International Journal of Science, Environment and Technology, Vol. 8, No 1, 2019, 124 – 131

ISSN 2278-3687 (O)
2277-663X (P)

FORECASTING WEEKLY TEMPERATURE USING ARIMA MODEL: A CASE STUDY FOR TRINCOMALEE IN SRI LANKA

Mohamed Cassim Alibuhtto[1] and Himash Rashmika Ariyarathna[2]
[1]Department of Mathematical Sciences, Faculty of Applied Sciences,
South Eastern University of Sri Lanka, Sri Lanka
[2]Department of Actuarial, Janashakthi Insurance PLC, Colombo, Sri Lanka
E-mail: mcabuhtto@seu.ac.lk (*Corresponding Author*)

Abstract: Modelling and forecasting temperature is not only scientific challenging but is also important for planning and formulating agricultural strategies. In this paper, the ARIMA model is used to forecast the average weekly temperature in the Trincomalee district in Sri Lanka. The analysis reveals that the best time series model for forecasting the temperature is ARIMA (1,0,0). The forecasting performances of the selected ARIMA model produces better results due to RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error).This study will be useful for farmers and tourists to make decisions in advance.
Keywords: ARIMA, MAE, Modelling, Temperature, RMSE.

1. Introduction

Climate change will have a major impact on the environment, and socio-economic and related sectors, including water resources, agriculture and food security, human health and forest diversity. Increasing temperatures will cause changes in crop seasons that affect food safety and changes in the spread of diseases that increase the risk for people (UNFCCC, 2006). In addition, it is widely believed that developing countries in tropical regions of the world will suffer worse than developed countries (Shamsnia, Shahidi, Liaghat, Sarraf & Vahdat, 2011).

The regional differences observed in air temperature in Sri Lanka are mainly related to altitude, not latitude. The average monthly temperature varies slightly depending on the seasonal movements of the sun, with some changes caused by rain. The average annual temperature in Sri Lanka shows very homogeneous temperatures in the low lands and falling temperatures in the highlands (Department of Meteorology, 2018).

Trincomalee is the capital of the Eastern Province of Sri Lanka with commercial and historical importance. There are many tourist attractions places in this region, such as the Nilaveli beach, Arisimale beach and Kannia hot springs. In this area, mostly rainy, windy and cloudy dry seasons and it is hot and depressing all year. In addition, the prominent crop such

*Received Jan 2, 2019 * Published Feb 2, 2019 * www.ijset.net*

as paddy, chili, long beans and brinjal are the main vegetables in this area. Thus, temperature predictions are especially useful for farmers and tourists to make decisions in advance.

Many time series models have been used to predict temperature. Among them, Autoregressive Integrated Moving Average (ARIMA) model for forecasting is an effective method (Tektas, 2010; Kelikume & Salami, 2014; Mondal, Shit, & Goswami, 2014; Goswami, & Hazarika, 2017).The purpose of this study is to develop a suitable time series ARIMA model for forecasting weekly temperature of Trincomalee in Sri Lanka.

2. Materials and Methods

2.1 Data Collection

The daily mean temperature data for the last two years (from January 1, 2016 to December 31, 2017) were obtained from the Department of Meteorology for Trincomalee station in Sri Lanka. This daily temperature data is converted into weekly data through averaging method.

2.2 Statistical Analysis

2.2.1. Stationary Time Series

Stationary time series is a random process whose joint probability distribution does not change with time. There are several methods that can be used to convert non-stationary series to stationary series. However, the most widely used variance stabilization method is Box-Cox transformation. The Augmented Dickey Fuller (ADF) test, Phillip -Perron (P-P) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test can be used to check whether the series is stationary or not.

2.2.2. Autoregressive Integrated Moving Average (ARIMA) Model

1). Autoregressive (AR) model of order p-AR (p)

Autoregressive model represents current value of time series as combination of one or more previous values of the same series. It shows the dependency of one value with its nearest previous values. If Y_t {t=1,2,...,n} is the time series of AR model of order p without a constant term can be written as:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + e_t \quad (2.1)$$

Where, e_t is error term and $\phi_1, \phi_2, ..., \phi_p$ are autoregressive parameters to be estimated.

2). Moving Average (MA) model of order q- MA (q)

In this model, the current value of the time series Y_t is expressed linearly in terms of current and previous values of the white noise series (e_t). This white noise series constructed from the forecast errors or residuals when demand observation become available. The moving average model of order q without a constant term can be written as:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \quad (2.2)$$

Where, e_t is error term and $\theta_1, \theta_2, \ldots, \theta_p$ are parameters to be estimated.

3). Autoregressive Moving Average (ARMA) model of order p,q- ARMA(p,q)

Autoregressive Moving Average model of order p and q is formed by combining terms of AR of order p and MA of order q models. Autoregressive Moving Average model of order p and q is generally written as:

$$Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \quad (2.3)$$

The ARMA model assumes that the time series data is stationary. But the real data are not stationary in nature. Time series data is made stationary by differencing process. In general, the first order differencing process of time series Y_t becomes stationary. However, if ARMA time series is made stationary by differencing of order d, it is known as Integrated Autoregressive Moving Average and denoted by ARIMA (p,d,q).

2.2.3. ACF and PACF

Autocorrelation and partial autocorrelation function are a type of graphs that contain correlations of different time lags. ACF and PACF can be used to determine the behaviour of the series, whether stationary or not and to identify the number of components in an ARMA model. The number of significant spikes in the ACF indicates the number of MA terms in the model, while the number of significant spikes in PACF indicates the number of AR terms in the model.

3. Results and Discussion

3.1. Preliminary Analysis

The descriptive statistics of the temperature series are presented in Table 1. The mean and standard deviation of the temperaturewere28.91 and 1.94, while minimum and maximum value of the temperature were 22.50and 32.60 respectively.

Table 1: Descriptive statistics for the temperature data (1^{st} Jan 2016-31^{st} Dec 2017)

Variable	Mean	Std Dev	Minimum	Maximum
Temperature	28.91	1.94	22.50	32.60

Weekly temperature data of the Trincomalee station is used to study the behaviour of time series. Figure 1 shows the plot of weekly temperature and it is highly fluctuating.

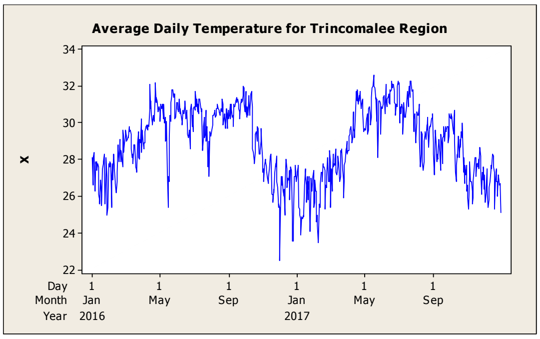

Figure 1: Time series plot of weekly average temperature

3.2. ARIMA (Box-Jenkins) Modelling

The daily temperature data from January 1, 2016 to December 31, 2017 were converted into weekly data (104 observations). In addition, the converted weekly data was divided into training set (80%), consisting 84 observations, and validation set (20%), consisting 20 observations. The Box-Jenkins modelling was performed using training data. The data must be checked, whether stationary or not using KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test, before it can be used to estimate and develop a model. The corresponding results are shown in Table 2.

Table 2: Result of unit test for temperature

Null Hypothesis: Temp is stationary	
KPSS test statistic	0.1418
Critical values 1% level	0.7390
5% level	0.4630
10% level	0.3470

Table 2 shows that the null hypothesis that series is stationary cannot be rejected since the test statistic of KPSS test is less than the critical values at 5% significance level. Hence, the temperature series is stationary.

As the stationary was achieved at the level it is required to search models of the family of ARIMA (p, 0, q), where p and q are possible order of AR and MA components. The plot of sample ACF and PACF of temperature series (Fig.2) was considered for identification of suitable values for p and q.

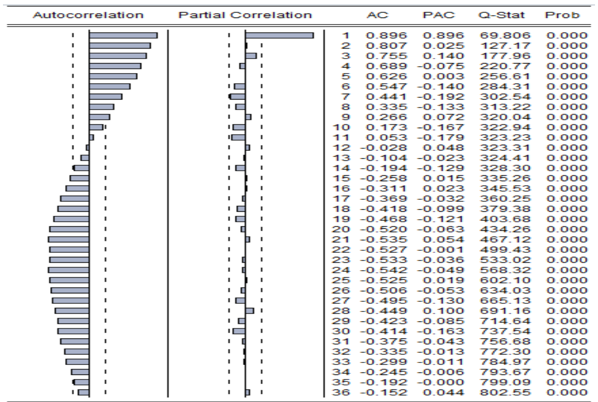

Figure 2: Sample ACF and PACF plot of temperature series

According to Figure 2, the sample PACF has one significant autocorrelation at lag 1. Thus, it can be hypothesised in the ARMA model to be fitted AR order to be 1. Therefore, the following models were considered as possible models ('parsimonious' models) to represent the original series. They are: (i) ARIMA(1,0,0), (ii) ARIMA(1,0,1), (iii) ARIMA(1,0,2), (iv) ARIMA(2,0,0) (v) ARIMA(2,0,1) and (vi) ARIMA(2,0,2)

Table 3: Results of model estimation

Model	Log likelihood	AIC	SIC	DW
ARIMA(1,0,0)	-101.237	2.515	2.602	2.048
ARIMA(1,0,1)	-102.463	2.535	2.651	1.931
ARIMA(1,0,2)	-101.888	2.545	2.689	1.979
ARIMA(2,0,0)	-102.515	2.536	2.652	1.966
ARIMA(2,0,1)	-102.324	2.555	2.700	1.914
ARIMA(2,0,2)	-101.525	2.560	2.734	1.955

The results in Table 3 show that of the six models the maximum log likelihood estimates and the lowest AIC and SIC values were obtained by ARIMA (1,0,0) model. Thus, it can be concluded the best model of the six is ARIMA(1,0,0). In addition, Table 4 indicates that constant and AR(1) termsare significant in the selected model.

Table 4: Parameter estimation of ARIMA (1,0,0)

Variable	Coefficient	Standard Error	t-Statistic	P-Value
C	28.982	0.786	36.888	0.000
AR(1)	0.892	0.047	18.972	0.000

3.3 Model Diagnostics

1). Randomness

The ACF plot of the residuals of selected model (Fig.3) shows that the residuals are relatively small and not statistically significant. Therefore, it can be considered that the residuals of the fitted model is randomly distributed.

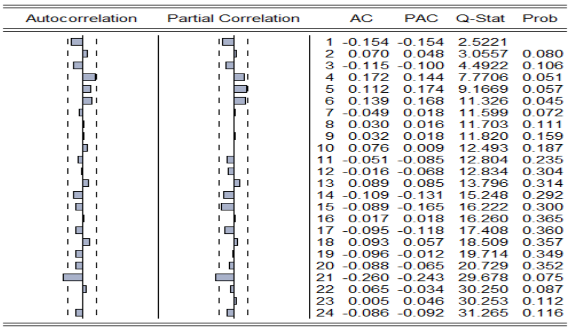

Figure 3: ACF and PACF plot of residuals

2). Normality

The normal probability plot of the residuals was carried out to check whether residuals are normal or not.

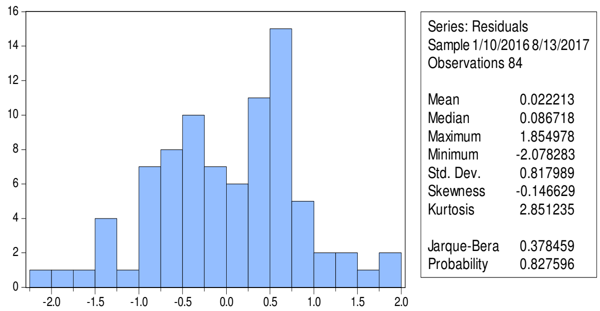

Figure 4: Normal probability plot of residuals

Figure 4 shows that the respective Jarque-Bera statistic (JB=0.379, p=0.828). Thus, it is confirmed that the residual series is normally distributed.

3). Heteroscedasticity

The ARCH LM test was conducted to observe the heteroscedasticity of the developed model. The results of heteroscedasticity of residuals are shown in Table 5 and it shows that the heteroscedasticity exist at 5% significance level (p=0.448).

Table 5: Results for heteroscedasticity of residuals

ARCH LM Test			
F-statistic	0.56	Prob	0.454
Obs* R-squared	0.575	Prob	0.448

Based on the above detailed analysis of residuals, it can be confirmed that the developed ARIMA (1,0,0) model satisfies all the diagnostic tests. Hence, the ARIMA (1,0,0) is considered as the best fitted model for forecasting the weekly temperature of Trincomalee.

3.4. Forecasting temperature using ARIMA (1,0,0) Model

Before forecasting the values, it is useful to validate the present model with observed data (training set) as well as an independent data set (validation set). For the validation purpose 20 weeks observed temperature data from August 20, 2017 to December 31, 2017is used. The forecasts errors of validation set of observed data are shown in Table 6.

Table 6: Results of forecast performances of ARIMA (1,0,0) model

Type of data	Period	RMSE	MAE
Validation set	Aug 20, 2016 –Dec 31,2017	0.814	0.709

Table 6 indicates that the RMSE and MAE for validation set from ARIMA (1,0,0) model deviates from the observed data are 0.814 and 0.709, respectively, which would be considered as being within acceptable range. The estimated and observed values are shown in Figure 5.

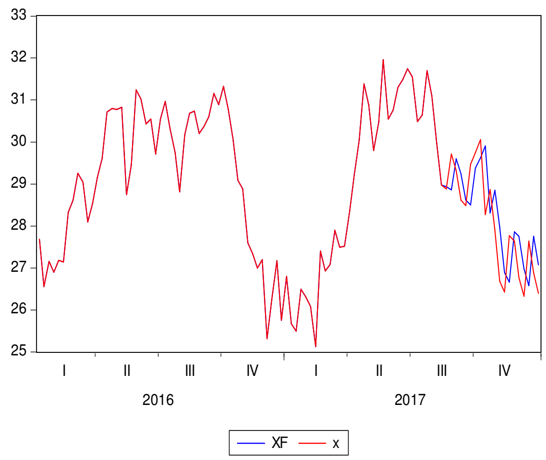

Figure 5: Time series plot for observed and forecasted values

4. Conclusion

Time series analysis is an important technique in modelling and forecasting temperature. In this study, the ARIMA model was used to forecast weekly temperature of Trincomalee station in Sri Lanka. The best fitting model for temperature was identified as ARIMA (1,0,0). Although, the developed model was verified using diagnostic tests. Consequently, this developed model can help to determine possible future strategies for agricultural and tourist sector in the respective region. However, this study needs to be expanded due to fluctuations in the data series.

References

[1] Department of Meteorology. (2018, April 5). Climate of Sri Lanka. Retrieved December 2, 2018, from http://www.meteo.gov.lk/index.php=com_content&view=article&id=94& Itemid=310&lang=en.

[2] Goswami, K. & Hazarika, J. (2017). Monthly Temperature Prediction Based on Arima Model: a Case Study in ibrugarh Station of Assam, India. International Journal of Advanced Research in Computer Science, (8), 292–298. https://doi.org/10.26483/ijarcs.v8i8.4590.

[3] Kelikume, I. & Salami, A. (2014). Time Series Modeling and Forecasting 10.1. International Journal of usiness and Finance Research, 8(2), 41–52.

[4] Mondal, P., Shit, L., & Goswami, S. (2014). Study of e ffectiveness of t ime s eries modeling (ARIMA) in forecasting s tock. *International Journal of Computer Science, Engineering and Applications* (*IJCSEA*), 4(2), 13–29.

[5] Shamsnia, S.A., Shahidi, N., Liaghat, A., Sarraf, A., & Vahdat, S.F. (2011). Modeling of Weather Parameters Using Stochastic Methods (ARIMAModel)(Case Study : Abadeh Region, Iran). *International Conference on Environment and Industrial Innovation IPCBEE,* 12, 282-285.

[6] Tektaş, M. (2010). Weather forecasting using ANFIS and ARIMA models. *Environmental Research, Engineering and Management*, 1(51), 5–10.

[7] UNFCCC(2006, July 25). Climate change: Impacts, vulnerabilities and adaptation in developing countries. Retrieved December 2, 2018, from https://unfccc.int/resource/docs/publications/impacts.pdf