International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Predict Long-Term Average Monthly Temperature in Saudi Arabia's Al-Baha Region using SARIMA Model

Salem Alzahrani

Al-Baha University, Faculty of Science & Arts, dep of Math, Al-Mandag, KSA
Email: *salem.bb[at]hotmail.com*

Abstract: ***In this paper, we predict average monthly temperature in Al-Baha region, Saudi Arabia, from June 2003 to December 2019, using SARIMA (Self-Regressive Seasonal Moving Average) techniques. A detailed explanation of model selection and prediction accuracy is provided. The results of the model analysis show that the proposed research approach obtains good predictive accuracy.***

Keywords: Temperatures, Box-Jenkins, S ARIMA, Evaluation criteria

1. Introduction

The international warming is now an assumptive fact that is influencing the lower atmosphere as well as the oceans of the globe, and that the climate parameters remain in dynamic equilibrium. In addition, the dynamics of the climate system is chaotic.

So for a better planning, civilian planners need to have a clear insight of the situation in future.

This requires plausible forecasts of future population and other urban parameters. Fortunately, over the years, researchers have successfully developed advanced tools for obtaining better forecasts to support urban planning and management.

In recent years, increasing number of studies has appeared dealing with the impact of intense heat on health of urban dwellers

This study exact the general model of mean temperatures recorded in Al-Baha Region from 2003 to 2019 and developed Seasonal Autoregressive Integrated Moving Average (SARIMA) forecast model for predicting 2025 mean monthly temperature. Temperature is a physical amount that is a measure of hotness and coldness on a numerical scale. It is a measure of hotness and coldness on a numerical scale. It is a measure of the thermal energy per particle of matter or radiation and it is measured by a thermometer, which may be inspect in any of different temperature scales

1.1 Study Problem and Purpose

Predicting temperatures in the Al-Baha region in Saudi Arabia requires finding a model that can reasonably represent it. In the literature, several methods have been proposed for constructing time series models, however, the suitability of any of these methods for a given time series data must be judged on the basis of their suitability for that data. In this study, a time series modeling method, namely Box-Jenkins, will be applied to the data that represent the temperature in Al-Baha city in the Kingdom of Saudi Arabia, b

1.2 Hypothesis

The study assumes that the modeling and prediction of Box-Jenkins methodology is more accurate when each of them is used to model and predict the temperature of Al-Baha region in the Kingdom of Saudi Arabia.

1.3 Methodology

The study uses the augmented Dickey test as well as Box-Jenkins' methods of data representing temperature energy in order to fit a suitable model for temperature modeling and prediction. Criteria used in comparison such as mean absolute error and mean square error BIC.

2. Literature Review

Time series analysis of mean temperature based on air temperature data obtained from State Meteorological Service in Turkey between 1950-1994 was investigated by. The researcher said changes in average temperatures in Turkey from 1950 to 1994. The researcher also noticed a statistically significant cooling trend in 21 stations, as well as the direction of temperature rise in one station and the absence of direction in 36 stations.

Applied SARIMA on hourly bicycle count and temperature data and modeled Vancouver Bicycle Traffic Chang et al. (2012) applied the seasonal ARIMA model to the time series precipitation data from 1961 to 2011, for forecasting monthly precipitation in Yantai, China. They found that the model SARIMA (1, 0, 1) (0, 1, 1)12 fitted the past data and could be used successfully for forecasting. Asamoah Boaheng (2014) forecasted temperature by using SARIMA model in Ashanti region by analyzing the past data from 1980 to 2013, and The study concluded that the best predictor model was SARIMA (2,1,1) (1,1,2) (12). Mills (2014) successfully modeled the monthly temperatures of Cephalonia using the seasonal multiplexed ARIMA model. Kibunjaet al. (2014) forecasted precipitation of Mt. Kenya region by SARIMA model. The modelling was carried out based on Box Jenkins methodology, which involved model identification, parameter estimation and diagnostic check. Their model was found to be adequate for forecasting, as it also passed the normality tests of residuals. Dabralet al.(2017) developed the "Seasonal Auto Regressive Integrative Moving Average" models (SARIMA) for monthly, weekly and daily monsoon rainfall time series and predicted monsoon time series for 14 years, from 2014 to 2027.

D. K. DWIVEDI, G. R. SHARMA & S. S. WANDRE Forecasting of temperature can be done by combination of several mathematical models, using time series analysis. ARIMA (Auto Regressive Integrated Moving Average) model considers past data and prediction errors and relates its present data to obtain forecast. If time series has seasonality inherent in it, then Season ARIMA models are utilized to make forecast. Mean temperature forecasting for Junagadh city was carried out using SARIMA model, by using the past data from the period of 1984 to 2015. The orders The researcher estimated the model from autocorrelation diagram and partial autocorrelation diagram. Several nominee models have been developed to foretell mean temperature. The model with the lowest value was chosen for the Akaike Information Standard (AIC) as a suitable model for foretell mean temperature. SARIMA (1, 0, 1) (1, 1, 1) (12) was selected as the best model for foretell mean temperature. A normal residue test was performed to verify the adequacy of the specified form. allow to model diagnostics, the model was reliable for foretell mean temperatures.

3. Materials and Methods

Temperature data recorded from 2003 to 2019 were obtained for AL-BAHA city, From the Statistics Yearbook - Department of Agriculture and Water, in Al-Baha city. Monthly data in time series was analyzed by SARIMA model for forecasting means temperature using XLSTAT. SARIMA model requires the data to be stationary. Unit root test can be performed on the data to test whether the data is stationary or not. If it is not found to be stationary, then the data should be converted into stationary data. On shifting the time position to any arbitrary period, if the joint does not change, then the data is said to be stationary data.

Data Used Secondary data From the Statistics Yearbook - Department of Agriculture and Water from the period of 2003 to 2019 was used to develop a forecast model (SARIMA) in predicting future mean temperature values in AL-BAHA Region of KSA. The data was used since it is a time series data and the surveys were collected sequentially in time (monthly). Data was analyzed with R-Box-Jenkins methodology

In econometrics, The Box-Jenkins methodology consists of an iterative procedure of four steps: initial identification, estimation, diagnostic examination, and prediction.

The first step in developing the Box-Jenkins model is to determine if the time series is consistent and if there is any significant seasonality that needs modeling.

Stability assessment of the autocorrelation scheme. Non-static is often indicated by an automatic correlation diagram with very slow decay. In time series models, a linear random process has a unit root if 1 is the root of the equation characteristic of the process. The process will be inconsistent. If the other roots of the characteristic equation are inside the unit circuit, then the first difference in the process will be constant.

The augmented Dickey - Fuller test (ADF) is a test for unit root in a time series sample. The improved Dickey - Fuller (ADF) statistic used in the test is a negative number. The more negative it is, the more it rejects the premise that there are roots for loneliness at a certain level of confidence. In the modeling stage, the goal of the Dickey - Fuller test is to detect seasonality, if any, and to determine the order of the terms of the seasonal self-regression and the seasonal moving average. For many chains, the period is known.

Also, it is better to apply a seasonal difference to the data and to renew the auto-link and partial auto-link schemas. Because it helps to define a pattern for the non-seasonal component of the pattern. In some cases, seasonal variation may eliminate most or all of the effects of seasonality

The Autoregressive Model

In the autoregressive model, the current value X_t in the time series is expressed as a linear combination of the previous values, and an unexplained portion e_t. A typical autoregressive model of order p takes the form:

$$\left(1-\phi_1 B^1 - \phi_2 B^2 - \dots\dots\dots - \phi_p B^p\right)\tilde{X}_t = e_t$$

where the term m is a constant which representing the mean of the process, $\phi\ j$ (j=1,2,……,p) is the jth autoregressive parameter and e_t are the error term at time t.

The $e_t s$ are assumed to be independent normal distribution random variable with mean zero and variance $\sigma^2_{e_t}$.

The Moving Average Model:

In the moving average model of order q denoted by $MA(q)$ the current observation X_t is expressed as a linear combination of the random disturbances going back q periods, it is equation is written as:

$$\tilde{X}_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots\dots\dots - \theta_q e_{t-q}$$

The Mixed Autoregressive – Moving Average Models:

For stationary time series $X_t, X_{t-1}, X_{t-2}, \dots\dots\dots$ the mixed ARMA model is expressed as follows:

$$\phi(B)\tilde{X}_t = \theta(B)e_t$$

where δ is the constant mean of the series.

$\phi\ j$ are the autoregressive part parameters.

θj are the moving average part parameters. e_t is the error term at time t.

The ARIMA models:
ARIMA model takes the form:

$$\phi(B)(1-B)^d X_t = \theta(B)e_t$$

where $(1-B)^d$ is the dth order difference.
This is the model which summons the difference in the dth command of the time series to make it fixed. Doing exercise d is $0,1,$ or at most 2.

The seasonal model:
Seasonal ARIMA model of order $(p,d,q)\times(P,D,Q)^S$ can be written as:

$$\phi(B)\Phi(B^s)\nabla_d \nabla_s^D X_t = \theta(B)\Theta(B^s)e_t.$$

where
$(p,d,q) \equiv$ No seasonal part of the model.
$(P,D,Q) \equiv$ Seasonal part of the model.
S is the number of periods.

4. SARIMA Modelling

SARIMA model is the product of seasonal and non-seasonal polynomials and is designated by SARIMA (p, d, q) x (P, D, Q)S, where (p, d, q) and (P, D, Q) are non-seasonal and seasonal components, respectively with a seasonality's'. SARIMA model was defined at Equation 1 [5[
Φ(BS) φ(B)(1 – BS)D (1 – B)d yt = Θ(BS) θ(B) εt (1)

Where: Φ and φ = autoregressive (AR) parameters of seasonal and non-seasonal components, respectively; Θ and θ = moving average (MA) parameters of seasonal and non-seasonal components, respectively; B = backward operator, B(yt) = yt–1; (1–BS)D = Dth seasonal difference of season s; (1 – B)d = dth non-seasonal difference; εt = an independently distributed random variable; P and p = the orders of the AR components; Q and q= the orders of MA components; D and d are difference terms. Four sequential steps [6] as described below were followed for SARIMA modelling and forecasting.

4.1 Criteria

Numerous important quantitative error measurements used evaluation of time series models such as mean absolute error, mean absolute percentage error, Akaike's information criteria, Schwartz Basian criteria, and Theil U statistics. The formulation of any of these measures is expressed as follows:

Mean Absolute Error

$$\text{MAE} = \sum_{t=1}^{n} \frac{|e_t|}{n}$$

Sum Squared Error

$$\text{SSE} = \sum_{t=1}^{n} e_t^2$$

Mean Squared Error

$$\text{MSE} = \sum_{t=1}^{n} \frac{e_t^2}{n}$$

$$\text{AIC} = \ln \hat{\sigma}^2 + \frac{2k}{T}$$

$$\text{SB} = \ln \hat{\sigma}^2 + \frac{k}{T}\ln T$$

Where:
$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T}(\varepsilon_t - \mu)^2$, T is the number of observations and k is the number of model parameters. Among these measures the mean absolute error AIC and BSC are used to determine which of the two models: Box-Jenkins model or Exponential smoothing model is appropriate in modeling and forecasting Electricity energy sales in the kingdom of Saudi Arabia data.

4.2 Data

The data used in this study are the average Monthly Temperatures in Al-Baha Region from January 2003 to December 2019. The original temperature data are from Department of Agriculture and Water in Al-Baha Region, collected on an hourly basis, and there are no missing values. Meanwhile, the monthly mean temperature data are from the original observed temperature data. The time series of the monthly mean temperature is plotted in Figure 1.

4.3 Empirical Result

This section discusses the empirical analysis results such as augmented dickey fuller test of unit root, correlogram and the construction of SARIMA models for modeling the average monthly Temperatures recorded in Al-Baha Region in the Saudi Arabia as well as the comparison results among the methods under consideration.

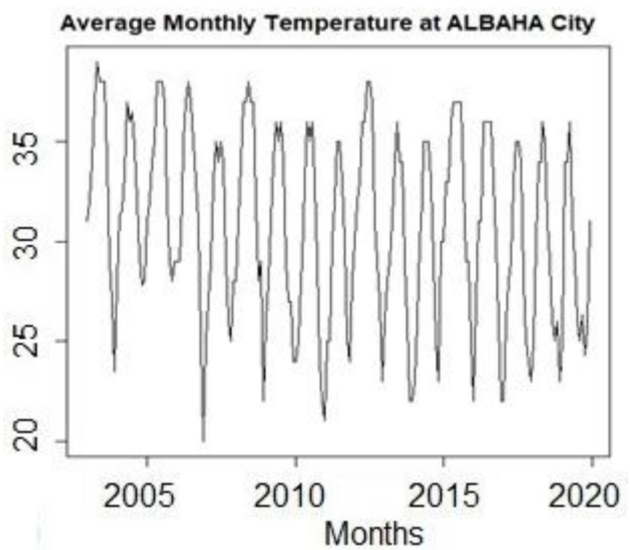


Figure 1: Sequence chart of average monthly Temperatures recorded in Al-Baha Region in Saudi Arabia from the periods 2003 to 2019

ADF Test
We apply ADF test for stationarity (Ho: There is no Stationarity H1: The series is stationary) table (1) indicted

that the average monthly temperatures series is stationary this mean that the non-seasonal differencing (d =0) furthermore, Fig1 show seasonal pattern in order to determine the orders/values of the non-seasonal and the seasonal AR and MA parts. Figure 2 & figure 3 gives the ACF and the PACF of our series.

Table (1) ADF test of average Monthly Temperatures in Al-Baha Region in Saudi Arabia from the periods 2003 to 2019.

Unit root test results / Diabetes series		
Test Type	Level	
	Test value	Prb
ADF test	-8.0345	0.01

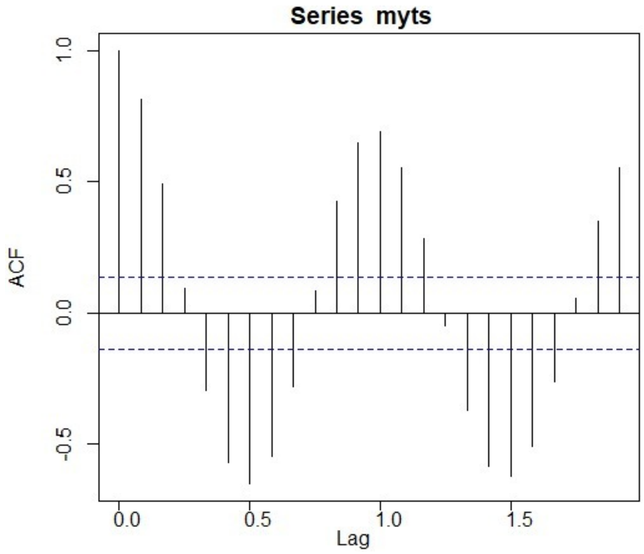


Figure 2: The ACF of average Monthly Temperatures in Al-Baha Region in Saudi Arabia from the periods 2003 to 2019

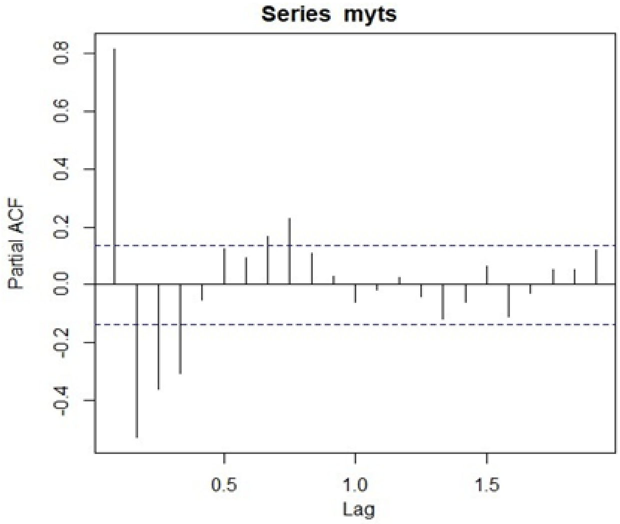


Figure 3: The PACF of average Monthly Temperatures in Al-Baha Region in Saudi Arabia from the periods 2003 to 2019.

We analyzed data using the ACF function to determine the seasonal trend. (Fig. 2) Showed that the average monthly temperature data have a seasonal trend with a return period 12 = ω. It was also shown that by taking seasonal differencing at lag 1, the seasonal trend was eliminated by choosing d = 1 as degree of differencing; in other words, seasonal differencing of the order 1 had the lowest level of ACF function violated from the confidence intervals. Other degrees of the SARIMA model, including autoregressive and seasonal and non-seasonal moving averages, were chosen based on the ranging from 0 to 3 using the trial and error method and the results of the evaluation have been presented in Table 2.

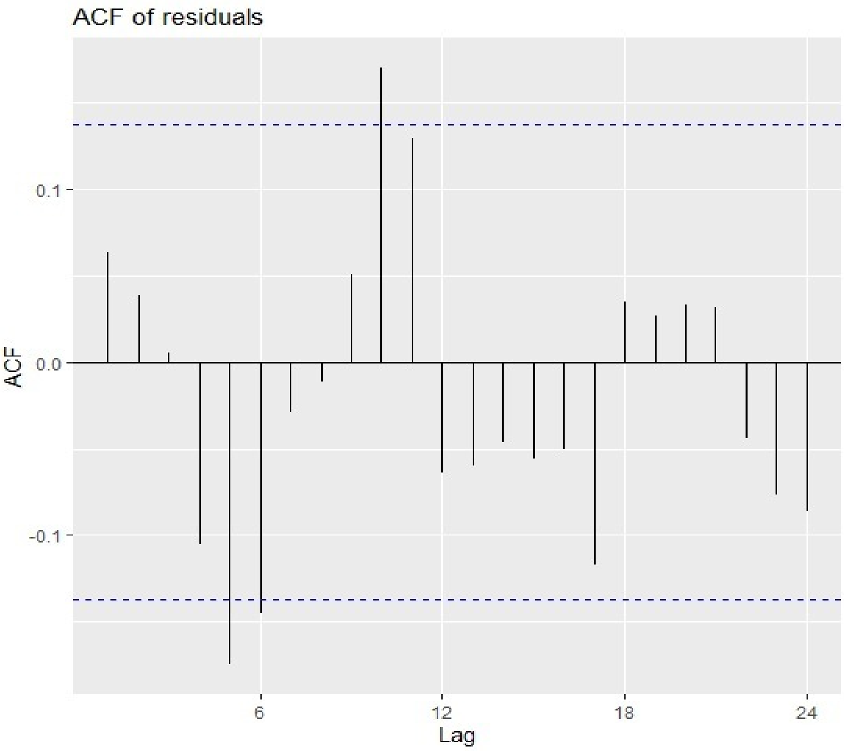


Figure 4: The correlogram of average Monthly Temperatures in Al-Baha Region in Saudi Arabia from the periods 2003 to 2019.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

To determine the ARIMA configuration, the auto.arima function in R is used, below the result:

Fitting models using approximations to speed things up...

ARIMA(2,0,2)(1,1,1)[12] with drift : 817.8924
ARIMA(0,0,0)(0,1,0)[12] with drift : 991.2744
ARIMA(1,0,0)(1,1,0)[12] with drift : 849.7656
ARIMA(0,0,1)(0,1,1)[12] with drift : 842.4043
ARIMA(0,0,0)(0,1,0)[12] with drift : 987.3255
ARIMA(2,0,2)(0,1,1)[12] with drift : 812.6509
ARIMA(2,0,2)(0,1,0)[12] with drift : 899.442
ARIMA(2,0,2)(0,1,2)[12] with drift : 815.8069
ARIMA(2,0,2)(1,1,0)[12] with drift : 860.2638
ARIMA(2,0,2)(1,1,2)[12] with drift : 821.5096
ARIMA(1,0,2)(0,1,1)[12] with drift : 808.2884
ARIMA(1,0,2)(0,1,0)[12] with drift : 893.5857
ARIMA(1,0,2)(1,1,1)[12] with drift : 817.6389
ARIMA(1,0,2)(0,1,2)[12] with drift : 811.9875
ARIMA(1,0,2)(1,1,0)[12] with drift : 858.4423
ARIMA(1,0,2)(1,1,2)[12] with drift : 821.1995
ARIMA(0,0,2)(0,1,1)[12] with drift : 824.0335
ARIMA(1,0,1)(0,1,1)[12] with drift : 804.4615
ARIMA(1,0,1)(0,1,0)[12] with drift : 888.3736
ARIMA(1,0,1)(1,1,1)[12] with drift : 814.2723
ARIMA(1,0,1)(0,1,2)[12] with drift : 808.442
ARIMA(1,0,1)(1,1,0)[12] with drift : 854.5845
ARIMA(1,0,1)(1,1,2)[12] with drift : 817.4156
ARIMA(1,0,0)(0,1,1)[12] with drift : 800.2494
ARIMA(1,0,0)(0,1,0)[12] with drift : 883.3181
ARIMA(1,0,0)(1,1,1)[12] with drift : 810.8101
ARIMA(1,0,0)(0,1,2)[12] with drift : 804.4005
ARIMA(1,0,0)(1,1,2)[12] with drift : 813.6989
ARIMA(0,0,0)(0,1,1)[12] with drift : 915.7929
ARIMA(2,0,0)(0,1,1)[12] with drift : 805.4834
ARIMA(2,0,1)(0,1,1)[12] with drift : 809.6309
ARIMA(1,0,0)(0,1,1)[12] with drift : 797.1627
ARIMA(1,0,0)(0,1,0)[12] with drift : 878.1244
ARIMA(1,0,0)(1,1,1)[12] with drift : 808.2484
ARIMA(1,0,0)(0,1,2)[12] with drift : 801.2578
ARIMA(1,0,0)(1,1,0)[12] with drift : 844.7271
ARIMA(1,0,0)(1,1,2)[12] with drift : 811.7645
ARIMA(0,0,0)(0,1,1)[12] with drift : 927.9215
ARIMA(2,0,0)(0,1,1)[12] with drift : 803.5265
ARIMA(1,0,1)(0,1,1)[12] with drift : 801.7973
ARIMA(0,0,1)(0,1,1)[12] with drift : 847.0589
ARIMA(2,0,1)(0,1,1)[12] with drift : 808.7176

Now re-fitting the best model(s) without approximations.
ARIMA(1,0,0)(0,1,1)[12]: 830.9945
Best model: ARIMA(1,0,0)(0,1,1)[12]

Estimating the selected Box-Jenkins model (SARIMA (1,0,0)(0,1,1) [12] with drift**):**

After a tentative average Monthly Temperatures in Al-Baha Region in Saudi Arabia model has been identified, its parameters will estimates. Bellow the estimated parameters of the selected Box Jenkins model using least square method.

Table (2) SARIMA (1,0,0)(0,1,1) [12] with drift model parameters estimation and other related statistics

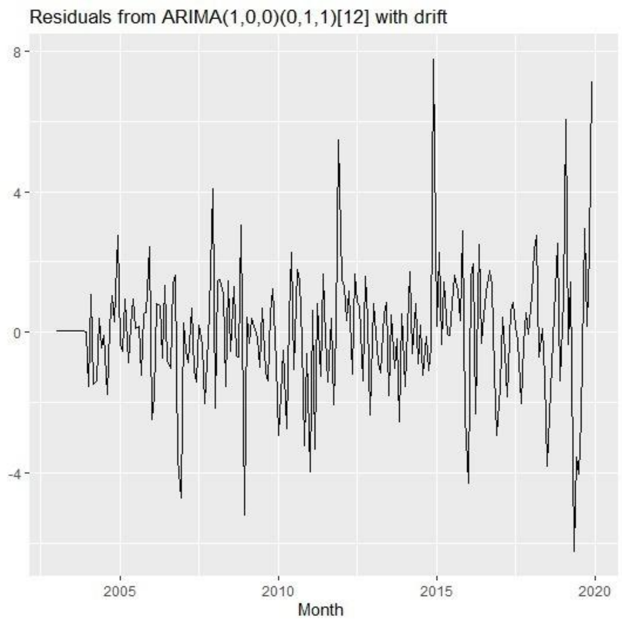
		Estimate	SE	t	Sig
average Monthly Temperatures in Al-Baha Region in Saudi Arabia model	ar1	-.089	- 0.8632	-.402-	.002
	sma1	-0.8632	0.0708	1.995	.023
	drift	-0.0162	0.0088		

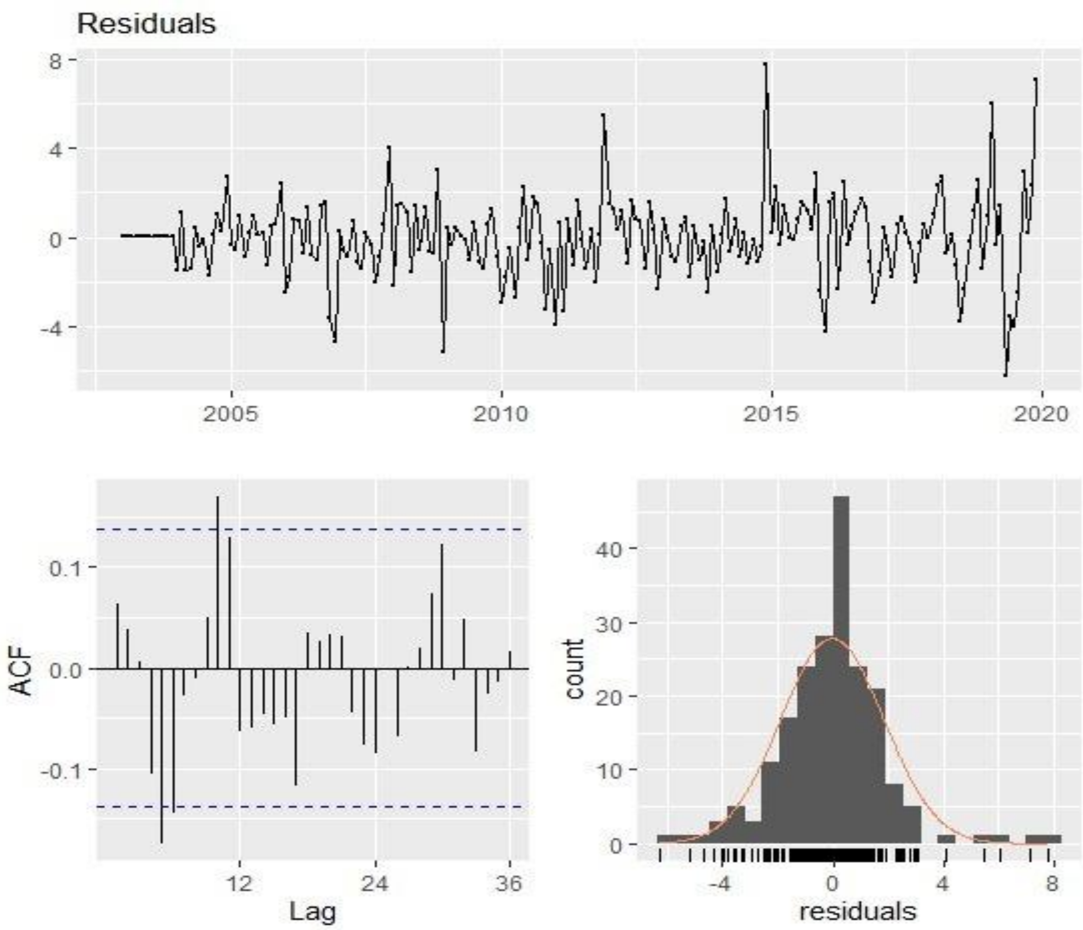
Table (2) shows SARIMA (1,0,0)(0,1,1) [12] with drift model parameter estimates of Electricity sales energy in Saudi Arabia data

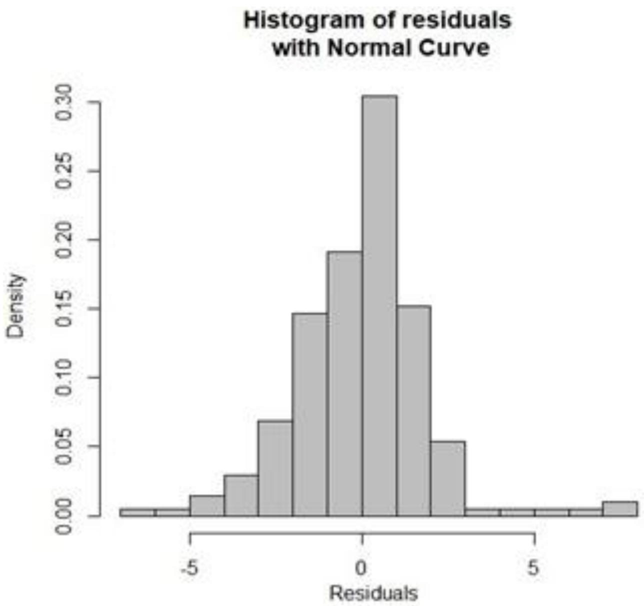
Model Diagnostic Checking:

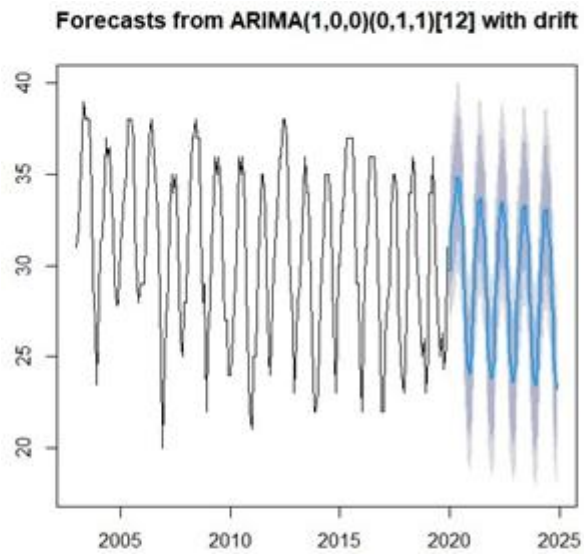
Normality Test for Residuals

Figure (5) gives the normality test of the residuals; it can be shown that most of the points pass through the straight line. This display that the residuals are normally distributed. And distribution of the residuals can be clearly seen as normal, Therefore SARIMA (1, 0, 0) (0, 1, 2) (12) with drift satisfies all the model assumptions indicating that the model is very good for forecasting









Forecasting using SARIMA (1, 0, 0)(0, 1, 1) [12] with drift

Using the derived model, the following Prediction was made for the year 2020 up to 2025 as shown in Figure (7).

5. Conclusion

In This paper we forecast average Monthly Temperatures in Al-Baha Region in Saudi Arabia using SARIMA model and also to determine the precision of the SARIMA model in prediction. To avoid fitting over parametrized model, AIC and BIC were working in choosing the best model. The model with a minimum value of these information criterions is considered as the best (Akaike (1979); Akaike (1974)). Furthermore, ME, MSE, RMSE, MAE, MPE, MAPE were also employed. The residuals of the model were examined to see whether if they were white noise

6. Declaration of Competing Interest

The author declare that they have no known competing financial

Interests or personal relationships that could have appeared to influence the work reported in this paper

Conflicts of Interest

The author declare no conflicts of interest regarding the publication of this paper.

References

[1] A. Meyler, G. Kenny and T. Quinn, (1998) 'Forecasting Irish Inflation using ARIMA Models ' Research paper Economic Analysis, Research and Publications department, central bank of Ireland, Dublin.
[2] Al-Kubaisi, Q. Y., and M. M. Gardi. 2012. Dust storm in Erbil City as a result of climatic change in Kurdistan Region Iraq. Iraqi Journal of Science 53 (3C):40–44.
[3] Asamoah - Boaheng, M. 2014. Using SARIMA to forecast monthly mean surface air temperature in the Ashanti Region of Ghana. International Journal of Statistics and Applications (6):292–98. doi:10.5923/j.statistics.20140406.06.
[4] Babazadeh, H., and S. A. Shamsnia. 2014. Modeling climate variables using time series analysis in arid and semi arid regions. African Journal of Agricultural Research 9 (26):2018–27.
[5] Box, G. E., and G. M. Jenkins. 1970. Time series analysis: Forecasting and control. San Francisco, CA: Holden-Day
[6] Box, G. E., G. M. Jenkins, and G. C. Reinsel. 1994. Time series analysis: Forecasting and control, 3rd ed. Englewood Cliffs, NJ: Prentice Hall.
[7] Box, G. E., Jenkins, G. M., Reinsel, G. C., &Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
[8] Can, A., and A. T. Atimtay. 2002. Time series analysis of mean temperature data in Turkey. Applied Time Series 4:20–23.
[9] Chang, X., Gao, M., Wang, Y., &Hou, X. (2012). Seasonal autoregressive integrated moving average model for precipitation time series. Journal of Mathematics & Statistics, 8(4).
[10] Chatfield, C. 2004. The analysis of time series: An introduction, 6th ed. London, UK: Chapman & Hall/CRC.
[11] Dabral, P. P., & Murry, M. Z. (2017). Modelling and Forecasting of Rainfall Time Series Using SARIMA. Environmental Processes, 1-21.
[12] El-Kadi, A. 2001. Variation of rainfall and drought conditions in Gaza–Palestine: On a regional and global context. Journal of the Islamic University 9 (2):41–66.
[13] Grieser, J., S. Tromel, and C. D. Schonwiese. 2002. Statistical time series decomposition into significant components and application to European temperature. Theoretical and Applied Climatology 71:171–83
[14] Hamilton. J., (1994). 'Time Series Analysis', Princeton University press, New Jersey, USA
[15] Jeffrey, J. 1990. Improving forecasts by decomposing the error. Journal of Business Forecasting Methods & Systems 9 (1):12–15.
[16] Kaushik, I., & Singh, S. M. (2008). Seasonal ARIMA model for forecasting of monthly rainfall and temperature. Journal of Environmental Research and Development, 3(2).
[17] Khalid, N. J. 2014. Urban heat island in Erbil City. Student thesis series, INES, NGEM01 20131, Report No. 308, 1–57. https://doi.org/lup.lub.lu.se/student-papers/record/4449039 (accessed July 2, 2015).
[18] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin (1992) 'Testing the null of stationarity against the alternative of a unit root: How sure are we that the economic time series have a unit root? Journal of Econometrics 54: 159-178.
[19] Seasonal Autoregressive Integrated Moving Average Model for Precipitation Time Series Xinghua Chang, Meng Gao, Yan Wang, Xiyong HouMathematics2012