International Journal of Academic Research and Development
ISSN: 2455-4197
Impact Factor: RJIF 5.22
www.academicjournal.in
Volume 3; Issue 6; November 2018; Page No. 27-32

Outbreaks of Leptospirosis in selected districts of Sri Lanka

S Reyalt Gnanapragasam
Department of Mathematics, The Open University of Sri Lanka, Nawala, Nugegoda, Sri Lanka

Abstract
Leptospirosis is a severe infectious disease caused by bacteria. In Sri Lanka, it is one of the main communicable diseases. From 2010 to 2017, more than 4000 outbreaks of leptospirosis per year were reported to the epidemiology unit. Although outbreaks were reported in all the districts throughout the year, Colombo, Kalutara, Gampaha, Ratnapura, Galle, Matara, Kegalle, Matale and Kurunegala were named as leptospirosis endemic districts in Sri Lanka. In the first half of the year 2018, more than 2000 outbreaks had been reported island wide. Nearly half of the total outbreaks in the island were from five districts namely, Colombo, Gampaha, Kalutara, Ratnapura, Kurunegala and considered them, in this study, as the "selected five districts". Therefore, the objective of this study is to model the outbreaks reported in these five districts of Sri Lanka. The relevant techniques and the appropriate tests were employed to develop model to forecast the outbreaks of leptospirosis in those districts. This study concludes that, leptospirosis outbreaks in Colombo, Gampaha, Kurunegala districts are decreasing while in Kalutara and Ratnapura districts the outbreaks are in increasing trend among the selected five districts of Sri Lanka in the study period. Over 92% accuracy of the best fitted model was developed. Hence the estimated outbreaks only in the last nine months from the selected five districts of Sri Lanka in the year 2018 will be 1200.

Keywords: outbreaks of leptospirosis, districts of Sri Lanka, time series modelling

1. Introduction
Leptospirosis is a severe infectious disease caused by a type of bacteria and different names are used in different countries but it is commonly known as rat fever in Sri Lanka. Though the leptospirosis occurs worldwide, it is more prevalent in tropical regions in the world and areas with heavy rainfall. It is considered as the most widespread disease that is transmitted by some animals in the world such as cattle, pigs, horses, dogs, rodents, and some other wild animals. This disease can occur when people are exposed to water that is contaminated with urine from infected animals after following extreme rainfall or flooding. It mostly affects the people those who are involved in jobs associated with mud and water such as farm and agricultural workers.

In the context of Sri Lanka, leptospirosis is one of the main communicable diseases and it was prevalent with occasional outbreaks in the past. Sri Lanka is considered as a leptospirosis high prevalent country as it is having one of the highest outbreaks of leptospirosis [1]. Leptospirosis was first investigated in Sri Lanka in 1953 and the first leptospirosis case was confirmed in 1959 [2]. In 1991, leptospirosis was made a notifiable disease in Sri Lanka. [1] Further says that a steady increase of reporting of leptospirosis until 2007 and it was based on the disease notification data from 1991. Nevertheless, an alert as an emerging disease was created in 2003 among health professionals on leptospirosis [3]. In 2008, Sri Lanka experienced the largest documented outbreaks of leptospirosis [4]. From 2008 to 2014, in Sri Lanka on the average of more than 5000 outbreaks were reported annually to the epidemiology unit [5]. The unusual clinical features observed during the year 2011 leptospirosis outbreaks in Sri Lanka could be due to an uncommon strain that arose in the dry context rather than wet season epidemiology [6].

According to the literature review to the study, models were developed in [8] to analyse leptospirosis outbreaks in Sri Lanka and its relation to rainfall during the period from 2006 to 2010. Alco spatial variables associated with suspected leptospirosis risk during endemic and outbreak periods from 2005 to 2009 were investigated in [9] and in which space–time scan statistics were combined with regression modeling to test associations during endemic and outbreaks periods. Further in [10], the spatial and seasonal patterns of human leptospirosis from 2004 to 2013 in Gampaha district of Sri Lanka were analysed and predictive model provided an evidence base for reducing disease burden by improving the understanding of the high risk areas and seasonal patterns of the disease in the district. Nevertheless, statistical studies of district-wise leptospirosis outbreaks are less in Sri Lanka. The data until 2008 in the epidemiology unit of Sri Lanka listed that although all the other districts report outbreaks throughout the year, Colombo, Kalutara, Gampaha, Ratnapura, Galle, Matara, Kegalle, Matale and Kurunegala are named as leptospirosis endemic districts in Sri Lanka. However, notification data after 2008 also show that outbreaks are reported from all districts each year.

The results of [7] showed that, Western province of Sri Lanka was the mostly affected part of the island by human leptospirosis. Moreover it says that, Gampaha, Kalutara and Colombo districts in the Western province were ranked among the first 5 districts (Ratnapura and Kurunegala are the other two districts) of Sri Lanka based on average number of recorded outbreaks. As per the records in the epidemiology

unit of the ministry of health, nutrition and indigenous medicine in Sri Lanka [5], the number of outbreaks is increasing continually. In the first half of the year 2018, more than 2100 outbreaks have been reported and in which some districts, including Ratnapura, Gampaha, Kalutara and Colombo of Sri Lanka have reported over 100 outbreaks. Nearly half of the total outbreaks in the island are from the five districts namely, Colombo, Gampaha, Kalutara, Ratnapura, Kurunegala and considered them in this study as the "selected five districts" of Sri Lanka. Therefore, the objective of this study is to model the outbreaks reported in these five districts. Since the forecasting the outbreaks in this regards is to plan the necessary future activities to control the outbreaks, one can use this model to forecast the future outbreaks to take action accordingly when it is needed.

2. Materials and Methods

The monthly data of outbreaks of leptospirosis released by the epidemiology unit of Sri Lanka are extracted for this study. The following techniques are carried out in order to check the behaviour of the series in terms of stationary condition of the series:

Plot of time series is generally used to get an idea about the series and its behaviour. Also it is to inspect for extreme observations, missing data, and elements of non-stationary such as trend or seasonality or cyclic pattern or irregular variations. Statistically, *Augmented Dickey- Fuller (ADF)* test is used to confirm that the stationary of the series in terms of trend availability. At the same time, *Kruskal- Wallis* test is used to confirm the existence of seasonal pattern in the series. *Autocorrelation function (ACF)* and *partial autocorrelation function (PACF)* are examined to determine the nature of the process under consideration. If both ACF and PACF decay exponentially then the series is stationary.

The differencing method is employed to transform the non-stationary series into a stationary series such that to improve the forecasting performance as follows: If the series has a trend then by taking *regular differencing* the trend can be eliminated from the series. In the method of *seasonal differencing*, differences are taken at seasonal lags. If the spikes appear repeatedly in the ACF graph at particular lags, then it can be assumed that there is a seasonal pattern in the series. After the transformation, the stationary data is used to develop the models. A model with combinations of *autoregressive (AR) terms* and *moving average (MA) terms* are generally called as auto regressive moving averages (ARMA) model. If the series has a trend, it can often be converted to a stationary series by differencing and it is generally denoted as $\mathrm{ARIMA}(p, d, q)$, where p and q are the order of autoregressive and moving average processes respectively and d indicates the amount of differencing. In some cases, the series shows a repeating or cyclic behaviour. These seasonal patterns can be very effectively used to further improve the forecasting performance. A seasonal ARIMA (SARIMA) model or $\mathrm{ARIMA}(p, d, q)(P, D, Q)_S$ usually contains: regular $\mathrm{AR}(p)$ and $\mathrm{MA}(q)$ terms that account for the correlation at low lags. Seasonal $\mathrm{AR}(P)$ and seasonal $\mathrm{MA}(Q)$ terms that account for the correlation at the seasonal lags where d, D and S indicate the amount of regular differencing, seasonal differencing and seasonality respectively. Diagnostic tests are performed to determine the adequacy of the model. After identifying the tentative models, the following tests are applied to check whether the underline conditions of diagnostic checking are satisfied by the fitted model. The necessary conditions of diagnostic checking are the residuals of the fitted models should be distributed normally and independently with constant variance. In this study, the following tests are employed to check those conditions.

Anderson Darling (AD) test is used to test, if a sample of data comes from a population with a specific distribution. One of the most important tests for detecting serial correlation is *Lagrange's Multiplier (LM)* test which is applied to test the independency of the residuals of fitted models. *White's general* test is used in order to check the constant variance of residuals.

In some situations, two or more models could satisfy all the conditions of the diagnostic checking. In such situations, to select the best model among those significant models, the following criterions are applied:

Akaike Information Criterion (AIC) is often used for model selection. *Schwartz's Bayesian Criterion (SBC)* is another mostly used technique for model selection. Generally, the best model is the one which gives the lowest AIC and SBC values.

To check the forecasting accuracy of the fitted model, the *mean absolute percentage error (MAPE)* is employed. It is generally used for evaluation of the forecast against the validation sample. To compare the average forecast accuracy of different models, MAPE statistic is used.

3. Results and Discussion

Before develop the model, the original data are inspected and the relevant information is reported as descriptive statistics. In this descriptive statistics part, the monthly leptospirosis outbreaks data from January 2010 to December 2017 are considered.

Figure 1(a) and Figure 1(b) provide the island-wide annual leptospirosis outbreaks and the annual leptospirosis outbreaks in the selected five districts of Sri Lanka reported from January 2010 to December 2017 respectively.

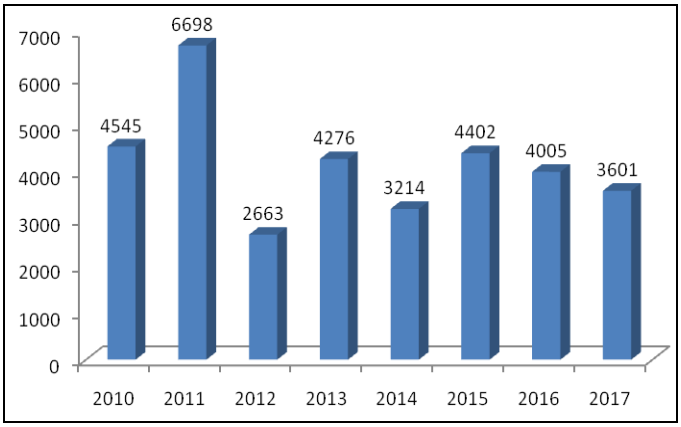

Fig 1(a): Annual outbreaks in Sri Lanka

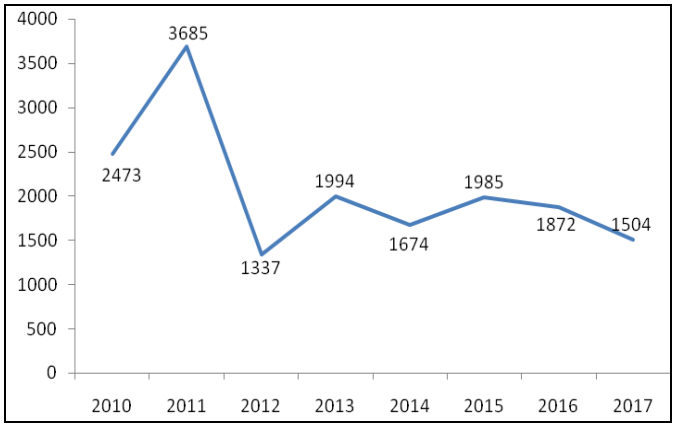

Fig 1(b): Annual outbreaks in selected districts

As per Figure 1(a), the highest number of outbreaks (6698) is reported in the year 2011 and it is extremely high related to the records in all other years. But at the very next year the total outbreaks reduces to the lowest number. Also the average outbreaks during this 8 years period are about 4176. Thus more than 4000 outbreaks per year are reported island wide in Sri Lanka in this period.

According to [7], on the average, leptospirosis outbreaks are considered, Western province, Subaragamuwa province, Southern province and North Western province are ranked as first 4 provinces of Sri Lanka. At the same point, Ratnapura, Kurunegala, Gampaha, Kalutara and Colombo districts are having first 5 ranks as the average number of leptospirosis outbreaks reported in Sri Lanka. Therefore this study is interested to observe the outbreaks reported in those five districts and considered them as the "selected five districts" of Sri Lanka in further analysis.

It is very clear from Figure 1(b) too, the highest number of leptospirosis outbreaks are reported in the year 2011 and lowest are reported in the subsequent year 2012. It can be stated from Figure 1(b) that, on the average, 2065 outbreaks per year are reported from 2010 to 2017. Thus, during this period, nearly the half (49.47%) of the total outbreaks are reported from the selected five districts of Sri Lanka. In Figure 2(a) and 2(b), the selected district-wise annual and proportion Outbreaks are respectively summarised as follows.

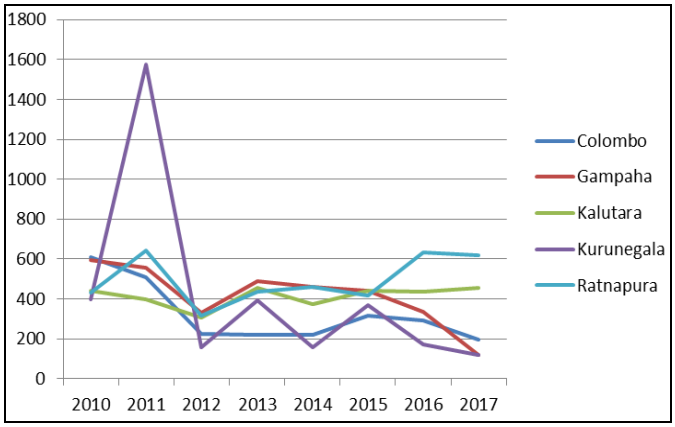

Fig 2(a): District-wise annual outbreaks

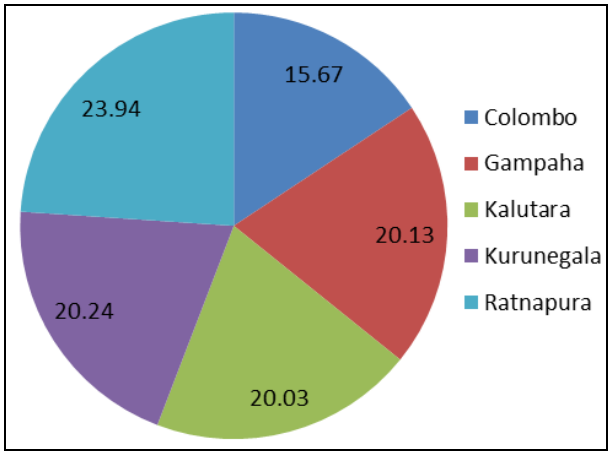

Fig 2(b): District-wise proportion outbreaks

It can be seen from Figure 2(a) that, annual leptospirosis outbreaks in Colombo, Gampaha, Kurunegala districts are decreasing while in Kalutara and Ratnapura districts, the outbreaks are in increasing trend. Furthermore, it is clearly observed a peak in the year of 2011 in the district of Kurunegala and at the inspection of the raw data, it is noted that 745 outbreaks were reported only in the month of March in 2011 in Kurunegala district of Sri Lanka. It is the highest outbreak in a month in any district of Sri Lanka from the years 2010 to 2017. However, the outbreaks are reducing in every other year and it is noted in the years 2016 and 2017 that the total outbreaks are relatively less when compare with other 4 districts. Nevertheless, in Ratnapura district getting increase outbreaks over this 8 year period and in the years 2016 and 2017, which has the highest number of outbreaks among selected five districts. Moreover, it is very clear from Figure 2(b) that, highest proportion is reported from Ratnapura (23.94%) and proportion from Colombo district is the lowest (15.67%) whilst in other 3 districts almost the same proportion (approximately 20%) of outbreaks is reported.

The empirical results of model development are reported as follows. Figure 3 represents the plot of the time series of the leptospirosis outbreaks reported in those five districts of Sri Lanka from January 2010 to December 2017.

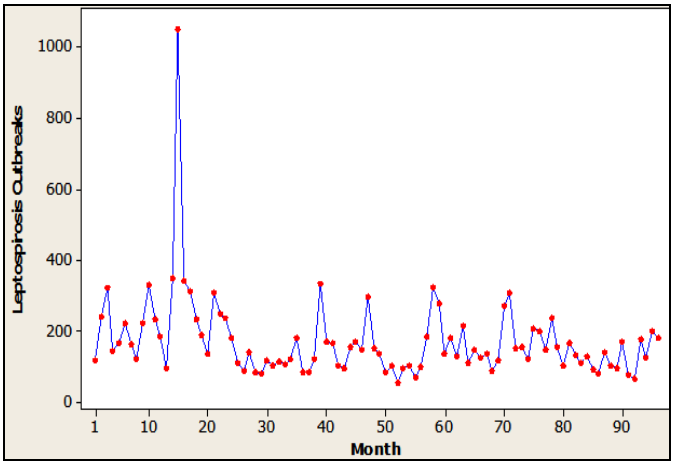

Fig 3: Plot of time series of leptospirosis outbreaks

It cannot be seen a clear trend or seasonal pattern in the series of leptospirosis outbreaks in those 5 districts in Figure 3. Further it is noted that there is an extreme observation at the 15th index of the month of March 2011. Since this extreme observation (1050) will definitely effect further calculation, the value is replaced by the average outbreaks (315) reported in the months of March from 2010 to 2017. Then the new time series for the adjusted outbreaks is plotted in Figure 4(a). It is noted that in the rest of the work the adjusted leptospirosis outbreaks are taken into account.

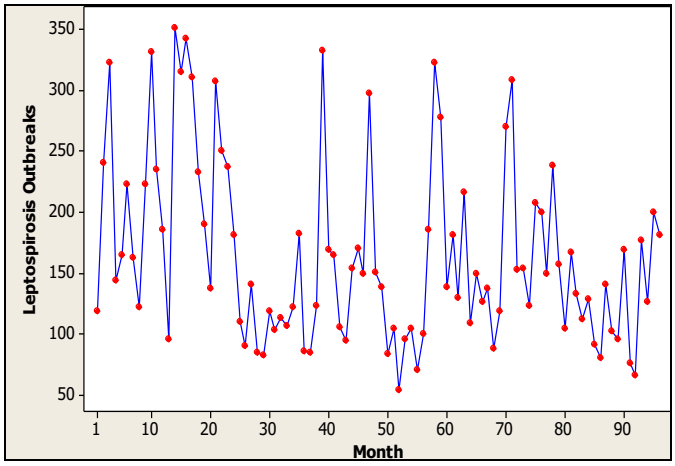

Figure 4(a): Plot of adjusted outbreaks

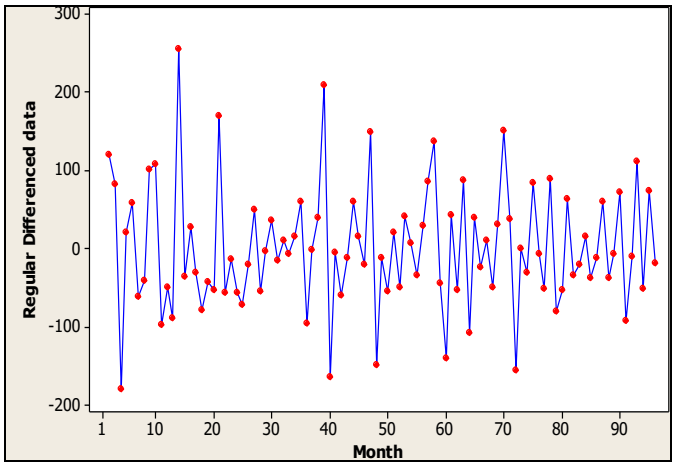

Fig 4(b): Plot of regular differed series

It seems there is no clear trend or seasonal or any systematic patterns in the outbreaks in Figure 4(a). Nevertheless, it can be observed that after 24 months there is a slightly increase in the outbreaks. It suggests that there can be a trend in the series but there is no clear picture on the seasonal pattern from Figure 4(a). However, it has to be verified through statistical tests. Therefore, ADF and Kruskal- Wallis tests are performed to check the existence of trend and seasonality respectively.

Based on the p- value (0.064) of ADF test, it can be confirmed (at moderate level) that, the series has a trend. However, the p-value (0.007) of Kruskal- Wallis strongly suggests that the series has seasonality. Therefore, it can be concluded with 95% confidence that, the original series is non-stationary. Therefore, that seasonality has to be removed prior to identifying the process of ARIMA models. At the first stage to remove the trend, the regular difference is taken. Then corresponding time series plot of regular differenced is presented in Figure 4(b).

Now it seems from Figure 4(b) that, there is no trend and the p- value (0.00) of ADF test also confirms the same. Therefore, it can be concluded with 95% confidence that, there is no trend in the regularly differenced series. Now the graph of ACF is plotted in Figure 5(a) to observe the behaviour of the regular differenced series.

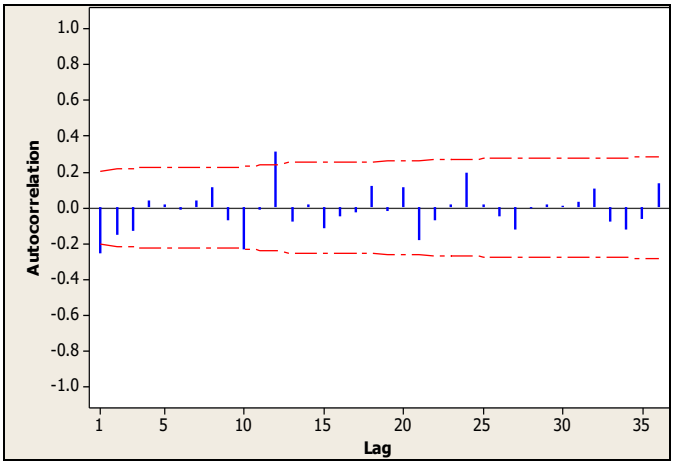

Fig 5(a): ACF of regular differenced series

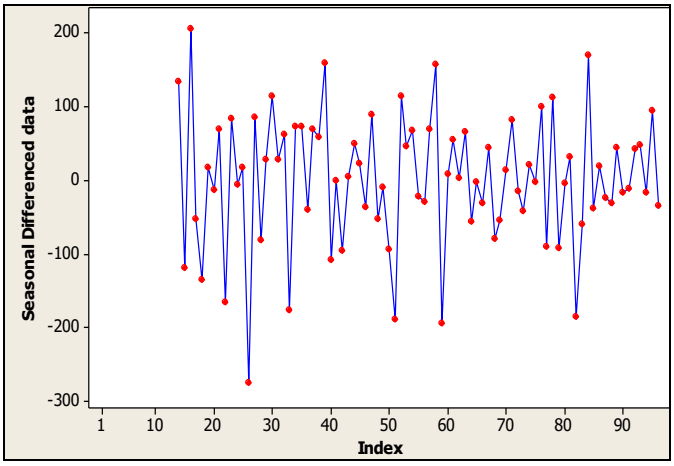

Fig 5(b): Plot for seasonal differed series

It is clearly observed from ACF plot of regular differenced series in Figure 5(a) that, ACF at lag 12 is non-significant and further, relatively, high spikes appear at 24th and 36th lags. Thus it is an indication of the existence of seasonality and also the p-value (0.00) of Kruskal-Wallis strongly confirms that the series has a seasonal pattern. Therefore, it concludes that, the regular differenced series has seasonality with length 12. To remove this seasonality, again the 12th seasonal difference is taken and relevant plots of time series, ACF and PACF are presented in Figure 5(b), Figure 6(a) and Figure 6(b) respectively.

As per the plot appears in Figure 5(b), it can be seen that, trend or seasonal or any systematic patterns do not exists in the seasonal differenced series. Further, the p- values of ADF

(0.00) and Kruskal- Wallis (0.996) tests strongly confirm that there is no trend or seasonal pattern. Therefore, it can be concluded with 95% confidence that the seasonal differenced series is stationary and now it can be used to develop the seasonal ARIMA (SARIMA) models.

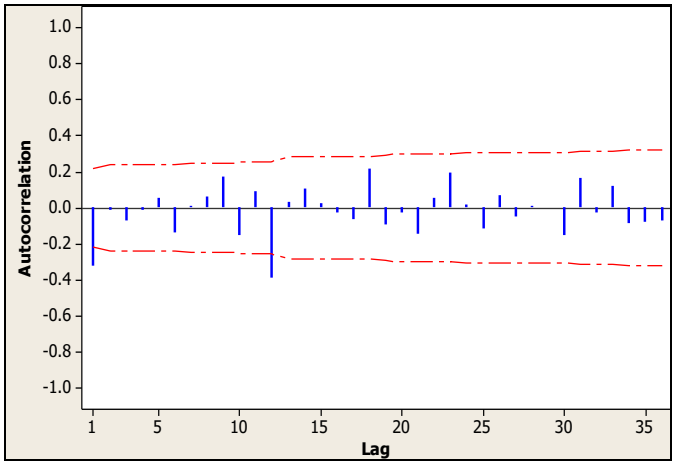

Fig 6(a): ACF of seasonal differenced series

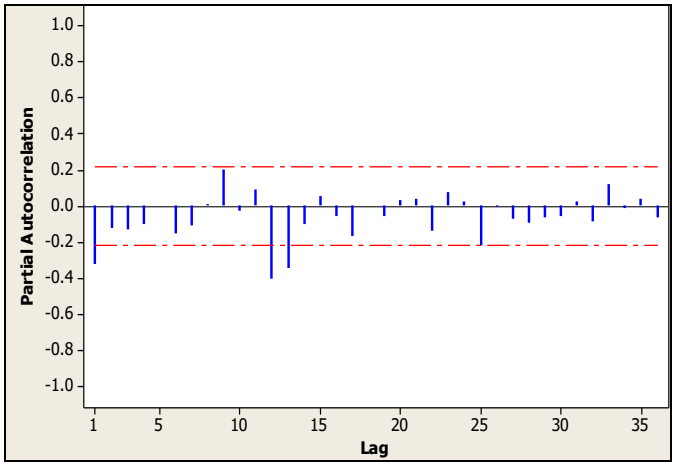

Fig 6(b): PACF of seasonal differenced series

From the plot of ACF of seasonal differenced series in Figure 6(a), all the ACF are significant except at the lags of 1 and 12. Thus it can be stated that the seasonal ARIMA model can have 2 MA terms such as MA (1) and MA (12). Similarly from the plot of PACF of seasonal differenced series in Figure 6(b), it can be claimed that, the model can have 2 AR terms such as AR (1) and AR (12) within the 1st circle of 12 lags. Accordingly, all possible models with these terms are tried out. The significant models with the p- values of the relevant tests and statistic values are summarized in Table 1.

Table 1: Significant models

Model	AD test	LM test	White's General test	AIC	SBC
SARIMA $(0,1,1)(1,1,0)_{12}$	0.18	0.12	0.92	11.23	11.34
SARIMA $(0,1,1)(1,1,1)_{12}$	0.35	0.20	0.79	11.28	11.32
SARIMA $(0,1,1)(0,1,1)_{12}$	0.56	0.28	0.30	11.34	11.40

As per the results appear in Table 1, all three models satisfy the underlined assumptions of diagnostic checking. Accordingly, all the models in Table 1 satisfy the normality test as the p-values of AD test are significant at 5% significant level. Similarly p-values of LM test and Whites general test of all the models in Table 1 are significant at 5% significant level. These confirm that all three models satisfy the conditions of independency and heteroscedasticity. Therefore, it can be concluded with 95% confidence that residuals of all three models are normally and independently distributed with constant variance. Hence, to select the best model among all 3 significant models, AIC and SBC values are taken into account.

Based on the AIC and SBC selection criterions, the third model has the highest values. Thus the third model is not taken for further calculations. However, those values in first two models in Table 1 are relatively similar. Further, on the one hand, AIC value recommends the first model is better, on the other hand, SBC value recommends the second model is better. Therefore, both models are taken to calculate the ex-ante forecast and based on the MAPE value the best model will be selected. The relevant ex-ante forecasted values of months of January, February and March in the year 2018 with observed values in parenthesis are reported in Table 2.

Table 2: Ex-ante forecast

Model	Forecasted values			MAPE
	Jan (110)	Feb (88)	Mar (175)	
SARIMA $(0,1,1)(1,1,0)_{12}$	121.0	97.4	172.5	**7.37**
SARIMA $(0,1,1)(1,1,1)_{12}$	102.6	116.2	190.3	15.84

The first model in Table 2 has the lowest MAPE value as 7.37 and hence it can be stated that the model SARIMA (0, 1, 1) $(1,1,0)_{12}$ has the highest accuracy as over 92%. Therefore SARIMA (0, 1, 1) (1, 1, 0)$_{12}$ [with the coefficients of the terms MA=0.62, SAR=-0.63] is selected as the best model to forecast the future leptospirosis outbreaks in the selected five districts of Sri Lanka. The monthly wise ex-post forecasted values of the last 3 quarter of the year in 2018 are presented in Table 3.

Table 3: Ex-post forecast

Month in 2018	Forecasted outbreaks	Month in 2018	Forecasted outbreaks	Month in 2018	Forecasted outbreaks
April	151.19	July	114.46	October	117.74
May	116.78	August	77.31	November	131.77
June	199.88	September	158.01	December	134.86

According to the monthly wise forecasted leptospirosis outbreaks in Table 3, it can be expected that in the month of June nearly 200 outbreaks and over 1200 outbreaks (1202) will be reported only in the last 9 month in the year 2018.

4. Conclusion

It can be concluded that from January 2010 to December 2017, more than 4000 outbreaks of leptospirosis per year are reported island wide in Sri Lanka. Also, during this period, nearly the half of the total outbreaks is reported from the selected five districts of Sri Lanka. This study further concludes that, leptospirosis outbreaks in Colombo, Gampaha, Kurunegala districts are decreasing while in Kalutara and

Ratnapura districts the outbreaks are in increasing trend among the selected five districts of Sri Lanka in the study period.

Over 92% accuracy of the best fitted model to forecast the future leptospirosis outbreaks in the selected five districts is SARIMA (0, 1, 1) $(1, 1, 0)_{12}$ with the relevant coefficients of the terms for MA is 0.62 and for SAR is -0.63 respectively. Hence the estimated outbreaks only in the last 9 months from Colombo, Gampaha, Kalutara, Kurunegala and Ratnapura districts of Sri Lanka in the year 2018 will be 1200.

5. References

1. Agampodi S. Peacock SJ, Thevanesam V. The potential emergence of leptospirosis in Sri Lanka. Lancet Infect Dis. 2009; 9:524-526. doi: 10.1016/S1473-3099(09)70211-7,
2. Rajasuriya K, Somasunderam M, Nagaratnam N. Leptospirosis in Ceylon. J Trop Med Hyg. 1959; 62:205-210.
3. Warnasekara JN, Agampodi S. Leptospirosis in Sri Lanka. Sri Lankan Journal of Infectious Diseases. 2017; 7(2):67-75. doi:10.4038/sljid.v7i2.8155
4. Agampodi SB, Peacock SJ, Thevanesam V, Nugegoda DB, Smythe L, Thaipadungpanit J, *et al.* Leptospirosis outbreak in Sri Lanka in 2008: lessons for assessing the global burden of disease Am J Trop Med Hyg. 2011; 85(3):471-478.
5. Epidemiology Unit, Ministry of Health, Nutrition and Indigenous Medicine. 2018. Retrieved from www.epid. gov lk
6. Agampodi SB, Dahanayaka NJ, Bandaranayaka AK, Perera M, Priyankara S, Weerawansa P, *et al.* Regional Differences of Leptospirosis in Sri Lanka: Observations from a Flood-Associated Outbreak in 2011. PLoS Negl Trop Dis. 2014; 8(1):e2626. Doi: 10.1371/journal. pntd. 0002626
7. Gnanapragasam SR. An Empirical Study on Human Leptospirosis Cases in the Western Province of Sri Lanka. OUSL Journal. 2017; 12(1):109-127. doi:10.4038/ ouslj.v12i1.7354
8. Plouffe CCF. Space-time modelling of emerging infectious diseases: Assessing leptospirosis risk in Sri Lanka. Theses and Dissertations (Comprehensive).Paper 1809. Wilfrid Laurier University. 2016; Retrieved from http://scholars.wlu.ca/cgi/viewcontent.cgi?article=2913&context=etd
9. Robertson C, Nelson TA, Stephen C. Spatial epidemiology of suspected clinical leptospirosis in Sri Lanka. Epidemiology and Infection. 2012; 140(4):731-743.
10. Denipitiya DTH, Chandrasekharan NV, Abeyewickreme W, Viswakula S, Hapugoda MD. Spatial and seasonal analysis of human leptospirosis in the District of Gampaha, Sri Lanka. Sri Lankan Journal of Infectious Diseases. 2016; 6(2):83-93.