Modeling Monthly Average Temperature of Dhahran City of Saudi-Arabia Using Arima Models

Nurudeen Ayobami Ajadi¹, *, Jimoh Olawale Ajadi², Adams Saddam Damisa³, Osebekwen Ebenezer Asiribo¹, Ganiyu Abayomi Dawodu¹

¹Department of Statistics, College of Physical Sciences, Federal University of Agriculture, Abeokuta, Nigeria
²Department of Mathematics and Statistics, King Fahad University of Petroleum and Minerals, Dhahran, Saudi Arabia
³Department of Statistics, Ahmadu Bello University, Zaria, Nigeria

Email address: ajadinurudeen2014@yahoo.com (N. A. Ajadi)
*Corresponding author

To cite this article:
Nurudeen Ayobami Ajadi, Jimoh Olawale Ajadi, Adam Saddam Damisa, Osebekwen Ebenezer Asiribo, Ganiyu Abayomi Dawodu. Modeling Monthly Average Temperature of Dhahran City of Saudi-Arabia Using Arima Models. International Journal of Data Science and Analysis. Vol. 3, No. 5, 2017, pp. 40-45. doi: 10.11648/j.ijdsa.20170305.12

Received: August 30, 2017; Accepted: September 19, 2017; Published: November 1, 2017

Abstract: Temperature is the coldness and hotness of the body and its unit is measured in Celsius. The data used for this research work is the average monthly temperature of Dhahran city which is located in the Kingdom of Saudi Arabia. The data range is from 1951 to 2010, and sample data of 1951 to 2008 was used for the estimation to choose the best model and the sample data from 2009 through 2010 was left for the forecast. Different models were tried but ARIMA (2, 1, 1) (0, 1, 1) is selected as the best model because of its low sic and aic criteria and also the forecast error, the best model is used for forecasting.

Keywords: ARMA, ARIMA AR, MA, SMA

1. Introduction

Temperature as it is commonly known is the degree of hotness or coldness of a body or region. This means that there is a need to monitor the variation in the temperature of different regions from time to time. According to [1-2], the world is warming 0.6±0.2°C over 100 years. So there is need to predict future climate.

The monthly mean, maximum and minimum temperatures of countries with 37% global land mass were analyzed by [3]. According to [4], there is change in climatic condition of many countries and it is one of the major environmental threats to food production and livelihoods. Homogeneity of annual mean temperature in given stations was analyzed using Cumulative deviation test [5] and first order ACF test [6].

Dhahran’s climate is characterized by hot, humid summers, and cold long winters. Temperatures can rise to more than 40°C (100°F) in the summer, coupled with extreme humidity (85 & ndash 100%), given the city’s proximity to the Persian Gulf. The highest recorded temperature in Dhahran is 51.1°C (124.0°F). In winter, the temperature rarely falls below −2°C (28°F), with the lowest ever recorded being −5°C (23°F) in January 1964. The Shamal winds usually blow across the city in the early months of the summer, bringing dust storms that can reduce visibility to a few metres. These winds can last for up to six months.

Several authors have studied the climate variability in different countries, [7] studied the climate of Bahrain during the past six decades, principally the temperature and rainfall trends. The study [7] demonstrated enormous climate variability, represented by alternate hot-dry and cool-wet events. The Mean, maximum and minimum surface air temperatures recorded at 70 climatic stations in Turkey during the period 1929–1999 were evaluated by [8]. Also Climate change has the potential to affect all natural systems, thereby becoming a threat to human development and survival socially, politically, and economically [10].

2. Methodology

Time series models have been very useful in studying the
behavior of process over a period of time. It has wider applications which include; sales forecasting, weather forecasting, inventory studies etc. In decisions that involve factor of uncertainty of the future, time series models have been found one of the most effective methods of forecasting.

2.1. Autoregressive Moving Averages (ARMA)

An ARMA process of order \( p, q \) is a stationary process \( X_t \) that satisfies the relation

\[
X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}
\]

Where \( \{\varepsilon_t\} \) is a white noise.

In lag form, equation (1) becomes;

\[
X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}
\]

\[
\phi(B)X_t = \theta(B)\varepsilon_t
\]

\[
X_t = \theta(B)\varepsilon_t \phi^{-1}(B).
\]

Also,

\[
\varepsilon_t = \phi(B)X_t \theta^{-1}(B)
\]

Also, equation (2) and (3) can also be written as;

\[
X_t = \theta(B)\varepsilon_t
\]

Where \( \theta(B) = \theta(B)\varepsilon_t \phi^{-1}(B) \).

\( \theta(B) \) is called the phi-weight.

Also,

\[
\varepsilon_t = \pi(B)X_t
\]

Where, \( \pi(B) = \theta(B)\varepsilon_t \phi^{-1}(B) \).

\( \pi(B) \) is called the pi-weight.

2.2. Autoregressive Integrated Moving Average Process (ARIMA)

A process \( X_t \) is said to be an autoregressive integrated moving average process of order \( (p, d, q) \) if its \( d \)th difference is an ARMA \( (p, q) \) process. An ARIMA \( (p, d, q) \) model can be defined by:

\[
\phi(\beta)\nabla^d X_t = \theta(\beta)\varepsilon_t
\]

Where \( p, d, q \) are non-negative integers.

Note: When \( d = 0 \), the ARIMA \( (p, d, q) \) becomes ARMA(\( p, q \)).

2.3. Diagnostic Checking

The Box-Jenkins [9] methodology required examining the residuals of the actual values minus those estimated through the model. The model is assumed to be appropriate if its residuals are random, but if the residuals are not random, another model will be entertained, then its parameters will be estimated, in order to check for randomness. Several tests (e.g., the Box-Pierce, Box Ljung test, Shapiro test e.t.c) have been suggested to help users determine if overall the residuals are indeed random. Although it is a standard statistical procedure not to use models whose residuals are not random, it might be interesting to test the consequences of lack of residual randomness on post-sample forecasting accuracy.

3. Result and Discussion

The data used in this paper is the average monthly temperature of the Dhahran city which is located in the Kingdom of Saudi Arabia. The data collected range from 1951 to 2010, the sample data of 1951 to 2008 was used for the estimation to choose the best model and the sample data 2009 through 2010 was left for the forecast.
The data-sets in figure 1 will be reduced for the period of 1951 to 1960, in order to reveal its monthly seasonal pattern in better details. There is no need to model trend here since there is no trace of it.

The reduced data-set shows that the graph is seasonal and some cyclical variation could be noticed. Now, the seasonality using dummy variables will be modeled.

Table 1. Model for the Seasonality with Dummy Variables.

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|-------|
| D1       | 15.60678    | 0.179038   | 87.17034    | 0.0000|
| D2       | 17.14576    | 0.179038   | 95.76620    | 0.0000|
| D3       | 20.92203    | 0.179038   | 116.8582    | 0.0000|
| D4       | 25.93220    | 0.179038   | 144.8421    | 0.0000|
| D5       | 31.24828    | 0.180575   | 173.0492    | 0.0000|
| D6       | 35.85085    | 0.179038   | 200.2419    | 0.0000|
| D7       | 35.31186    | 0.179038   | 197.2314    | 0.0000|
| D8       | 28.45593    | 0.179038   | 182.0183    | 0.0000|
| D9       | 35.85085    | 0.179038   | 200.2419    | 0.0000|
| D10      | 35.31186    | 0.179038   | 197.2314    | 0.0000|
| D11      | 32.58814    | 0.179038   | 182.0183    | 0.0000|
| D12      | 22.71186    | 0.179038   | 158.9382    | 0.0000|
| R-squared| 0.965633    |            |             |       |
| Adjusted R-squared| 0.965633| S. D. dependent var | 26.47327 |
| S. E. of regression| 1.375215 | Akaike info criterion | 7.360239 |
| Sum squared resid | 1314.395 | Schwarz criterion | 3.569339 |
| Log likelihood | -1222.395 | Hannan-Quinn criter. | 3.521836 |

Dependent Variable: TEMPERATURE
Method: Least Squares
Sample: 1951M01 2009M12
Included observations: 707

Table 1 shows the estimation results. The 12 seasonal dummies account for more than 96 percent of the variation in Dhahran monthly temperature, as $R^2 = 0.965633$. At least few of the remaining variation are cyclical, which will be designed to capture in the final model. Also the low Durbin-Watson stat shows the evidence of serial correlation which needs to be removed.

The residual of the model in Figure 3 shows that the data are random along the mean zero. There are still some non-random pattern in the residual of the model between the period of 1970 - 1977 which shows that the model still need little adjustment, also there is need to check the corellogram of the output of the model whether it is white noise or not.

Table 2. The Corellogram of the Model.

| S/N | Autocorrelation | Partial Autocorrelation | Q-Stat | Prob |
|-----|----------------|-------------------------|--------|------|
| 1   | 0.451          | 0.451                   | 141.96 | 0    |
| 2   | 0.361          | 0.198                   | 233.26 | 0    |
| 3   | 0.298          | 0.102                   | 295.49 | 0    |
| 4   | 0.279          | 0.097                   | 350.24 | 0    |
| 5   | 0.259          | 0.073                   | 397.5  | 0    |
| 6   | 0.265          | 0.088                   | 447.03 | 0    |
| 7   | 0.259          | 0.07                    | 494.39 | 0    |
| 8   | 0.257          | 0.065                   | 541.18 | 0    |
| 9   | 0.301          | 0.123                   | 605.29 | 0    |
| 10  | 0.288          | 0.066                   | 664.08 | 0    |
| 11  | 0.246          | 0.006                   | 706.99 | 0    |
| 12  | 0.277          | 0.082                   | 761.31 | 0    |
| 13  | 0.261          | 0.04                    | 809.9  | 0    |
| 14  | 0.227          | -0.003                  | 846.73 | 0    |
| 15  | 0.203          | -0.009                  | 876.24 | 0    |
| 16  | 0.199          | 0.006                   | 904.45 | 0    |
### Table 3. Testing for Stationarity.

| Variable          | Coefficient | Std. Error | t-Statistic | Prob.  |
|-------------------|-------------|------------|-------------|--------|
| TEMP(-1)          | -0.155149   | 0.058680   | -2.643970   | 0.0084 |
| D(TEMP(-1))       | -0.419506   | 0.064393   | -6.514760   | 0.0000 |
| D(TEMP(-2))       | -0.376226   | 0.061403   | -6.127201   | 0.0000 |
| D(TEMP(-3))       | -0.454223   | 0.056430   | -8.049315   | 0.0000 |
| D(TEMP(-4))       | -0.488096   | 0.050892   | -9.590896   | 0.0000 |
| D(TEMP(-5))       | -0.516201   | 0.046732   | -11.04602   | 0.0000 |
| D(TEMP(-6))       | -0.516197   | 0.043415   | -11.88969   | 0.0000 |
| D(TEMP(-7))       | -0.551983   | 0.039998   | -13.80038   | 0.0000 |
| D(TEMP(-8))       | -0.614759   | 0.037812   | -16.25824   | 0.0000 |
| D(TEMP(-9))       | -0.562388   | 0.037685   | -14.92339   | 0.0000 |
| D(TEMP(-10))      | -0.485414   | 0.038239   | -12.69410   | 0.0000 |
| D(TEMP(-11))      | -0.322482   | 0.036435   | -8.850916   | 0.0000 |
| C                 | 4.113473    | 1.553740   | 2.647465    | 0.0083 |

Dickey fuller (Table 3) shows that the data is not stationary since it is not significant, hence there is need to check the correlogram of the series.
Table 4. The Output of ARMA Models.

| Table 4. The Output of ARMA Models. | Seasonal Model only | Seasonal Model with ARMA (1, 2) | Seasonal Model with ARMA(1, 1) | Seasonal Model with ARMA (0, 2) | ARMA (0, 2) and SAR (12) |
|------------------------------------|---------------------|-------------------------------|-------------------------------|-------------------------------|--------------------------|
| R-Square                           | 0.9659              | 0.9752                        | 0.9741                        | 0.9935                        | 0.9544                   |
| Adjusted R-Square                  | 0.9653              | 0.9747                        | 0.9741                        | 0.988                         | 0.9543                   |
| Durbin Watson                      | 1.1022              | 1.9709                        | 1.79871                       | 1.8910                        | 1.9912                   |
| Aic                                | 3.4824              | 3.1697                        | 3.1909                        | 2.7685                        | 3.7442                   |
| Sic                                | 3.5608              | 3.269                         | 3.2836                        | 3.4286                        | 3.7519                   |

After series of models were tried, we arrived at ARIMA (2, 1, 1) (0, 1, 1)_{12} as the best model because of its low Sic and Aic criteria and forecast error. The output is given below:

Dependent Variable: DLSTEMP  
Method: Least Squares  
Sample (adjusted): 1952M03 2008M12  
Included observations: 682 after adjustments  
Convergence achieved after 10 iterations

| Variable               | Coefficient | Std. Error | t-Statistic | Prob.   |
|------------------------|-------------|------------|-------------|---------|
| AR(1)                  | 0.226564    | 0.041769   | 5.424168    | 0.0000  |
| MA(1)                  | -0.931651   | 0.015684   | -59.40130   | 0.0000  |
| SMA(12)                | -0.952084   | 0.009168   | -103.8529   | 0.0000  |
| R-squared              | 0.645415    | Mean dependent var | 0.004106   |
| Adjusted R-squared     | 0.644371    | S. D. dependent var | 1.968256   |
| S. E. of regression    | 1.173763    | Akaike info criterion | 3.616295   |
| Sum squared resid      | 935.4711    | Schwarz criterion | 3.182600   |
| Log likelihood         | -1075.479   | Human-Quinn criter. | 3.170399   |
| Durbin-Watson stat     | 2.020834    | .23        |             |         |
| Inverted AR Roots      | 1.00        | .93        | .86+.50i    | .86-.50i|
| Inverted MA Roots      | 0.50+.86i   | .50-.86i   | .00+1.00i   | .00-1.00i|
|                        | -0.50+.86i  | -.50-.86i  | -.86+.50i   | -.86-.50i|
|                        | -1.00       |            |             |         |

The dependent variable was DLSTEMP (after taking the first trend difference and first seasonal difference) while the independent variables were AR(1), MA(1) and SMA(12). All the independent variables were significant because there probabilities were nearly equal to zeros. The R square value was only about 64.4 percent of the variance of the dependent variable (DLSTEMP) by the variables included in the model (AR(1), MA(1) and SMA(12)). The R-Square measures the in-sample success of the regression equation in forecasting DLSTEMP. Durbin Watson value is about 2.020834 which indicate that there is no serial correlation in the series. The values of SIC and AIC are the lowest compare to other model tried. Now let check its residual and corellogram before using it for the forecast.

3.2. Interpretation of the Output of the Preferred Model

The equation of the model is given as follows

\[ y_t = y_{t-12} + (1 + \varphi_1) y_{t-1} + (\varphi_2 - \varphi_1) y_{t-2} - (1 + \varphi_1) y_{t-13} + \varphi_2 y_{t-3} + (\varphi_1 + \varphi_2) y_{t-14} + y_{t-15} + \epsilon_t + \theta \epsilon_{t-1} + \eta \epsilon_{t-12} + \theta \eta \epsilon_{t-13} \]

the mean zero, which is good for our model. There is no serial correlation in the model but the presence of white noise in the model need to be checked before using it for the forecast.

![Figure 4. The Residual of The Model.](image-url)
The condition of the invertibility of the MA (1) and SMA (12) process also hold. The inverses of all of the roots must be inside the unit circle.

The forecast graph is very good since all the data and their estimate values lies in the 95% confidence band and also the actual values and their estimates are not too far from each other.

4. Conclusion

After series of models were tried, we arrived at ARIMA (2,1,1) (0,1,1)_{12} as the best model because of its low Sic and Aic criteria and forecast error. It has been observed that if you fit a seasonal model, the estimate of our forecast is predictable because of its low Durbin Watson. Dhahran average monthly temperature is better fitted with seasonal ARIMA (2,1,1) (0,1,1)_{12} in many ways, from its advantages mentioned above, it also has no serial correlation and its forecast is good as well not predictable beforehand.

References

[1] Nicholls N, Collins D. 2006. Observed climate change in Australia over the past century. Energy & Environment 17(1): 1–12.

[2] Folland CK, Karl TR, Christy JR, Clark RA, Gruza GV, Jouzel J, Mann ME, Oerlemans J, Salinger MJ, Wang SW. 2001. Observed Climate Variability and Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Johnson CA (eds). Cambridge University Press: Cambridge; 99–181.

[3] Karl TR, Janes PD, Knight RW, Kukla J, Plummer N, Razuvaev V, Gallo KP, Lindesay J, Charlson RJ, Peterson TC. 1993. A symmetric trends of daily maximum and minimum temperatures: empirical evidence and possible causes. Bulletin of the American Mathematical Society 74: 1007–1023.

[4] Nury, A. H., Koch, M. & Alam, M. J. B. (2013), Time Series Analysis and Forecasting of Temperatures in Sylhet Division of Bangladesh, 4th International Conference on Environmental Aspects of Bangladesh (ICEAB), August 24-26.

[5] Buishand JA. 1982. some methods for testing the homogeneity of rainfall records, 11-57 pp.

[6] Mitchell JM. 1966. Climatic Change, Technical note 79, WMO No. 195, 2-5, 60 pp. Rasouli AA. 2003. Preliminary time series analysis of Tabriz air temperature, IRIMO, NIVAR No. 46, 47, 16-20 pp.

[7] Elagib NA, Addin Abdu AS. 1997. Climate variability and aridity in Bahrain. Journal of Arid Environment 36: 405–419.

[8] Turkes M, Sumer UM, Demir I. 2002. Re-evaluation of trends and changes in mean, maximum and minimum temperatures of Turkey for the period 1929–1999. International Journal of Climatology 22: 947–977.

[9] Box, G. E. P. & Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control. Revised Edition, Holden Day: San Francisco, CA.

[10] Olufemi, S. O., Femi, J. A. & Oluwatosin, T. D. (2010), Time Series Analysis of Rainfall and Temperature in South West Nigeria, the Pacific Journal of Science and Technology, Vol. 11 No. 2 pp 552-564.