Analysing the temperature of Sri Lanka: SARIMA approach

D. M. T. N. Jayawardana[a]*, A. Jahufer[b]

Department of Mathematical Sciences, Faculty of Applied Sciences,
South Eastern University of Sri Lanka

([a]thanu.nawo.95@gmail.com, [b]jahufer@seu.ac.lk)

Keywords: AIC, Box-Jenkins, forecasting, modeling, SARIMA, SIC, temperature.

Introduction

The main objective of this research study is to fit a suitable time series Seasonal Autoregressive Integrated Moving Average (SARIMA) model and forecast Monthly Average Temperature (MAT) of Sri Lanka.

Sri Lanka is a small Island between 6^0 N and 10^0 N latitude and 80^0 E and 82^0 E longitude in the Indian ocean. The Average Temperature (AT) of Sri Lanka usually ranges from 28^0 C to 32^0 C. The MAT of Sri Lanka differs slightly depending on the seasonal movement of the sun, with some modified influence caused by rainfall. The coldest months according to the mean MT are December and January while the warmest months are April and August.

SARIMA model was used to analysis MAT in Sri Lanka. Many time series methods can be used to predict temperature among them, many researches have used SARIMA model for forecasting temperature.

[1] applied the SARIMA model for the AT and rainfall and forecasted Monthly Temperature (MT) and rainfall in Ambo Area, Ethiopia.

[2] forecasted mean temperature of Junagadh city, Gujarat using past data from the period of 1984 to 2015. They used the Box-Jenkins time series SARIMA approach for forecast mean temperature.

Materials and Methods

1. Data collection. The MAT data recorded from January 1990 to December 2020 were obtained for Sri Lanka, from the World Bank report. The temperature data were measured in Celsius (^{0}C).

2. Statistical Analysis

Stationary test. The ADF unit root test is carried out to determine whether the series is stationary or non-stationary.

3. Model Specifications

ARMA model. ARMA processes are a combination of Autoregressive (AR) and Moving Average (MA) processes. ARMA model of order p and q process is given by:

$$Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (1)$$

Autoregressive Integrated Moving Average (ARIMA) model. ARIMA model is a generalization of an ARMA model. The ARMA models can be used for stationary time series data. If the series is non-stationary, it can be stationary using by differencing and hence the term "integrated" can be used.

The ARIMA (p, d, q) model can be written as:

$$Y'_t = c + \varphi_1 Y'_{t-1} + \varphi_2 Y'_{t-2} \ldots + \varphi_p Y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_p \varepsilon_{t-p} + \varepsilon_t \qquad (2)$$

where, p and q are order AR and MR component respectively, d is the number of differencing, c is constant term and Y'_t is the differenced time series value of at time t and ε_t is the error term.

SARIMA model.

The SARIMA model can be written as:

$$\text{SARIMA}\ \underbrace{(p, d, q)}_{\uparrow\ \boxed{\text{Parameter of Non-seasonal Part}}}\ \underbrace{(P, D, Q)}_{\uparrow\ \boxed{\text{Parameter of Seasonal part}}}S$$

where, p and q are Non-Seasonal (NS) AR and MA order respectively, d = NS differencing, P and Q are seasonal AR and moving average order respectively, D = seasonal differencing and S is number of periods per season.

The SARIMA model is given by:

$$\varphi_p(B)\,\Phi_P(B^S)(1-B)^d\,(1-B^S)^D\,Y_t = \delta + \theta_q(B)\,\Theta_Q(B^S)\varepsilon_t \qquad (3)$$

4. Lag length selection

Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF). ACF and PACF plots of the series are necessary to determine the order of AR and MA terms. ACF can be used to determine the optimal lags of MA(q) terms and PACF can be used to determine the optimal lags of AR(p) terms.

5. Model selection. Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) can be used to compare different possible models and determine which one is the best model of among tentative models using their least values.

Results and Discussion

1. Preliminary analysis. The time series plot for the MAT data for Sri Lanka is presented in Figure 1. It shows that the temperature data does not indicate trend pattern.

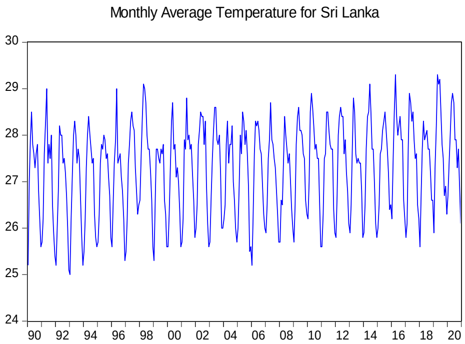

Figure 1. Plot for MAT from Jan. 1990 to Dec. 2020.

To identify the behavior of the temperature data, the seasonal graph (Figure 2) of data was generated. According to the figure 2, each month has a different average behavior, which is what characterizes a seasonal series.

2. Stationarity test. The Augmented Dicky Fuller (ADF) test is used to check whether the series is stationary or not at level. The results are shown in Table 1.

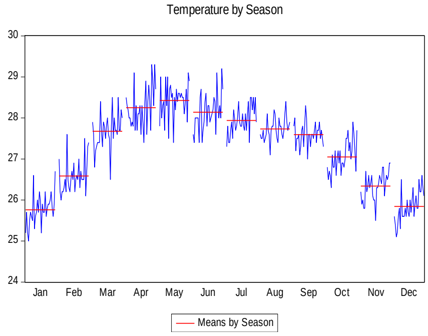

Figure 2. Seasonal plot of the temperature data.

Table 1. Results of unit root test for AT.

Null Hypothesis: The MAT is non-stationary		
	t-Statistics	Prob.
ADF test statistic	-4.8615	0.0004
Test critical values: 1% level 5% level 10% level	-3.9840 -3.4225 -3.1341	

Table 1 shows that, the ADF test value -4.861488 is less than the critical values of the significance level of 0.01. Also, the p-value is less than the 0.05 significant level. Therefore, that the null hypothesis can be rejected. Hence, the temperature series is stationary at level.

3. Model identification. The first step of the Box-Jenkins method is identification of appropriate model. The SARIMA models can be used to forecast MAT of Sri Lanka. The orders of the SARIMA model parameters can selected using the plots of ACF and PACF. It is confirmed that the ACF at lags 1, 2 and 3 are significant since the bars passes out of the confidence limits. Therefore, the order of the NS MA term is 3. The seasonal term, in ACF only lag 12 bar is significant. So, the order of seasonal MA term is 1. Similarly, a significant bar at lag 1 and 2 in the PACF indicates possible NS AR terms. The seasonal term, in PACF lag 12, 24 and 36 bars are significant. Hence, the order of seasonal AR term is 3.

The results of Table 2 show that the tentative models for mean temperature. Out of the ten models, SARIMA (2 0 1)(0 1 1)$_{12}$ model generated the lowest AIC and SIC values, so, it can be confirmed that among the ten models the best model for AT is SARIMA(2 0 1)(0 1 1)$_{12}$.

4. Parameter estimation. Table 3 shows that the estimated coefficients of AR, MA, SAR, and SMA of SARIMA (2 0 1)(0 1 1)$_{12}$ model.

According to the results of Table 3, there are four model parameters are significant, except NS AR (1) of seasonal ARIMA model.

Table 2. Tentative models for AT.

Model	AIC	SC
SARIMA (1 0 1)(1 1 1)$_{12}$	0.716867	0.781635
SARIMA (1 0 1)(2 1 1)$_{12}$	0.712600	0.788163
SARIMA (1 0 2)(2 1 1)$_{12}$	0.711623	0.797981
SARIMA (1 0 2)(3 1 0)$_{12}$	0.826248	0.912606
SARIMA (1 0 3)(1 1 1)$_{12}$	0.713006	0.799364
SARIMA (2 0 1)(0 1 1)$_{12}$	0.706340	0.771108
SARIMA (2 0 1)(1 1 1)$_{12}$	0.709079	0.784642
SARIMA (2 0 2)(2 1 0)$_{12}$	0.874286	0.960644
SARIMA (2 0 2)(3 1 1)$_{12}$	0.713323	0.821270
SARIMA (2 0 3)(2 1 0)$_{12}$	0.871604	0.968756

Table 3. Parameter of SARIMA (2 0 1)(0 1 1)$_{12}$ model.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.019350	0.004097	4.723233	0.0000
AR(1)	-0.264043	0.230922	-1.143432	0.2536
AR(2)	0.362703	0.082714	4.385008	0.0000
MA(1)	0.628569	0.242264	2.594561	0.0099
SMA(1)	-0.930539	0.041974	-22.16954	0.0000

5. Model diagnostics

Randomness of the residuals. According to the plots of ACF and PACF of the residuals it was observed that the p-values are almost more than 0.05 and it indicates that at 5% significance level the residuals are not significant. Therefore, the errors are white noise.

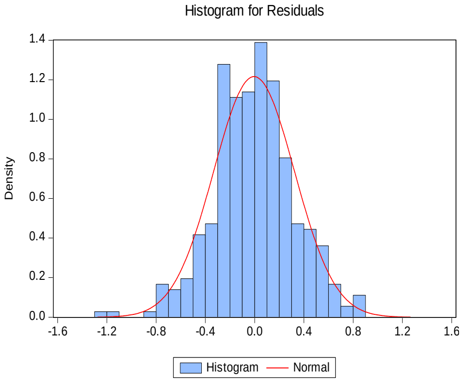

Figure 3. Histogram for residuals.

Normality of residuals. Figure 3 shows the histogram of the residuals. It can be seen that it follows bell shape. Hence, the SARIMA (2 0 1) (0 1 1)$_{12}$ model of AT follows normal distribution.

Heteroscedasticity test. The heteroscedasticity was tested by the ARCH test and results are given in Table 4. From these results it is confirmed that no heteroscedasticity in the residual.

Table 4. Results of residuals ARCH LM test

Heteroskedasticity Test: ARCH			
F-statistic	0.042949	Prob. F(1,357)	0.8359
Obs*R-squared	0.043184	Prob. Chi-Square(1)	0.8354

According to the above tested results it can be said that the selected SARIMA (2 0 1)(0 1 1)$_{12}$ model satisfied all the diagnosis tests for MAT of Sri Lanka. Therefore, the SARIMA (2 0 1)(0 1 1)$_{12}$ model can be used as the best model for forecasting the MAT of Sri Lanka.

Forecasting temperature using SARIMA (2 0 1)(0 1 1)$_{12}$ model. Data from January 1990 to December 2020 used to fit the model. Whereas, data from January 2021 to December 2025 used for forecasting period as shown Figure 4. The best selected model was used to forecast the future value of the temperature in Sri Lanka. In Figure 4, the forecasted value (Red line) and actual value (Blue line) have similar pattern.

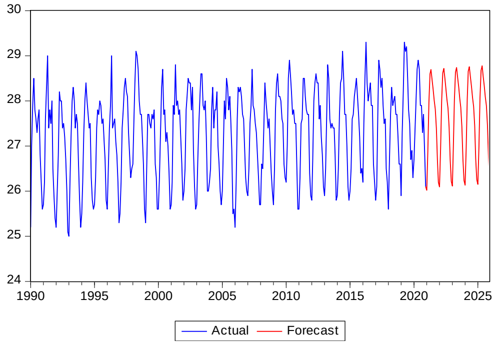

Figure 4. Plot for observed and forecasted values.

Conclusion

In this study, the SARIMA model was used to analyze MAT of Sri Lanka from January 1990 to December 2020. The minimum AIC and SIC values were used to select the best model. The best fitted model for MAT of Sri Lanka was identified as SARIMA (2 0 1)(0 1 1)$_{12}$. Finally, residual diagnostic tests for best fitted model carried out and it was satisfied all the diagnosis test. The fitted SARIMA model is very useful to forecast future temperature value of Sri Lanka.

References

[1] Abebe, T. H., Time series analysis of monthly average temperature and rainfall using seasonal ARIMA model (in Case of Ambo Area, Ethiopia). International Journal of Theoretical and Applied Mathematics, 2020. 6(5): 76-87.

[2] Dwivedi D.K, Sharma G.R., and Wandre S.S., Forecasting mean temperature using SARIMA model for Junagadh city of Gujarat. International Journal of Agricultural Science and Research (IJASR), 2017. 7 (4): pp.183-194.