Analyzing the leptospirosis incidence and its association with climatic factors in Sri Lanka

1 Introduction

Leptospirosis is the world’s most important emerging bacterial zoonotic disease caused by pathogenic Leptospira species. The pathogenic Leptospira's survival depends on the soil moisture, humidity, temperature, and surface water (Ehelepola, Ariyaratne and Dissanayake, 2019). Leptospirosis is known to have an association with climatic factors, and thus it was a suitable case for developing predictive statistical models with the association of the climatic factors. Several studies have assessed the effects of different climatic factors (e.g., temperature, rainfall, humidity) on leptospirosis transmission and incidence (Ashford *et al.*, 2000)(Chadsuthi *et al.*, 2012b).

The main objective of this study is to determine the relationship between climatic factors and the leptospirosis cases in Sri Lanka and develop a statistical model to predict the future leptospirosis cases. In this paper, the emerging areas of leptospirosis were investigated by using a spatial context.

2 Materials and Methods

The data set consisted of 1836 observations with nine variables: Leptospirosis cases, rainfall, temperature, humidity, cultivation area, population density, year, month, and district. Leptospirosis cases were obtained from Epidemiology department, climatic data were obtained from meteorological department, Population and agricultural area values were obtained from department of census and statistics. The district shapefile of Sri Lanka was extracted from the GDM online database, version 2.5.

Under the descriptive analysis, firstly the leptospirosis cases, rainfall, temperature, and humidity were only subjected to outlier tests and adjustments. The outliers were checked by plotting the Boxplots, and the adjustments were made by the “Winsorization” method. Then the normality was checked for the variables leptospirosis cases, rainfall, temperature, and humidity. The probability density plots and Shapiro Wilk statistic were used to check the normality of the data.

The leptospirosis incidence values were calculated as it needed for the hot spot analysis.

A Spearman’s rank correlation test was conducted on the dependent (leptospirosis cases) and independent (climatic factors) variables to determine a relationship. The Spearman’s correlation test was used because the independent and dependent variables are not normally distributed. Seventeen districts were considered here.

A hotspot analysis was conducted to find the hot and cold spots for leptospirosis in Sri Lanka. The analysis was done by ArcGIS 10.7 version.

Finally, generalized linear mixed models (GLMM) for the relationship between leptospirosis cases and climatic factors were constructed using different likelihood approximation methods namely

Laplace, penalized quasi and adaptive Gaussian quadrature. The variable month and year used as random effects & district as both fixed and random. All the meteorological variables (rainfall, temperature & humidity), population density, and cultivation area were considered as fixed variables. The modeling was done using the Poisson model with a log link function in this study. Four different Generalized linear mixed models were constructed for each likelihood estimation method by changing the fixed and random effects. Next, each estimation method's best model was selected by considering the predictive accuracy value of the models and performed a residual analysis by plotting the residuals against the predicted values' exponent for the best models. Then, the selected models were validated using accuracy evaluation criteria by Lewis(Lewis, 1982).

3 Results and Discussion

3.1 Descriptive Analysis

The probability density plots of Rainfall, leptospirosis patients, Temperature and Relative humidity don't have a bell-shaped curve. So, all the considered variables are not normally distributed. The results of the Shapiro-Wilk test also further confirm the results of the above plots. According to the box plots obtained, there are 152 outlier's leptospirosis cases, 71 total outliers for rainfall, 145 outliers for temperature, and 30 outliers for humidity. The outliers for each variable was significantly reduced after using the winsorization method.

3.2 Correlation Analysis

The results didn't have a clear statistically significant correlation pattern between climatic factors, namely the rainfall, mean temperature, and mean humidity, with district-level Leptospirosis incidence. However, High incident districts have shown a better correlation than low incidence districts.

3.3 Hot Spot Analysis

When considering the overall hotspot results, all the years have at least one hot spot, but only three years have cold spots. From 2010 to 2017, hot spots were gathered in Western, Sabaragamuwa, Southern, and North Western province districts. The hotspots were situated in the Northern province in 2018.

When considering the cold spots, they were gathered into the Northern province only. So, the Leptospirosis threat in the Northern province was lesser than in the other regions, but in 2018 the hot spots were shifted into the Northern province.

4 Analysis using Generalized linear mixed models

The final best models and their accuracy values using three likelihood estimator methods were displayed below.

Table 4:1 Estimation methods with best model equation and predictive accuracy values.

Estimation method	Model equation	Accuracy (%)
Adaptive Gaussian Quadrature	$Log(\mu_{ijk}) = +0.0936(s.Humidity)_{ijk} + 0.7793(Anuradhapura) + 0.6491(\text{Hambanthota}) + 0.6483(\text{Monaragala}) + 0.657(\text{Polonnaruwa}) + 0.8913(\text{Rathnapura})$	81.008
Penalized Quasi	$Log(\mu_{ijk}) = -0.1703 + 0.0232(s.Rainfall)_{ijk} + 0.1(s.Humidity)_{ijk} - 0.1353(s.Temperature)_{ijk} + 0.2823(s.pop\_density)_{ijk} + 0.1639(s.area)_{ijk}$	81.156
Laplace	$Log(\mu_{ijk}) = -0.15926 + 0.13064(s.Humidity)_{ijk} + 0.23783(s.pop\_density)_{ijk} + 0.28291(s.area)_{ijk}$	81.365

All the above accuracy values are approximately the same. But according to the results of the residual analysis the model with adaptive gaussian quadrature likelihood estimator method is the best model. The graphical results (residual vs predicted) of the residual analysis is shown below.

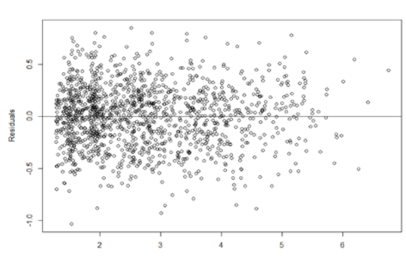

Figure 3.1: Graph of the adaptive Gaussian quadrature likelihood estimation method model residuals vs. exp (predicted value)

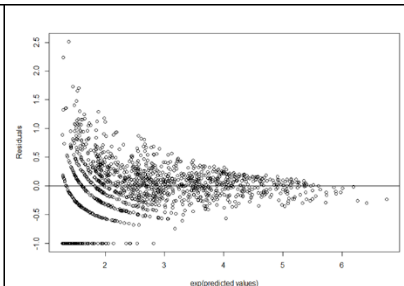

Figure 3.2:Graph of the penalized quasi-likelihood estimation method model residuals vs. exp (predicted value)

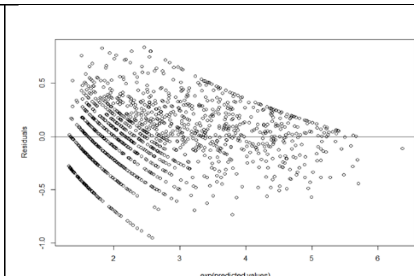

Figure 3.3: Graph of the Laplace likelihood estimation method model residuals vs. exp (predicted value)

Only the figure 3.1 has the random pattern. Therefore, it's the best fitted model out of all the models.

5 Model Validation

According to table 2, the sMAPE values are between 10% to 20% (Accuracy in between 80%-90%). Therefore, the predictive accuracy of the above models has Good forecasting.

6 Conclusions

In general, the present study did not find a clear statistically significant correlation pattern between climatic factors, with district-level Leptospirosis incidence. Western, Sabaragamuwa, and Southern province districts are most likely to be leptospirosis hotspots except in 2018. In 2018, hotspots were shifted to the Northern province. So, leptospirosis prevention precautions should be initiated in those districts. The cold spots were only significant in 2011, 2012, and 2013 in the Northern province. The models built by using all the methods had approximately the same predictive accuracy values. According to the models' residual analysis, the model with the adaptive Gaussian quadrature likelihood estimation method had the best performance. The model with the adaptive Gaussian quadrature likelihood estimation method is considered as the best model to predict future leptospirosis cases. Resources could be allocated effectively to control the leptospirosis incidence by using the model developed.

7 References

Ashford, D. A. *et al.* (2000) 'Asymptomatic infection and risk factors for leptospirosis in Nicaragua', *American Journal of Tropical Medicine and Hygiene*, 63(5–6), pp. 249–254. doi: 10.4269/ajtmh.2000.63.249.

Chadsuthi, S. *et al.* (2012a) 'Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses', *Asian Pacific Journal of Tropical Medicine*. Hainan Medical College, 5(7), pp. 539–546. doi: 10.1016/S1995-7645(12)60095-9.

Chadsuthi, S. *et al.* (2012b) 'Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses', *Asian Pacific Journal of Tropical Medicine*, 5(7), pp. 539–546. doi: 10.1016/S1995-7645(12)60095-9.

Ehelepola, N. D. B., Ariyaratne, K. and Dissanayake, W. P. (2019) 'The correlation between local weather and leptospirosis incidence in Kandy district, Sri Lanka from 2006 to 2015', *Global Health Action*. Taylor & Francis, 12(1). doi: 10.1080/16549716.2018.1553283.

Lewis, C. (1982) 'International and Business Forecasting Methods Butterworths: London'.

Pappachan, M. J., Sheela, M. and Aravindan, K. P. (2004) 'Relation of rainfall pattern and epidemic leptospirosis in the Indian state of Kerala [1]', *Journal of Epidemiology and Community Health*. BMJ Publishing Group Ltd, 58(12), p. 1054. doi: 10.1136/jech.2003.018556.