Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 33, Issue 4, 2022 DOI: http://doi.org/10.23851/mjs.v33i4.1168

Research Article **Open Access**

Estimation of the Daily Maximum Air Temperature for Baghdad City Using Multiple Linear Regression

Hayder M. Al-Samarrai[1]*, Monim H. Al-Jiboori[2]

[1]Iraqi Meteorological Organization and Seismology, Baghdad, IRAQ.
[2]Atmospheric Sciences Department, College of Science, Mustansiriyah University, Baghdad, IRAQ.

*Correspondent contact: hydermuslem@gmail.com

Article Info

Received
04/06/2022

Accepted
03/07/2022

Published
30/12/2022

ABSTRACT

In this paper, we relied on historical observations for the period between (2005-2020) for the Baghdad meteorological station, which is characterized by a hot, dry climate in summer and cold and rainy in winter, as it is an example of a subtropical region. The multiple linear regression equation was developed and improved to produce a formula that predicts maximum air temperature. This was done by relying on climatic elements, namely minimum air temperature, wind speed, and relative humidity, and entering them into the formula as independent inputs that have a direct impact on estimating the maximum air temperature also calculating the correlation coefficients for each of them. The bias of the model was calculated and its value was entered as the correction for errors that accompany the application of the model.

KEYWORDS: Maximum air temperature; bias; subtropical climate; Iraq; multiple linear regression.

الخلاصة

تم الاعتماد في هذه الورقة على المشاهدات التاريخية للفترة بين (**2020-2005**) لمحطة ارصاد جوي بغداد التي تمتاز بمناخ حار جاف صيفا وبارد ممطر شتاءاً فهي مثالاً للمناطق شبه الاستوائية. تم تطوير معادلة الانحدار الخطي المتعدد وتحسينها لأنتاج صيغة تتنبأ بدرجة حرارة الهواء العظمى. تم ذلك بالاعتماد على عناصر مناخية وهي درجة حرارة الهواء الصغرى وسرعة الرياح والرطوبة النسبية وأدخالها في الصيغة كمدخلات مستقلة لها تأثير مباشر على تقدير درجة حرارة الهواء العظمى كما تم حساب معاملات الارتباط الخاصة بكل منها. تم حساب التحيز الخاص بالنموذج وادخال قيمته كحد تصحيح الاخطاء التي ترافق تطبيق الموديل.

INTRODUCTION

Predicting the daily maximum air temperature is one of the main and very important topics that have attracted the attention of scientists specialized in many fields, including those working in the field of meteorology (it is considered one of the basics in the study of global warming and climate changes) [1] and the environment (the effect of cold waves and frost, forest fires, drought and water shortage, melting glaciers in addition to other problems that threaten the ecosystem) [2], geology (high temperatures that promote evaporation and transpiration, reduce ground and surface water, floods and earthquakes) [3], as well as industrial, agricultural and commercial workers (scientists, researchers or craftsmen) [4] because of their impact directly on making their own decisions, as well as those working in the field of the housing system and trying to make it more suitable and in line with rapid climate change [5], as well as scientists working in the fields of medicine and human comfort (the spread of many diseases, mortality rates and human health) [6], and in a related context, this topic has gained attention Many of the general public and for the same reason, many of them associate the rise or fall of the daily air temperature with what do it today. Accordingly, meteorological researchers have given increasing attention to finding a formula that is closer to reality for predicting the daily air temperature. This is what we will try to reach in this study.

The daily maximum air temperature is defined as the recorded value of the maximum air temperature over a continuous 24-hour period of time for a given location, according to the manual published by the World Meteorological Organization, usually recorded in the afternoon [7].

Climate and weather are a typical model of high-dimensional and complex phenomena convergence. The atmospheric system is the result of very complex interactions with different degrees of freedom or multiple patterns [8]. Therefore, trying to understand the dynamic behavior of the atmosphere has become an urgent necessity in order to gain insight into it. where, atmospheric scientists have developed methods for displaying enormous amounts of atmospheric data in time and space in a chronological technique, making use of predictive statistics methods [9].

The properties of statistical model error, such as biases and mean square errors, are rarely mentioned, despite the fact that they are widely utilized in atmospheric research to assess the quality of results (MSE), for example, in an accurate SST prediction model, statistical bias correction methods were used to correct for biases, and prediction outputs before and after correction were evaluated by Nordiati *et al.* (2021) [10], the findings mention that correcting for statistical bias can also improve prediction. Predictions for atmospheric studies can be enhanced by using these statistical variables, which process the greatest amount of errors related with predictive model operation.

Louise *et al.* (2016) [11], studied biases from seasonal precipitation forecasting results in 16 watersheds in France, improving overall forecasting skills when bias correction was applied. Two bias-correction methods show the best performance for the studied watersheds, with each method being more successful in improving certain attributes of the predictions.

Al-Jiboori *et al.* (2020) [12] worked on developing a non-linear regression equation to predict the maximum temperature of semi-arid regions on summer days using the daily temperature range with the calculation and addition of the error adjustment limit.

MATERIALS AND METHODS

Study Area and Data Collection

The study area, Baghdad Meteorological Station, is situated in Baghdad International Airport 16 km west side of Baghdad city, central Iraq. geographically situated at latitude 33.14 north and longitude 43.34 east of Greenwich line at an altitude of 33 m above sea level [13]. subtropical environments, like the middle and south of Iraq, have a dry and hot climate with usually clear skies, low relative humidity, extremely high temperatures and minimal precipitation [14]. The clear sky allows the greatest solar radiation to reach the Earth's surface during the day. The daily temperature increases to 48 ℃ in the summer, while they are moderate to cold in the winter, with a daily value of 16 degrees Celsius, and sometimes decreases to be close to zero degrees Celsius at night [15].

Method of Data Analysis

Linear regression is a statistical tool that allows researchers to model the link between a dependent variable's response and one or more independent variables with known values while also understanding the direction and degree of the association. If the independent variable is one, linear regression is simple; if the independent variables are two or more, linear regression is multiple [16].

The aforementioned multiple linear regression idea has been developed based on the elements of this study, considering wind speed WS, minimum air temperature $Tmin$ and relative humidity RH as independent variables, and considering maximum air temperature as a dependent variable, and the following formula was proposed:

$$Tmax = b_0 + b1Tmin(Tmin) + b2WS(WS) + b3RH(RH) + E \quad (1)$$

where b_0 is the multiple regression equation constant, and $b1$, $b2$, $b3$ are the slopes of the independent $Tmin$, WS and RH respectively.

E is the errors caused by the device and model, t is the time of day = 1, 2, 3, …, n days. Which can be calculated from the following formula [17]:

$$E(t) = Tmax,p - Tmax,o \quad (2)$$

where, $Tmax,p$ is the Predicted maximum air temperature, $Tmax,o$ is the observed maximum air temperature. The Bias value was calculated as a form of prediction investigation as follows [18]:

$$Bias = \frac{1}{n}\sum_{i=1}^{n} E(t) \quad (3)$$

where, n is the number of errors of the forecasted values.

The quality of models is usually checked using common accuracy measures, but many of them are not reliable for a variety of reasons that depend on the quality and indications of the values used, so the most reliable and highest accuracy measures are the

Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 33, Issue 4, 2022 DOI: http://doi.org/10.23851/mjs.v33i4.1168

mean squared deviation (MSE) and the square root of the mean square deviation (RMSE). which can be calculated as [19]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \varepsilon i^2(t) \quad (4)$$

$$RMSE=\sqrt{MSE} \quad (5)$$

A p-value was calculated based on the null hypothesis to assess the acceptance of our hypothesis, with a smaller probability value indicating more evidence to support the paper's hypothesis, ($p<0.05$) and a 95% confidence level was selected to ensure greater accuracy of the model.
The total number of data analyzed and interpreted in this paper with moderate conditions was 4388 observations.

RESULTS AND DISCUSSION

Study and Analysis of The Relationship Between Maximum Temperature and Climatic Variables

The value of the constant b_0 was calculated by entering the observed values of minimum temperature, wind speed and relative humidity as coordinates and $Tmax$ a limitation, which is the value at which the regression line cuts the y-axis, and its value (24.66) was placed in equation 1.

Minimum Air Temperature with Maximum Temperature

The first term in equation 1 refers to the minimum air temperature, which is the lowest temperature recorded at (0300 GMT) during the past 24 continuous hours. Data for $Tmax$ and $Tmin$ observations along the study period are represented in Figure 1.

we can see, a high positive correlation with a value of (0.95). It is clear that the increase in $Tmax$ is associated with an increase $Tmin$.

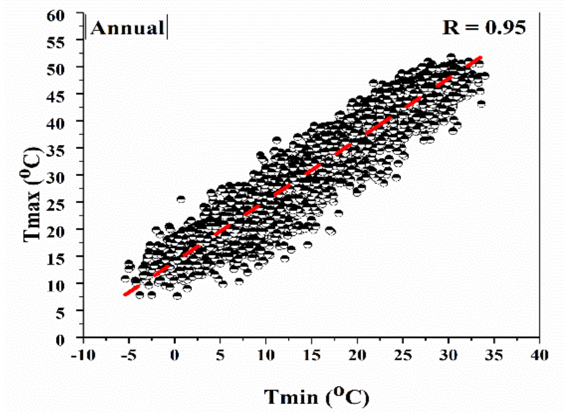

Figure 1. Relationship of Maximum Air Temperature and Minimum Air temperature.

Wind Speed with Maximum Temperature

Wind speed has the impact of strengthening or weakening the predicted recorded $Tmax$ value, here this impact will be interpreted relying on observations of wind speeds, which are shown in Figure 2.

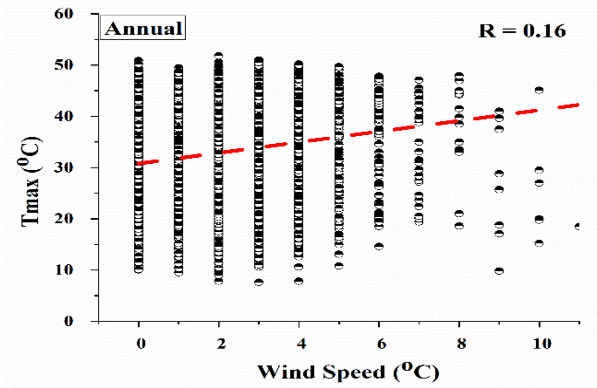

Figure 2. Relationship of maximum air temperature and wind speed.

The majority of the data is focused in the (2-4) m/sec range, the value of this correlation is (0.16) which represents no liner correlation between them.

Air Humidity with Maximum Temperature

In this subsection, we will study the impact of variation in relative humidity on the maximum air temperature expected to be recorded in the afternoon. Figure 3 shows the distribution of relative humidity data and its role in varying with the $Tmax$.

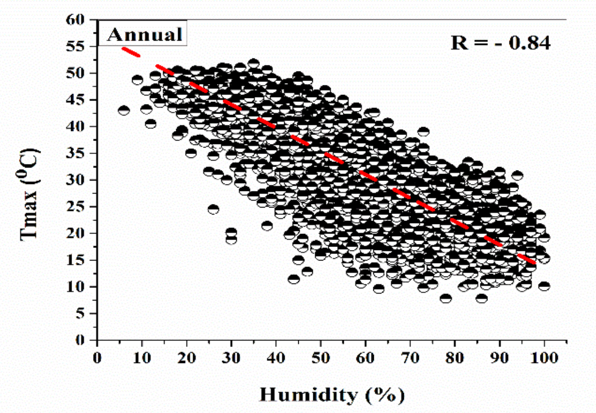

Figure 3. Relationship of maximum air temperature and humidity.

It is noticeable that relative humidity data has a wide spread between (15-95%) and that there is a downward trend for $Tmax$ with a high negative correlation (0.8), which explains the dilution in the maximum temperature when the relative humidity increases.

Multiple Linear Regression Model

In this subsection, Multiple Linear Regression for the selected elements in the model were calculated based on the statistical program Excel and SPSS, and they are referred to in the multiple linear regression equation proposed in 1 with the symbols $(b1Tmin, b2WS, b3RH)$ and it was as in the Table 1.

Table 1. Regression Coefficients values (Slope) for the weather variables selected for the model.

Regression Coefficient (Slope)	$b1Tmin$	$b2WS$	$b3RH$
Value	0.96	- 0.8	- 0.11

Root Mean Square Error (RMSE), and Bias

Now we can enter the values of the coefficients that were calculated in the previous section $(bTMIN, bWS, bRH)$ to estimate the maximum air temperature, and as an expected result, the predicted $Tmax$ results were not close to their recorded values because the error correction limit resulting from the model was not added, so it was studied and analyzed these errors are calculated by statistical methods and calculated as a value and a positive or negative sign. The prediction errors (residuals) were calculated by subtracting the recorded value from the expected value in equation 2, then these errors were arranged as categories in statistical forms to show their distribution as shown in Figure 4.

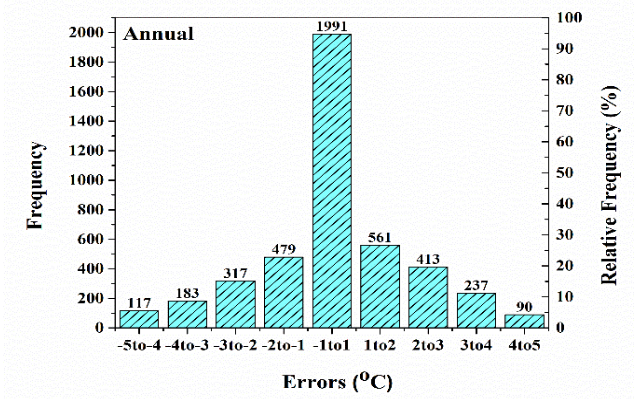

Figure 4. Number of Frequency and Relative Frequency of Errors.

The important thing that can be easily noticed when running the model is that the largest number of errors (1991) and percentage (45.4%) occurred at the interval of ±1°C, which can be considered without errors, which can be treated as without errors as it is within the acceptable range of an accurate forecast. While the largest errors at the interval of ±5°C had a small number (207) and a small percentage (4.7%). Equation 3 was run after calculating the daily errors for the length of the study period in Equation 2 to calculate the total bias resulting from the model and it was presented as a time series in Figure 5, where the amount of bias was (0.15°C), which is considered evidence of the accuracy of the model due to its small value. This value has been added in place of the last term in Equation 1.

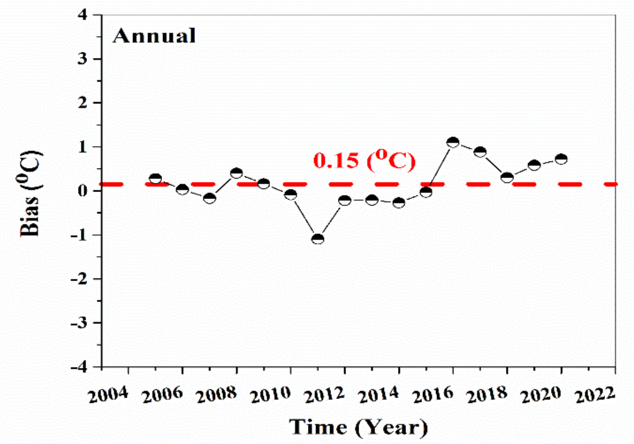

Figure 5. Annual Variance of Biases.

To evaluate prediction accuracy, the mean square error (MSE) was calculated to extract the root mean square error (RMSE) value [equation 4 and 5, respectively] to serve as the main index to consider the size of errors associated with the predictions, which amounted to (RMSE = 2.3).

ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 33, Issue 4, 2022 DOI: http://doi.org/10.23851/mjs.v33i4.1168

Finally, after completing the calculation of all the constants and parameters of the proposed model and calculating the bias, the developed equation can be written to predict the maximum air temperature for Baghdad city based on the selected weather variables as follows:

$$Tmax = 24.66 + 0.96 * Tmin - 0.8 * WS - 0.11 * RH + 0.15 \quad (6)$$

Representation of the observed ($Tmax$) data and the predicted ($Tmax$) value

The relationship between the daily maximum temperature recorded at the Baghdad meteorological station and the daily forecast coming from the model given in this study and aforementioned in Equation 6 is depicted in Figure 6. Despite the dispersion of values, we see the close association and very high positive correlation with a value of (0.98).

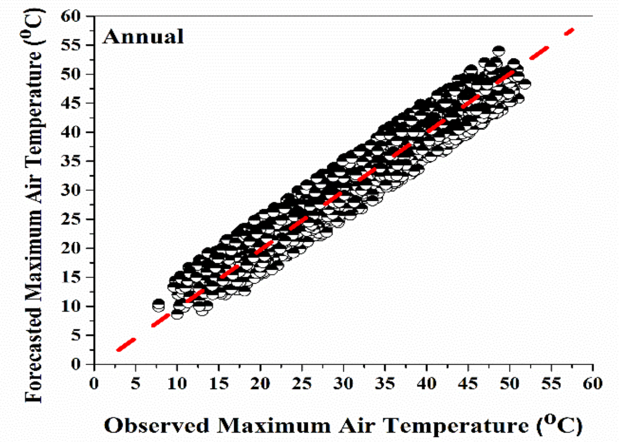

Figure 6. Observed and Expected Observations of ($Tmax$).

We will refer in Table 2. to some of the statistical values of the model to show some of its statistical properties. The value of (R) indicates a high positive correlation and perfectly matches the trend. In addition, the model explains a large part of the collected data, and other tests also reinforce the presence of indications impressive statistics. The most considered value in Table 2 is the probabilistic value (p-value), which shows that there is substantial evidence supporting the paper's hypothesis (rejecting the null hypothesis and accepting the hypothesis of the study, which includes a strong relationship between the variables selected for the model) from its small value. We can now confidently use the prediction model proposed in this study to predict $Tmax$ over the course of the year.

Table 2. Some statistical values of model.

No. of data	R	R^2	Standard error	Variance	P-value
4388	0.98	0.95	2.3	1.000	$6 * 10^{-34}$

CONCLUSIONS

In this paper, we can deduce the following correlations after building a multivariate linear regression equation and conducting analyses and testing on meteorological variables and investigating their influence on daily forecasting of maximum air temperature:

1. The relationship between $Tmax$ and $Tmin$ is characterized as having a high positive correlation.
2. The nature of the correlation between $Tmax$ and wind speed was highly negative.
3. Similar to wind speed, relative humidity had a low negative correlation.
4. For all study periods, daily prediction errors were calculated and then categorized. According to the analysis and investigation, these errors occurred at a rate of approximately 50% in the interval of ±1°C, which is classified as without errors.
5. A comparison between the recorded and expected daily high-temperature observations showed a very high positive correlation between the values, showing the relationship between them.

ACKNOWLEDGMENT

The authors thank the Iraqi Meteorological Organization and Seismology for initialization the climatic data necessary to complete this work. The authors would also thank the editors and reviewers for their important comments and suggestions for improving this paper.

Disclosure and conflict of interest: The authors declare that they have no conflicts of interest.

REFERENCES

[1] M. Kanda, "Progress in urban meteorology: A review," J. Meteorol. Soc. Japan. Ser. II, vol. 85, pp. 363-383, 2007. https://doi.org/10.2151/jmsj.85B.363

[2] W. Li, Q. Cao, K. Lang, and J. Wu, "Linking potential

heat source and sink to urban heat island: Heterogeneous effects of landscape pattern on land surface temperature," Sci. Total Environ., vol. 586, pp. 457-465, 2017. https://doi.org/10.1016/j.scitotenv.2017.01.191

[3] T. Ahmed, M. Scholz, F. Al-Faraj, and W. Niaz, "Water-related impacts of climate change on agriculture and subsequently on public health: A review for generalists with particular reference to Pakistan," Int. J. Environ. Res. Public Health, vol. 13, no. 11, p. 1051, 2016. https://doi.org/10.3390/ijerph13111051

[4] W. Schlenker and M. J. Roberts, "Nonlinear temperature effects indicate severe damages to US crop yields under climate change," Proc. Natl. Acad. Sci., vol. 106, no. 37, pp. 15594-15598, 2009. https://doi.org/10.1073/pnas.0906865106

[5] D. Cho, C. Yoo, J. Im, Y. Lee, and J. Lee, "Improvement of spatial interpolation accuracy of daily maximum air temperature in urban areas using a stacking ensemble technique," GIScience Remote Sens., vol. 57, no. 5, pp. 633-649, 2020. https://doi.org/10.1080/15481603.2020.1766768

[6] R. Bansal, M. Ochoa, and D. Kiku, "Climate change and growth risks," National Bureau of Economic Research, 2017. https://doi.org/10.3386/w23009

[7] G. S. Tudorache, "SELECTIVE PARAMETRIC SPECIFICITY OF CLUJ-NAPOCA METEOROLOGICAL STATION ACCORDING TO ANNEX 1B OF WMO NO. 8 GUIDE," Aerul si Apa. Compon. ale Mediu., pp. 43-50, 2018. https://doi.org/10.24193/AWC2018_06

[8] Y. Zhang, I. Held, and S. Fueglistaler, "Projections of tropical heat stress constrained by atmospheric dynamics," Nat. Geosci., vol. 14, no. 3, pp. 133-137, 2021. https://doi.org/10.1038/s41561-021-00695-3

[9] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson, "Empirical orthogonal functions and related techniques in atmospheric science: A review," Int. J. Climatol. A J. R. Meteorol. Soc., vol. 27, no. 9, pp. 1119-1152, 2007. https://doi.org/10.1002/joc.1499

[10] S. Nurdiati, E. Khatizah, M. K. Najib, and L. L. Fatmawati, "El nino index prediction model using quantile mapping approach on sea surface temperature data," Desimal J. Mat., vol. 4, no. 1, 2021. https://doi.org/10.24042/djm.v4i1.7595

[11] L. Crochemore, M.-H. Ramos, and F. Pappenberger, "Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts," Hydrol. Earth Syst. Sci., vol. 20, no. 9, pp. 3601-3618, 2016. https://doi.org/10.5194/hess-20-3601-2016

[12] M. H. Al-Jiboori, M. J. Abu Al-Shaeer, and A. S. Hassan, "Statistical forecast of daily maximum air temperature in arid areas in the summertime," J. Math. Fundam. Sci., vol. 52, no. 3, 2020. https://doi.org/10.5614/j.math.fund.sci.2020.52.3.8

[13] V. Sissakian, N. Al-Ansari, and S. Knutsson, "Origin of some transversal linear features of NE-SW trend in Iraq, and their geological characters," Nat. Sci., vol. 6, no. 12, pp. 996-1011, 2014. https://doi.org/10.4236/ns.2014.612091

[14] A. A. Al Khudhairy, "Spatio-temporal analysis of Maximum temperature over Iraq," Al-Mustansiriyah J. Sci., vol. 29, no. 1, 2018. https://doi.org/10.23851/mjs.v29i1.257

[15] R. M. Shubbar, H. H. Salman, and D. Lee, "Characteristics of climate variation indices in Iraq using a statistical factor analysis," Int. J. Climatol., vol. 37, no. 2, pp. 918-927, 2017. https://doi.org/10.1002/joc.4749

[16] H. A. Pahlavan, B. Zahraie, M. Nasseri, and A. Mahdipour Varnousfaderani, "Improvement of multiple linear regression method for statistical downscaling of monthly precipitation," Int. J. Environ. Sci. Technol., vol. 15, no. 9, pp. 1897-1912, 2018. https://doi.org/10.1007/s13762-017-1511-z

[17] R. Nuzzo, "Statistical errors," Nature, vol. 506, no. 7487, p. 150, 2014. https://doi.org/10.1038/506150a

[18] C. Piani, J. O. Haerter, and E. Coppola, "Statistical bias correction for daily precipitation in regional climate models over Europe," Theor. Appl. Climatol., vol. 99, no. 1, pp. 187-192, 2010. https://doi.org/10.1007/s00704-009-0134-9

[19] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Clim. Res., vol. 30, no. 1, pp. 79-82, 2005. https://doi.org/10.3354/cr030079

How to Cite

H. M. Al-Samarrai and M. H. . Al-Jiboori, “Estimation of the Daily Maximum Air Temperature for Baghdad City Using Multiple Linear Regression”, *Al-Mustansiriyah Journal of Science*, vol. 33, no. 4, pp. 9–14, Dec. 2022.