



Time series analysis of leishmaniasis incidence in Sri Lanka: evidence for humidity-associated fluctuations

Tharaka Wijerathna¹ · Nayana Gunathilaka¹

Received: 21 September 2021 / Revised: 25 June 2022 / Accepted: 4 November 2022 / Published online: 15 November 2022
© The Author(s) under exclusive licence to International Society of Biometeorology 2022

Abstract

Leishmaniasis is a vector-borne disease of which the transmission is highly influenced by climatic factors, whereas the nature and magnitude differ between geographical regions. The effects of climatic variables on leishmaniasis in Sri Lanka are poorly investigated. The present study focused on time-series analysis of leishmaniasis cases reported from Sri Lanka with selected climatic variables. Variance stabilized time series of leishmaniasis patients of entire Sri Lanka and major districts from 2014 to 2018 was fitted to autoregressive integrated moving average (ARIMA) models. All the possible models were generated by assigning different values for autoregression and moving average terms using a function written in R statistical program. The top ten models with the lowest Akaike information criterion (AIC) values were selected by writing another function. These models were further evaluated using RMSE and MAPE error parameters to select the optimal model for each area. Cross-autocorrelation analyses were performed to assess the associations between climate and the leishmaniasis incidence. Most associated lags of each variable were integrated into the optimal models to determine the true effects imposed. The optimal models varied depending on the area. SARIMA (0,1,1) (1,0,0)₁₂ was optimal for the country level. All the forecasts were within the 95% confidence intervals. Humidity was the most notable factor associated with leishmaniasis, which could be attributed to the positive impacts on sand fly activity. Rainfall showed a negative impact probably as a result of flooding of sand fly larval habitats. The ARIMA-based models performed well for the prediction of leishmaniasis in the short term.

Keywords Leishmaniasis · Modelling · Time series · ARIMA · Climate · Temperature · Humidity · Sri Lanka

Introduction

Leishmaniasis is an infectious disease caused by Protozoan parasites in the genus *Leishmania*, which transmits by the bite of the infected female sand flies. This disease is more prevalent in developing countries in the tropical region including India, Bangladesh, Iran, Brazil, Afghanistan, Algeria and Ethiopia (WHO 2022). Leishmaniasis has three main clinical manifestations, namely, cutaneous leishmaniasis (CL), mucocutaneous leishmaniasis (MCL), and visceral leishmaniasis (VL). Among these, CL is considered the least complex, but the most common type (CDC 2022). In Sri Lanka, all three clinical forms of the disease have been

reported, whereas the CL is the predominant manifestation (Wijerathna et al. 2017).

Outbreaks in infectious diseases have interactions with spatial and temporal dimensions (Sattenspiel 2009). It is known that spatial heterogeneity is subject to differences in the distribution of vectors and other risk factors. Further, it influences shifting the patterns of vector parasite interactions, vector-host contact, and susceptibility of the population (Werneck 2008). In Sri Lanka, climatic conditions have been proven to affect the occurrence of vector-borne diseases such as Malaria and Dengue (Briët et al. 2008; Ehelepola et al. 2015).

Some studies have evaluated the spatial and temporal distribution of leishmaniasis cases with associated risk factors (Assunção et al. 2001; Shimabukuro et al. 2010; Gálvez et al. 2011; de Souza et al. 2012; Karagiannis-Voules et al. 2013; Gomez-Barroso et al. 2015). Research from other countries shows both positive and negative correlations between climate factors and disease incidence (Toumi et al. 2012; Rosales et al. 2017; Sharafi et al. 2017). All these

✉ Nayana Gunathilaka
n.gunathilaka@kln.ac.lk
Tharaka Wijerathna
tharaka.wijerathna90@gmail.com

¹ Department of Parasitology, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

studies suggest that the incidence of leishmaniasis is influenced by climatic variables, but the nature and magnitude of these effects may differ from one geographical region to another. Therefore, prediction approaches developed for different geographical regions or countries are needed to achieve a better outcome in disease forecasting.

Despite new developments in disease control and advanced treatment methods, leishmaniasis is still one of the most prevalent tropical diseases in the world (WHO 2022). Leishmaniasis disease control programs mainly depend on the availability of knowledge on various aspects of the disease such as vector ecology, human-vector-parasite interactions, epidemiology, risk factors, disease trends, and possible climate dependencies (Wijerathna et al. 2017). Hence, the predictions through time series analyses are important to identify trends and possible disease outbreaks, which may ultimately facilitate control programs through proper precautionary interventions (Huang et al. 2011).

The autoregressive integrated moving average (ARIMA) is a linear model used in time series forecasting (Brockwell et al. 2002). The ARIMA models are usually denoted as ARIMA(p, d, q), where p , d , and q are integers. The parameter p is the order (number of time lags) of the autoregressive model, d is the degree of differencing or the number of times the past values of each reading of the data set should be subtracted to achieve the stationarity, and q is the order of the moving-average model (Brockwell et al. 2002). If the data set indicates a seasonal pattern, a seasonal component must be included in the model. This includes the order of the seasonal autoregressive model (P), the degree of seasonal differencing (D), and the order of the seasonal moving average model (Q). The values of p , d , q , P , D , and Q are determined based on the autocorrelation and partial autocorrelation plots and subsequent validation methods.

The ARIMA-based models have been used to predict the prevalence of various infectious diseases with high accuracy in the recent past (Ture and Kurt 2006; Huang et al. 2011; Liu et al. 2011; Sharafi et al. 2017). Climatic factors are also known to influence the predictive power of such models significantly (Huang et al. 2011; Roger et al. 2013; Anwar et al. 2016). The development of accurate prediction systems and identification of climatic factors associated with a disease incidence are critical for a developing country like Sri Lanka to plan and implement control activities in a cost-effective manner. Recent studies using spatio-temporal modelling (Karunaweera et al. 2021) and univariate time series analyses (Galgamuwa et al. 2018) with relatively low data points have added significantly to the knowledge of leishmaniasis progression so far and how much it will be in the future. However, further studies using different approaches such as multivariate time series analysis with a relatively long period of data may uncover different future scenarios. Furthermore, ARIMA modelling with external regressors will

reveal any lagged climatic dependencies of leishmaniasis in Sri Lanka. The present study was conducted to develop an ARIMA-based multivariate time series model for the short-term forecasting of the leishmaniasis disease incidence and to evaluate climatic factors associated with leishmaniasis incidence in Sri Lanka.

Materials and methods

Study area

Sri Lanka ($6^{\circ} 56' N 79^{\circ} 52' E$) is one of the endemic countries for cutaneous leishmaniasis in the world. It covers 65,610 km² land area. Approximately, 21 million inhabitants live in the country (Department of Census Statistics 2019). The country has tropical warm climatic conditions with a mean temperature ranging from 17 to 33 °C (Department of Meteorology 2019). The country receives the highest rainfall during the monsoon periods. Wet zone areas (Southern, South Western and Central parts of the country) receive rainfalls as high as 2,500 mm, whereas dry zone (Northern, Eastern and South Eastern parts) receive rainfalls from 1200 to 1,900 mm (Department of Meteorology 2019).

Patient information

Recorded cases of leishmaniasis patients from each district were retrieved by assessing the open-source records available at the Epidemiology Unit, Ministry of Health, Sri Lanka.

Meteorological data

Monthly mean values of the six climatic parameters (rainfall, average temperature, minimum temperature, maximum temperature, and relative humidity) from January 2014 to December 2019 were recorded at agrometeorological stations in Sri Lanka which were obtained from the data repository at the Department of Meteorology, 383, Bauddhaloka Road, Colombo 07, Sri Lanka. Monthly climatic data surfaces were created using kriging-based spatial prediction and the average values for selected climatic factors were extracted for each month.

Diagnosis criteria

The diagnosis of the disease is done only for patients seeking medical attention for their symptoms. No program is conducted for active case detection at the early stages of infection (Wijerathna et al. 2018). The diagnosis is mainly achieved through microscopic observations of GIEMSA-stained smears, while advanced techniques such as polymerase chain reaction (PCR), enzyme-linked immunosorbent

assay (ELISA), and other molecular diagnostic facilities are limited (Karunaweera et al. 2007; Siriwardana et al. 2010; Wijerathna et al. 2018).

Statistical analysis

The variances introduced by trends of the number of patients' time series were first stabilized using log transformation. The autocorrelation function (ACF) and partial autocorrelation function (PACF), which are measures of correlation between an observation and its past values of a time series, were plotted to determine whether there is a seasonality in the data set, indicated by similar values at regular intervals (For instance, higher number of patients after every 12 months, which is indicated by the tall bars of correlation plots). Seasonality was also modelled, where the seasonality is present in the time series.

In order to determine the p , d , and q values, initially, the stationarity of the time series was evaluated using the Augmented Dickey-Fuller (ADF) test (Said and Dickey 1984) to check whether the time series changes with time. If results indicated a non-stationarity pattern, the first difference of the time series was tested. This was repeated until the time series is stationary. The number of differences required to make the data set stationary was taken as the value for d (the integration component of the model), and seasonal difference values were used to determine D (the seasonal counterpart of d). The PACF and ACF plots were used to get multiple possible estimations for p , P , q , and Q values based on the number of lines that exceed the correlation line in ACF and PACF plots.

The effects of climatic variables on the number of patients were first analyzed using Spearman's correlation. Furthermore, the cross-autocorrelation analyses were performed for each climatic variable to check how long it takes for a change in climate conditions to exert a change in the number of patients (the lag with the highest correlation coefficient value). The lagged time series of each selected environment parameter was generated. After stabilizing the variance (deviation from the mean value, which differs with time if not stabilized) by log transformation, the remaining irregular part was used as an external regressor in the ARIMA model. Akaike information criterion (AIC) was used to compare the goodness-of-fit after integrating climatic variables, if the AIC value is lower than the original model (before adding external regressors) once a particular climatic factor is integrated, that factor significantly affects the number of patients.

Evaluation of the prediction model

All the possible ARIMA models were created as a list by assigning all the possible estimated values for p , d , P , and Q

using a code written in R programming language. Using the same method, AIC values were calculated for all the models and the models were arranged in the ascending order based on the AIC values. The top ten models with the lowest AIC values were selected for final evaluations using error values. The models were developed based on the information from January 2014 to December 2018 (training data set). The developed prediction models were assessed for validity and accuracy by predicting the disease incidence from January to December 2019 (testing data set). The error measures such as root mean square error (RMSE) and mean absolute percentage error (MAPE) were considered in assessing the accuracy. The residuals (the difference between observed and predicted values) of the selected models were assessed by the Ljung–Box test to check the statistical significance of the selected model (Box and Pierce 1970). Basic R, tidyverse (which include tidyr, dplyr and ggplot2), tseries, and forecast packages of R (R Core Team, Vienna) were used in data clearing, statistical analysis, data visualization, and time series modelling.

Results

Univariate ARIMA modelling

The Augmented Dickey Fuller (ADF) test indicated the time series of leishmaniasis incidence in Sri Lanka does not show stationarity ($P > 0.05$). This was the same for district level data as well. All the district data sets and all-island data showed stationarity after the first differencing. Thus, the value for non-seasonal integration term (d) was 1 for each of the optimal models (Table 1). The type of the model varied depending on the area (Table 1). The collective data set for the entire country showed a seasonality of 12 months. A total of 140 models were generated for leishmaniasis incidence in Sri Lanka. Among them ARIMA (0,1,1) (1,0,0)₁₂ was selected as the best model based on the goodness of fit and error estimates (Table 1). The analysis of residuals indicated that this model is statistically suitable to predict leishmaniasis incidence ($P_{\text{Box-Ljung}} > 0.05$). Only five districts with the highest cutaneous leishmaniasis incidence were selected to develop separate models. There was a seasonality in the disease incidence in the Matara district. Thus, the seasonality was also modelled. Other districts did not show a seasonality. Optimal models for Hambanthota, Anuradhapura, Kurunegala, Polonnaruwa, and Matara were (2,1,3), (0,1,2), (3,1,1) and (3,1,0), (1,1,2) (0,1,1)₁₂, respectively. All the models for district levels were statistically suitable for disease incidence prediction at 95% confidence intervals ($P_{\text{Box-Ljung}} > 0.05$). The AIC values and the key error parameters for each model are mentioned in Table 1.

Table 1 Best predictive ARIMA models for Sri Lanka and main endemic districts

	Sri Lanka	Hambanthota	Anuradhapura	Kurunegala	Polonnaruwa	Matara
Model	(0,1,1) (1,0,0) ₁₂	(2,1,3)	(0,1,2)	(3,1,1)	(3,1,0)	(1,1,2) (0,1,1) ₁₂
AIC ^a	32.02	124.73	70.39	81.85	434.48	91.93
RMSE ^b training	0.2966	0.5989	0.4115	0.4497	8.5550	0.4270
RMSE ^b testing	0.1417	0.4732	0.2377	0.1312	9.8192	0.3172
MAPE ^b training	4.7012684	14.8965971	9.6893507	13.9757074	68.236	11.6400673
MAPE ^b testing	1.9322937	9.0641847	5.3365764	2.2430627	34.257001	7.1763089
<i>P</i> _{Box-Ljung} ^c	0.503	0.8641	0.9179	0.9992	0.9997	0.4584

^aModel with the lowest AIC values is better

^bModels with error values are low and the difference between training and testing error values are better

^cIf the *P* value is higher than 0.05, the model is statistically significant

Disease forecasts

All the tested models were suitable for forecasting within 95% confidence intervals as indicated by forecast plots (Fig. 1). The annual forecasts for the whole country are an underestimation, while the same was observed for Hambanthota, Anuradhapura and Polonnaruwa (Table 2). The predicted numbers were overestimations for Kurunegala and Matara. However, all the predictions are well within the 95% confidence level. The forecasts for the coming years after the study period indicate that the number of patients at the end of 2020, 2021, and 2022 is 4335, 4387, and 4395 respectively. The trends indicate that the number of patients per year will approximately be at the same level for Hambanthota, Anuradhapura, Polonnaruwa, and Kurunegala, where Kurunegala will have the highest number of patients. A slight increase can be expected for the country level. Nevertheless, the Matara district shows a significant increase in the number of patients in the coming years.

Cross-autocorrelation analysis

The partial autocorrelation analyses were performed using Spearman's rank correlation test and none of the climatic parameters showed direct correlations to the leishmaniasis incidence neither for Sri Lanka nor any of the districts at 95% confidence intervals ($P > 0.05$). However, the cross autocorrelation revealed that relative humidity is positively associated with the leishmaniasis incidence in the entire country with a lag of 3 months ($P < 0.05$, $r = 0.230$). In Hambanthota, both the maximum temperature ($r = 0.555$, $P < 0.05$) and relative humidity ($r = 0.255$, $P < 0.05$) were positively associated with the number of leishmaniasis patients. This was also with a 3-month lag period for both the variables. Relative humidity, which exclusively had only positive impacts, was also a determining factor in Matara

($r = 0.489$, $P < 0.05$) and Anuradhapura ($r = 0.267$, $P < 0.05$) with 1- and 3-month lags, respectively. In Matara, relative humidity was the only climate variable with a correlation, while in Anuradhapura, average temperature ($r = -0.267$, $P < 0.05$), minimum temperature ($r = -0.281$, $P < 0.05$), and maximum temperature ($r = -0.319$, $P < 0.05$) were negatively associated with the leishmaniasis incidence with lag periods of 9,9 and 4 months, respectively. Rainfall showed an impact on the disease incidence only for Polonnaruwa ($r = -0.239$, $P < 0.05$). This was a negative correlation with a 6-month lag. No significant correlations were observed for any of the climate variables for the Kurunegala district.

Multivariate ARIMA (ARIMAX) models

The appropriately lagged time series of the climate variables that showed the highest cross-autocorrelation values to the leishmaniasis incidence (Table 3) was integrated as an external regressor. The predictive power of the model created for the entire country was increased after integrating the relative humidity as indicated by the AIC value. The same was observed for relative humidity in Hambanthota and Anuradhapura. Furthermore, the maximum temperature resulted in increased predictive power in Hambanthota. Rainfall which showed a negative impact on the leishmaniasis incidence in the Polonnaruwa district also increases the forecasting power of the optimal model. Nevertheless, other associated climatic factors did not increase the predictive efficacy of the models.

Discussion

The occurrence of vector-borne disease in tropical regions may be partially attributed to various environmental and biological factors, which are favorable for pathogens and vectors to thrive and spread diseases (Hotez et al. 2007;

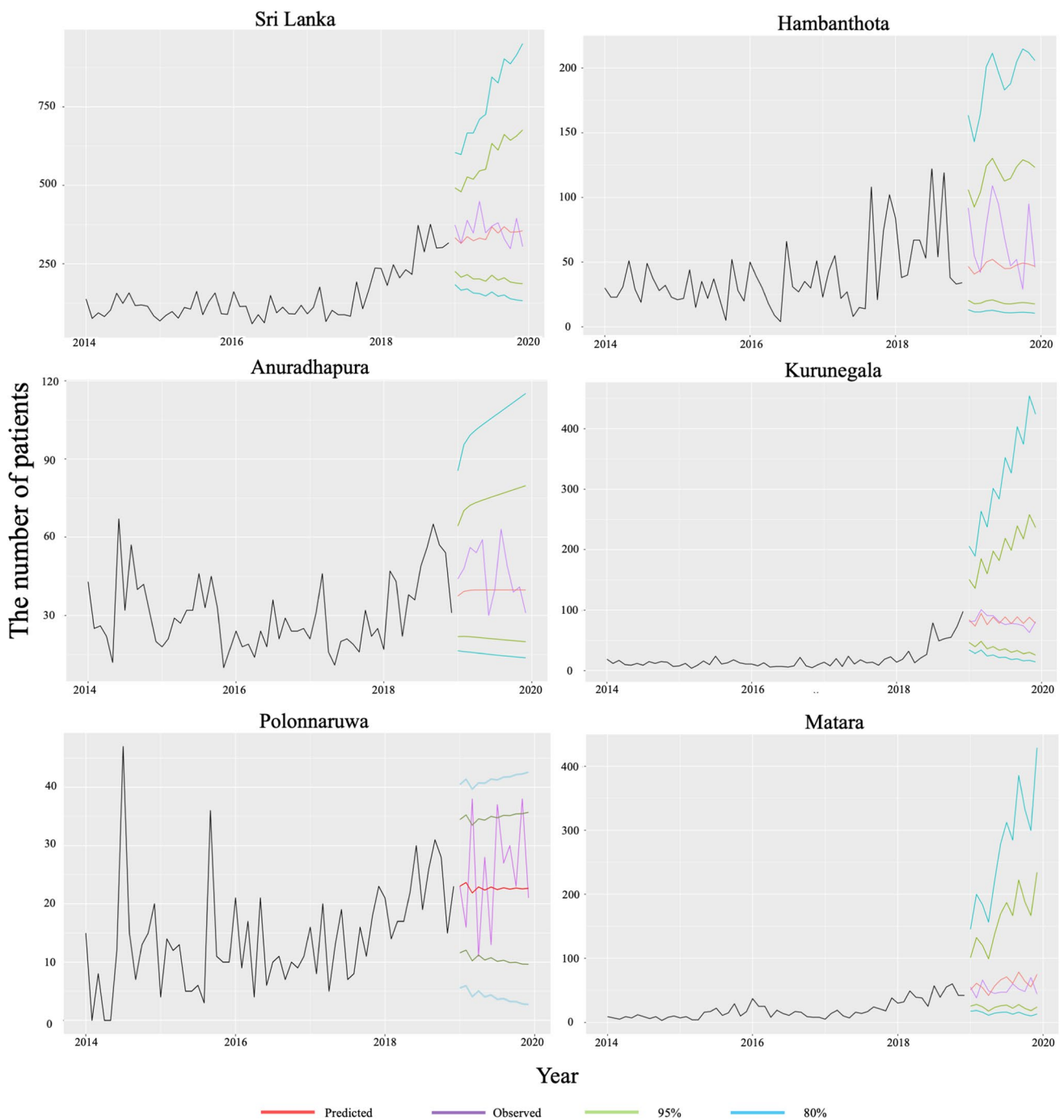


Fig. 1 Forecast plots generated from the optimal model for each area

WHO 2019). Climatic conditions are one of the major contributors to the prevalence of vector-borne diseases (Roger et al. 2013; Anwar et al. 2016). Therefore, a clear view of disease trends with climatic and environmental variables is an important aspect to facilitate control programs and the decision-making process. There are some attempts to predict disease incidence based on epidemiological and climatic factors. The ARIMA-based approach is one of

the useful tools widely used in epidemiological studies and disease control activities. However, such tools have been sparsely used to forecast vector-borne diseases in Sri Lanka and no proper attempt has been made for leishmaniasis. Therefore, this study highlights the first investigation in Sri Lanka using ARIMA based approach to correlate disease occurrence of leishmaniasis with climatic variables.

Table 2 Comparison of the predicted and actual number of patients for 2019 and the predictions for the following 3 years

Area	Number of patients				
	2019		2020	2021	2022
	Predicted	Actual			
Sri Lanka	4111	4319	4335	4387	4395
Hambanthota	564	810	567	565	564
Anuradhapura	477	554	504	509	518
Kurunegala	995	976	1000	999	998
Polonnaruwa	274	305	276	281	291
Matara	735	621	1090	1614	2393

The type and the model parameters varied depending on the area. The selection of these models was based on the AIC values as well as RMSE and MAPE values, which indicate the predictive efficiency of the models. The selected models showed an independent pattern according to the Box-Ljung test, which reflects that this model is suitable for short-term predictions of leishmaniasis incidence. Moreover, the forecast plot indicated that the actual values were always within the 95% confidence interval range forecasted by the model, further confirming the validity of each model.

In the use of disease forecasting for decision making, over-estimations are usually better than under-estimations as it provides an opportunity to have better preparedness. However, even though the resulted models give under-estimations for several areas including the whole country data, the actual values are always within the 95% confidence intervals. Therefore, in the decision-making, the consideration of the margin of error reflected by the confidence intervals as shown in forecast plots is necessary. Furthermore, the forecasts indicate a significant and continuous increase in the number of patients in Matara. This is also continuous

with the general trends observed in other areas in the past few years. Especially, the Kurunegala district, which initially showed a low number of patients and recently became the district with the highest leishmaniasis incidence in the country. Our results signify that the Matara district must be given the priority in coming years to minimize the disease burden. Kurunegala district is also a noticeable district in terms of prioritization for control activities. Although the predicted number of patients does not indicate an increasing trend, the number of patients stays at a higher level close to 1000 patients per year.

The associations between climatic factors and leishmaniasis incidence are characterized by numerous studies from various geographical regions in the world such as North and East Africa, Middle-East and South Asia, and South America (Thomson et al. 2003; Chaves and Pascual 2006; Toumi et al. 2012; Amin et al. 2013; Bounoua et al. 2013; Roger et al. 2013; Akbari et al. 2014; Shirzadi et al. 2015; Azimi et al. 2017; Moradiasl et al. 2018). In this study, the cross-autocorrelation analyses indicated that several climatic factors were associated with the leishmaniasis incidence at different monthly lags. However, only a few of the factors increased the prediction power of the selected models.

Temperature is a factor often linked with the incidence of leishmaniasis (Chaves and Pascual 2006; Azimi et al. 2017). Some studies suggest that there is a positive correlation between temperature and leishmaniasis incidence (Bounoua et al. 2013; Roger et al. 2013; Shirzadi et al. 2015; Moradiasl et al. 2018), while some studies indicate a negative impact (Thomson et al. 2003; Chaves and Pascual 2006; Amin et al. 2013; Akbari et al. 2014; Azimi et al. 2017). On the other hand, in some areas, the temperature has no apparent impact on leishmaniasis incidence (Toumi et al. 2012). In the present study, the number of patients in the Hambanthota district showed a moderate positive

Table 3 Cross-autocorrelation of climate variables with leishmaniasis cases and integration to the optimal model

Area	Climate variable	Most associated lag (month)*	<i>r</i>	<i>P</i>	AIC of the multivariate model
Sri Lanka	Relative humidity	-3	0.230	<0.05	30.80[#]
Hambanthota	Maximum temperature	-3	0.555	<0.05	124.24[#]
	Relative humidity	-3	0.255	<0.05	122.32[#]
Anuradhapura	Average temperature	-9	-0.267	<0.05	70.41
	Relative humidity	-3	0.267	<0.05	70.34[#]
	Minimum temperature	-9	-0.281	<0.05	71.02
	Maximum temperature	-4	-0.319	<0.05	70.48
Polonnaruwa	Rainfall	-6	-0.239	<0.05	432.00[#]
Matara	Relative humidity	-1	0.489	<0.05	101.04

*Number of months taken to reflect the climate condition in the number of patients (for instance, if the relative humidity is higher in the current month, we can expect an increase in the number of patients three months later)

[#]AIC values are lower than the univariate models. Thus, the integration of these climate factors increases the model accuracy, thus likely to affect disease incidence

correlation with the maximum temperature with a 3-month lag period. This means that the maximum temperature in the current month may lead to an increase in the number of patients 3 months later. This was also evident in multivariate ARIMA modelling as the predictive power of the model was increased when the maximum temperature with a lag of 3 months was integrated into the optimal model selected for the Hambanthota district. Observed associations of temperature parameters and leishmaniasis incidence in other districts did not increase the predictive efficacy of the selected models. Thus, such associations are more likely to be coincidental correlations.

Rainfall is another climatic factor known to associate with leishmaniasis incidence in tropical areas (Thomson et al. 2003; Amin et al. 2013; Bounoua et al. 2013; Roger et al. 2013; Akbari et al. 2014; Shirzadi et al. 2015; Azimi et al. 2017; Moradiasl et al. 2018). In the countries with dry climates, the disease incidence was positively associated with the rainfall (Toumi et al. 2012; Bounoua et al. 2013; Azimi et al. 2017). In areas with relatively wet climates, the rainfall inversely correlates with leishmaniasis incidence (Amin et al. 2013; Roger et al. 2013). Rainfall was not a determining factor for the disease occurrence in most of the areas. However, in Polonnaruwa, which has dry climatic conditions, the rainfall showed a negative relationship with a lag of 6 months. The use of rainfall as an external regressor improved the predictive power of the model further validating the observed associations. These effects are more likely to be due to the effects of rain on vector breeding. Studies from Sri Lanka suggest that the main breeding habitats proffered by sand flies are moist soil in draining irrigational tanks and paddy fields (Wijerathna and Gunathilaka 2020). The rain increases the water levels of the irrigational tanks and the muddiness of the paddy fields and other breeding sites, which will reduce the preferable conditions for sand fly larval development. This will subsequently lead to a decrease in the sand fly populations, which will be reflected in a decrease in the number of patients after 6 months. In other countries, the relationship is opposite this, where the rainfall in dry areas positively affects the leishmaniasis prevalence. This may be attributed to the differences in land usage in Sri Lanka. In these agricultural areas of Sri Lanka, the main breeding grounds of sand flies were known to be the drying irrigational tanks (with small water puddles), rice paddies and cattle huts (Wijerathna and Gunathilaka 2020), which have microhabitats with high humidity levels even in the dry months due to continuous water supply through man-made sources. During the rainy season, these places get muddy or filled with water creating unfavorable conditions for sand flies, while in other countries the rainfall makes the soil moist, creating perfect microhabitats for sand flies.

Humidity, on the other hand, almost always has had only positive impacts on the occurrence of the disease

according to previous studies (Thomson et al. 2003; Toumi et al. 2012; Amin et al. 2013; Akbari et al. 2014; Azimi et al. 2017). In the current study, the most notable effect of environmental factors is the effect of the relative humidity, which showed a positive correlation to the disease occurrence. Interestingly, almost always the effect was seen with a lag of three months. The positive impacts can be attributed to its impacts on sand flies. Sand flies require humid places to develop from larvae to adult flies through pupal stages. Increased relative humidity may create preferable conditions for sand flies. Nevertheless, one other approach to explaining this effect is the increased activity of sand flies under humid climates and subsequent increases in sand fly bites. Sand flies tend not to come out during dry and less humid conditions to seek blood meals and prefer to be active when the humidity levels are high (Wijerathna et al. 2020b). Considering the parasite's incubation period and the time from the onset of symptoms to diagnosis and notification, the latter is more likely. However, further investigations may shed light on the exact reasons for the observed impact of humidity on the leishmaniasis incidence.

The main factors affecting the transmission of vector-borne diseases are population density of vector population, vectorial capacity, and host immunity (Kaddu 1986; Warburg et al. 1991; Locksley and Louis 1992; Felicangeli and Rabinovich 1998). Any climatic variable with impacts on these factors may indirectly influence the disease incidence. The leishmaniasis vector sand flies are highly sensitive to environmental conditions. The physical and population changes have been observed in the same species under different environmental conditions (Mann and Kaufman 2010; Lawyer et al. 2017; Tiwary et al. 2017; Wijerathna et al. 2019, 2020a). Therefore, fluctuation of the leishmaniasis cases with relation to changes in climatic factors may also be an indication of a change in the sand fly population size or/and fitness, which in turn affects their vectorial capacity.

The time between the occurrence of a certain climatic event and the consequent changes in the reported number of leishmaniasis patients depends on the incubation period of the parasite and the delay between the onset of disease and seeking treatments. The cross-correlation analysis captures any delayed effect of the environmental conditions on the leishmaniasis incidence. The incubation period of CL usually ranges between 6 weeks and 6 months, but maybe as low as 10 days (Berberian 1944; WHO 2010). It is important to note that the incubation period may be affected by host immunity (Locksley and Louis 1992). There is a chance that the patient's immunity against disease is also affected by climatic factors (Hedlund et al. 2014). Therefore, the observed correlations can be partially accounted for the reduced host immunity that may be resulted from the changes in climatic conditions, especially the humidity levels.

The ARIMA-based models are useful for the short-term prediction of future values of a time series (Lavenbach 2017). However, the exact number of future values that could be predicted by a particular model is uncertain. The training of the model with more recent data improves the accuracy of the model (Lavenbach 2017). Thus, these models may predict future values for 1–2 seasons at highly accurate levels. The selection of a suitable time series forecasting technique from ARIMA, SARIMA and, ETS (Error, trend, seasonal model) is done based on the trend, seasonal, cyclical, and irregular components of the series (Brockwell et al. 2002). Traditional time series techniques such as ARIMA, SARIMA, and ETS are designed to handle single seasonality in a time series, though they are easy to perform (Naim et al. 2018). In the presence of a data set with multiplicative seasonality, advanced techniques like BATS (exponential smoothing state-space model with Box-Cox transformation, ARMA errors, trend and seasonal components) and TBATS (trigonometric exponential smoothing state-space model with Box-Cox transformation, ARMA errors, trend and seasonal components) are more suitable (de Livera et al. 2011). The data sets with seasonality in the current study showed an additive seasonal pattern indicating that the SARIMA is the suitable forecasting model for the present data.

This study was conducted based on a few assumptions. The first one is that the notification of patient information did not change throughout the time and all patients are reported to the RDHS office. However, this is not always true. There were some misidentifications of leishmaniasis cases as leprosy during the documentation. Furthermore, some cases had remained unreported for an unknown reason. Another issue associated with this assumption is that all patients seek treatment from government hospitals. However, some patients had visited traditional healers, while some had completely ignored them without seeking treatment and follow-up. These patients had not reported to the general notification system. Therefore, the number of patients is likely to be an underestimation of the actual case number. Hence, the diagnosis facilities must be further improved and the awareness levels among the communities living in endemic areas must be enhanced. Community-based studies are required to identify what proportion of cases are left unnotified or misnotified. Those values must be integrated with forecast values to get a more accurate final estimation. Nevertheless, the currently developed model provides an overestimation when considering the upper limit of the 95% confidence intervals, which may somewhat compensate for the errors created by the underestimation of the actual case number.

In addition, the use of patients that were seeking the medical attention as the incidence of CL could be a limitation in this study, because changes in the frequency of case detection may be influenced by many factors such as access to health services, the severity of the disease, and variations in

the incubation period, which may impair the ability to assess the association with climate variables. However, this limitation cannot be compensated since the case detection of leishmaniasis in Sri Lanka is passive and only the patients seeking medical attention for their symptoms are screened for the presence of parasites (Wijerathna et al. 2018). This is less likely to be an issue for the current approach as the inhabitants in endemic areas are well aware of the disease during the last few years, which results in an immediate seeking of medical attention following the onset of first symptoms. Another limitation of this study is that the climatic variables were recorded from agro-meteorological stations located at several places in each district. The value obtained for the district may not completely represent the climatic factors of a selected sub-region of the districts. The predicted cases were not able to compare with the actual reported cases in 2020 and 2021 due to incompleteness and under-report of disease incidence due to COVID-19 pandemic. Therefore, predicted cases versus actual cases were compared only for 2019. The present model closely predicted the number of patients in 2019 (Predicted: 4111, Actual: 4319). A recent study conducted in Sri Lanka based on a mixed spatiotemporal regression-auto-regression model has also predicted the number as 4064 cases for 2019 (Karunaweera et al. 2021). Hence, this may evidence the accuracy in the model since the present work predicted more closely to the actual cases in 2019.

Despite the limitations, the current study provides a good indication of how the future leishmaniasis incidence will vary in Sri Lanka and major endemic districts. All the ARIMA models constructed here predict the leishmaniasis incidence well within the 95% confidence intervals and a clear view of how the climatic variables affect the leishmaniasis incidence in Sri Lanka is presented.

Conclusions

The ARIMA-based linear state-space models performed well for the prediction of leishmaniasis incidence in Sri Lanka, which showed no or additive seasonality for all the considered districts and whole country data. If a multiplicative seasonality is present, more advanced modelling algorithms must be used. The forecasts indicate an increase in the total number of patients in Sri Lanka in the coming years. Especially a significant increase can be expected in the Matara district. Kurunegala district must also be noted as the case number is likely to stay at a higher level. The relative humidity is the best predictive variable for leishmaniasis in Sri Lanka. It is also important to establish proper data reporting/recording systems catering to the essential data which support further optimization of forecasting models. Further, it is recommended to assess the need for the establishment of

a leishmaniasis control unit under the Ministry of Health to conduct such activities in a systematized manner.

Acknowledgements The National Research Council, Sri Lanka (Grant No. NRC 16-142), is acknowledged for financial support.

Author contribution TW and NG conceived and designed the study. TW conducted the data collection, data analysis, and manuscript preparation. All authors read and approved the final version of the manuscript.

Funding This study was funded by Sri Lanka National Research Council, Sri Lanka, Grant No. NRC 16–142.

Data availability The data generated during the study will be available from the corresponding author upon reasonable request.

Declarations

Ethics approval The study was approved by the Ethics Review Committee, Faculty of Medicine, University of Kelaniya (Ref. No. P/204/12/2016).

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Akbari E, Mayvaneh E, Entezari A, Nazari M (2014) Survey of the role of bioclimatic factors in the outbreak of Cutaneous Leishmaniasis. *Iran J Epidemiol* 10:65–74
- Amin MR, Tareq SM, Rahman SH, Uddin MR (2013) Effects of temperature, rainfall and relative humidity on Visceral Leishmaniasis prevalence at two highly affected Upazilas in Bangladesh. *Life Sci J* 10:1440–1446
- Anwar MY, Lewnard JA, Parikh S, Pitzer VE (2016) Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malar J* 15:566. <https://doi.org/10.1186/s12936-016-1602-1>
- Assunção RM, Reis IA, Oliveira CDL (2001) Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Stat Med* 20:2319–2335. <https://doi.org/10.1002/sim.844>
- Azimi F, Shirian S, Jangjoo S et al (2017) Impact of climate variability on the occurrence of cutaneous leishmaniasis in Khuzestan Province, southwestern Iran. *Geospat Health* 12:15–22. <https://doi.org/10.4081/gh.2017.478>
- Berberian DA (1944) Cutaneous Leishmaniasis (Oriental Sore). *Arch Derm Syphilol* 50:231. <https://doi.org/10.1001/archderm.1944.01510160003001>
- Bounoua L, Kahime K, Houti L et al (2013) Linking Climate to Incidence of Zoonotic Cutaneous Leishmaniasis (L. major) in Pre-Saharan North Africa. *Int J Environ Res Public Health* 10:3172–3191. <https://doi.org/10.3390/ijerph10083172>
- Box GEP, Pierce DA (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J Am Stat Assoc* 65:1509. <https://doi.org/10.2307/2284333>
- Briët OJ, Vounatsou P, Gunawardena DM et al (2008) Temporal correlation between malaria and rainfall in Sri Lanka. *Malar J* 7:77. <https://doi.org/10.1186/1475-2875-7-77>
- Brockwell PJ, Davis RA, Davis RJ (2002) Introduction to time series and forecasting, 1st edn. Springer, New York
- CDC (2022) CDC - Leishmaniasis. <https://www.cdc.gov/parasites/leishmaniasis/index.html>. Accessed 12 Jun 2020
- Chaves LF, Pascual M (2006) Climate Cycles and Forecasts of Cutaneous Leishmaniasis, a Nonstationary Vector-Borne Disease. *PLoS Med* 3:e295. <https://doi.org/10.1371/journal.pmed.0030295>
- de Livera AM, Hyndman RJ, Snyder RD (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. *J Am Stat Assoc* 106:1513–1527. <https://doi.org/10.1198/jasa.2011.tm09771>
- de Souza VAF, de Cortez LRP, B, Dias RA, et al (2012) Análise de aglomerados espaço-temporais da leishmaniose visceral americana em Bauru, São Paulo, Brasil. *Cad Saude Publica* 28:1949–1964. <https://doi.org/10.1590/S0102-311X2012001000013>
- Department of Census Statistics (2019) District statistical hand book. Department of Census & Statistics - Sri Lanka. <http://www.statistics.gov.lk/ref/HandbookDictionary>. Accessed 23 Apr 2019
- Department of Meteorology (2019) Climate of Sri Lanka. http://www.meteo.gov.lk/index.php?option=com_content&view=article&id=94&Itemid=310&lang=en. Accessed 23 Apr 2019
- Ehelepola NDB, Ariyaratne K, Buddhadasa WMNP et al (2015) A study of the correlation between dengue and weather in Kandy City, Sri Lanka (2003–2012) and lessons learned. *Infect Dis Poverty* 4:42. <https://doi.org/10.1186/s40249-015-0075-8>
- Feliciangeli DM, Rabinovich J (1998) Abundance of Lutzomyia ovallesi but not Lu. gomezi (Diptera: Psychodidae) correlated with cutaneous leishmaniasis incidence in north-central Venezuela. *Med Vet Entomol* 12:121–131. <https://doi.org/10.1046/j.1365-2915.1998.00072.x>
- Galgamuwa LS, Dharmaratne SD, Iddawela D (2018) Leishmaniasis in Sri Lanka: spatial distribution and seasonal variations from 2009 to 2016. *Parasit Vectors* 11:60. <https://doi.org/10.1186/s13071-018-2647-5>
- Gálvez R, Descalzo MA, Guerrero I et al (2011) Mapping the current distribution and predicted spread of the leishmaniasis sand fly vector in the madrid region (Spain) based on environmental variables and expected climate change. *Vector Borne Zoonotic Dis* 11:799–806. <https://doi.org/10.1089/vbz.2010.0109>
- Gomez-Barroso D, Herrador Z, San Martín JV et al (2015) Spatial distribution and cluster analysis of a leishmaniasis outbreak in the south-western Madrid region, Spain, September 2009 to April 2013. *Eurosurveillance* 20:11–20. <https://doi.org/10.2807/1560-7917.ES2015.20.7.21037>
- Hedlund C, Blomstedt Y, Schumann B (2014) Association of climatic factors with infectious diseases in the Arctic and subarctic region – a systematic review. *Glob Health Action* 7:1–16. <https://doi.org/10.3402/GHA.V7.24161>
- Hotez PJ, Molyneux DH, Fenwick A et al (2007) Control of Neglected Tropical Diseases. *N Engl J Med* 357:1018–1027. <https://doi.org/10.1056/NEJMra064142>
- Huang F, Zhou S, Zhang S et al (2011) Temporal correlation analysis between malaria and meteorological factors in Motuo County. *Tibet Malaria J* 10:54. <https://doi.org/10.1186/1475-2875-10-54>
- Kaddu JB (1986) *Leishmania* in Kenyan phlebotomine sandflies—III. Advances in the investigations of vectorial capacity and vector-parasite relationships of various species of sandflies in Kenya. *Int J Trop Insect Sci* 7:207–212. <https://doi.org/10.1017/S1742758400008973>
- Karagiannis-Voules D-A, Scholte RGC, Guimarães LH et al (2013) Bayesian Geostatistical Modeling of Leishmaniasis Incidence in Brazil. *PLoS Negl Trop Dis* 7:e2213. <https://doi.org/10.1371/journal.pntd.0002213>
- Karunaweera ND, Pratlong F, Siriwardane HVYDYD et al (2007) Sri Lankan cutaneous leishmaniasis is caused by *Leishmania*

- donovani zymodeme MON-37. *Trans R Soc Trop Med Hyg* 97:380–381. [https://doi.org/10.1016/s0035-9203\(03\)90061-7](https://doi.org/10.1016/s0035-9203(03)90061-7)
- Karunaweera ND, Senanayake S, Ginige S et al (2021) Spatiotemporal distribution of cutaneous leishmaniasis in Sri Lanka and future case burden estimates. *PLoS Negl Trop Dis* 15:e0009346. <https://doi.org/10.1371/JOURNAL.PNTD.0009346>
- Lavenbach H (2017) Short-Term Forecasting with ARIMA Models. In: Lavenbach H (ed) *Change & Chance Embraced: Achieving Agility with Smarter Forecasting in the Supply Chain*, 1st edn. Delphus Publishing, New Jersey, USA, pp 226–282
- Lawyer P, Killick-Kendrick M, Rowland T et al (2017) Laboratory colonization and mass rearing of phlebotomine sand flies (Diptera: Psychodidae). *Parasite* 24:42. <https://doi.org/10.1051/parasite/2017041>
- Liu Q, Liu X, Jiang B, Yang W (2011) Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infect Dis* 11:218. <https://doi.org/10.1186/1471-2334-11-218>
- Locksley RM, Louis JA (1992) Immunology of leishmaniasis. *Curr Opin Immunol* 4:413–418. [https://doi.org/10.1016/S0952-7915\(06\)80032-4](https://doi.org/10.1016/S0952-7915(06)80032-4)
- Mann RS, Kaufman PE (2010) Colonization of *Lutzomyia shannoni* (Diptera: Psychodidae) utilizing an artificial blood feeding technique. *J Vector Ecol* 35:286–294. <https://doi.org/10.1111/j.1948-7134.2010.00084.x>
- Moradiasl E, Rassi Y, Hanafi-Bojd AA et al (2018) The Relationship between Climatic Factors and the Prevalence of Visceral Leishmaniasis in North West of Iran. *Int J Pediatr* 6:7169–7178
- Naim I, Mahara T, Idrisi AR (2018) Effective Short-Term Forecasting for Daily Time Series with Complex Seasonal Patterns. *Procedia Comput Sci* 132:1832–1841. <https://doi.org/10.1016/j.procs.2018.05.136>
- Roger A, Nacher M, Hanf M et al (2013) Climate and Leishmaniasis in French Guiana. *Am J Trop Med Hyg* 89:564–569. <https://doi.org/10.4269/ajtmh.12-0771>
- Rosales JC, Blas OJA, Yang HM (2017) Monte Carlo simulation of American Tegumentary Leishmaniasis epidemic. The case in Oran, Salta, Argentina, 1985–2007. *Appl Math Sci* 11:1–14. <https://doi.org/10.12988/ams.2017.610255>
- Said SE, Dickey DA (1984) Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* 71:599. <https://doi.org/10.2307/2336570>
- Sattenspiel L (2009) *The Geographic Spread of Infectious Diseases: Models and Applications* on JSTOR. Princeton University Press, Princeton
- Sharafi M, Ghaem H, Tabatabaee HR, Faramarzi H (2017) Forecasting the number of zoonotic cutaneous leishmaniasis cases in south of Fars province, Iran using seasonal ARIMA time series method. *Asian Pac J Trop Med* 10:79–86. <https://doi.org/10.1016/j.apjtm.2016.12.007>
- Shimabukuro PHF, Da Silva TRR, Ribeiro FOF et al (2010) Geographical distribution of American cutaneous leishmaniasis and its phlebotomine vectors (Diptera: Psychodidae) in the state of São Paulo. *Brazil Parasites Vectors* 3:121. <https://doi.org/10.1186/1756-3305-3-121>
- Shirzadi MR, Mollalo A, Yaghoobi-Ershadi MR (2015) Dynamic relations between incidence of Zoonotic Cutaneous Leishmaniasis and climatic factors in Golestan Province Iran. *J Arthropod-Borne Dis* 9:148–160
- Siriwardana HVYD, Thalagala N, Karunaweera ND (2010) Clinical and epidemiological studies on the cutaneous leishmaniasis caused by *Leishmania (Leishmania) donovani* in Sri Lanka. *Ann Trop Med Parasitol* 104:213–223. <https://doi.org/10.1179/136485910X12647085215615>
- Thomson MC, Bendall A, Osman ME et al (2003) Risk mapping of visceral leishmaniasis: the role of local variation in rainfall and altitude on the presence and incidence of kala-azar in Eastern Sudan. *Am J Trop Med Hyg* 68:10–17. <https://doi.org/10.4269/ajtmh.2003.68.1.0680010>
- Tiwary P, Singh SK, Kushwaha AK et al (2017) Establishing, expanding, and certifying a closed colony of phlebotomus argentipes (Diptera: Psychodidae) for xenodiagnostic studies at the Kala Azar Medical Research Center, Muzaffarpur, Bihar, India. *J Med Entomol* 54:1129–1139. <https://doi.org/10.1093/jme/tjx099>
- Toumi A, Chlif S, Bettaieb J et al (2012) Temporal Dynamics and Impact of Climate Factors on the Incidence of Zoonotic Cutaneous Leishmaniasis in Central Tunisia. *PLoS Negl Trop Dis* 6:0–7. <https://doi.org/10.1371/journal.pntd.0001633>
- Ture M, Kurt I (2006) Comparison of four different time series methods to forecast hepatitis A virus infection. *Expert Syst Appl* 31:41–46. <https://doi.org/10.1016/J.ESWA.2005.09.002>
- Warburg A, Montoya-Lerma J, Jaramillo C et al (1991) Leishmaniasis vector potential of *Lutzomyia* spp. in Colombian coffee plantations. *Med Vet Entomol* 5:9–16. <https://doi.org/10.1111/j.1365-2915.1991.tb00514.x>
- Werneck GL (2008) Forum: Geographic spread and urbanization of visceral leishmaniasis in Brazil Introduction. *Cadernos De Saude Publica* 24:2937–2940
- WHO (2010) Control of the leishmaniasis. Report of a meeting of the WHO Expert Committee on the Control of Leishmaniases. World Health Organisation, Geneva
- WHO (2019) Neglected tropical diseases. https://www.who.int/neglected_diseases/en/. Accessed 9 Nov 2019
- WHO (2022) Leishmaniasis. In: WHO. <http://www.who.int/leishmaniasis/disease/en/>. Accessed 9 Nov 2019
- Wijerathna T, Gunathilaka N (2020) Diurnal adult resting sites and breeding habitats of phlebotomine sand flies in cutaneous leishmaniasis endemic areas of Kurunegala District. *Sri Lanka Parasit Vectors* 13:284. <https://doi.org/10.1186/s13071-020-04154-7>
- Wijerathna T, Gunathilaka N, Gunawardena K (2018) The Economic Impact of Cutaneous Leishmaniasis in Sri Lanka. *Biomed Res Int* 2018:1–9. <https://doi.org/10.1155/2018/3025185>
- Wijerathna T, Gunathilaka N, Gunawardena K, Rodrigo W (2020) Population dynamics of Phlebotomine sand flies (Diptera: Psychodidae) in cutaneous leishmaniasis endemic areas of Kurunegala District. *Sri Lanka Acta Tropica* 230:106406
- Wijerathna T, Gunathilaka N, Gunawardena K, Rodrigo W (2017) Potential Challenges of Controlling Leishmaniasis in Sri Lanka at a Disease Outbreak. *BioMed Res Int* 2017:Article ID 6931497. <https://doi.org/10.1155/2017/6931497>
- Wijerathna T, Gunathilaka N, Gunawardena K, Rodrigo W (2019) Bionomics of Leishmaniasis Vector Sand Flies (Psychodidae:Phlebotominae) in Kurunegala District, Sri Lanka. In: 12th International Research Conference 2019 of Kotelawala Defence University. Sir John Kotelawala Defence University, Ratmalana, p 157
- Wijerathna T, Gunathilaka N, Gunawardena K (2020a) Establishment of a Colony of *Phlebotomus argentipes* under Laboratory Conditions and Morphometric Variation between Wild-Caught and Laboratory-Reared Populations. *J Trop Med Article ID* 7317648. <https://doi.org/10.1155/2020/7317648>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.