Global Warming in Cameron Highlands: Forecasting its Temperature Level via ARIMA vs ARAR

Nurzawanah Raihah Zamri¹, Nurul Nisa’ Khairol Azmi²*

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Cawangan Selangor, Kampus Shah Alam, 40450 Shah Alam, Selangor.
²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, 70300 Seremban, Negeri Sembilan.

*nurulnisa@uitm.edu.my

Abstract. The average global temperature has increased at a rapid rate over the past 50 years leading to global warming. The impact of climate change can be felt across the continents. In this research, analysis was conducted to model and forecast the monthly temperature of Cameron Highlands in 2020 and 2021, against its historical monthly average temperature from January 1990 until December 2019. Two (2) methods namely (i) Seasonal Autoregressive Integrated Moving Average (SARIMA) model and (ii) Autoregressive Autoregressive (ARAR) algorithm were compared to determine the best model to forecast the monthly temperature of Cameron Highlands. SARIMA (1,1,2)(1,1,1)½ was found to be the best at forecasting the monthly temperature in Cameron Highlands as RMSE and MAPE values were lower than ARAR. In year 2021, the temperature in Cameron Highlands is estimated to increase by 1.6 °C. The result of the forecast showed that its monthly temperature was expected to increase in the next two (2) years. Hence, this calls for serious action to be taken by higher authorities.

Keyword: ARIMA, ARAR, MAPE, RMSE, SARIMA

1. Introduction
The average global temperature has increased at a rapid rate over the last 50 years, leading to global warming. The impact of climate change is felt across the continents. According to the 2019 Global Climate report by the National Oceanic and Atmospheric Administration (NOAA), the combined land and ocean temperature have increased at an average rate of 0.07°C per 100 years since 1880. The average rate of world temperature on the other hand has increased by 0.18°C since 1981 and twice the former’s temperature surge [1]. The rise in global temperature can cause a serious, bad phenomenon where the sea temperature and ocean increase, affecting the rise of sea level [2].

Activities like businesses and agriculture generally depend on daily temperature. In Malaysia, climate change has found to be worsen and unfortunately, only a few Malaysians are paying serious attention to the matter. Malaysia has a tropical weather that is quite humid with an average temperature between 20°C and 30°C for its major cities. Whereas the average temperature of high lands in Malaysia does not exceed 25°C even during hot season. By 2050, projections show that Malaysia is going to be hotter with a temperature rise up to 1.5°C [3]. Similarly, [4] asserted the maximum, minimum, and mean temperature of Cameron Highlands to be respectively estimated to rise around 3.8°C, 1.8°C, and 2.8 °C in 100 years. Within 30 years of inaction, the climate of Kuala Lumpur will likely reach that of
Palembang today. The Air Pollutant Reading (API) hit a new high figure of 921 last months and the average temperature rose by 2.3°C [5].

Forecasting temperature at any location is crucial as it can serve as a clue to foresee any unfortunate future events. Many outdoor activities such as agriculture and businesses depend on the weather and temperature. To illustrate, farmers plan their timeframe when planting by referring to forecasts. This is one of the many instances asserting the agriculture sector’s need for a better temperature prediction, taking into account that farmers need to schedule their time with consideration of the suitable temperature for their crops. The same also goes to the agriculture activities in Cameron Highlands, a place known to have a cold temperature; popular for its agriculture activities involving fresh vegetables; and favorable for the cultivation of tea, fruits and flowers. This study hence attempted to forecast the monthly temperature of Cameron Highlands.

There are many methods of forecasting temperature available, [6-8], but lack of applications on forecasting temperature for Malaysia’s data. Some of the research focus only on trend analysis and the movement of temperature increase on the average such as multiple linear regression [9]. Furthermore, there is no research found on modelling and forecasting monthly temperature for Cameron Highland using Auto Regressive Integrated Moving Average (ARIMA) model and AutoRegressive AutoRegressive (ARAR) algorithm. The ARIMA model is an established forecasting tool that has an ability to capture the trend and seasonality. On the other hand, ARAR approach was chosen to determine whether it can perform better than seasonal ARMA model which possesses stable forecast-performance and reasonably good forecasting accuracy, [10, 11]. Hence, this paper employed the ARIMA and ARAR algorithm for modelling and forecasting the future temperatures of 2 years ahead for Cameron Highland.

2. Methodology
This section describes the research process, description of data, and method of data analysis. This research employed a time series data involving monthly temperature of Cameron Highlands. The procedure of modelling and forecasting for ARIMA and ARAR algorithm are discussed extensively in this section.

2.1. Data Description
This data consisted of the temperature of Cameron Highlands in the last 30 years (1990 - 2020 in monthly form), which was requested from the Malaysian Meteorological Department. The temperature obtained was in degree Celsius (°C) and the selection of the location was due its geographical area.

The data was divided into two (2) parts namely (i) estimation and (ii) evaluation. Estimation consisted of 70% of the data series for modelling purpose, whereas the remaining 30% was for the performance evaluation of the proposed model [12].

2.2. Method of Analysis
There were two (2) methods used to model and forecast the monthly temperatures of Cameron Highlands. Since the data was highly volatile due to seasonality, seasonal ARIMA and ARAR algorithm were chosen as both models were deemed to be suitable for the data series based on their characteristics.

2.2.1. Seasonal Autoregressive Integrated Moving Average (SARIMA). Seasonal ARIMA or SARIMA is a probabilistic model that is used to analyse non-stationary data caused by seasonality, like monthly temperature. SARIMA model procedure comprises of three (3) stages: (i) Identification, (ii) Estimation, and (iii) Diagnostic Checking and Forecasting. In order to understand the model, the basic models of Autoregressive (AR) and Moving Average (MA) are explained.

Autoregressive (AR), the current value of the variable in AR model, is defined as a function of previous values plus error term. The function of time lagged value said as independent variable which is \( y_t \) [13]. AR model \( (p) \) is as shown in the following equation:
\[ y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + e_t \]  

where \( y_t \) is a monthly temperature at time \( t \), \( c \) is a constant term, \( \phi_1, \phi_2, \ldots, \phi_p \) are AR coefficients of the lagged values of \( y_{t-i} \) (i varies from 1, 2, \ldots, \( p \)), \( y_{t-1}, y_{t-2}, \ldots, y_{t-p} \) are lagged of temperature up to period \( p \) and \( e_t \) is the error terms.

Moving Average (MA) is a method that is related to averaging current and previous period of noise. It is also a method used in a data collection to get an observation of the trends where there is an average of some sub-set of numbers. It is usually used for predicting long-term patterns. Any time of period can be measured by moving average. MA model which is developed from the current value of time series depends on the current and previous values of residuals that are acquired from AR model. MA(\( q \)) model can be expressed as the equation presented below [12]:

\[ y_t = c + \sum_{j=1}^{q} \theta_i e_{t-j} + e_t \]  

where \( y_t \) is monthly temperature at time \( t \), \( c \) is constant term, \( \theta_1, \theta_2, \ldots, \theta_q \) are MA coefficients of the lagged values of \( y \) (i varies from 1, 2, \ldots, \( q \)) and \( e_{t-1}, e_{t-2}, \ldots, e_{t-q} \) are the lagged of temperature up to period \( q \) and \( e_t \) is the error terms.

ARMA or autoregressive moving average is denoted as ARMA \((p,q)\). It provides time series model and framework upon studying stationary. As a part of ARIMA, its process can be split in three (3) parts namely (i) autoregressive (AR), (ii) integrated, and (iii) moving average (MA) part [14]. The combination of these parts make up ARMA model, which is ARMA \((p,q)\). The model can be described as follows:

\[ y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \]  

where \( e \) and \( c \) are error term and constant respectively, \( \phi_j \) is the coefficient for each lag variable of \( y_{t-i} \) with \( i=1,2,\ldots,p \) and \( \theta_j \) is the coefficient for each lag variable of \( e_{t-j} \) where \( j=1,2,\ldots,q \).

If the data is not stationary, one of the methods to rectify the problem is by differencing the data. The ARMA model with differencing is called as ARIMA model. In general, ARIMA model or ARIMA \((p,d,q)\) can be explained as follows: \( p \) is the number of autoregressive values, \( d \) is the order of differencing for bringing the series to produce stationary series, and \( q \) is the number of moving average values [15].

Seasonal ARIMA can be applied to the data that consist of seasonal pattern that can cause stationary assumption to be violated. Seasonal pattern refers to repeating behavioral patterns in a season, normally for 12 months in one (1) year for monthly data. While monthly data series is often in a seasonal period of 12 months, quarterly data series is always in a period of four (4) quarters. Seasonality can be determined by examining whether the autocorrelation function of the data series with a specified seasonal order is significantly different from zero. The general form of Seasonal ARIMA \((p,d,q)\) \((P,D,Q)\) is shown as follows:

\[ \phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = c + \theta(B)\Theta(B^s) e_t \]  

where \( e_t \) is Gaussian white noise. \( \phi(B)\) and \( \theta(B) \) are ordinary autoregressive and moving average components respectively. \( \Phi(B^s) \) and \( \Theta(B^s) \) are seasonal autoregressive and moving average components respectively. \((1-B)^d\) is an ordinary differencing component while \((1-B^s)^D\) is a seasonal differencing component.

### 2.2.2. Box-Jenkins Methodology

Box-Jenkins Methodology for modelling and forecasting the monthly temperature of Cameron Highlands consists of five (5) steps, (i) model identification, (ii) model estimation, (iii) diagnostic checking, (iv) model evaluation and (v) forecasting [16].
Model identification, the first step in the procedure, explains the possible models of ARIMA that can be listed out by observing the time series plot of data series and correlogram plot of Autocorrelation (ACF) and Partial Autocorrelation Function (PACF). Some information regarding on possible trend and seasonality can be obtained from the time series plot. Based on the plot, possible trend can be identified based on the measurement on average of the tendency to increase and decrease over time. Seasonality on the other hand can be observed through the regular repetition in a pattern of high and low data. Data is assumed to be stationary if they fluctuate around constant value with constant variance. Input time series for an ARIMA model needs to be stationary. The time series should have a constant mean, variance, and autocorrelation through time. Therefore, the stationarity of the data series needs to be identified first. If the data shows an increasing or decreasing pattern, the data is then suspected not to be stationary data. If not, the non-stationary time series is then required to be stationary by performing first order differencing. If ACF shows a slowly decaying pattern with all spikes being highly significant, the data is then considered not stationary. A wavelength pattern asserts the effect of seasonality on data, making it not stationary.

In model estimation, the Box-Jenkins models are usually estimated based on the sample statistics that must be tested to ensure their validity as the estimate of the true population parameter values. In this case, the parameter estimation used the approach of Maximum Likelihood Estimation (MLE). MLE is a method of determining the parameters of the probability distribution by optimising the likelihood function, so that the data observed is most likely under the assumed statistical model.

The adequacy of the estimated seasonal ARIMA model is diagnosed by using Ljung Box test where the absence of serial autocorrelation among errors or the error terms are white noise was tested up to specified lag k [17]. The significant p-value of Ljung Box test indicates that the error terms are not white noise which imply the model is not well specified and the accuracy of the predicted model is questionable.

The performance of the seasonal ARIMA models is determined by AIC and BIC criteria. The values of AIC and BIC are decided by the lowest value and declared as the best model for the forecast. AIC or Akaike Information Criterion is applied to determine an estimator of prediction error and relative quality of statistical models of data set. Bayesian Information Criterion (BIC), also known as Schwarz information criterion, is based on Bayesian modelling. BIC is frequently used for short and long sample sizes (N) time-series data [18].

2.2.3 Autoregressive Autoregressive (ARAR). The ARAR algorithm is one of the proposed methods to forecast monthly temperature of Cameron Highlands. ARAR algorithm consists of three (3) phases: (i) memory shortening process, (ii) fitting autoregressive model, and (iii) forecast.

From [19], the first step of ARAR algorithm is making decision whether the underlying process is “long-memory,” and if so to apply a memory-shortening transformation before attempting to fit an autoregressive model. The memory shortened series can be described as follows:

\[ \tilde{Y}_t = Y_t - \hat{\phi}(\tau)Y_{t-\tau} \]  \hspace{1cm} (5)

\[ \tilde{Y}_t = Y_t - \hat{\phi}_1Y_{t-1} - \hat{\phi}_2Y_{t-2} \]  \hspace{1cm} (6)

The following are the three (3) courses action based on equation (5) and (6):

- L- Declare \{Y_t\} to be long-memory and form \{\tilde{Y}_t\} using Equation (5)
- M-Declare \{Y_t\} to be moderately long-memory and form \{\tilde{Y}_t\} using equation (6)
- S-Declare \{Y_t\} to be short-memory

With the aid of the five-step algorithm described below, the course action can identify by:

1. For each \( \tau = 1,2,\ldots,15 \), find the value \( \hat{\phi}(\tau) \) of \( \phi \) that minimizes...
\[ ERR(\phi, \tau) = \sum_{i=1}^{n} \left[ Y_i - \phi \cdot Y_{i-\tau} \right]^2 \]

then define and choose the lag \( \hat{\tau} \) to be the value \( \tau \) that minimizes \( ERR(\tau) \).

2. If \( ERR(\hat{\tau}) \leq \frac{8}{n} \), go to \( L \).

3. If \( \hat{\phi}(\hat{\tau}) \geq 0.93 \) and \( \hat{\tau} > 2 \), go to \( L \).

4. If \( \hat{\phi}(\hat{\tau}) \geq 0.93 \) and \( \hat{\tau} = 1 \) or 2, determine the values \( \hat{\phi}_1 \) and \( \hat{\phi}_2 \) of \( \phi_1 \) and \( \phi_2 \) that minimize \( \sum_{i=3}^{n} (Y_i - \phi_1 Y_{i-1} - \phi_2 Y_{i-2})^2 \); then go to \( M \).

5. If \( \hat{\phi}(\hat{\tau}) < 0.93 \), go to \( S \).

By fitting a subset of autoregression, let \( \{ S_i, t = k + 1, ..., n \} \) denote the memory-shortened series derived from by algorithm and let \( \bar{S} \) denote the sample mean of \( S_{k+1}, ..., n \). To fit an autoregressive process to mean-corrected series procedure is

\[ X_t = S_t - \bar{S}, \quad t = k + 1, ..., n \]  

(8)

The fitted model has the form of

\[ X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \phi_4 X_{t-4} + e_t \]  

(9)

where \( \{ e_t \} - WN(0, \sigma^2) \) and for given lags, \( l_1, l_2 \) and \( l_3 \), the coefficients \( \phi_j \) and the white noise variance \( \sigma^2 \) are found in the Yule-Walker equations.

\[
\begin{bmatrix}
1 & \hat{\phi}(l_1-1) & \hat{\phi}(l_2-1) & \hat{\phi}(l_3-1) \\
\hat{\phi}(l_1-1) & 1 & \hat{\phi}(l_2-l_1) & \hat{\phi}(l_3-l_1) \\
\hat{\phi}(l_2-1) & \hat{\phi}(l_2-l_1) & 1 & \hat{\phi}(l_3-l_2) \\
\hat{\phi}(l_3-1) & \hat{\phi}(l_3-l_1) & \hat{\phi}(l_3-l_2) & 1 \\
\end{bmatrix}
\begin{bmatrix}
\phi_1 \\
\phi_2 \\
\phi_3 \\
\phi_4 \\
\end{bmatrix}
= 
\begin{bmatrix}
\hat{\phi}(l_1) \\
\hat{\phi}(l_2) \\
\hat{\phi}(l_3) \\
\end{bmatrix}
\]

(10)

If the memory-shortening filter found in the first step has coefficients \( \psi_0(=1), \psi_1, ..., \psi_k (k \geq 0) \), the memory-shortened series is expressed as:

\[ S_t = \psi(B) Y_t = Y_t + \psi_1 Y_{t-1} + ... + \psi_k Y_{t-k} \]  

(11)

where \( \psi(B) \) is the polynomial in the backward shift operator,

\[ \psi(B) = 1 + \psi_1 B + ... + \psi_k B^k \]  

(12)

If the coefficients of the subset autoregression found in the second step are \( \phi_1, \phi_2, \phi_3 \) and \( \phi_4, \phi_5, \phi_6 \), then the subset AR model for the mean-corrected series \( \{ X_t = S_t - \bar{S} \} \) is...
\[
\phi(B)X_t = e_t
\]  \hfill (13)

where \( \{e_t\} \sim WN(0, \sigma^2) \) and

\[
\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3
\]  \hfill (14)

From Equations (13) and (14) obtained as

\[
\xi(B)Y_t = \phi(1)S + e_t
\]  \hfill (15)

where \( \xi(B) = \psi(B)\phi(B) = 1 + \xi_1 B + \ldots + \xi_{k+1} B^{k+1} \).

Assuming that the fitted model is appropriate and that the white noise term \( e_t \) is uncorrelated with \( \{Y_j, j < t\} \) for each by determining the minimum mean squared error linear predictors:

\[
P_n Y_{n+h} = \sum_{j=1}^{k+1} \xi_j P_n Y_{n+h-j} + \phi(1)S, \quad h \geq 1
\]  \hfill (16)

with the initial conditions \( P_n Y_{n+h} = Y_{n+h} \) for \( h \leq 0 \).

3. Results and Discussion

The monthly temperature of Cameron Highlands was modelled by employing SARIMA model and ARAR algorithm. The performance of these two models was compared. The best model was then used to forecast the monthly temperature for the next two (2) years.

3.1. Description of Data on the Monthly Temperature of Cameron Highlands

Figure 1 shows the pattern of the monthly temperature of Cameron Highlands in the last 30 years. The graph shows the trends to be slightly increasing. There was a constant repetition in the monthly temperature from 1990 to 1997, with a difference. The lowest peak of 16.7°C was recorded in January 1996. From April 1998 onward, the monthly temperature increased, and the highest peak was recorded at 20°C. Huge changes in the temperature can be spotted in 1998.

![Figure 1. Monthly temperature of Cameron Highlands](image_url)

3.2. Modelling and Forecasting the Temperature of Cameron Highlands

This section presents the results of the monthly temperature forecast of Cameron Highlands. The procedure and results of time series analysis are discussed in the following subsection.
3.2.1 Data Partition.
The data was split into two parts of (i) Estimation and (ii) Evaluation. The rule of thumbs used to split the data comprised 70% for estimation and the remaining 30% for evaluation, [20]. The data on the monthly temperature was recorded in Celsius from January 1990 until December 2019.

Table 1. Data partition of Cameron Highlands

| Part            | Model                          |
|-----------------|--------------------------------|
| Estimation (70%)| January 1990 – December 2010  |
| Evaluation (30%)| January 2011 – December 2019  |

Referring to table 1, 70% of the data comprised the monthly temperature from January 1990 until December 2020. This was categorised as the estimation part that helped to check the stationarity of the data and build the model. Whereas the rest of the data consisted of the evaluation part, which recorded temperature from January 2011 until December 2019. This was then used to verify upon determining the best model to forecast.

3.2.2 Model Identification.
In this part, the model was identified by observing the correlogram of Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). ACF and PACF were also used to determine whether the data meet the stationary assumption. Data is generally assumed to be stationary if ACF and PACF show no clear oscillatory pattern and a spectral density plot that does not show the sign of large and high spikes in frequency domain [21]. If ACF and PACF spikes exceed the confidence limit, the correlation between the lags is significant.

Figure 2 shows a wavelength pattern in ACF for the monthly temperature of Cameron Highlands, oscillating at a regular frequency that was suspected not to be stationary due to seasonal effect. In addition, most of the lags were significant and exceeded the 95% confidence interval limit. Looking at PACF plot in figure 3 which shows the monthly temperature of Cameron Highlands, the pattern was also in wave and regular frequency, confirming the non-stationary series. Hence, seasonal differencing of order one was performed to remove the seasonal component in order to achieve stationary process [22].

Figure 4 shows the correlogram of ACF after the seasonal differencing of order one. The lags slowly decayed to zero and many spikes exceeded the 95% confidence limit. This was a sign of non-stationarity figure 5 depicts PACF after the first order of seasonal differencing. There were several significant spikes at lags 1,2,12,13, and 20, hence providing a conclusion that the data was still not stationary. Thus, first
order differencing was performed.

![Figure 4. ACF after the first order of seasonal differencing](image.png)

![Figure 5. PACF after the first order of seasonal differencing](image.png)

Figure 4 and 5 show the ACF and PACF after the first order differencing. Both ACF and PACF fell off after a few significant spikes, indicating the monthly temperature of Cameron Highlands after seasonal and the first order differencing. ACF was used to determine the order of Moving Average (MA) and seasonal Moving average (SMA), while PACF was to identify the order of autoregressive (AR) and Seasonal Autoregressive (SAR).

![Figure 6. ACF after the first order differencing](image.png)

![Figure 7. PACF after the first order differencing](image.png)

Hence, from the ACF and PACF plot, SARIMA(2,1,2)(1,1,1)_{12} was determined as the proposed model.

The ACF in figure 6 shows three significant spikes at lag 1, lag 2, and lag 13. This represented the order of MA and SMA. Whereas PACF indicated that there were five significant spikes at lag 1, lag 2, lag 11, lag 12, and lag 13, representing the order of AR and SAR.

Hence, from the ACF and PACF plot, SARIMA(2,1,2)(1,1,1)_{12} was determined as the proposed model. As for the purpose of comparison, the other three (3) models suggested were SARIMA(1,1,2)(1,1,1)_{12}; SARIMA(1,1,1)(1,1,1)_{12}; and SARIMA(1,1,0)(1,1,1)_{12}.

3.2.3 Diagnostic Checking and Model Evaluation.

In this section, the results pointed out diagnostic check via Ljung Box test and determine the best model to be used as the candidate to forecast the future monthly temperature in Cameron Highlands. The model is well specified if the error terms are white noise. The Ljung Box test was used to verify whether the models are well specified or not. Since the p-values for all proposed models as depicted in table 2 are not significant, it can be concluded that the error terms are white noise and well specified. Hence, all proposed models can be a candidate of best model that fit the monthly temperature of Cameron Highlands.

The best model was determined based on Akaike Information Criterion (AIC) and Bayesian information criterion (BIC). While AIC generally deals with both risk of overfitting and risk of underfitting, BIC is used for the selection of model among a finite set. The model with the lowest AIC and BIC was the model that fit best the study.
Table 2. Performance of SARIMA on the Monthly Temperature of Cameron Highlands based on AIC and BIC values with Ljung Box test p-value

| Model                     | AIC  | BIC  | p-value (Ljung Box test) |
|---------------------------|------|------|--------------------------|
| SARIMA(2,1,2)(1,1,1)\_12  | 94.77| 119.10| 0.7922                   |
| SARIMA(1,1,2)(1,1,1)\_12  | 92.83| 113.68| 0.7339                   |
| SARIMA(1,1,1)(1,1,1)\_12  | 98.64| 116.02| 0.8837                   |
| SARIMA(1,1,0)(1,1,1)\_12  | 93.23| 114.09| 0.2468                   |

Table 2 presents the AIC and BIC values for the proposed seasonal ARIMA, based on 70% of estimation data series. SARIMA(1,1,2)(1,1,1)\_12 was found to be the best model to forecast the monthly temperature of Cameron Highlands as its AIC and BIC values were the lowest.

3.3 Analysis of ARAR algorithm in Cameron Highlands

ARAR analysis was carried out using R software. All the steps including memory shortening, fitting a subset of autoregression, and forecasting were summarized in one result. After the data run in the software turned stationary, the second step in the algorithm was applied- that was to fit the autoregressive model. Based on the results from R software, the suggested optimal lags for the fitted model were 1, 11, 12, and 13.

3.4 Model Performance Evaluation

This section on model performance evaluation discusses about the results of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), upon determining the best forecast model (between ARAR and SARIMA) for this research. Both models were tested using 30% of the test data from January 2011 until December 2019.

Table 3. Model performance of the monthly temperature of Cameron Highlands

| Method of Forecasting | RMSE | MAPE |
|-----------------------|------|------|
| SARIMA(1,1,2)(1,1,1)\_12 | 0.2867 | 1.1169 |
| ARAR                 | 2.6553 | 13.194 |

Table 3 shows the model performance evaluation values in Cameron Highlands namely (i) RMSE and (ii) MAPE for both SARIMA and ARAR method. The respective values of RMSE for SARIMA and ARAR as presented in the table were 0.2867 and 2.6553. The value of MAPE for SARIMA (1.1169%) was categorised as highly accurate as it was less than 10%, while the value RMSE for ARAR (13.194%) was categorised as a good accuracy.

Al-Qaness, Ewees, Fan, and Abd El Aziz [23] stated that the best model is chosen based on the lowest value of RMSE and MAPE between the models tested. In the case of this research, the results show the smallest RMSE and MAPE values to be 0.2867 and 1.1169 respectively. Therefore, seasonal ARIMA method was found to be the best method to model the monthly temperature of Cameron Highlands during the evaluation.

3.5 Forecasting Future Temperature using SARIMA

Since SARIMA (1,1,2) (1,1,1)\_12 was chosen as the best model to forecast the monthly temperature in Cameron Highlands, the model was also used to predict future temperature. Figure 8 depicts the forecasted values of the monthly temperatures in Cameron Highlands from January 2020 to December 2021 by applying SARIMA model.
Figure 8. The temperature of Cameron Highlands in 2020 and 2021

The model predicted the temperatures in Cameron Highlands to increase as compared to the temperature in the previous year. The first temperature predicted in January 2020 was 17.4 °C, showing a decrease by 0.2 °C as compared to December 2019. Then, the temperature was predicted to continuously increase to reach the maximum point of temperature of 19°C in year 2019. The maximum predicted temperature in 2019 can however be categorised low as it was not higher than the temperature in the previous years that reached 20°C. It is to be pointed out nonetheless that the concern was the fact that the minimum predicted temperature was not quite average, in comparison to the temperature of 17.2°C in the previous years. Despite the forecasted temperature in 2021 showed the same pattern as 2020, this year’s maximum temperature was predicted to slightly increase to 19.4°C with a minimum temperature of only 17.5 °C. It can hence be concluded that though the predicted temperature in 2020 and 2021 was forecasted to majorly increase, it still did not reach the maximum points. The minimum temperature nevertheless increases with an average of 1.6°C by year end. This situation has therefore affected the agriculture activities in Cameron Highlands as some plantations need low temperature to grow.

4. Conclusion and Recommendations

This research aims to determine the best model to forecast monthly temperature in Cameron Highlands. To forecast monthly temperature level at the place, this research employed two (2) types of forecasting methods- SARIMA model and ARAR algorithm.

In order to determine the best method to forecast the temperature in Cameron Highlands, the data was divided into two (2) parts namely (i) estimation and (ii) evaluation. Results from the analysis found that the best model to be used to forecast the monthly temperature of Cameron Highlands was SARIMA (1,1,2) (1,1,1)_12. In conclusion, though the temperature in Cameron Highlands in next two (2) years is expected to increase, it will still be colder than other places. As a matter of fact, its temperature in 2020 and 2021 were predicted not to reach the maximum record of temperature.

Future research can adopt more sophisticated models in comparison to monthly temperature data model such as Artificial Neural Network and Modified Holt’s method, to name a few. SARIMA and ARAR algorithm on the other hand can be applied in many areas of time series data as their applications, especially ARAR algorithm, are not widely applicable. The method of measuring the performance of the models can be extended by applying simulated data in the future.

References

[1] Rebecca L D. Climate Change: Global Temperature [Internet]. 2021 [updated 2021 July 15]. Available from: https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature#:~:text=According%20to%20the%20NOAA%202019,more%20than%20twice%20as%20great.

[2] Sarkar M, Kabir S, Begum R A, Pereira J J, Jaafar A H and Saari M Y 2014 Impacts of and adaptations to sea level rise in Malaysia Asian journal of water, Environment and Pollution 11 29-36
[3] Rahman H A 2018 Climate Change Scenarios in Malaysia: Engaging the Public. *International Journal of Malay-Nusantara Studies* **1** 55-77

[4] Tan K W and Loh P N 2017 Climate change assessment on rainfall and temperature in Cameron Highlands, Malaysia, using regional climate downscaling method *Journal of Earth and Environmental Sciences* **12** 413-21

[5] Insyiraah B. The Climate Change Problem: Tackling Malaysia's Transport Industry [Internet]. 2019. Available from: https://www.ukeconline.com/climate-change/.

[6] Khatib T, Mohamed A, Sopian K and Mahmoud M 2012 Estimating ambient temperature for Malaysia using generalized regression neural network *International journal of green energy* **9** 195-201

[7] Naing, W Y N and Htike Z Z 2015. Forecasting of monthly temperature variations using random forests. *ARPN journal of Engineering and Applied Sciences* **10** 10109-12

[8] Zahroh S, Hidayat Y, Pontoh R S, Santoso A and Bon A T (2019, November). Modeling and forecasting daily temperature in Bandung. In *Proceedings of the International Conference on Industrial Engineering and Operations Management Riyadh, Saudi Arabia* (pp. 406-12).

[9] Chu, F L 2008 Analyzing and forecasting tourism demand with ARAR algorithm. *Tourism management* **29** 1185-96

[10] Miswan, N H, Hussin N H, Said R M, Hamzah K and Ahmad E Z 2016 ARAR algorithm in forecasting electricity load demand in Malaysia *Glob. J. Pure Appl. Math.* **12** 361-7

[11] David Adelia S N 2018 Box-Jenkins’ Methodology in Predicting Maternal Mortality Records from a Public Health Facility in Ghana *Open Journal of Applied Sciences* 189-202

[12] Makridakis S and Hibon M 1997 ARMA models and the Box–Jenkins methodology *Journal of Forecasting* **16** 147-63

[13] Wang H R, Wang C, Lin X and Kang J 2014 An improved ARIMA model for precipitation simulation *Nonlinear Processes in Geophysics* **21** 1159-68

[14] Akhter M, Hassan D and Abbas S 2020 Predictive ARIMA Model for coronal index solar cyclic data *Astronomy and Computing* **32** 100403

[15] Brockwell P J and Davis R A 2016 Introduction to time series and forecasting (3rd edn. ed.). In: Berlin: Springer

[16] Liu H and Cocea M 2017 Semi-random partitioning of data into training and test sets in granular computing context *Granular Computing* **2** 357-86

[17] Mahmud I, Bari S H and Rahman M 2017 Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method *Environmental Engineering Research* **22** 162-8

[18] Camara A, Feixing W, and Xiaojin L 2016 Energy consumption forecasting using seasonal ARIMA with artificial neural networks models *International Journal of Business and Management* **11** 231

[19] Al-Qassem M A, Ewees A A, Fan H and Abd El Aziz, M 2020 Optimization method for forecasting confirmed cases of COVID-19 in China *Journal of Clinical Medicine* **9** 674