Statistical Forecast of Daily Maximum Air Temperature in Arid Areas in the Summertime

Monim H. Al-Jiboori[1*], Mahmoud J. Abu Al-Shaeer[2] & Ahmed S. Hassan[1]

[1] Atmospheric Sciences Department, College of Science, Mustansiriyah University, Iraq.
[2] Al-Rafidain University College, Waziriya, Baghdad, Iraq.
[*]Correspondence: mhaljiboori@gmail.com

Abstract. Based on historical observations of daily maximum temperature, minimum air temperature and wind speed during the summertime for the period from 2004 to 2018, measured at time 0600 GMT, a non-linear regression hypothesis was developed for forecasting daily maximum air temperature (T_{max}) in arid areas with a hot climate and no rain events or cloud cover, for example around Baghdad International airport station. Observations with dust storm events were excluded, so this hypothesis could be used to predict daily T_{max} at any day during the summertime characterized by fair weather. Using the mean annual daily temperature range, the daily minimum temperature and the trend of maximum temperature with wind speed, T_{max} values were forecasted and then compared to those recorded by meteorological instruments. To improve the accuracy of the hypothesis, daily forecast errors, biases and mean absolute error were analyzed to detect their characteristics by calculating relative frequencies of occurrence. Based on this analysis, a value of -0.45 ºC was added to the hypothesis as a bias term.

Keywords: *bias; daily temperature range; maximum air temperature; mean absolute error; non-linear regression equation.*

1 Introduction

In recent years, many studies have analyzed extreme heat waves based on surface air temperature, especially daily maximum temperature (T_{max}), e.g. Gönencgil & Deniz [1] and Campbell, *et al*. [2]. These events have serious negative effects on ecosystems, human health and mortality rates [3-4], and heat-stress associated human thermal discomfort [5-6]. Furthermore, they have a significant relationship with the occurrence of wildfire on hot and sunny days [7].

Daily surface maximum air temperature is defined as the highest temperature recorded in the course of a continuous time period of 24 hours for a given location. According to the guide published by the World Meteorological

Received May 25th, 2019, Revised June 8th, 2020 Accepted for publication December 22th, 2020
 ISSN: 2337-5760, DOI: 10.5614/j.math.fund.sci.2020.52.3.8

Organization [8], classical, sensor and electronic instruments for surface temperature measurement exposed to the air must be in a place sheltered from direct solar radiation (1.2 to 2 m high above the ground). The time of the highest temperature reading normally occurs in the afternoon.

Statistical forecasting is commonly used to operationally predict weather (short) and climate (seasonal, long-time) variables. The most used statistical methods in weather forecasting include dynamical forecast information [9], which is essential as a guidance to assist weather forecasters. These methods are important to provide forecasts for scalars and locations.

Despite some statistical techniques being frequently used in research of atmospheric science (meteorology, weather) to verify the accuracy of the results, the characteristics of errors resulting from these models, e.g. biases and mean absolute error (MAE), are seldom described. For example, bias and MAE characteristics resulting from model output data of temperature forecasts were calculated for over 200 locations in the United States by Taylor & Leslie [10]. These statistical parameters could improve the forecasts and the T_{max} forecast errors exhibited less variability during summer months than they did over the rest of the year. Lin & Hubbard [11] investigated the instrumental bias in T_{max} resulting from surface temperature sensors due to sampling rates and average algorithms. When comparing the climate and standard liquid-in-glass T_{max} observations they found that the resulting biases made the diurnal temperature range more biased in papers on extreme climates.

The main objective of this study was to develop a non-linear regression equation for estimating daily T_{max} in the summertime for arid environments with clear sky conditions. This could be applied in nowcasting as well as for interpolating missing data. The accuracy of the method was evaluated by comparison with observed T_{max} data from weather stations. The daily errors and MAEs in long-term records from a given station were also analyzed. Lastly, verification of the estimation was done by testing the hypothesis using the p-value and Pearson's correlation coefficient.

2 Methodology

Multiple linear regression has been particularly used since 1950 and is a powerful tool for numerically forecasting air temperature [12]. Our hypothesis was to develop a particular regression to forecast daily T_{max} for arid areas under clear sky conditions in the summertime when there are no rain events or clouds cover and with a low frequency of dust storms.

The factors that determine air temperature magnitudes from one place to another are: amount of solar radiation reaching the surface, latitude, land and water distribution, ocean currents elevation and wind speed [1,13-14].

Globally, arid environments such as central Iraq are characterized by a hot dry climate with the highest temperatures in the summertime and no rain events or cloud cover. Clear skies cause maximum radiation reaching the earth's surface during the daytime [15]. Here, our overall hypothesis was applied to a single station located on land away from water bodies and fixed elevation. This was done to decompose the variance in situ maximum temperature records into variable derivation based on minima and maxima over long-time records. The suggested general model below was employed for daily forecasting of T_{max} in the summertime:

$$T_{max}(t) = DTR(t) + T_{min}(t) \pm T_{max}(U) \pm \varepsilon(t) \quad (1)$$

where DTR is the daily temperature range, T_{min} is the daily minimum air temperature, $T_{max}(U)$ is the trend of maximum air temperature as a function of wind speed (U), ε is the error caused by instrument and model, t is the time of day = 1, 2, …. 92 days during the summertime (June, July and August).

The first term in Eq. (1), DTR, is defined as the difference between T_{max} and T_{min} which was calculated for each day by using the following expression:

$$DTR(t) = T_{max}(t) - T_{min}(t) \quad (2)$$

The second term, T_{min}, defines the lowest air temperature in the course of a continuous time interval of 24 hours based on the lowest temperature records for a given location during a given period. This variable will be the first one to be known by the forecaster when using our model for forecasting T_{max}. The third term in Eq. (1), wind speed, is taken into account for including synoptic and mechanical stirring effects, wherein turbulent eddies on windy days are able to mix and diffuse hot surfaces with cooler air moving over them. The last term of Eq. (1) represents the error resulting from the observational data, the suggested model, and the measurement instruments. For each day, the error in each forecast is calculated using the following equation:

$$\varepsilon(t) = T_{max,f}(t) - T_{max,o}(t) \quad (3)$$

where $T_{max,f}$ is the forecasted maximum temperature and $T_{max,o}$ is the observed maximum temperature. The bias statistics are then computed as a type of forecast verification, given as

$$Bias = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i(t) \quad (4)$$

where *n* is the total number of daily forecast errors of T_{max}. Another common accuracy measure for continuous $T_{max,f}$ is mean absolute error, MAE, given as follows [9]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\varepsilon_i(t)|. \quad (5)$$

Clearly, $MAE = 0$ if the T_{max} forecasts are perfect (each $T_{max,f} = T_{max,o}$), while it increases as discrepancies between $T_{max,f}$ and $T_{max,o}$ become larger.

Confidence intervals (CI) at the 95% level are calculated around each bias and MAE value by assuming a Gaussian distribution of errors:

$$CI = \overline{\varepsilon} \pm 1.96 \frac{\sigma}{\sqrt{n}} \tag{6}$$

where $\overline{\varepsilon}$ is the mean forecast error and σ the standard deviation of the errors given as

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(T_{max_i} - \overline{T_{max}}\right)^2} \tag{7}$$

Lastly, to test our hypothesis, the p-value was calculated according to the null hypothesis.

3 Meteorological Station and Data

A meteorological station is located at Baghdad International Airport, at geographical latitude 33.14 °N and longitude 43.34 °E. The elevation of the station is at 33 m above mean sea level. It is located in a suburb about 16 km west of downtown Baghdad in Baghdad province [16], which according to the Köppen climate classification system has a subtropical desert climate (i.e. BWh), featuring extremely hot and dry summers [17].

Some of the climatological characteristics can be summarized by averaging the annual values of meteorological variables during the summer. The mean maximum air temperature is 45.2 °C, rainfall is never observed, as can be seen in Figure 1, with the exception of very small amounts of rain at the beginning of June, and mean Shamal wind speed is about 2.2 m/s. In spring and summer, dust storm events occur sometimes, with visibility frequently less than 500 m.

Observational data of daily maximum and minimum (2 m) and wind speeds recorded at dawn were collected from Baghdad station for the period 2004 to 2018. Using the codes for present (ww) and past (W1W2) weather as well as visibility (VV), the data of T_max associated with dust storms were excluded. Thus, the total number of analyzed data in the present study associated with clear conditions was 1286.

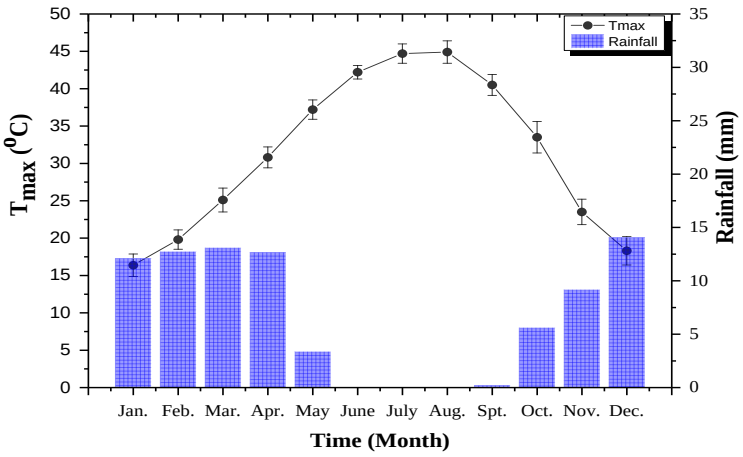

Figure 1 Monthly variations of T_{max} and rainfall for all years in the period from 2004 to 2018. The vertical lines in the T_{max} curve represent the standard deviation.

4 Results and Discussion

Before displaying our analysis of the statistical forecasting results for daily T_{max}, the stationarity of the observed daily T_{max} data used to execute this work was examined. Figure 2 displays the time series of daily T_{max}, T_{min} and DTR for the summer of 2011 as an example, in which the variability in T_{max} was roughly constant around the mean over the time period, except at the beginning of June. Furthermore, since the observational data of T_{max} were equally spaced, the series was considered to be strictly stationary [18].

Daily forecasting of T_{max} under the same conditions for the data used can confidently be considered, because meteorological data in general and temperatures in particular normally have persistence [13]. The degree of similarity between two successive time intervals for T_{max} time series was calculated for the year 2011 using an autocorrelation function. At lag = 1, the correlation coefficient was close to 0.74, indicating that the series had persistence and clustering, thus supporting our hypothesis.

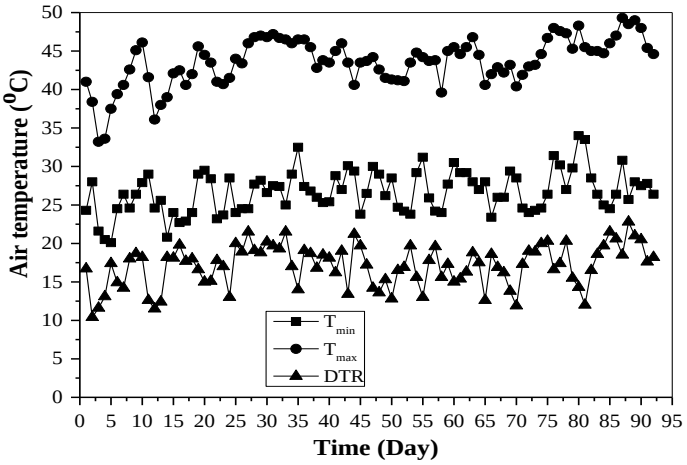

Figure 2 Time series of T_{max}, T_{min} and DTR for year 2011.

4.1 Daily Temperature Range

One of the terms used in the suggested model Eq. (1) is DTR, which was calculated with Eq. (2) for all data used in this work. Figure 3 illustrates the daily DTR values for the three months (June to August) of all summers for all years (2004-2018). There was no clear daily variation in DTR over the three months. The DTR values ranged approximately from 12 to 21 °C and had an average of 17.8 °C.

To confirm the abovementioned average value of DTR for all years, first the average of DTR for each year was calculated with the standard deviation (σ). Figure 4 presents the annual means of DTR with the standard deviation represented by vertical bars. It is interesting to see that the σ values were approximately the same for all years. Also, the annual values of DTR were all close to each other, except for the minimum value (16.3 °C) in 2012. Across all annual DTR values, the mean value was 17.8 °C, which is very similar to the 17.7 °C mentioned in the previous paragraph.

4.2 Wind Speed

In this subsection the hourly wind speed effects observed in the early morning (0600 GMT) and their role in diluting surface air temperature at noon (i.e. T_{max}) [19] are discussed. The observational data for these variables are displayed in Figure 5 with wind speed as abscissa and T_{max} as ordinate. As can be seen in the figure, most wind speed data are concentrated in the range of 2.2 to 3.1 m/s with an inverse relation, i.e. T_{max} can be seen as a function of wind speed. However, there is a decreasing trend in the T_{max} values when wind

speeds have high values. This trend has a value of -0.64 °C.s/m across all data points with a standard error of 0.045 °C. This trend value was substituted in the third term of Eq. (1).

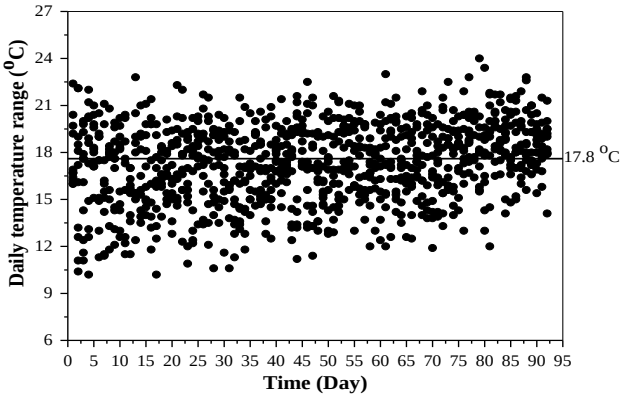

Figure 3 Distribution of all daily temperature ranges over the period studied in this paper.

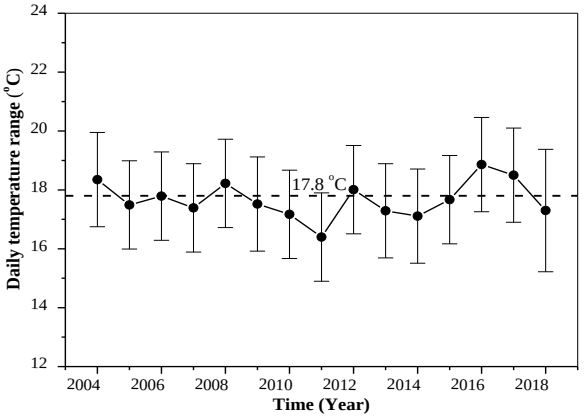

Figure 4 Annual variation of daily temperature range with the total average plotted as a dashed line.

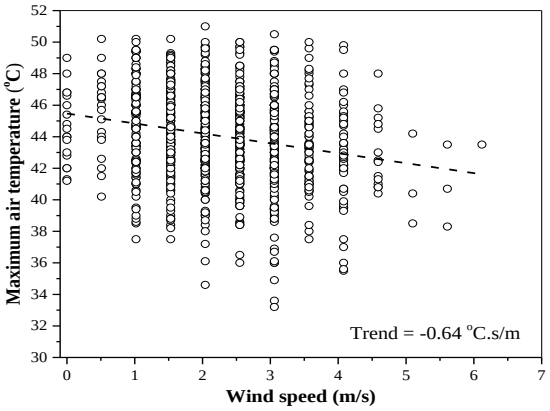

Figure 5 Relation of daily maximum air temperature with wind speed observed at 0600 GMT.

4.3 Bias and Mean Absolute Error

Using the meteorological records for daily T_{min} and the results obtained from the above subsections, i.e. DTR (17.8 °C) and T_{max}/wind speed (-0.64 °C.s/m), we could obtain the forecasting data of daily maximum air temperature ($T_{max,f}$) from Eq. (1) without the error term (ε). The results of $T_{max,f}$ were not in agreement with the actual data of maximum air temperature ($T_{max,o}$). Therefore, these errors were statistically analyzed to discover their magnitude and direction (i.e. their sign, plus or minus). In this way the outputs of Eq. (1) could hopefully be improved.

The daily errors between the T_{max} values produced from Model (1) and those measured at the Baghdad station were calculated using Eq. (2) at the same time over the entire period of this work. Then, these errors were arranged using several different statistical methods to examine the characteristics of their distribution at this station. The ε values were organized into several categories, as shown in the abscissa of Figure 6. Of special interest, the relative frequency percentage of the forecast has the highest magnitude (44%) with number 531 and an interval of ±1 °C. This can be considered error-free. Meanwhile the largest errors, i.e. ε < -5 °C and ε > 6 °C, had a very small percentage of 1% and 1.1%, respectively.

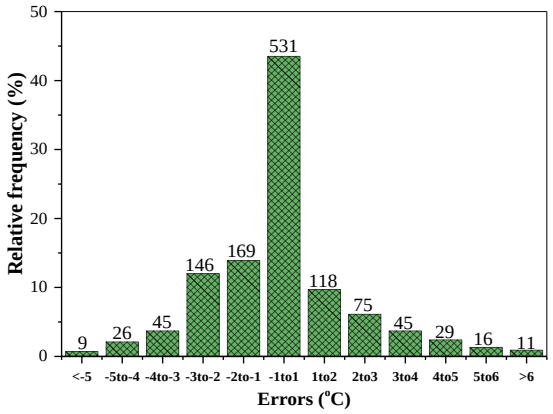

Figure 6 Percentage of relative frequencies of biases with their numbers over the bars.

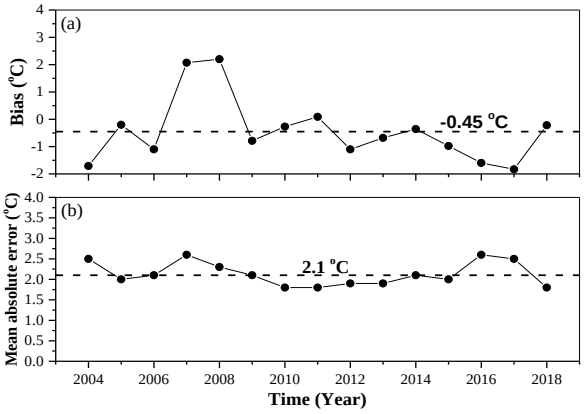

Figure 7 Annual variation of (a) bias and (b) mean absolute error.

Biases in the forecast errors for each summer season of each year were computed using Eq. (4). The time series of annual biases are presented in Figure 7(a), which illustrates that the annual biases were mostly less than zero, except in two years, 2007 and 2008. The positive large (hot) biases in these years are not surprising, because there was a change in climate, not only in the specified area but also in many other countries around the world, characterized by more periods with extreme heat. The dashed horizontal line plotted in Figure 7(a) illustrates the average annual value of the biases (-0.45 °C) over the entire

period in this work. Thus, to decrease the probable errors in $T_{max,f}$ this value was added to the model Eq. (1) as the last term.

The mean of absolute errors, MAE, tells us how large the average error is that can be expected from the forecast. The determination of this measure is in common use for examining the accuracy of daily T_{max} forecasts. It can also be concluded that MAE is the typical magnitude of the forecast errors in a given verification data set. MAE was computed for each year by use of Eq. (5). The results of MAE are presented in Figure 7(b). During the years studied, the minimum error was 1.8 °C in 2010, 2011 and 2018, while the highest error was 2.6 °C in 2016, with a mean annual value of 2.1 ± 0.29 °C, plotted as a dashed horizontal line in Figure 7(b).

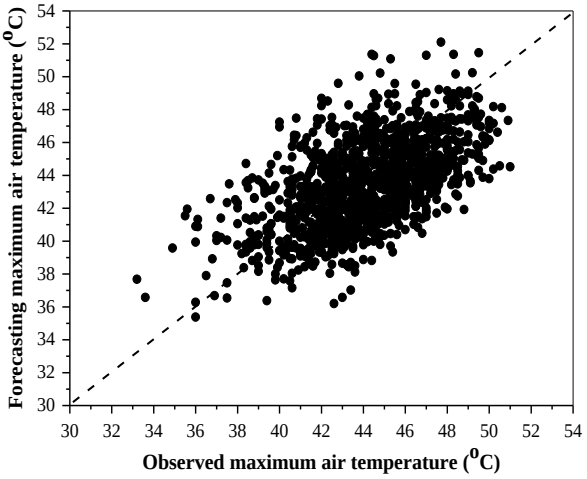

Figure 8 Forecasting daily $T_{max,f}$ versus observed $T_{max,o}$.

4.4 T_{max} Forecast Versus Observation

The relationship between the station's maximum air temperature measurements and forecasts produced by the model proposed in this paper, Eq. (1), is summarized in Figure 8. Some statistical tests such as t-student, standard error, Pearson's correlation coefficient (R) and variance are presented in Table 1.

Table 1: Statistical tests between Forecasted and observed maximum air temperature

No. of data	t-student	Standard error	R	Variance	p-value
1281	21.87	0.93	0.58	0.33	2.2×10^{-90}

Although there is scatter between daily $T_{max,f}$ and $T_{max,o}$, the linear relationship is almost clear. This scatter plot appears to be normally distributed based on bias, mean absolute and forecast error. According to the statistical measures above, the positive value of Pearson's correlation coefficient shows that $T_{max,f}$ is in the same direction as $T_{max,o}$. In addition, they are well correlated with each other.

The most important summary of the statistical analyses used in this paper is the p-value, which can be considered a test outcome for the model expressed by Eq. (1). According to the value reported in Table 1 for the p-test, the p-value was below 0.05. Therefore, the forecast model can be effectively used to predict T_{max} in the summer.

5 Conclusions

In this paper, a non-linear regression model was proposed to forecast daily maximum air temperature for summers located in arid areas, represented by a meteorological station in Baghdad, where cloud cover and precipitation events never occurred. Historical data of daily maximum, minimum air temperature measurements and wind speed at 0600 GMT for the period from 2004 to 2018 were analyzed, resulting in a number of empirical relations, such as daily temperature range (DTR) and T_{max} versus wind speed. The adapted model, Eq. (1), was used to forecast daily T_{max} on any day of summer. Bias and mean absolute error statistics were computed for each year to detect error patterns, which were used to improve our hypothesis.

DTR was studied and it was found that it ranged from 12 to 21 °C with an average of 17.7 °C. This wide range is a significant feature of arid areas and seems to be one of the reasons for the forecasting errors in this study. Wind speed measurements were found to be inversely related to T_{max} with a trend of -0.64 °C.s/m.

The daily errors in the forecasts were grouped into several categories. Most of these errors (44%) occurred with an interval of ±1 °C. Annual biases showed normal variation (around -0.45 °C) for the period studied with an extreme heat bias in 2007 and 2008. This value was added as a separate term to our model to improve the accuracy of the forecasts. Yearly calculation of MAE also produced information about the general accuracy and variability of the model. Its value was 2.1 °C over the entire period. Generally, notwithstanding the normal errors produced by the model, the forecast of the model showed a clear linear relationship with the observations (correlation coefficient = 0.6).

6 Acknowledgment

The authors would like to thank Dr. Thulfiqar Hussein Muhi, College of Arts, Mustansiriyah University for improving the English in this article by editing a draft of the manuscript. The authors would also like to thank the anonymous reviewers for their constructive comments for improvement of this paper.

References

[1] Bolger, A., *Science of Weather and Environment*, Oxford Book Company, pp. 15-49, 2010

[2] Campbell, S., Remenyi, T., White, C. & Johnston, F., *Heatwave and Health Impact Research: A Global Review*, Health and Place, pp. 210-218, 2018.

[3] Arbuthnott, K.G. & Hajat, S., *The Health Effects of Hotter Summers and Heat Waves in the Population of the United Kingdom: A Review of the Evidence*, Environmental Health, **16**(1), pp. 1-13, 2017.

[4] Scovronick, N.S., Acquaotta, F., Garzena, D., Fratianni, S., Wright, C.Y. & Gasparrini, A., *The Association Between Ambient Temperature and Mortality in South Africa: A Time-Series Analysis*, Environmental research, **161**, pp. 229-235, 2018.

[5] Yee Yong, L., Din, M.F., Pnraj, M., Noor, Z.Z., Kenzo, I. & Chelliapan, S., *Overview of Urban Heat Island (UHI) Phenomenon Towards Human Thermal Comfort*, Environment Engineer Management Journal, **16**(9), pp. 2097-2111, 2017.

[6] Ndetto, E. & Matzarakis, A., *Urban Atmospheric Environment and Human Biometeorological Studies in Dar Es Salaam, Tanzania*, Air Qual. Atmos. Health, **8**, pp. 175-191, 2015.

[7] Litschert, S., Brown, T. & Theobald, D., Effects of Climate Change and Wildfire on Soil Loss in The Southern Rockies Ecoregion, Forest Ecology and Management, **269**, pp. 124-133, 2012.

[8] WMO, *Guide to Meteorological Instruments and Methods of Observations*, World Meteorological Organization, Geneva, Switzerland, 2007.

[9] Wilks, D., *Statistical Methods in The Atmosphere Sciences*, Academic press, 2009.

[10] Taylor, A. & Leslie, L., *A Single-Station Approach to Model Output Statistics Temperature Forecast Error Assessment*, Weather and forecasting, **20**, pp. 283-294, 2005.

[11] Lin, X. & Hubbard, *What Are Daily Max and Minimum Temperature in Observed Climatology*, International Journal in Climatology, pp. 283-294, 2008.

[12] Panofsky, H. & Brier, G., *Some Applications of Statistical to Meteorology*, Pennsylvania State university, 1968.
[13] Ahren, C., *Meteorology Today: An Introduction to Weather, Climate, and the Environment*, 10 ed., Cengage Learning, 2013.
[14] Lazaridis, M., *Frist Principles of Meteorology and Air Pollution*, Springer, 2011.
[15] Jafari, M., *Reclamation of Arid Lands: Environmental Sciences and Engineering*, Springer International publishing, 2018.
[16] Sundus, H. & Al-Jiboori, M.H., *The Study of Refractive-Index Structure of Coefficient Behavior Derived from Two Weather Stations at Baghdad City 28(3)*, Al-Mustansiriyah Journal of Science, **29**(4), pp. 1-6, 2018.
[17] Geiger, R., *Überarbeitete Neuausgabe von Geiger, R. Köppen-Geiger/Klima der Erde*, Wandkarte, **1**, pp. 16, 1961. (Text in German)
[18] Y.R. & McGee, M., *Introduction to Time Series Analysis and Forecasting with Application of SAS and SPSS*, Academic press, Inc., 2000.
[19] Fujibe, F., *Relation Between Long-Term Temperature and Wind Speed Trends as Surface Observation Stations in Japan*, SOLA, **5**, pp. 81-84, 2009.
[20] Gönencgil, B. & Deniz, D., *Extreme Maximum and Minimum Air Temperature in Mediterranean Coasts in Turkey*, Environment, **1**(9), pp. 59-70, 2016.
[21] Silva, P.J., Vawdrey, E.L., Corbett, M. & Erupe, M., Silva, P.J., Vawdrey, E.L., Corbett, M. & Erupe, M., *Fine Particle Concentrations and Composition During Wintertime Inversions in Logan, Utah, USA*, Atmospheric Environment, **41**(26), pp. 5410-5422, 2007.