Temperature Forecasting during Ramadan and Hajj in Year 1438AH Using Seasonal-Trend Decomposition

M. Aljamal, T. Habebullah, E. Morsi, S. Munir, E. Falata
The Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research,
Umm Al-Qura University

Abstract

Nowadays, scientists devote their research to establish numerical weather prediction methods for critical weather phenomena, especially air temperature. The peak air temperature is one of the most important variables for predicting future climate changes. Therefore, this study focuses on forecasting (14:00 O'clock) air temperature at Al-Aziziah –Makkah during the seasons of Ramadan and Hajj, 1438 A.H.
The prediction was done by using Seasonal-Trend Decomposition based on Loess (STL), which has a high accuracy of 3.24%. This means that the prediction error was small and, therefore this method can be used in predicting air temperature during any hour of the day. This study is the first study to use this kind of forecasting model to predict Makkah's temperature.

1 Introduction

In big events monitoring weather is vital due to the plans based on it. In Mecca Saudii Arabia there are the big event: pilgrimage season, which is based on lunar calender[1]. This big event requires a lot of preparations of water quantity, electricity, food storage,...etc. In addition to that there is Umrah to Makkah all over the year, the crowd of Umrah reaches peak at Ramadan month (the month of fasting). The Proper planning for these events is crucial, especially if it comes in summer where the temperature reaches near 50 Degree Celsius. So, temperature forecasting is of great importance to us. In this paper, we employ the *Seasonal-Trend Decomposition based on Loess (STL)* to predict future temperature, especially at peak hour of the day (14:00 O'clock) to prevent thermal stress for many people.

First as we have missing values of the temperature in the data, we must preprocess the data by using imputation of missing data. Afterwards, we get a time series of the temperature to be used for prediction. The obtained time series was used as input to predict temperature in years 2016 and 2017.

At the time of writing this paper, we had data for the first seven months of year 2016 (213 days). The first test of the prediction technique was to compare those values against predicted values and the

A lunar year is about 354.367 days and drift each solar year by 11 to 12 days.[1]

error was 6.59%. The next step was to get the temperature of next Ramadan (27 May ~ 25 June) and Hajj from 8 Zul-Hijjah ~ 13 Zul-Hijjah (1 September ~ 6 September 2017).

2 Data Preprocessing

The data were taken from observatory Al-Aziziah, which had some missing observation values. In the next subsections, we exploit *Missing Data Imputation Techniques*.

2.1 Imputation of Missing Data

A direct approach to missing data is to exclude them from analysis. Another approach is to fill in or imputate missing values. A variety of imputation approaches can be used that range from extremely simple to rather complex approaches. These methods keep the full sample size, which can be advantageous for bias and precision; however, they can yield different kinds of bias.

There are many techniques of imputation, e.g., Mean substitution and Regression imputation.

2.1.1 Regression Imputation

Regression imputation has the opposite problem of mean imputation. A regression model is estimated to predict observed values of a variable based on other variables, and that model is then used to impute values in cases where that variable is missing. In other words, available information for complete and incomplete cases is used to predict whether a value on a specific variable is missing or not. Fitted values from the regression model are then used to impute the missing values. The problem is that the imputed data do not have an error term included in their estimation, thus the estimates fit perfectly along the regression line without any residual variance. This causes relationships to be over identified and suggest greater precision in the imputed values than is warranted. The regression model predicts the most likely value of missing data but does not supply uncertainty about that value [2]. Stochastic regression was a fairly successful attempt to correct the lack of an error term in regression imputation by adding the average regression variance to the regression imputations to introduce error. Stochastic regression shows much less bias than the above-mentioned techniques [6].

In this paper, the stochastic regression was used to impute the missing values.

3 Seasonal-Trend Decomposition based on Loess (STL)

The method of STL originally proposed by [1] is a filtering procedure for decomposing a seasonal time series Y_t into three components as in figure (1), i.e.,

$$Y_t = T_t + S_t + R_t, \qquad (1)$$

where T_t trend component, S_t seasonal component component, R_t remainder component. Basically, the STL procedure is composed of a sequence of applications of the locally weighted regression smoother (LOESS). LOESS is a useful polynomial for smoothing the data of time series. Let x_i and y_i are the measurements of independent and dependent variables respectively, where $1 \leq i \leq n$, one can fit the LOESS regression curve $h(x)$ by LOESS smoothing.

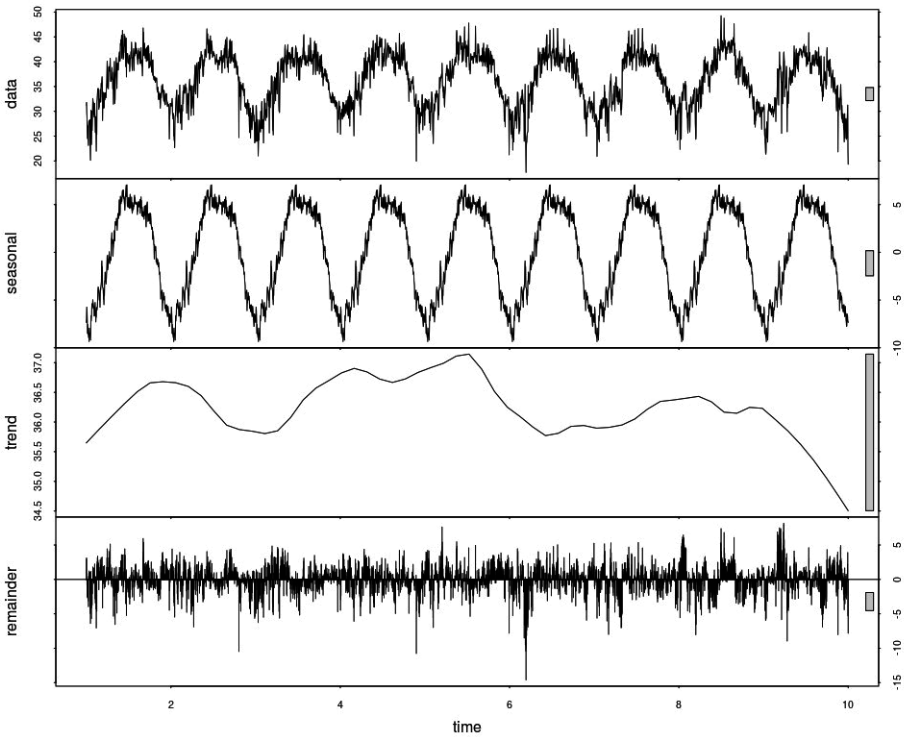

Figure 1: The temperature (top) and its three additive components obtained from STL decomposition.

The LOESS is a very important feature in STL and allows us to deal with missing values and extract the trend and the seasonal component of a time series in a straightforward way [1]. Two recursive procedures compose STL, namely outer loops and inner loops. The inner loop is nested inside an outer loop; each of the passes through the inner loop consists of a seasonal smoothing that updates the seasonal component, followed by a trend smoothing that updates the trend component once. There are six steps in the inner loop: detrending, cycle-subseries, low-pass filtering of a smoothed cycle-subseries, detrending of a smoothed cycle-subseries, deseasonalizing and trend smoothing. Each pass through the outer loop computes the robustness weight, which can be used in the next inner loop to reduce the influence of transient, aberrant behavior of trend and seasonal components, but is only needed if there are outliers. Therefore, all of the loops in the STL with a sequence of smoothing operator Loess is for the purpose of detrending, deseasonalizing, and reducing the influence of transient outliers on both the trend and seasonal components [3, 4].

STL has six parameters that determine the degree of smoothing in trend and seasonal components [5]:

1. n_p= the number of observations in each cycle of the seasonal component;

2. n_i= the number of passes through the inner loop;
3. n_o= the number of robustness iterations of the outer loop;
4. n_l= the smoothing parameter for the low-pass filter;
5. n_s= the smoothing parameter for the seasonal component; and
6. n_t= the smoothing parameter for the trend component.

In order to use STL method to model our temperature data set, four steps are needed:

1. use STL to separate S_T from the original temperature time series,
2. fit different conics for the remainders with different period lengths and evaluating the *p*-values of these conics to detect whether the T_t exists. If the conic exists, we choose the best one to model T_t according to the smallest *p*-value,
3. model the last component as the color noise using the AutoRegression(AR) model,
4. forecast the temperature values using the different component models which have been built.

To forecast a decomposed time series, we separately forecast the seasonal component, and the seasonally adjusted component. It is assumed that the seasonal component is unchanging, or changing extremely slowly, and so it is forecast by simply taking the last year of the estimated component. In other words, a seasonal naive method is used for the seasonal component. To forecast the seasonally adjusted component, any non-seasonal forecasting method may be used [4].

4 Simulation

The temperature data recorded every hour over a day i.e. 24 hours per every day and as we are interested in the highest hourly temperature at day time namely at 14:00 o'clock. It is clear that the data have seasonality every year, where every day is represented by hottest degree, so there are 365 values per year, total of nine years were used to do forecast of the next two years.

It might also be important to mention that the mean absolute percentage error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage, and is defined by the formula:

$$MAPE=\frac{100}{n}\sum_{i=1}^{n}\left|\frac{T_i-F_i}{T_i}\right| \qquad (2)$$

Where T_i is the actual value and F_i is the forecast value.

4.1 Two Years Prediction

In this subsection, we predict two years, i.e., 2016 and 2017 as next Ramadan and Hajj comes in year 2017. As shown in figure (1), the predicted values are evaluated using forecast intervals of 80% and 90%, respectively and the mean of either represents the point prediction. Figure (2) zooms the two years prediction and shows locations of Ramadan and Hajj seasons in a visualized manner.

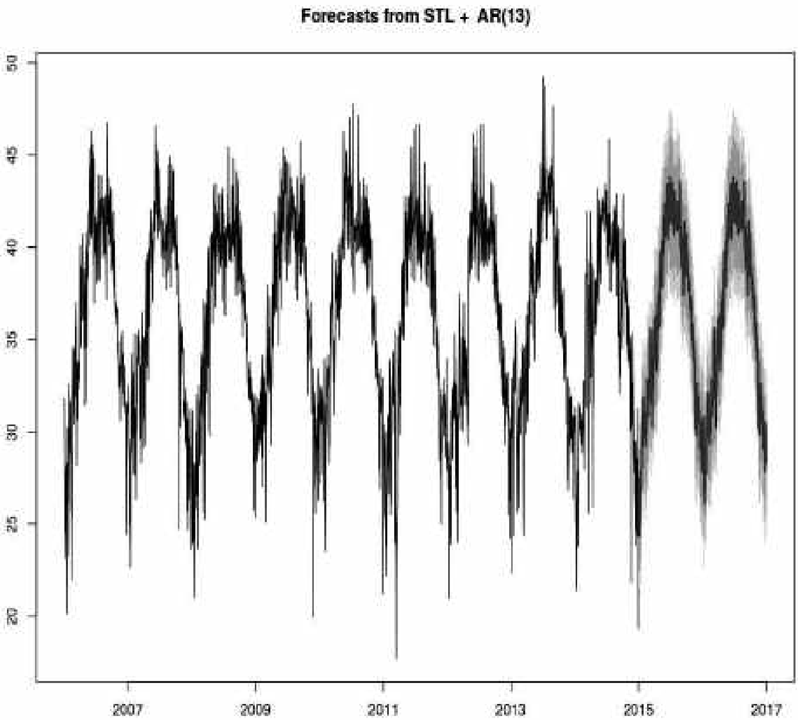

Figure 1: STL-prediction of years 2016 and 2017, with 80% and 95% forecast intervals respectively.

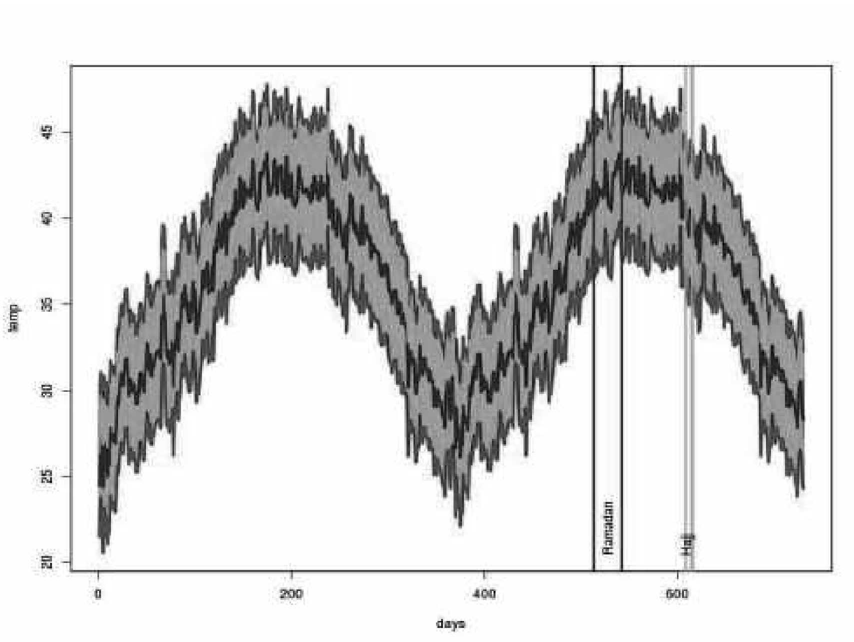

Figure 2: Ramadan and Hajj days on the prediction curve.

4.2 Prediction of year 2016

As mentioned above, we got 7 months of temperature data of year 2016, these real data was compared to the predicted and the result are shown in figure(3), MAPE error was 6.59%.

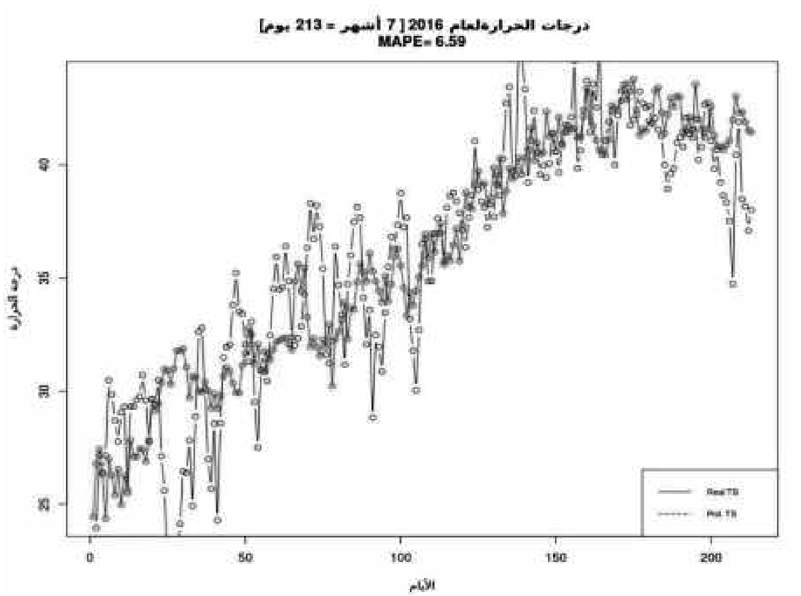

Figure 3: STL-prediction of 2016 available 7 months.

4.3 Prediction of Ramdan 1438 Temperature

From year 2017 the values of next Ramadan (27 May ~ 25 June) listed in table (1) with both 80% and 95% forecast intervals in addition to the mean. Figure (4) displays the predicted temperature values.

Forecast level / Day	2cmRamadan lower80	2cmRamadan upper80	2cmRamadan mean	2cmRamadan lower95	2cmRamadan upper95
1	38.64	43.88	41.26	37.26	45.27
2	38.74	43.98	41.36	37.36	45.37
3	37.99	43.23	40.61	36.6	44.62
4	39.51	44.75	42.13	38.12	46.13
5	38.25	43.49	40.87	36.86	44.87
6	39.17	44.41	41.79	37.79	45.8
7	38.85	44.09	41.47	37.46	45.48
8	39.04	44.28	41.66	37.66	45.67
9	38.99	44.23	41.61	37.6	45.61
10	38.58	43.82	41.2	37.19	45.2
11	38.65	43.89	41.27	37.26	45.28
12	39.55	44.79	42.17	38.16	46.17
13	40.79	46.03	43.41	39.41	47.42
14	40.58	45.83	43.2	39.2	47.21
15	39.12	44.36	41.74	37.73	45.74
16	38.48	43.72	41.1	37.09	45.11
17	38.04	43.28	40.66	36.65	44.67
18	38	43.24	40.62	36.61	44.63
19	37.82	43.06	40.44	36.43	44.44

20	38.49	43.73	41.11	37.1	45.12
21	40.04	45.28	42.66	38.65	46.66
22	39.91	45.15	42.53	38.53	46.54
23	39.82	45.06	42.44	38.44	46.45
24	40.22	45.46	42.84	38.84	46.85
25	40.7	45.94	43.32	39.32	47.33
26	40.28	45.52	42.9	38.9	46.91
27	40.66	45.9	43.28	39.27	47.28
28	41.18	46.43	43.81	39.8	47.81
29	39.61	44.85	42.23	38.23	46.24
30	38.7	43.94	41.32	37.31	45.33

Table 1: The predicted values of Ramadan 1438, with forecast intervals 80% and 95%, and mean.

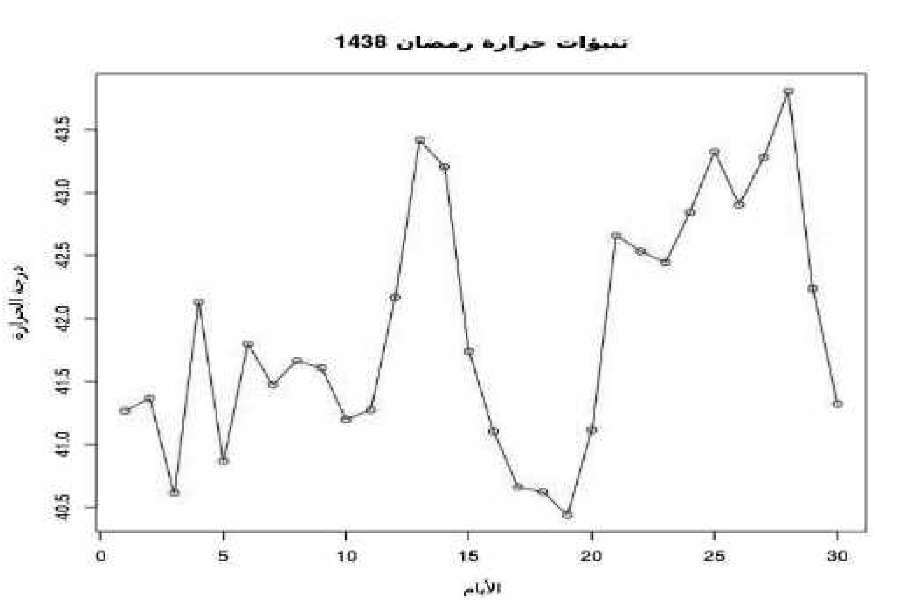

Figure 4: STL-prediction of Ramadan 1438.

4.4 Prediction of Hajj 1438 Temperature

Similarly the values of next Hajj(1 September ~ 6 September) extracted from year 2017 as listed in table(2) with both 80% and 95% forecast intervals in addition to the mean. Figure (5) displays the predicted temperature values.

Forecast level / Day(Zee Al-hegga)	2cmHajj lower80	2cmHajj upper80	2cmHajj mean	2cmHajj lower95	2cmHajj upper95
8	38.46	43.70	41.08	37.07	45.09
9	36.76	42.00	39.38	35.37	43.38
10	36.95	42.19	39.57	35.56	43.57
11	36.93	42.17	39.55	35.55	43.56
12	36.30	41.54	38.92	34.92	42.93
13	37.93	43.17	40.55	36.54	44.56
14	37.37	42.61	39.99	35.98	44.00
15	37.67	42.91	40.29	36.29	44.30

Table 2: The predicted values of Hajj 1438, with forecast intervals 80% and 95%, and mean.

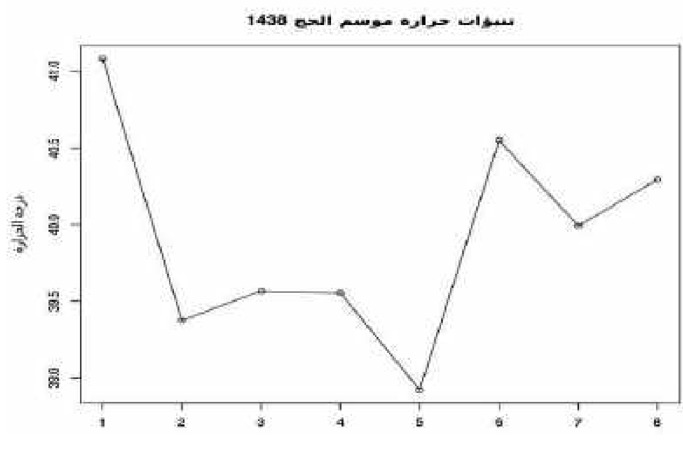

Figure 5: STL-prediction of Hajj 1438.

5 Conclusion

In this paper, we exploited STL decomposition to do predictions of 1438 Ramadan and Hajj temperatures at peak time day. The used technique achieved a good accuracy of 3.24%, therefore we used the available part of year 2016 temperature to compare it against the forecasted values, the MAPE error was 6.59%. Applying the same technique to predict next Ramadan and Hajj temperatures, the obtained results showed a consistency with the previous years.

References

[1] R. Cleveland, W. Cleveland, J. McRae, and I. Terpenning, *STL: A seasonal- trend decomposition procedure based on loess*, Journal of Official Statistics(1990), 6, 3âˆ'73.

[2] A. Baraldi, C. Enders, *An introduction to modern missing data analyses*, Journal of School Psychology 48 (2010) 5 â€" 37

[3] J. Foley, M. Costa, C. Delire, N. Ramankutty, and P. Snyder, *Green surprise? How terrestrial ecosystems could affect Earth's climate*, Frontiers in Ecology and the Environment(2003), 1, 38âˆ'44.

[4] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, Otexts 2014.

[5] B. Jiang, S. Liang, J. Wang, Z. Xiao, *Modeling MODIS LAI time series using three statistical methods*, Remote Sensing of Environment 114 (2010) P.1432â€"1444.

[6]M. Wallace, S. Anderson, and S. Mazumdar, *A stochastic multiple imputation algorithm for missing covariate data in tree-structured survival analysis*, Stat Med. 2010 Dec 20; 29(29).