017) to

Table 1: Parameter estimates of poisson regression model

Parameter	df	Estimate	Standard error	Wald 95%	Confidence limits	p value
Intercept	1	-3.7187 (0.0246)	0.7040	-5.0985	-2.3390	<.0001 **
LST	1	0.0577 (1.0594)	0.0151	0.0282	0.0872	0.0001 **
NDVI	1	-0.6678 (0.5128)	0.3241	-1.3031	-0.0326	0.0394 **
Air temperature	1	0.1435 (1.1543)	0.0346	0.0757	0.2112	<.0001 **
PET	1	-0.4541 (0.6350)	0.0529	-0.5578	-0.3505	<.0001 **
Relative humidity	1	0.0091 (1.0091)	0.0037	0.0017	0.0164	0.0158 *
Rainfall	1	-0.0910 (0.9130)	0.0221	-0.1343	-0.0477	<.0001 **
Soil moisture	1	-0.0466 (0.9544)	0.0116	-0.0694	-0.0238	<.0001 **
Wind speed	1	0.0257 (1.0260)	0.0575	-0.0870	0.1385	0.6546
Soil pH	1	0.1633 (1.1774)	0.0777	0.0111	0.3156	0.0355 **
Pearson Chi-Square				1.5050		

* and ** indicate significance of values at P=0.05 and 0.01, respectively (Value is the parenthesis are in the original scale)

Table 2: Criteria for assessing goodness of fit of the poisson model

Criterion	Values
AIC (smaller is better)	3575.4804
AICC (smaller is better)	3575.5318

Table 3: Parameter estimates of zero inflated poisson model

Parameter	df	Estimate	Standard error	Wald 95% confidence limits		p value
Intercept	1	-1.8511 (0.1571)	0.8595	-3.5358	-0.1665	0.0313 *
LST	1	0.0605 (1.0624)	0.0192	0.0229	0.0982	0.0016 **
NDVI	1	-0.6674 (0.4133)	0.4133	-1.4776	0.1427	0.1064
Air temperature	1	0.1250 (1.1331)	0.0444	0.0379	0.2121	0.0049 **
PET	1	-0.4084 (0.6647)	0.0648	-0.5353	-0.2814	<.0001 **
Relative humidity	1	0.0060 (1.0060)	0.0046	-0.0030	0.0149	0.1932
Rainfall	1	-0.0834 (0.9200)	0.0249	-0.1321	-0.0346	0.0008 **
Soil moisture	1	-0.9847 (0.3736)	0.3293	-1.6301	-0.3393	0.0028 **
Wind speed	1	0.0295 (1.0299)	0.0677	-0.1032	0.1622	0.6634
Soil pH	1	0.1218 (1.1295)	0.0951	-0.0645	0.3081	0.2002
Pearson Chi-Square				1.0028		

* and ** indicate significance of values at P=0.05 and 0.01, respectively (Value is the parenthesis are in the original scale)

Table 4: Criteria for assessing goodness of fit of the zero inflated poisson model

Criterion	Value
AIC (smaller is better)	3260.9991
AICC (smaller is better)	3261.0608

Table 5: Parameter estimates of zero inflated negative binomial model

Parameter	df	Estimate	Standard error	Wald 95% confidence limits		p value
Intercept	1	-1.8513 (0.1570)	0.8599	-3.5366	-0.1660	0.0313*
LST	1	0.0605 (1.0623)	0.0192	0.0229	0.0982	0.0016**
NDVI	1	-0.6675 (0.5129)	0.4135	-1.4779	0.1430	0.1065
Air temperature	1	0.1250 (1.1331)	0.0445	0.0379	0.2121	0.0049**
PET	1	-0.4084 (0.6647)	0.0648	-0.5354	-0.2814	<.0001**
Relative humidity	1	0.0060 (1.0060)	0.0046	-0.0030	0.0149	0.1934
Rainfall	1	-0.0834 (0.9199)	0.0249	-0.1321	-0.0346	0.0008**
Soil moisture	1	-0.9848 (0.3735)	0.3294	-1.6305	-0.3391	0.0028**
Wind speed	1	0.0295 (1.0299)	0.0677	-0.1033	0.1622	0.6634
Soil pH	1	0.1218 (1.1295)	0.0951	-0.0646	0.3082	0.2003
Pearson Chi-square				1.0021		

* and ** indicate significance of values at P=0.05 and 0.01, respectively (Value is the parenthesis are in the original scale)

Table 6: Criteria for assessing goodness of fit of the zero inflated negative binomial model

Criterion	Value
AIC (smaller is better)	3263.0664
AICC (smaller is better)	3263.1393

model the outbreak of Anthrax in Karnataka and the models were evaluated using AIC, AICC and goodness of fit was tested using Pearson Chi-square.

Zero Inflated Poisson (ZIP) :

This was used to develop a model which can forewarn the outbreak of Haemorrhagic Septicaemia disease, since the outbreak is count and the zero were found to be inflated in the data. This model was tried to see if the model fits well to the data. The values of the parameter estimates along with it its standard error and significance are presented in Table 3. Fitted model was found to be y =0.1571+1.1331 (Air temperature) + 1.0624 (LST) + 1.0060 (Relative humidity) + 0.9200 (Rainfall) + 0.6647 (PET) + 0.3736 (Soil moisture)

From the Table 3, it can be inferred that the expected count of the disease outbreak increase by 1.1331, 1.0624, 1.0060, 0.9200, 0.6647 and 0.3736 units for every one unit increase in Air Temperature, Land Surface Temperature, Relative Humidity, Rainfall, PET and Soil Moisture respectively. The corresponding standard error and Wald's 95 per cent Confidence Limits are presented in the Table 3.

The model indicated that the variables *viz.,* Air Temperature, Land Surface Temperature, Soil Moisture, Rainfall and PET found significant at 1 per cent level of significance and Relative humidity, Wind Speed, NDVI and Soil pH found non-significant. From this it can be inferred that the variables Air Temperature, Land Surface Temperature, Soil Moisture, Rainfall and PET contribute for the outbreak of disease. Further, the goodness of fit of the model developed was tested using the Pearson Chi-Square and the value was found to be 1.0028 which indicates the model is a good fit for the data. AIC and AICC were computed in order to compare this model with other similar models and values are presented in Table 3. Similar model was used by Ahmad *et al.* (2015) in case of pneumonia outbreak and the model was evaluated using AIC.

Zero Inflated Negative Binomial (ZINB) :

This was used to develop a model which can forewarn the outbreak of Haemorrhagic Septicaemia disease, since the outbreak is count and the zero were found to be inflated in ordered to check for over dispersion in the data. This model was tried to see if the model fits well to the data. Fitted zero inflated negative binomial model was observed as y =0.1570 +1.1331 (Air Temperature)+ 1.0623 (LST) + 0.9199 (Rainfall) + 0.6647 (PET)+0.3735 (Soil Moisture). The values of the parameter estimates along with it its standard error and significance are presented in Table 5.

From the Table 5, it can be inferred that the expected count of the disease outbreak increase by 1.1331, 1.0623, 0.9199, 0.6647 and 0.3735 units for every one unit increase in Air Temperature, Land Surface Temperature, Rainfall, PET and Soil Moisture, respectively. The corresponding standard error and Wald's 95 per cent Confidence Limits are given in the Table 5.

The model indicated that the variables *viz.,* Air Temperature, Land Surface Temperature, Soil Moisture, Rainfall and PET found significant at 1 per cent level of significance and Relative humidity, Wind Speed, NDVI and Soil pH found non-significant. From this it can be inferred that the variables Air Temperature, Land Surface Temperature, Soil Moisture, Rainfall and PET contribute for the outbreak of disease. Further, the goodness of fit of the model developed was tested using the Pearson Chi-Square and the value was found to be 1.0021 which indicates the model is a good fit for the data. AIC and AICC were computed in order to compare this model with other similar models and values are presented in Table 6. Ahmad *et al.* (2015) used Zero inflated Negative Binomial model in case of pneumonia outbreak and the model was evaluated using AIC.

Conclusion :

Among the different models fitted, Zero inflated models were performed to be better in forewarning the outbreak of Haemorrhagic Septicaemia disease with a smaller AIC and AICC values. The risk factors such as land surface temperature, air temperature, potential evapotransipiration, rainfall and soil moisture were found to played a significant role in the outbreak of the disease.

REFERENCES

Ahmad, A. W., Muhamad, W., Aisyah, A. S., Kasypi, M., Azlida, A. N., Halim, N. and Zalila, Ali (2015). Applications of zero inflated models for health sciences data. *J. Adv Sci. Res.*, **6** (2) : 39-44.

Ataei, S., Burchmore, R., Christopher, H. J., Finucane, A., Parton, R. and Coote, J.G. (2009). Identification of immunogenic proteins associated with protection against Haemorrhagic Septicaemia after vaccination of calves with a live-attenuated aroA derivative of Pasteurella multocida B:2. *Res. Vet. Sci.*, **87** (2) : 207-210.

Bain, R. V. S., Deauxis, M.C.L., Carter, G. R. and Gupta, B. K. (1982). Haemorrhagic Septicaemia, *Animal Production & Health,* pp. 33.

Bret, L. (2007). Poisson Regression. *Statistics*. 572.

Singh, V. P., Kumar, A. A., Srivastava, S. K. and Rathore, B. S. (1996). Significance of HS in Asia: India. International workshop on diagnosis and control of HS. Bali, Indonesia, Indonesian Department of Agriculture, 28- 30 May 1996, pp.16.

Suma, A. P., Suresh, K. P., Gajendragad, M. R. and Kavya, B. A. (2017). Outbreak prediction of anthrax in Karnataka using Poisson, Negative-Binomial and Zero Truncated Models. *Int. J. Sci. & Res.*, **6** (3) : 505-509.

Symonds, M. R. E. and Adnan Moussalli (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav Ecol Sociobiol.,* **65**: 13-21.

19th Year
★ ★ ★ ★ ★ of Excellence ★ ★ ★ ★ ★