Al-Samarrai and Al-Jiboori *Iraqi Journal of Science, 2023, Vol. 64, No. 4, pp: 2085-2094*
DOI: 10.24996/ijs.2023.64.4.42

ISSN: 0067-2904

Prediction of Daily Maximum Air Temperature for Transitional Seasons by Statistical Methods in Baghdad

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*

Received: 13/4/2022 Accepted: 14/8/2022 Published: 30/4/2023

Abstract
Predicting the maximum temperature is of great importance because it is related to various aspects of life, starting from people's lives and their comfort, passing through the medical, industrial, agricultural and commercial fields, as well as concerning global warming and what can result from it. Thus, the historical observations of maximum and minimum air temperature, wind speed and relative humidity were analyzed in this work. In Baghdad, the climatic variables were recorded on clear sky days dawn at 0300 GMT for the period between (2005-2020). Using weather station's variables multiple linear regression equation, their correlation coefficients were calculated to predict the daily maximum air temperature for any day during the transitional seasons (autumn, spring). By analyzing the results, a comparison was made between the expected and recorded maximum air temperature to improve the equation. The bias was tracked by analyzing the number of relative frequencies of the occurrence of these errors. (0.2 ℃) for the autumn season and (0.15 ℃) for the spring season was added to the multiple linear regression equation as a correction value.

Keywords: maximum air temperature forecast; climate change; correlation coefficient; multiple linear regression formula; bias.

التنبؤ بدرجة حرارة الهواء العظمى اليومية للمواسم الانتقالية بالطرق الإحصائية في بغداد

حيدر مسلم حسين السامرائي[1*] , منعم حكيم خلف[2]
[1] الهيأة العامة للأنواء الجوية والرصد الزلزالي العراقي.
[2] قسم علوم الجو, كلية العلوم, الجامعة المستنصرية, العراق.

الخلاصة
التنبؤ بدرجة الحرارة العظمى له أهمية كبيرة لأنه يتعلق بمختلف جوانب الحياة ، بدءاً من حياة الناس وراحتهم ، مروراً بالمجالات الطبية والصناعية والزراعية والتجارية ، فضلاً عن الاحتباس الحراري وما يمكن أن ينتج عنه. لذالك ، تم تحليل الملاحظات التاريخية لدرجة حرارة الهواء العظمى والصغرى وسرعة الرياح والرطوبة النسبية في هذا العمل. في بغداد ، تم تسجيل المتغيرات المناخية في فجر أيام السماء الصافية الساعة 0300 بتوقيت جرينتش للفترة ما بين (2005–2020). باستخدام معادلة الانحدار الخطي المتعددة لمتغيرات محطة الطقس ، تم حساب معاملات الارتباط الخاصة بهم للتنبؤ بدرجة حرارة الهواء العظمى اليومية لأي يوم خلال الفصول

*Email: hydermuslem@gmail.com

الانتقالية (الخريف ، الربيع). من خلال تحليل النتائج تم عمل مقارنة بين درجة حرارة الهواء العظمى المتوقعة والمسجلة لتحسين المعادلة. تم تتبع التحيز من خلال تحليل عدد التكرارات النسبية لحدوث هذه الأخطاء. تمت إضافة (0.2 ℃) لموسم الخريف و (0.15 ℃) لموسم الربيع إلى معادلة الانحدار الخطي المتعدد كقيمة تصحيح.

1 Introduction

Recently, many researchers have paid great attention to reaching more accurate models that meet the world's need to predict the maximum air temperature. These models have a direct impact in making appropriate decisions in various areas of life, as it became clear that global warming and climate fluctuations cause many problems for the system, so that they are considered one of the most important environmental problems in the world that threaten human survival on Earth [1]. These climatic fluctuations have directly affected people and the nature of their housing conditions during the succession of different periods of time. It forced them to change their housing system to suit these fluctuations[2]. The temperature change is one of the most important climatic variables affecting the development, growth, and productivity of agricultural crops [3]. The rise in temperatures on dry days and seasons increases the demand for water[4]. Not to forget the serious negative impact of seasonal heat waves, cold waves and frosts on ecosystems, mortality rates and human health [5].

Azad et al. (2020) investigated the possibility of reaching a developed method of adaptive neural fuzzy inference system (ANFIS) for estimating temperatures for 34 Iranian weather stations using this system with genetic algorithm (GA), differential evolution (DE) and others, with three parameters selected as variable inputs. The most accurate model was produced using ANFIS with GA to predict the maximum air temperature [6].

Astsatryan et al. (2021) used meteorological data for different ground stations along with ready-made satellite data with different accuracy and frequencies and used multiple neural networks to predict temperatures over the Ararat Valley of Armenia (the driest region of Armenia). This study implemented the machine learning mechanism to improve weather forecasting technology and to predict the air temperature for the next (3 - 24) hours with an accuracy of 87.31% [7].

Lim et al. (2022) calculated the air surface temperature(AST) for several selected stations in Iraq based on historical data from the air infrared probe for a period of 14 years (2003-2016). The effects of some atmospheric elements (O3, CH4, CO, H2O vapor, and outgoing longwave radiation)on AST were analyzed. The study concluded that there was a high correlation, with a rate of ($R = 0.9$), as well as a close match between the expected and observed values of surface air temperature with a small discrepancy, which indicates that the quality of the data entered in the regression model used and the accuracy of the model itself [8].

Forecasters in different atmospheric conditions do not dispense with statistical methods in preparing weather and climate reports and communicating them to decision makers. Statistical forecasting using statistical equations is one of the most common methods in weather forecasting based on dynamic forecasting facts, whether with regard to short-term, long-term or seasonal weather variables represented by various climatic factors [9].

Some statistical models are devoid of the limits of error correction and bias resulting from either the data used, the method of entering it, or other reasons, which leads to a decrease in

the accuracy of the produced models. The outputs of the models can be improved by calculating and adding their bias values [10].
Al-Jubouri et al. achieved a remarkable development in deriving a non-linear regression equation to predict the daily maximum air temperature for semi-arid environments and clear summer days using the daily temperature range (DTR) with the calculation of the error adjustment limit [11].

As a result of the variation in the values of some geographical and climatic parameters, including amount of incoming solar radiation, location relative to latitude, wind speed, distribution of water and land, and ocean currents, the value of the air temperature changes from one region to another.[12].

The main objective of this study is to derive a predictive technique for the daily maximum air temperature in the transitional seasons by developing a multiple linear regression equation. This can be applied in telecasting as well as for estimating missing data.

2 Methodology

Our hypothesis puts an advanced concept to produce a predictive regression formula for the maximum daily air temperature in semi-arid regions under temperate sky conditions. This was done for two different transitional seasons. This work was based on historical observations for 16 years. Data of weather elements related to air temperature and their seasonal distribution was collected and plotted by the Origin program. The prediction errors (residuals) were calculated by SPSS program to modify the predictions. Finally, a developed equation was reached that predicts the maximum air temperature for the two transitional seasons (spring and autumn).

3 model formulation

Linear regression is one of the statistical methods that enable analysts to model the relationship of the response of a dependent variable to be predicted to independent variables (one or more) with known values and knowing the direction and degree of this relationship. The type of regression depends on the number of independent variables. Linear regression is considered simple if the independent variable is one component only, and the regression is multiple if the independent variables are two or more [13]. The relationship is written in the general mathematical formula:

$$\hat{y} = a + b1X1 + b2X2 + b3X3 + \cdots + biXi + ei \tag{1}$$

Where: *(a)* is an equation constant, *b1,b2,b3, ... ,* bi is the slope of straight lines, *X1,X2,X3, ... ,* Xi are the independent variables, and ei is the estimation error.
This study assumes a multiple linear regression relationship between independent variables which are minimum air temperature, wind speed and relative humidity, and a dependent variable, which is the maximum air temperature. This relation is represented by the following formula:

$$Tmax = a + b1Tmin + b2U + b3RH + e \tag{2}$$

Where: (a) is a general constant, $(b1, b2, b3)$ are the model regression coefficients that represent the change in the predicted maximum air temperature $(Tmax)$ due to the effect of change in minimum air temperature $(Tmin)$, wind speed (U), and relative humidity (RH), respectively, and e is the model error.
Daily forecast errors are calculated by the following formula:

$$e = Tmax(o) - Tmax(p) \quad (3)$$

Where: $Tmax(o)$ is the observed air temperature , $Tmax(p)$ is the predicted air temperature.

By calculating the bias statistics and adding its value to the proposed formula, a more accurate model can be reached using the equation:

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

Where: n is the number of daily errors of the predicted value.

Common accuracy measures the quality of prediction by comparing the real values with the predicted values. Some of these measures are not reliable either because they do not represent large errors in an important way, or because they give small values in the case of equality of negative and positive values. More accurate measures were chosen such as the mean square errors (MSE) and the square root mean square errors (RMSE) [14], as shown by the following equations:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \varepsilon\text{i}^2 \text{ (t)} \quad (5)$$

$$RMSE = \sqrt{MSE} \quad (6)$$

4 Materials and Methods

The study area was the city of Baghdad in the middle of Iraq. Data were obtained from the Baghdad Meteorological Station, located at Baghdad International Airport. Geographically located at latitude 29º - 37º north and longitude 38º - 48º east at an altitude of 33 m above sea level [15].

By describing some climatic features of the study station, the average annual values of some weather variables during the transitional seasons is identified. In autumn, the average $Tmax$ was 34 ℃, the average wind speed was 2 m/s, and the relative humidity was 55%. In the spring, the average $Tmax$ was 31.7 ℃, the average wind speed was about 2 m / s, and the relative humidity was 57%. The total number of observations that were studied and analyzed in this study for the fall season was 1156 and for spring 970 observations.

5 Results and discussion

Initially, the format of the observed daily data used to carry out this work was tested. Figure 1 shows the time series for $Tmean$, $Tmax$, and $Tmin$ for autumn and spring from 2016 to 2020. The variance in $Tmax$ was mostly constant about the mean, over the stated time period. Furthermore, the series was considered stable, with the spacing of the $Tmax$ data being equal without anomalies.

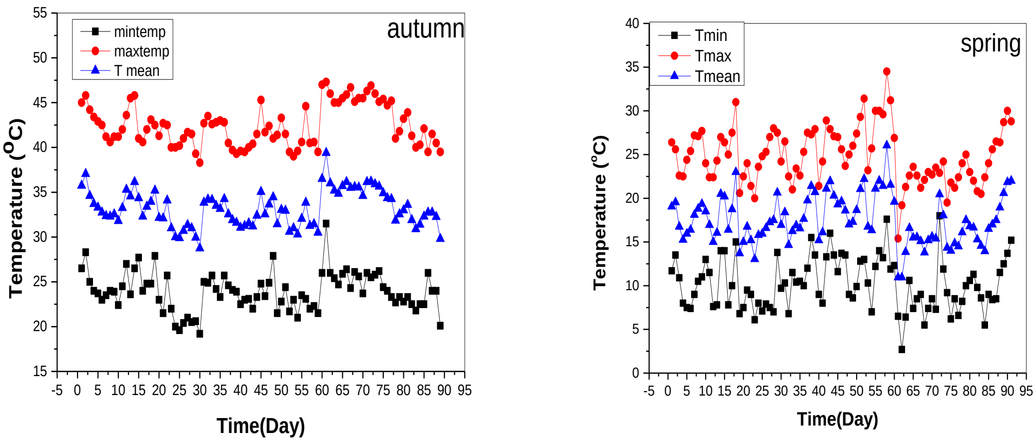

Figure 1: Time series of $Tmax$, $Tmin$ and $Tmean$ for the 2016 to 2020.

The constant (a) was calculated and substituted in Equation 2, which represents the point on the y-axis where the regression line crosses it, and its value was (27.55) for the autumn model, and (23.05) for the spring model.

Minimum air temperature ($Tmin$)

The first variable in Equation 2 refers to the minimum air temperature, which is the first temperature recorded by the observer at dawn. Observations data for $Tmax$ and $Tmin$ for the autumn and spring seasons are plotted in Figure 2. It is clear that the expected $Tmax$ increase is due to the increase in $Tmin$ recorded as a result of a high positive correlation with a value of 0.9 and 0.87 for the fall and spring seasons, respectively. These values were substituted into Equation 2 for ($b1$).

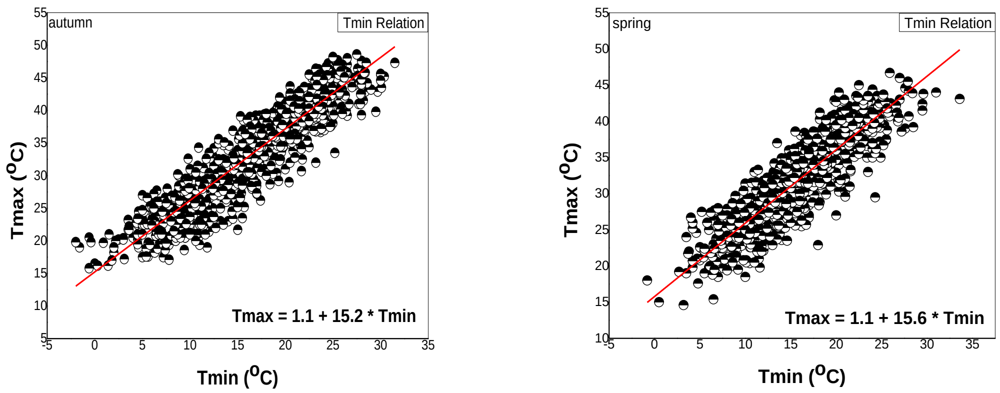

Figure 2: Relationship of $Tmax$ with $Tmin$.

Wind speed

In this subsection, the effect of wind speed on enhancing or mitigating the expected maximum temperature is discussed. The observations recorded at the station were drawn, as shown in Figure 3. For the autumn season, a high negative correlation of (-0.81) with most of the data between (1-3) m/s were noted, while it was a negative correlation with a value of (-0.84) and data concentration in (2-4) m/s for the spring season. The decrease in the expected air temperature can be explained to be a result of an increase in wind speed through the inverse relationships of the previous correlation coefficients. The values of these coefficients were substituted in Equation 2 for ($b2$).

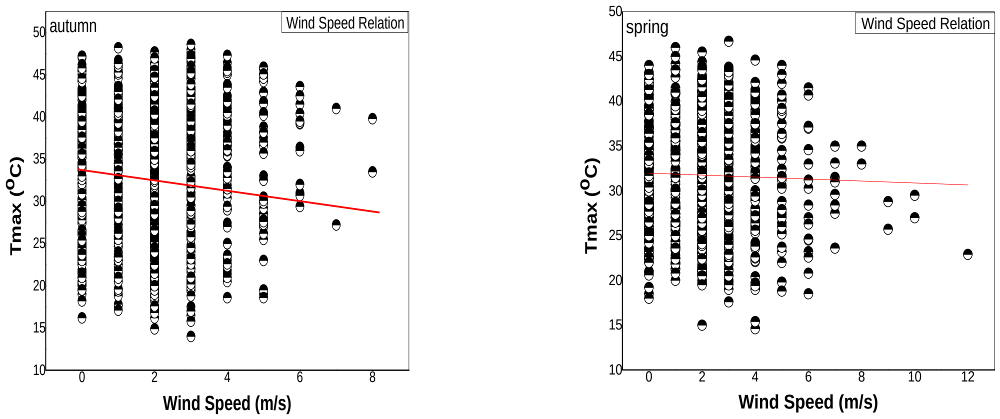

Figure 3: Relationship of $Tmax$ with wind speed.

Humidity

The change in relative humidity recorded at dawn and its effect on the expected air temperature is shown in Figure 4. The relative humidity observations are widely distributed in both seasons. There was a downward trend for $Tmax$ with a correlation coefficient (-0.13) in the autumn season, as for the spring season the correlation coefficient (-0.07). The correlation coefficients were substituted in Equation 2 for ($b3$).

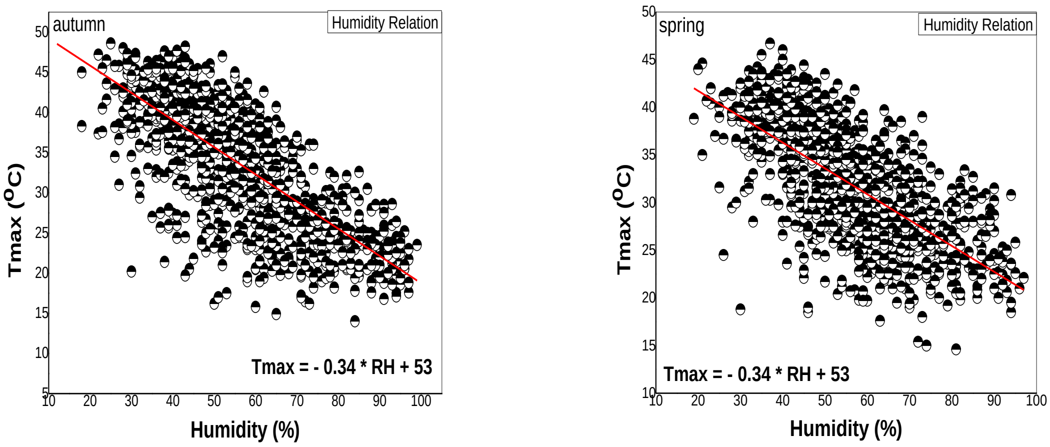

Figure 4: Relationship of $Tmax$ with Humidity.

Statistical analysis of Errors and Bias calculations.

Now the maximum air temperature can be predicted by entering the instantaneous values of minimum air temperature, wind speed and relative humidity as well as their correlation coefficients in the formula indicated by equation (2) without taking into account the model error term.

Because of differences, called residuals, between the recorded value of $Tmax$ and its expected values, the results were not close to reality. Therefore, those errors were analyzed in order to determine their values and their sign, positive or negative, and using Equation 3, the daily error of prediction was calculated, and then arranged as shown in Figure 5 as categories to show their distribution in statistical forms.

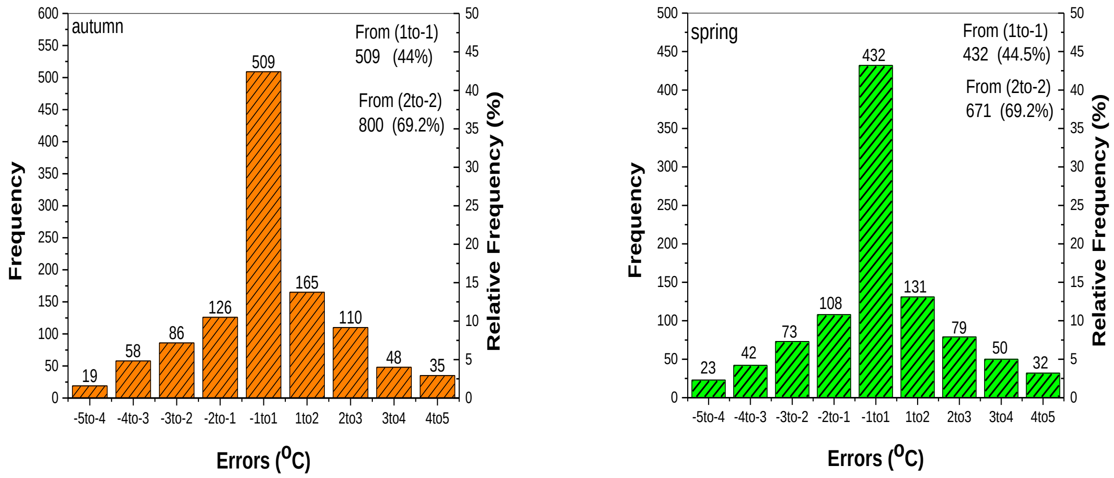

Figure 5: Frequency distribution and numbers of Errors.

The important result of this model is that the largest number of percentage errors were (44%,44.5%) for the autumn and spring seasons, respectively, at an interval of ± 1°C. These predictions can be considered correct. Meanwhile, the largest errors had a small percentage (1.6%, 3%) for the autumn season, and (2.4%, 3.3%)% for the spring season, i.e. at ± 5°C respectively.

The results of the calculated errors in Equation (3) were entered into Equation (4) to find the $Tmax$ biases for the two transitional seasons, which were of the values of (0.2, 0.15) °C for the autumn and spring seasons, respectively, which were indicative of the model's accuracy due to its small value. It represents the value of (e) in Equation (2). The time series of the annual variance of bias is shown in Figure 6.

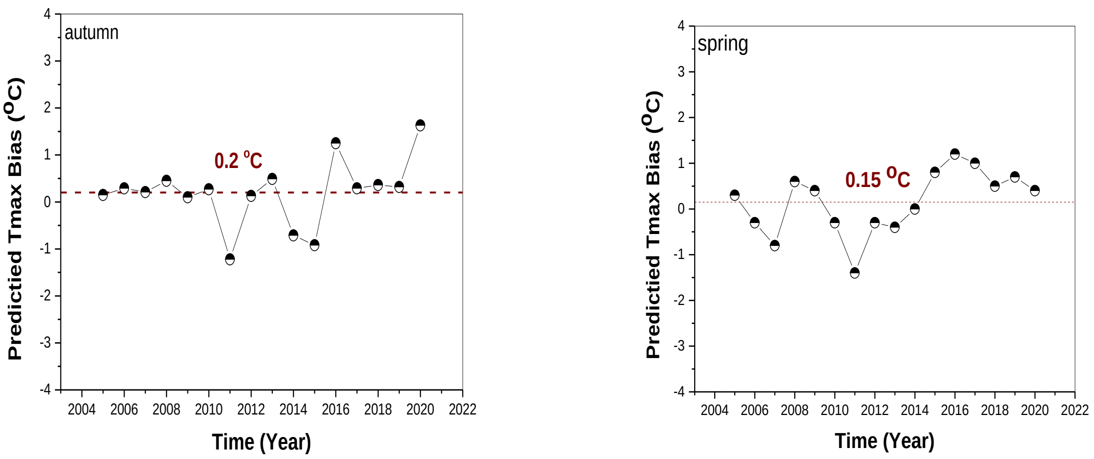

Figure 6: Variations of annual Biases.

The MSE value was calculated from Equation (5) to find the RMSE values of Equation (6) to evaluate the forecast quality, which indicates the amount of predictive errors, which was (2.3) °C for both seasons.

6 Representation of recorded and predicted Tmax values

For both the autumn and spring seasons, Figure 7 displays the relation between the daily maximum temperature recorded in Baghdad station and the daily forecast obtained from the model provided in this study as pointed out in Equation 2. Despite the scattering of the data, close convergence and high positive correlation can be seen.

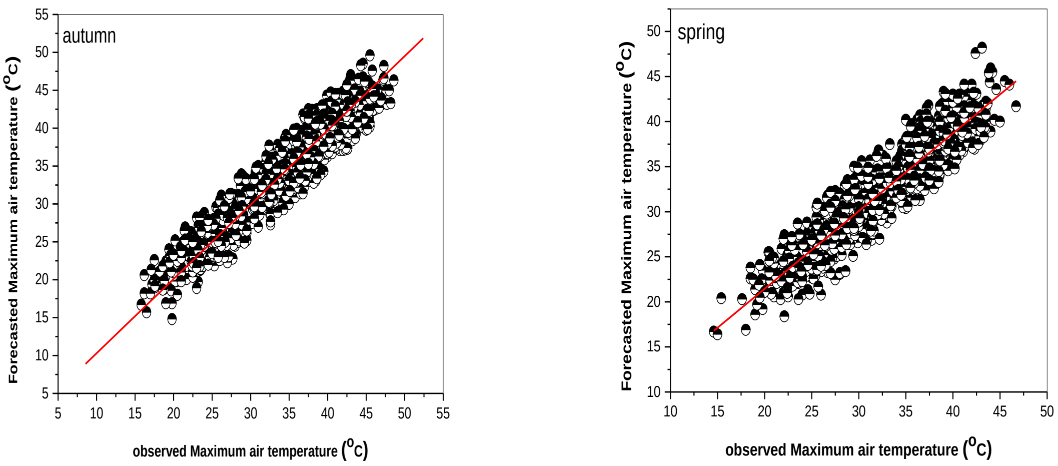

Figure 7: Simulation of the observed and predicted maximum air temperature.

Important statistical measures are presented in Table (1), which indicate the presence of a high positive correlation and perfect fit in the propagation direction. Also, the model explains a good percentage of the recorded data represented by a value of R2, in addition to other tests that confirm the existence of significant statistical significance, where the value of the Variance indicates the absence of a multicollinearity problem, and also the value of the t-test and the significance associated with it. The most important value in the table is the p-value, its small value (0.00) indicates the presence of strong evidence that supports the hypothesis on which the study was based. As a final result of what was previously mentioned, it is possible to work with the model mentioned in Equation (2) to predict the maximum air temperature for the transitional seasons of the study area with complete confidence.

Table 1: Some statistical measures of model

season	No. of data	R	R^2	Standard Error	Variance	t-test Seg.	P-value
autumn	1156	0.95	0.91	2.19	1.000	107.8 .000	0.00
spring	970	0.93	0.86	2.36	1.000	78.5 .000	0.00

7 conclusions

Daily atmospheric data from the archive of the Iraqi Meteorological Organization and Seismology for a period of 16 years (2005-2020) were used to develop and improve a multiple linear regression equation to be suitable for predicting the daily maximum air temperature for the transitional seasons (autumn and spring). Atmospheric variables such as maximum and minimum temperature, wind speed and air humidity were analyzed. The tests and the representation of these elements showed significant positive results represented in the presence of a direct impact of these variables on the prediction of daily $Tmax$. The results of the analysis and tests are summarized as follows:

1. The minimum temperature had a high positive correlation with a value of (0.9) for both transitional seasons summarizing the positive relationship between them. wind speed had an inverse relationship represented by a high negative correlation with a value of (0.8,0.8) for the autumn and spring seasons, respectively, that worked to mitigate the high $Tmax$. While, humidity had an inverse relationship with a negative correlation coefficient of (0.13,0.07) for the autumn and spring seasons,respectively.
2. Analysis of the daily forecast errors calculated for the length of the study period and grouped into categories showed that at ± 1℃ with a percentage of (44% , 44.5%) for the autumn and spring seasons, respectively, where this percentage is considered without errors.
3. To correct the model's work, the annual bias value was calculated and added with a value of (0.2,0.15) ℃ instead of the last term in Equation (2), which showed a natural variation for the length of the study period.
4. The comparison between the predicted and recorded $Tmax$ values showed a high positive correlation (R = 0.95, 0.93) for the autumn and spring seasons, respectively.
5. As proof of the model's accuracy and quality assessment, an RMSE value of (2.3) ℃ was calculated for both seasons.

In general, the results indicated the accuracy of verifying the influence of atmospheric variables on $Tmax$ over the study area, as well as the advantage of using those parameters and the ability of the proposed model to predict $Tmax$ efficiently.

8 References

[1] J. Cifuentes, G. Marulanda, A. Bello, and J. Reneses, “Air temperature forecasting using machine learning techniques: a review,” *Energies*, vol. 13, no. 16, p. 4215, 2020.

[2] 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.

[3] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, “Predicting solar generation from weather forecasts using machine learning,” in *2011 IEEE international conference on smart grid communications (SmartGridComm)*, 2011, pp. 528–533.

[4] M. Afzali, A. Afzali, and G. Zahedi, “Ambient air temperature forecasting using artificial neural network approach,” in *International Conference on Environmental and Computer Science IPCBEE*, 2011, vol. 19, pp. 176–180.

[5] D. Oudin Åström, F. Bertil, and R. Joacim, “Heat wave impact on morbidity and mortality in the elderly population: A review of recent studies,” *Maturitas*, vol. 69, no. 2. 2011, doi: 10.1016/j.maturitas.2011.03.008.

[6] A. Azad *et al.*, “Novel approaches for air temperature prediction: a comparison of four hybrid evolutionary fuzzy models,” *Meteorol. Appl.*, vol. 27, no. 1, p. e1817, 2020.

[7] H. Astsatryan *et al.*, “Air temperature forecasting using artificial neural network for Ararat valley,” *Earth Sci. Informatics*, vol. 14, no. 2, pp. 711–722, 2021.

[8] H. S. Lim, J. Rajab, A. Al-Salihi, Z. Salih, and M. MatJafri, “A statistical model to predict and analyze air surface temperature based on remotely sensed observations,” *Environ. Sci. Pollut. Res.*, pp. 1–11, 2021.

[9] T. T. K. Tran, S. M. Bateni, S. J. Ki, and H. Vosoughifar, “A review of neural networks for air temperature forecasting,” *Water*, vol. 13, no. 9, p. 1294, 2021.

[10] M. Newman, “An empirical benchmark for decadal forecasts of global surface temperature anomalies,” *J. Clim.*, vol. 26, no. 14, pp. 5260–5269, 2013.

[11] 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, doi: 10.5614/j.math.fund.sci.2020.52.3.8.

[12] K. Hashim, H. Al-Bugharbee, S. L. Zubaidi, N. S. S. Al-Bdairi, S. L. Farhan, and S. Ethaib, “Updated Moving Forecasting Model of Air Maximum Temperature,” in *IOP Conference Series:*

Earth and Environmental Science, 2021, vol. 877, no. 1, p. 12032.

[13] G. K. Uyanık and N. Güler, "A study on multiple linear regression analysis," *Procedia-Social Behav. Sci.*, vol. 106, pp. 234–240, 2013.

[14] T. Peng, X. Zhi, Y. Ji, L. Ji, and Y. Tian, "Prediction skill of extended range 2-m maximum air temperature probabilistic forecasts using machine learning post-processing methods," *Atmosphere (Basel).*, vol. 11, no. 8, p. 823, 2020.

[15] A. H. A. Suliman, T. A. Awchi, M. Al-Mola, and S. Shahid, "Evaluation of remotely sensed precipitation sources for drought assessment in Semi-Arid Iraq," *Atmos. Res.*, vol. 242, p. 105007, 2020.