



Time Series Analysis of Rainfall Using Seasonal ARIMA (SARIMA) and Sama Circular Model (SCM): Study from Vadamaradchi, Jaffna, Sri Lanka

T. Sellathurai^{1*}, T. Sivananthawerl², S.S. Sivakumar³, T. Mikunthan⁴ and T. Karunainathan⁵

¹Faculty of Graduate studies, University of Jaffna, Sri Lanka.

²Faculty of Agriculture, University of Peradeniya, Sri Lanka.

³Faculty of Engineering, University of Jaffna, Sri Lanka.

⁴Faculty of Agriculture, University of Jaffna, Sri Lanka.

⁵Agriculture Research Station, Thirunelveli, Jaffna, Sri Lanka.

ABSTRACT

ARTICLE INFO

Article history:

Received: 12 June 2021

Revised version received: 21 October 2021

Accepted: 01 November 2022

Available online: 1 April 2022

Keywords:

Rainfall

SARIMA

SCM

Time series analysis

Citation:

Sellathurai, T., Sivananthawerl, T., Sivakumar, S.S., Mikunthan, T. and Karunainathan, T. (2022). Time Series Analysis of Rainfall Using Seasonal ARIMA and Sama Circular Model: Study from Vadamaradchi, Jaffna, Sri Lanka *Tropical Agricultural Research*, 33(2): 234-246.

DOI: <http://doi.org/10.4038/tar.v33i2.8556>

Sellathurai T. 
<https://orcid.org/0000-0001-7576-8581>



*Corresponding author:
thusyanthi_sella@yahoo.com

The time series analysis was performed with Seasonal Autoregressive Integrated Moving Average (SARIMA) and Sama circular model (SCM) for the rainfall of Ampan, Karaveddi, and Puloly regions of Jaffna to understand the behaviour of rainfall and forecast it with a suitable model. Minitab 17 software was used to run the model with the available monthly data from 2013 to 2019. Time series plots were used for pattern recognition, the independence of the residuals was checked using autocorrelation function (ACF), and Ljung-Box Q statistics (LBQ). The normality of residuals was checked using probability plot. The model with the lowest predicting errors was selected to forecast the future values. The monthly rainfall fluctuates around the mean of 41.6, 71.9, and 35.3 mm for Ampan, Karaveddi, and Puloly respectively. The models SARIMA (0,0,0) (0,1,1)₆, SARIMA (1,2,1) (0,1,1)₆, and SARIMA (1,1,0) (0,1,1)₆ were found as most appropriate for Ampan, Karaveddi, and Puloly respectively and $Y_t = Y_{t-1} - 0.18 + 23.5 \sin 2\omega t + 28.5 \cos 1.5\omega t + 20.10 \cos 2\omega t - 26.47 \cos 5.5\omega t$, $Y_t = Y_{t-1} - 2Y_{t-2} - 5.9 + 73.5 \sin 4.5\omega t$ and $Y_t = Y_{t-1} - 2Y_{t-2} + 0.69 + 23.17 \cos 5.5\omega t$ were found as most appropriate SCM for Ampan, Karaveddi, and Puloly respectively. Among these models, SCM predicts reliable data with minimum error and it finds the seasonal and cyclic pattern of the rainfall. A five-month seasonal pattern and cyclic behaviour at 13 - months interval was noted in Ampan. Similarly, 10 - months seasonal pattern was observed in Karaveddi. The Puloly region expressed the cyclic pattern of rainfall only at 13-month interval. AD value is 0.40, 0.63 and 0.68 for Ampan, Karaveddi and Puloly respectively. The decreasing trend of estimated rainfall in Ampan (0.21 mm/year) and increasing trend in Puloly (1.15 mm/year) and Karaveddi (0.61 mm/year) is an alarming sign to the agriculture sector.

INTRODUCTION

Rainfall is a key element to meet the water needs of the world. Understanding the behaviour of rainfall is a fundamental component in climate change and the management of the demand for water (Abebe, 2018). The imbalance in the water supply is caused by increasing water demand by various stakeholders, lack of knowledge of water resources, and climate change (Uba and Bakari, 2015). Prediction of the future climate scenarios will help to reduce the undesired effect on agriculture and enable effective water resources management. Various mathematical models such as time series models have been used for predicting climate in the literature.

Time series can be defined as a set of values that occurs sequentially in time (Box et al, 2008). These models are used to detect the pattern of change in statistical information over regular intervals of time identifying the trend, seasonal, cyclical, and irregular components in a time series data. The trend is the overall tendency of the data series, and it can be either upward or downward. According to Konarasinghe, 2020, the time series, which follows a wavelike pattern and consists of less than 12 monthly data shows a seasonal variation, and if this wave oscillation consists of greater or equal to 12 monthly data will show a cyclic variation.

Several techniques are available for modelling and forecasting rainfall. Decomposition technique and Autoregressive Integrated Moving Average (ARIMA)/ Seasonal Autoregressive Integrated Moving Average (SARIMA) are some of the widely used techniques in climate-related studies. These models have been applied in various water resources and environmental management (Nail and Momani, 2009) studies. The ARIMA introduced by Box and Jenkins in 1970 is a relatively popular technique used for modelling the time series and rainfall forecasting due to easiness in its development and implementation. Several studies have been conducted to understand the rainfall and temperature using ARIMA and SARIMA. Dimri et al (2020) studied the time series of temperature and precipitation data and performed the seasonal analysis of monthly mean temperatures and precipitation in the Bhagirathi river basin in Uttarakhand, India. Partheepan et al 2005 used this modelling technique to evaluate the dynamics of weather patterns and to model the environmental consequences of the Batticaloa district of Sri

Lanka. In these models, the future value of a variable is modelled as a linear combination of past values and present errors.

These linear processes include Auto Regressive (AR), Moving Average (MA), Auto Regressive Moving Average (ARMA) and ARIMA. It can be expressed as ARIMA (p, d, q). Where, p, d and q are the autoregressive, differencing and moving average component respectively. The SARIMA model can be expressed as ARIMA (p, d, q) (P, D, Q) s. Where, P, D, Q, s represents the seasonal autoregressive, seasonal differencing, seasonal moving average parameters, and number of periods per season respectively.

The Sama circular model (SCM) technique introduced in the year 2018 can detect the trend, seasonal and cyclic variation, and differentiate among several seasonal and cyclic patterns in a data series. It is based on the Fourier Transformation (FT), Least Square Transformation (LST), and differencing technique (Konarasinghe, 2020). The model expresses the movement of a particle with an angular speed of ω in a horizontal circle with the radius of α (Figure 1). Equation 1 explains the trigonometric function of this motion.

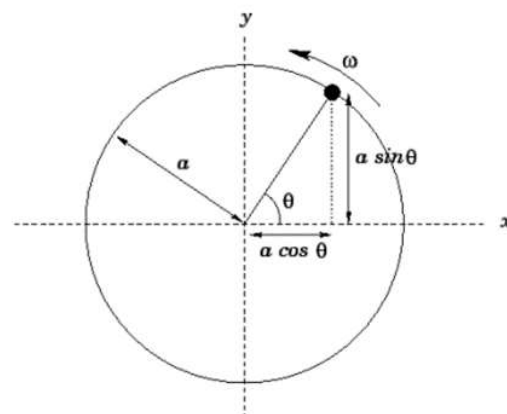


Figure 1: Motion of particle in horizontal circle (source: Konarasinghe, 2020).

$$(1 - B)^d Y_t = \sum_{k=1}^n (a_k \sin k\omega t + b_k \cos k\omega t) + \varepsilon_t \quad (1)$$

Where;

B - Back shift operator, d - dth order of difference, Y_t - Random variable at time t, a - Radius of the circle, ω - angular speed of the

particle, k - variable, t – time and ε - random error. The model assumes Y_t is a continuous random variable, $t \geq 0$, k is positive variable, $\sin \omega t$ and $\cos \omega t$ are independent and the ε is normally distributed and independent.

The study area, the Jaffna peninsula falls under the dry zone of Sri Lanka and covers 1025km² (Department of Agriculture, 2003). The average temperature of this area varies from 28.3°C to 32.2°C during the dry season and 25°C to 27.7°C during the wet season. The bimodal rainfall pattern received in this area is used to meet the water needs and recharge the aquifers. Vadamaradchi aquifer is one of the major aquifers which consists of three agrarian services divisions (ASD) namely Puloly, Ampan, and Karaveddi represents the administrative divisions Vadamaradchi North, Vadamaradchi East, and Vadamaradchi southwest, respectively. Sellathurai, T (2020) reported that the three ASDs dominated by the southwest monsoon (SWM) and followed by the Northeast monsoon (NEM). Nearly 34 – 42% of the rainfall was received during SWM 25 - 31% of the rainfall was received during NEM.

The location and the livelihood of the people who live in this area may create undesirable effects on the groundwater. Over extraction of groundwater to meet the higher demand of agriculture may create saline water, induce seawater intrusion and reduce water availability. Therefore, understanding the current and future rainfall patterns is prime important for this region to sustainable utilization of groundwater.

This study aims to understand the behaviour of short-term rainfall patterns of ASDs of Vadamaradchi using Box-Jenkins SARIMA and SCM modelling technique and compare the results to choose a suitable model for forecasting.

METHODOLOGY

Data Collection and analysis

The analysis consists of rainfall data collected from three ASDs of Vadamaradchi from 2013 to 2019 were used in this study. The monthly rainfall details were processed from the daily data collected. Altogether, 84 monthly observations were considered from each center. The Minitab 17 software was used to analyse the data. The outliers of the data were identified using the Box plot method and replaced by the

mean of order three. The first sixty observations were used for estimation, and the last twenty-four observations were used for verification of the model through forecasting.

Modelling of rainfall

The fitting of the ARIMA model consists of three phases; model identification, parameter estimation, and diagnostic checking (Karim et al, 2018). The steps in the model fitting, validation, and forecasting of future rainfall data using the ARIMA model are given in Figure 2. The first step of ARIMA is checking the stationarity of the data, where values of variables vary around the constant means and variance over time. The differencing technique was applied to make the data series stationary with the consideration of not over differencing the data. The correlograms of the data were used to check the stationarity of the rainfall. Based on the autocorrelation function (ACF), these plots are called autocorrelation functions. Because they show the degree of correlation with past values of the series as a function of the number of periods in the past (that is, the lag) at which the correlation computed; and Partial autocorrelation function (PACF) the autoregressive (AR) and moving average (MA) values were placed in the model (using trial and error method), and the best-fitted model was selected for testing. Table 1 express the behaviour of ACFs and PACFs of the ARIMA process.

The model fitting using the SCM method also consists of three phases: pattern recognition, model fitting, and validation. The time series plot and ACF were used for pattern recognition. The steps involved in this analysis are shown in Figure 3. After removing outliers of the data, the differencing was done to the data series to process SCM. This helps to find out the cycles of oscillation easily and facilitates counting the number of peaks. In this study, the SCM was applied on the first differenced data of rainfall for Ampan, and the second differenced data for Karaveddi and Puloly. The number of peaks was counted above the reference line (in this case it was zero).

These peak values are applied in equation 2 below and used to find out the ω of the oscillation. The 12 trigonometric series for $\sin k\omega t$ and 12 trigonometric series for $\cos k\omega t$ were obtained and used in the regression analysis.

$$\omega = (2\pi \times \text{number of peak}) / (\text{number of observation}) \quad (2)$$

Table 1: Behaviour of ACFs and PACFs in ARIMA process.

Process	ACFs	PACFs
ARIMA(0,0,0)	No significant lags	No significant lags
ARIMA (0,1,0)	Linear decline at lag 1 with many lags significant	Significant spike at lag 1
ARIMA (1,0,0) $\phi_1 > 0$	Exponential decline, with first two or more lags significant	Single significant spike at lag 1
ARIMA (1,0,0) $\phi_1 < 0$	Alternating exponential decline, with a negative ACF(1)	Single significant negative spike at lag 1
ARIMA (0,0,1) $\theta_1 > 0$	Single significant negative spike at lag 1	Exponential decline, with first few lags significant
ARIMA (0,0,1) $\theta_1 < 0$	Single significant positive spike at lag 1	Alternating exponential decline at lag 1

Three error measures (predicting error); Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Standard Deviation (MSD) were used to determine the adequacy of the model. The two models forecasting performance were compared with some frequently used error measures, MSE and MAD (Petris et al., 2009). The model selection criterion states that the model with the smallest (minimum) MAPE and MAD is considered to be the best (Sedghi et al, 2018). The ACF and Ljung – Box Q statistics (LBQ) were used to test the independence of the residuals. The probability plot and the Anderson Darling test were used to test the normality of residuals.

RESULTS AND DISCUSSION

Modelling and analysis using SARIMA

The time series analysis of the data shows that rainfall fluctuates around the mean of 41.6 ± 51.97 , 71.9 ± 83.15 and 35.3 ± 43.94 mm for Ampan, Karaveddi and Puloly respectively (Figure 4). The time series plot expressed that Puloly and Karaveddi area shows the increasing rate of rainfall while Ampan shows the

decreasing rate of rainfall. The higher standard deviation expresses the high range of variability of rainfall. Precipitation reaches its maximum in the Maha season (October to February). The sinusoidal pattern of the ACF plots that the series under study is non-stationary and first lag significant in all three ASDs in Figure 4, and there is seasonality. Therefore they need to be stationaries, and the seasonality needs to be removed. The differencing technique was applied to make the series stationary. The first-order difference ($d=1$) for Ampan and the second-order difference ($d=2$) for Karaveddi and Puloly made the data series which distributed around the mean (stationary).

By considering ACF and PACF plots of differenced data series, the AR, MA terms were determined. Afterward, seasonality (P, D, and Q) was removed like the p, d, and q. The models SARIMA (0,0,0) (0,1,1)₆, SARIMA (1,2,1) (0,1,1)₆, and SARIMA (1,1,0) (0,1,1)₆ were found as most appropriate for Ampan, Karaveddi and Puloly respectively. After determining the forecast model parameters, the applicability of the model was checked, and the stationarity of the residuals was confirmed.

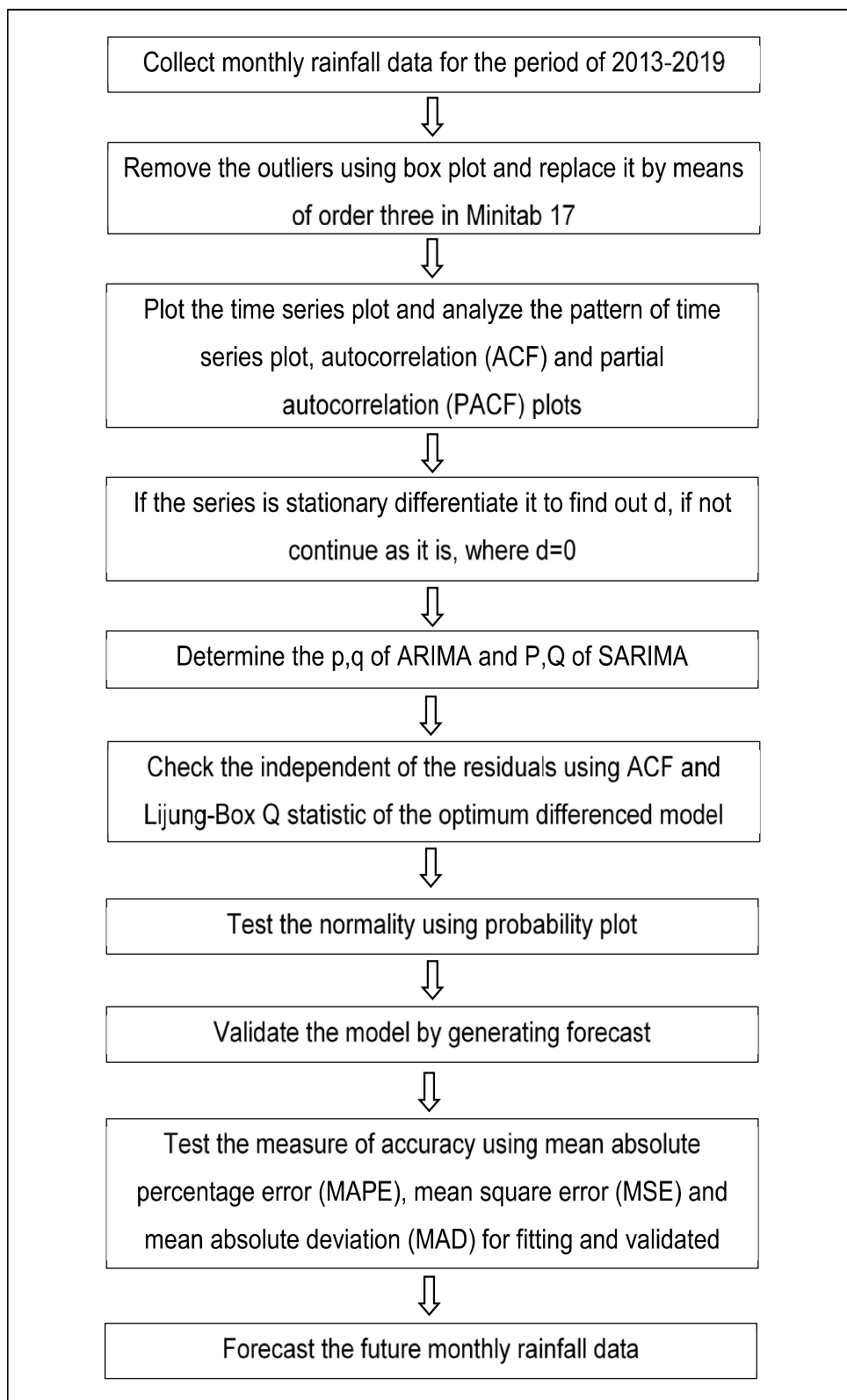


Figure 2: The detail methodology of ARIMA/SARIMA model fitting, validation and forecasting.

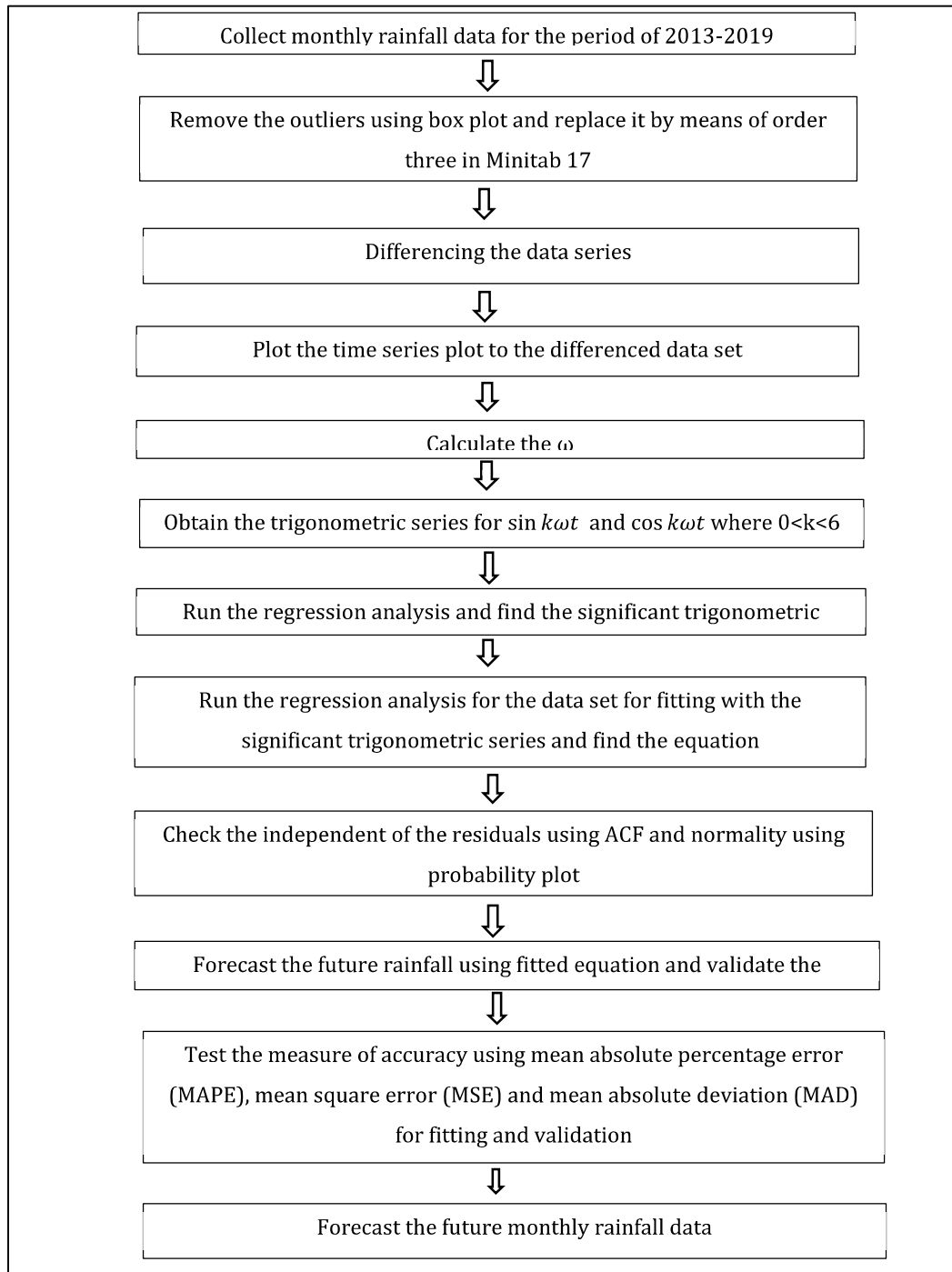


Figure 3: The detail methodology for SCM fitting, validation and forecasting.

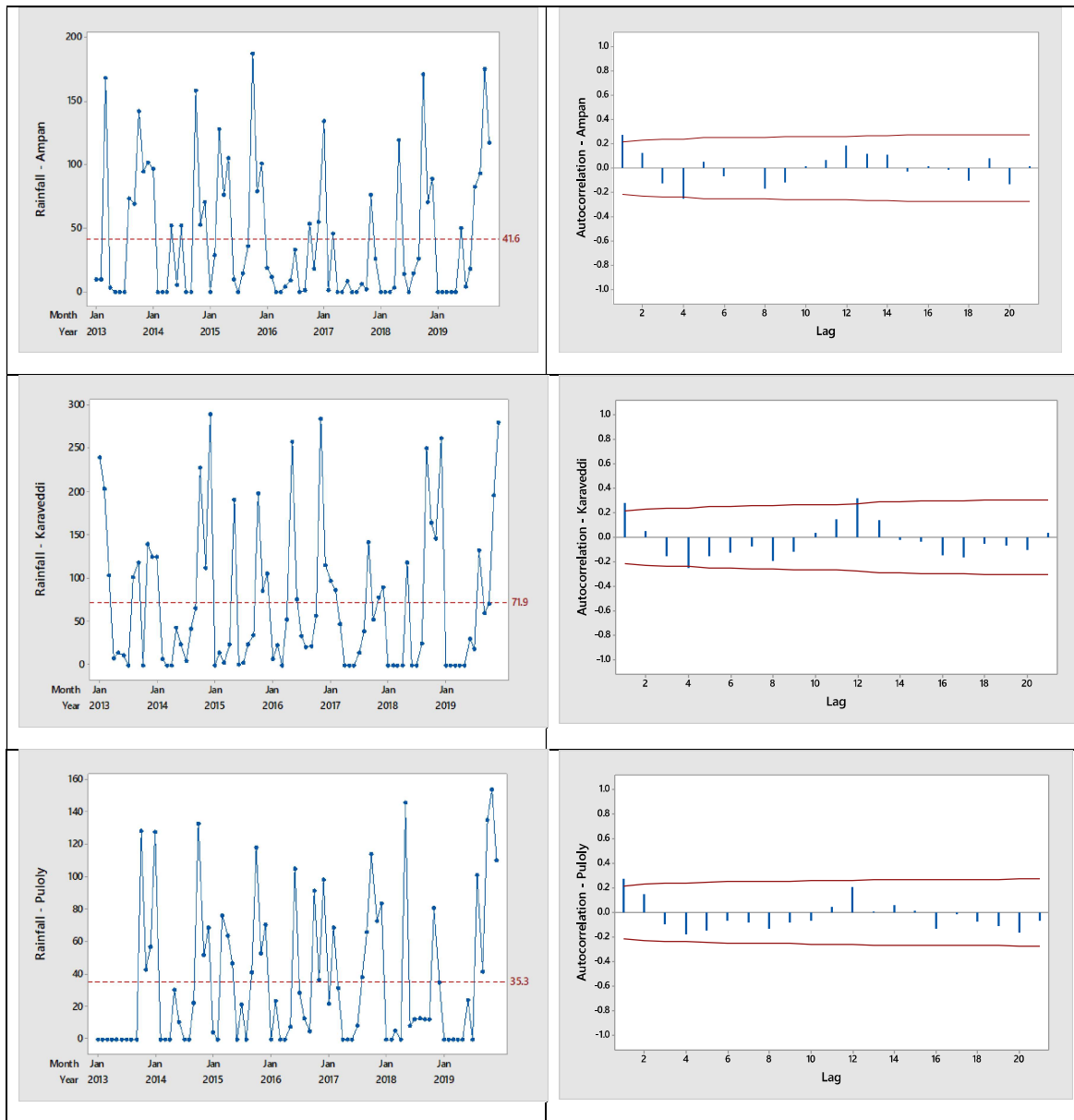


Figure 4: The time series plots and Correlograms of three ASDs.

Figure 5 expresses the time series plots for rainfall and SARIMA forecasted fits and the probability plot of residuals. The data sets developed through the model gave the independent and normally distributed residuals for fitting for three ASDs. However, in some periods, the predicted rainfall deviated from the actual pattern. It was noted in Karaveddi and Puloly area during the end of the Maha season. It is confirmed by the probability plots where the means are distributed close to the linear line implying the stationarity at a 95% confidence level.

Fitting values are close to the actual value of the rainfall data, and they follow a similar pattern of actual data in most of the events. The Anderson-Darling (AD) goodness of fit value is less, and the P-value is greater than 0.05 in all cases. The measurement of errors is also satisfactory small (Table 2). Therefore, the SARIMA can be used to forecast future values.

Modelling and analysis using SCM

Figure 6 shows the time series plot of the second differenced data of Puloly. Where 13 peaks were observed in the complete waves in the series. The ω value obtained was 0.99, 1.15, and 1.82

for Puloly, Karaveddi, and Ampan respectively. By using this value, the ωt was estimated and the 12 trigonometric series was developed; $0.5 \sin \omega t$ to $6 \sin \omega t$ with the increment of 0.5° . Similarly, $\cos \omega t$ also estimated and significant trigonometric series were found using the P-value of each series in the regression analysis.

The model was run with significant trigonometric series and best fitting SCM was selected as indicated in equations 3, 4 and 5 for Ampan, Karaveddi, and Puloly respectively. The SCM considered four trigonometric functions $\sin 2\omega t$, $\cos 1.5\omega t$, $\cos 2\omega t$ and $\cos 5.5\omega t$ for Ampan, $\sin 4.5\omega t$ for Karaveddi and

$\cos 5.5\omega t$ for Puloly to fit the model. The measurements of errors for all three models are satisfactory small (Table 2), and the residuals are random and normally distributed. Therefore, the model can be applied to forecast future values.

$$Y_t = Y_{t-1} - 0.18 + 23.5 \sin 2\omega t + 28.5 \cos 1.5\omega t + 20.10 \cos 2\omega t - 26.47 \cos 5.5\omega t \quad (3)$$

$$Y_t = Y_{t-1} - 2Y_{t-2} - 5.9 + 73.5 \sin 4.5\omega t \quad (4)$$

$$Y_t = Y_{t-1} - 2Y_{t-2} + 0.69 + 23.17 \cos 5.5\omega t \quad (5)$$

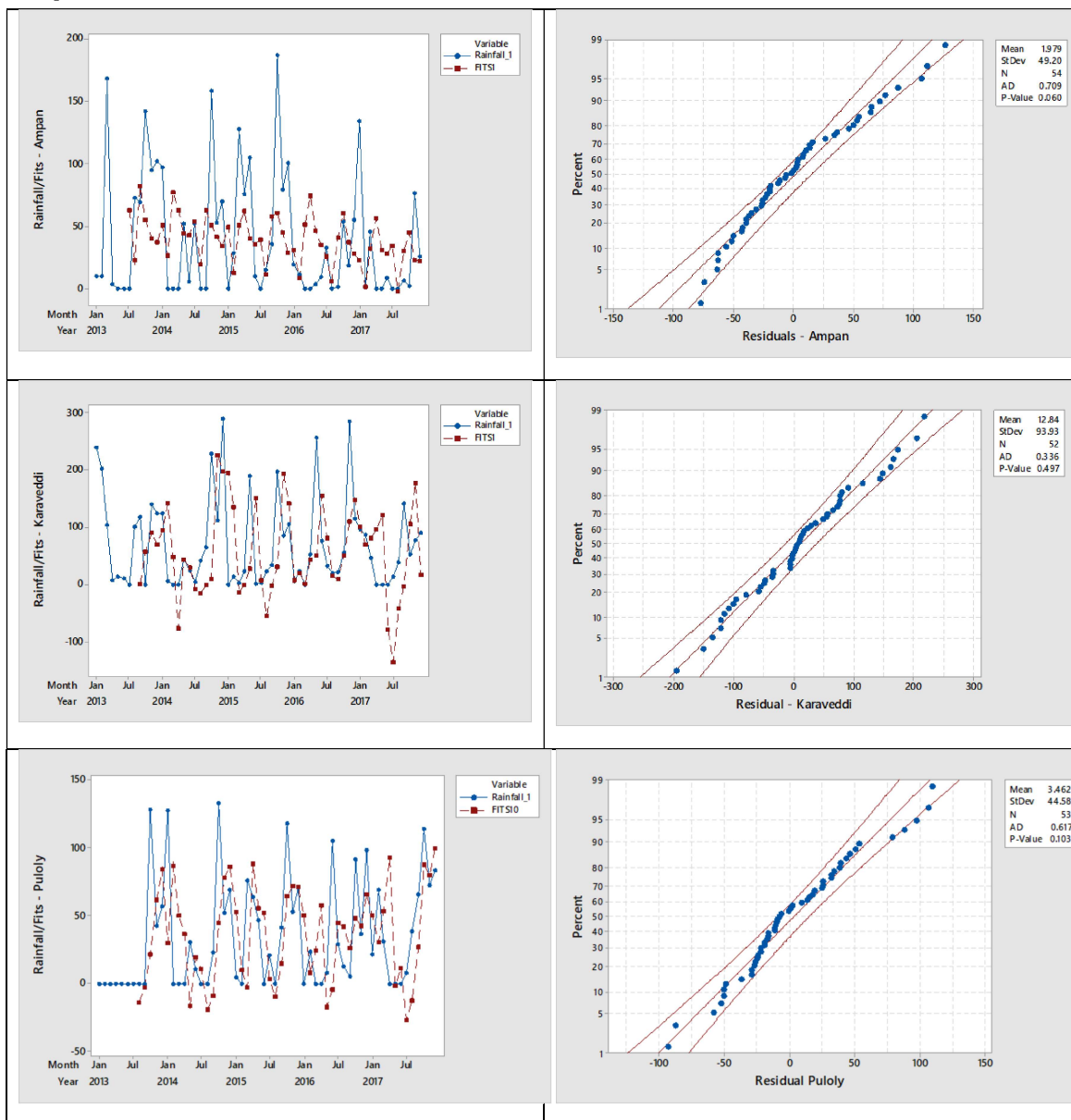


Figure 5: The time series plot for rainfall and fits (Left side) and the probability plot of residuals (right side) of SARIMA.

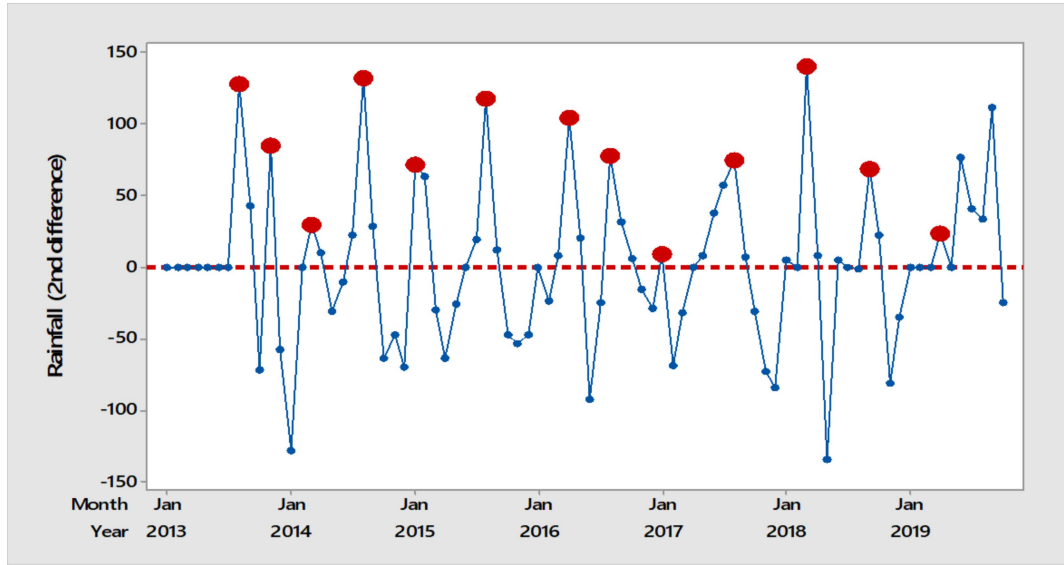


Figure 6: The time series plot of 2nd differenced data of Puloly.

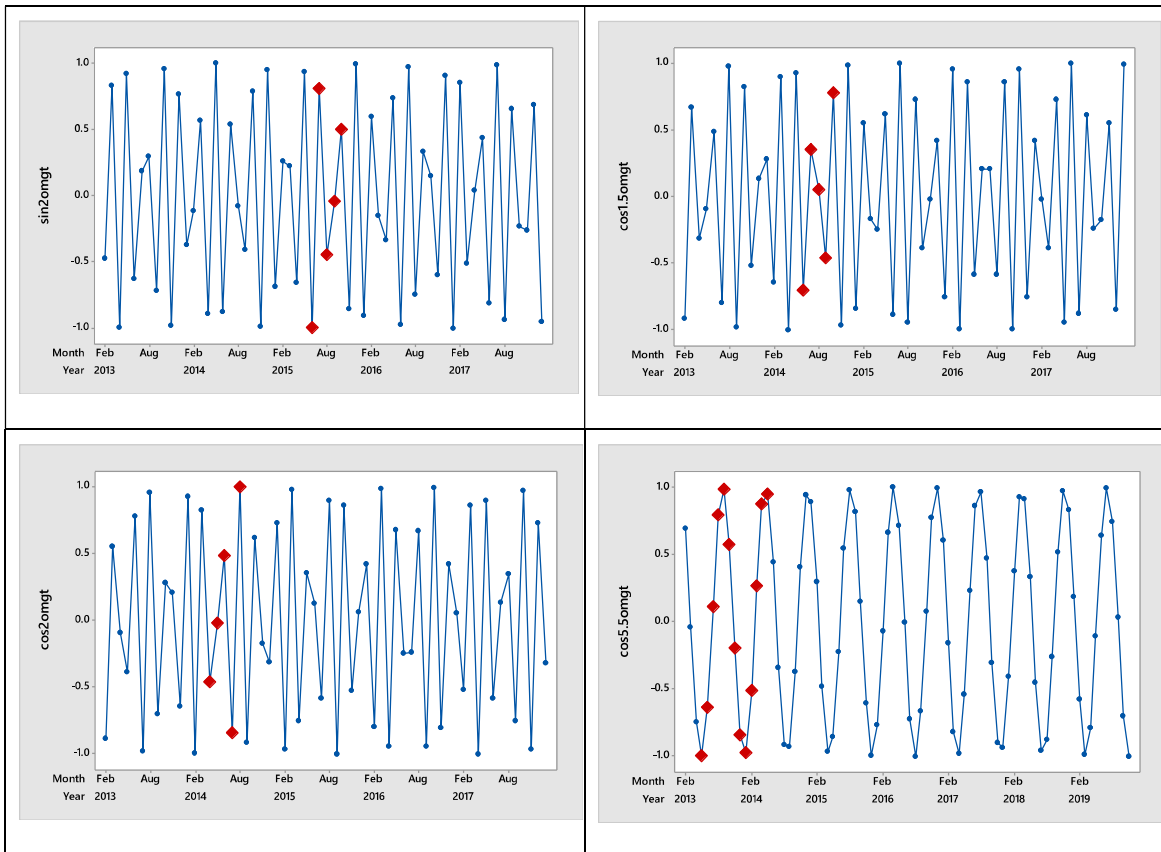


Figure 7: The time series plot of significant trigonometric series.

As indicated in Figure 7, the period of oscillation was counted as 5 and 13 months in Ampan. That means rainfall shows a seasonal rainfall pattern at 5-months and shows a cyclic pattern at the 13-month interval. The two major monsoon

seasons are clear, where the peak of the graph was observed. The cyclic pattern expresses the change in the onset of rainfall in this region from the Sri Lankan standard conditions.

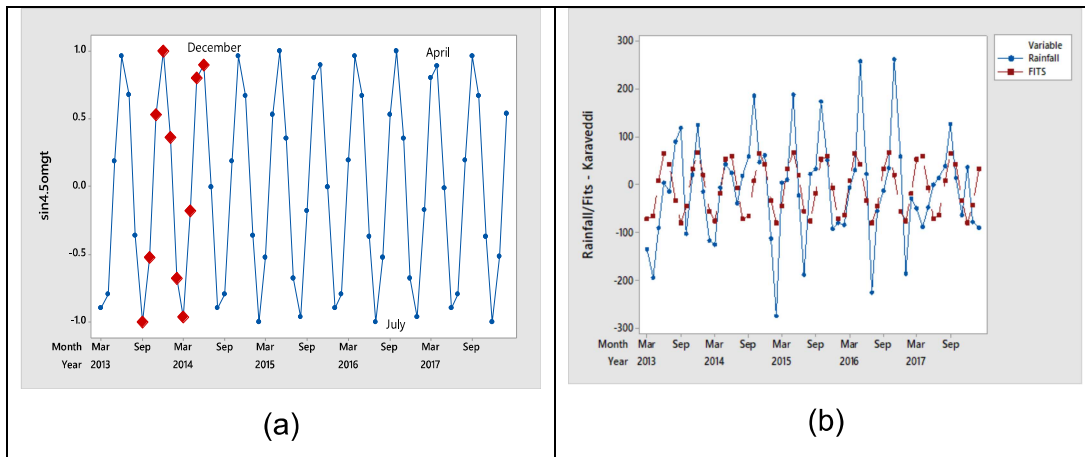


Figure 8: The time series plot of $\sin 4.5\omega t$ (a) and actual and fits of rainfall (b) for Karaveddi.

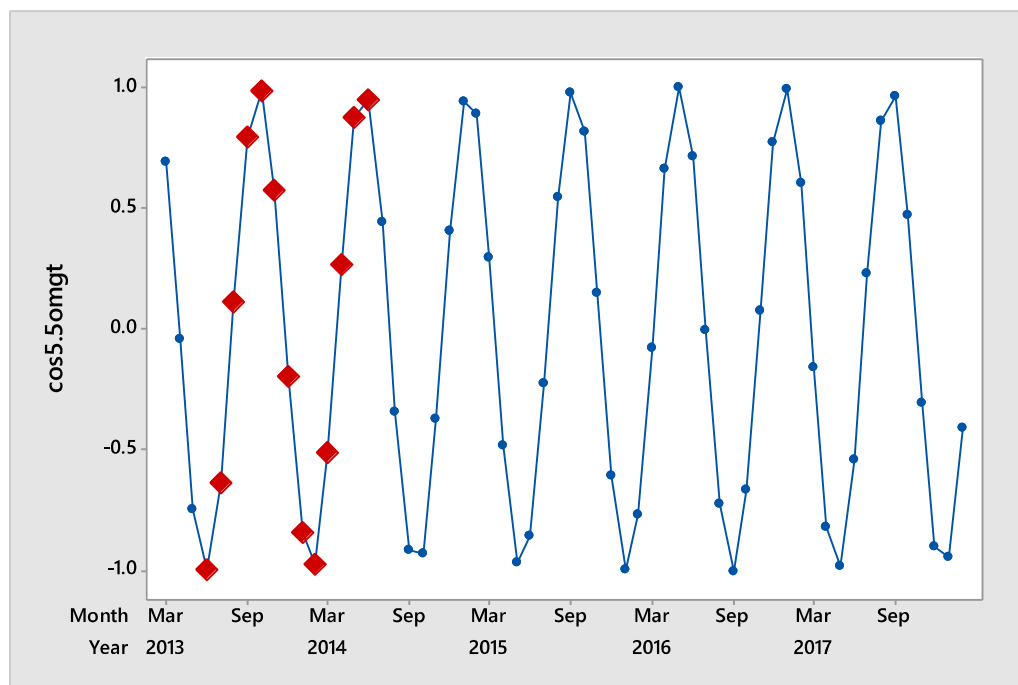


Figure 9: The time series plot of significant trigonometric $\cos 5.5\omega t$.

Table 2: Fitted model and measurement of accuracy for SARIMA and SCM.

Division	SARIMA				SCM			
	Fitting		Validation		Fitting		Validation	
	MSE	MAD	MSE	MAD	MSE	MAD	MSE	MAD
Ampan	2379.8	38.6	3899.7	43.5	2405.5	39.6	3601.9	49.0
Karaveddi	8818.1	72.3	34891.3	462.3	8663.8	74.1	17279.5	101.1
Puloly	1962.3	34.6	4125.1	60.4	2691.6	38.8	3107.5	38.6

The second differenced data were used in modeling the Karaveddi center with SCM. The significant trigonometric series is the $\sin(4.5\omega t)$ for Karaveddi. The seasonal changes were noted in the 10-month interval in this region (Figure 8 (a)). July month received less amount of rainfall while a higher amount of recorded during the month of December. Sellathurai, T (2020) observed a similar trend. The model has

graphically confirmed its suitability, and Figure 8 (b) expresses a similar peak and trough of rainfall with the fits of actual rainfall of the Karaveddi region. The Puloly region with no seasonal behavior, but the rainfall event has the cyclic pattern at the 13-month interval. The cyclic pattern may be due to the change or shift in the onset of rainfall.

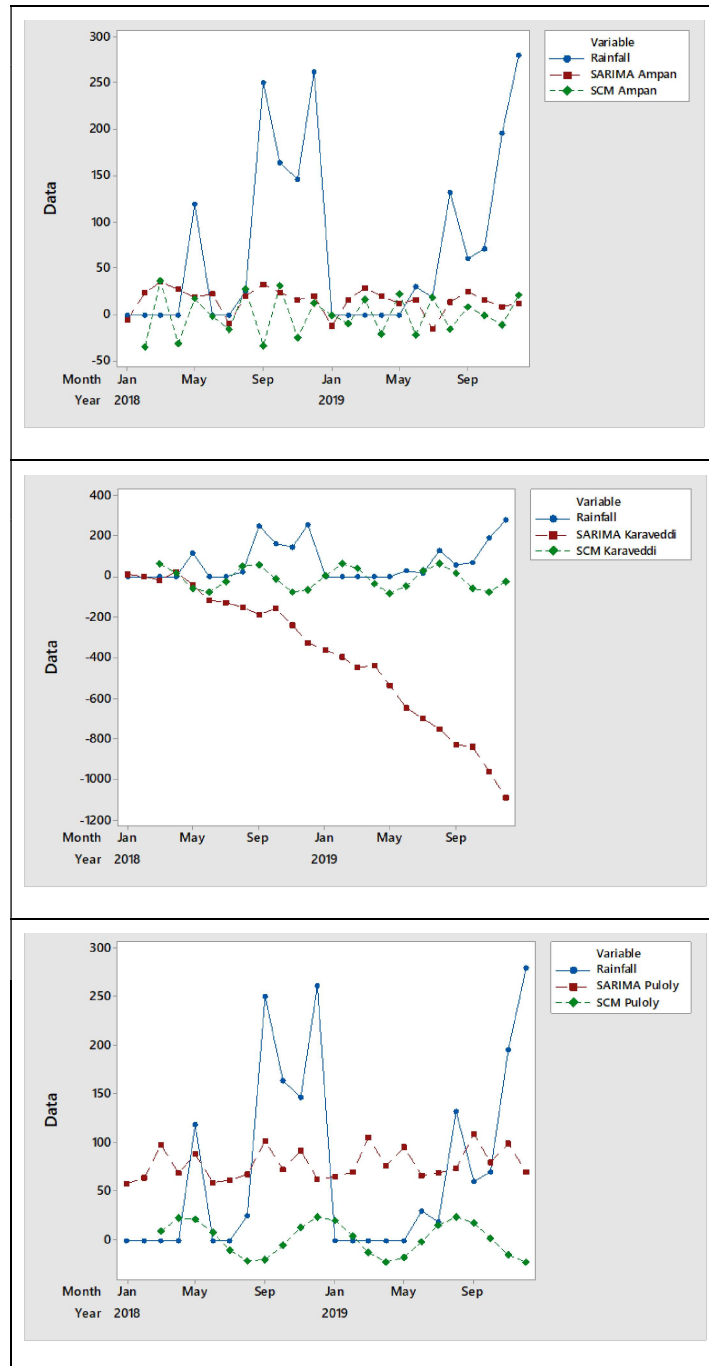


Figure 10: The Forecasting ability of SARIMA and SCM.

Comparison of SARIMA and SCM

The rainfall values were forecasted for the period between January 2018 and December 2019 with the determinant models as denoted by equations 3, 4, and 5 respectively for each ASDs.

The measurement of errors for fitting in SARIMA and SCM is nearly equal in Ampan and Karaveddi regions, but for validation, it is less in SCM in all cases (Table 2). Therefore, the SCM is more suitable than SARIMA in forecasting the future values of these regions' rainfall data.

The observed rainfall data for the period March 2018 to December 2019 and the estimated value to the same period as indicated in Figure 10. Except for Karaveddi's SARIMA, both models

follow the same pattern of original data. However, the SARIMA is somewhat overestimating the future value, and SCM's forecasting is nearly equal in all cases. The over-estimation may be due to the over differencing of the data, especially in Karaveddi and Puloly. The estimated rainfall shows a decreasing trend in Ampan 0.21 mm/year and an increasing trend in Puloly 1.15 mm/year and Karaveddi 0.61 mm/year.

P-Value of correlation between actual and forecasted rainfall is 0.56 for Ampan, 0.15 for Karaveddi, and 0.21 for Puloly. The Probability plot and the Anderson Darling (AD) test were used to test the normality of the modeled rainfall using SCM; the AD value for Ampan is 0.40, Karaveddi is 0.63, and Puloly is 0.68.

CONCLUSIONS

From this study the following conclusions can be made;

1. Among the SARIMA models tested, the model SARIMA (0,0,0) (0,1,1)₆, SARIMA (1,2,1) (0,1,1)₆, and SARIMA (1,1,0) (0,1,1)₆ gave the independent and normally distributed residuals for fitting for Ampan, Karaveddi, and Puloly respectively.
2. Since the data used in the study contain the non-stationary pattern at a 95% confidence level, the differencing is needed for the SARIMA process. For Ampan, first differencing and for Puloly and Karaveddi second differencing gave the stationary data.

3. The models, $Y_t = Y_{t-1} - 0.18 + 23.5 \sin 2\omega t + 28.5 \cos 1.5\omega t + 20.10 \cos 2\omega t - 26.47 \cos 5.5\omega t$, $Y_t = Y_{t-1} - 2Y_{t-2} - 5.9 + 73.5 \sin 4.5\omega t$ and $Y_t = Y_{t-1} - 2Y_{t-2} + 0.69 + 23.17 \cos 5.5\omega t$ are suitable to forecast the values of Ampan, Karaveddi, and Puloly respectively using SCM.

4. Using SCM is the most appropriate model to predict the rainfall of Ampan, Karaveddi, and Puloly agrarian centers. The center Ampan and Puloly have the cyclic pattern of rainfall, while 5-months and 10-months seasonal patterns are noted in Ampan and Karaveddi regions respectively.

5. The decreasing trend in the Ampan at the rate of 0.21 mm/year is an alarming sign to agriculture. Meantime the Puloly 1.15/year and Karaveddi 0.61/year regions have an increasing trend of rainfall in the following years. Therefore, the farmers need to plan their crops accordingly.

ACKNOWLEDGEMENTS

This work was financially supported by WaSo project of University of Jaffna.

REFERENCES

- Abebe, S. A. (2018). Application of Time Series Analysis to Annual Rainfall Values in Debre Markos Town, Ethiopia. *Computational Water, Energy, and Environmental Engineering*, 7, 81-94. <https://doi.org/10.4236/cweee.2018.73005>
- Box, G. E. P.; Jenkins, G. M., & Reinsel, G.C. (2008): *Time Series Analysis: Forecasting and Control*; 4th Edition; John Wiley & Sons; Inc. U.K.
- Dimri, T., Ahmad, S., & Sharif, M. (2020). Time series analysis of climate variables using seasonal ARIMA approach. *Journal of Earth system science*. Pp 129-149. <http://doi.org/10.1007/s12040-020-01408-x>
- Karim, H.M., Hossein, S., Abdolrasoul, T., & Hossein, B. (2018). Modeling climate variables of rivers basin using time series analysis (case study: Karkheh river basin at Iran). *Civil Engineering* 4(1), .78-92
- Konarasinghe, S. (2020). Comparison of forecasting ability of SAMA circular model, ARIMA and decomposition techniques.

- Journal of new frontiers in mathematics and statistics. Vol1, Issue 1. Pp1-15.
- Distric Sri Lanka; Water Professionals' Day Symposium.
- Machekposhti, K.H., Sedghi, H., Telvari, A., & Babazadeh, H. (2018). Modeling climate variables of rivers basin using time series analysis (Case study: Karheh river basin at Iran). *Civil Engineering Journal*. Vol. 4 <http://dx.doi.org/10.28991/cej-030970>
- Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic Linear Models* with R. Springer, New York. pp 93
- Naill, P. E., & Momani, M. (2009). Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis, *American Journal of Environmental Sciences* 5(5): 599 -604, ISSN 1553-345X
- Sellathurai, T., Mikunthan, T., Sivakumar, S.S., & Karunainathan, T. (2020). Seasonal variation of rainfall in Vadamaradchi area in Jaffna District Sri Lanka. 6th International conference on dry zone agriculture (ICDA 2020). pp30
- Partheepan, K., Jeyakumar, P., & Manobavan, M. (2005). Development of a time-series model to forecast climatic data in the Batticaloa
- Uba, E.S., & Bakari, H.R. (2015). An application of time series analysis in modeling monthly rainfall data for Maiduguri, North Eastern Nigeria. *Mathematical theory and modeling*. 5(11), 2015