Forecasting Monthly Maximum Temperatures in Kerbala Using Seasonal ARIMA Models.

Adnan K. Shathir   **Layla Ali Mohammed Saleh**   **Sumayah Amal Al-Din Majeed**

Department of Civil Engineering, University of Kerbala, Kerbala, Iraq.

Adnanshathir@gmail.com   Lazu200919 @yahool.com   eng.sumayah@gmail.com

Abstract

Weather forecasting is an important issue in meteorology and scientific research.In this research, the Seasonal Auto Regressive Integrated Moving Average (ARIMA) model which is based on Box-Jenkins method was adopted to build the forecasting model. The max. Monthly temperature data for Kerbala city for the period (Jan.1980 to Dec.2016) was employed. The autocorrelation and partial autocorrelation functions for time series data from years 1980 to 2015 were used to identify the most appropriate orders of the ARIMA models. The validation test of these models were performed using the monthly max. Temperature of the year 2016. To calculate the model's accuracy and compare among them, statistical criteria such as MAE, RMSE, MAPE, and R^2 were used. The model $(2, 1, 2) \times (1, 1, 1)_{12}$ gave the most accurate results and used to forecast the monthly max. Temperature for the period (2017 to 2021) for study region.

Key words: Auto Regressive, Box - Jenkins, Forecasting, Moving Average, Temperature.

1. Introduction

Time series forecasting is considered a major tool in meteorology and environmental applications such as humidity, rainfall, temperature, stream flow, etc. In this technique, the past data must be adopted to develop the best model for forecasting future values, i.e. it is a process for calculating future values through understanding the past [1]. Over the time, numerous intelligent methods have been developed to improve the efficiency and accuracy of time series forecasting. The most commonly used technique is the Autoregressive.Integrated.Moving,Average (ARIMA) models. The acceptance of this method is owing to its statistical properties in addition to the powerful Box–Jenkins approach [2]. In order to select a fitting ARIMA model, it must be prepared an initial model from historical data, estimated the model parameters, and check the accuracy of the chosen model [3]. Box-Jenkins methodology has been widely applied in various environmental and water management fields. Several studies have been conducted to improvement and development the time series models that forecast the climate change in the world. Time series analysis was used by Langu to detect the variations in the runoff and rainfall patterns and to find the significant changes in a number of rainfall time series components [4]. Momani used the Box-Jenkins process to set up ARIMA model for monthly rainfall, data collected from Amman airport site for the time from 1922 to 1999 [5]. Nury et al. developed ARIMA models to predicate monthly min. and max. temperatures in the Moulvibazar and Sylhet districts in Bangladesh using data between year 1977 and 2011[6]. Sarraf et al. used 20-year -old data of the average monthly temperature and Relative humidity of Ahvaz station, Iran. They applied the ARIMA model using time series analysis program (ITSM) [7]. Chen et al. analyzed the mean monthly temperature in Nanjing city, China, during 1951–2017. The data from 1951 to 2014 was used as training, while that through 2015–2017 considered as the testing set [8].

Temperature possess undeniable effects on the environment, agricultural, water consumptions and human efforts [7]. Moreover, it affects nearly all other climatic variables, such as relative humidity, evaporation rate, wind direction, wind speed and precipitation patterns. The present paper aims to develop a time series seasonal ARIMA model using the max. Monthly temperature from 1980 to 2016 for Kerbala station, Iraq.

2. Methodology

2.1. Study Area

Generally, the temperature can be measured with a higher degree of accuracy comparative to other weather elements [9]. Past monthly temperature data covering Kerbala city in Iraq has been collected from the Kerbala meteorological station The study data was for a period of 37 years, from Jan.1980 to Dec. 2016. The location of the study area in relative to Iraq is shown in Fig.1.
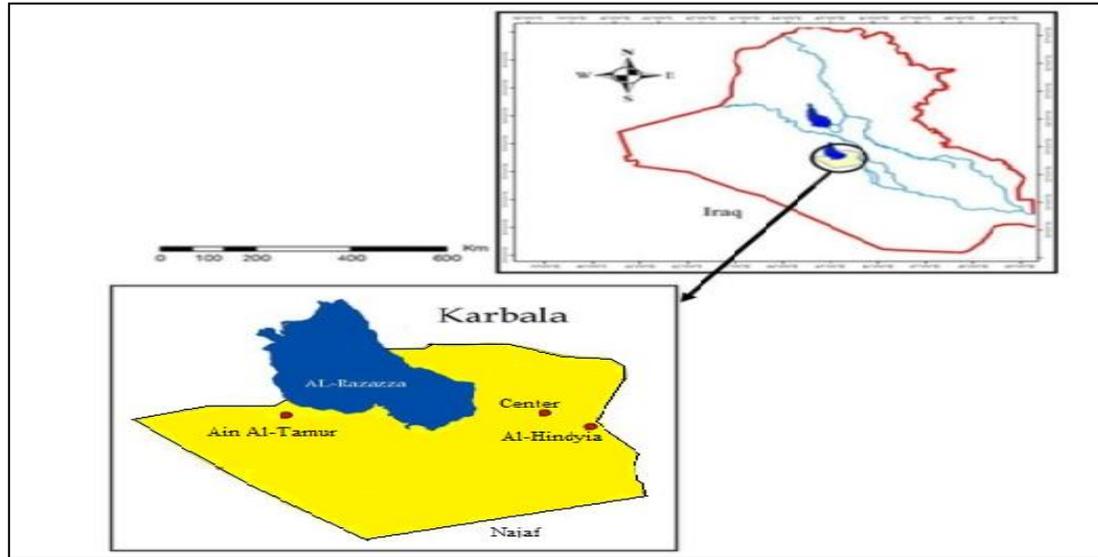
Fig.1: Location of study area relative to Iraq

2.2 Box-Jenkins Methodology

The term ARIMA stands for the acronym of "Auto-regressive. Integrated Moving Average" which are sometimes also named Box-Jenkins models [10]. The auto-regressive model of term p is usually classified as AR (p) while a moving average model of q order is identified as MA (q). The combined model which holds both the AR(p) and MA(q) terms is so-called an ARMA (p, q) model. In order to achieve the stationary for the time series, time is shifted with d differences order and the model is then turn into ARIMA (p, d, q).

Generally, a non-seasonal time series, ARIMA (p, d, q) model, can be modeled as a merger of previous values and errors, as explained below [3] [11][12]

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (1)$$

Where:

$\emptyset_t \ and \ \theta_t$ are coefficients, $p \ and \ q$ are the orders of the Auto-regressive and Moving Average models. Likewise, the seasonal ARIMA model can be expressed as (p, d, q) (P, D, Q).

The following steps are adopted to build Box-Jenkins ARIMA approach [12] [13]:

1. Test stationary: a stationary time series possess statistical attributes like the mean and the autocorrelation construction are constant with time. In order to create the stationary series, the proper degree of differencing (d) must be obtained. This step can be implemented by examining the autocorrelation function (ACF) and the partial autocorrelation function (PACF) or using unit root test.

2. Identification: the next step is to determine the appropriate p and q orders using the ACF and PACF. The result correlograms schemes i.e. the plots of ACF and PACF against the lag length are applied as a key to select one or further appropriate ARIMA models. In this stage, the estimated ACF and PACF are compared with different theoretical ACF and PACF. Later, a model is conditionally accepted whose theoretical ACF and PACF closely simulate the estimated ACF and PACF of the data series. The chosen model is treated only temporarily and it is a candidate for the final model.

3. Estimation: once a tentative model is specified, the estimation of the model parameters become straightforward. The parameters are determined such that the overall errors are minimized. Commonly, this can be done by applying the least squares procedure.

4. Diagnostic test: the residuals produced from the fitted model must be tested against adequacy. This is usually achieved by correlation analysis through the ACF and PACF residual plots and the goodness of fit test through Chi-square statistics. The model should be re-defined as in the first step when the residuals are correlated. Else, the autocorrelations are white noise and the model is appropriate to represent the time series.

3. Results & Discussions

The practical application of the Box and Jenkins technology contains three main steps which are data preparation, developing the mathematical model and apply the suitable forecasting model[3]. In this study, the time series of max. monthly temperature from Jan. 1980 to Dec. 2016 was used to represent an ARIMA model for forecasting future values in Kerbala city, Iraq. At first the data was checked, and it seems some missing data for some months in the year 2003, these missing data filled by the average max. monthly temperature for each month to make the time series is continuous and sufficient for analysis. The sketch for these data illustrated in Fig.2. Time series data from 1980 to 2015 was employed at the first stage to find the suitable ARIMA models and their parameters. For model validation stage, the data of the year 2016 were used. As shown in Fig.2, no abrupt variation occurred in the series i.e. no trend component, so the series will be considered stationary as a first assumption. The Autocorrelation and the Partial autocorrelation functions are the wisdom tool to choose the order of the ARIMA model as box and Jenkins suggest. The ACF & PACF plotted for a different order of simple and seasonal differencing with a lag time equal to 48 months as shown in Figures 3 to 10. This tool is a powerful way to choose the order of parameter for the simple model (p & q) and seasonal model (P&Q).

From the Figures 3& 4 where d=0 and D=0, the seasonal component is seen clearly which mean seasonal differencing is required. The sine wave for ACF and the cut of PACF means that there is an autoregressive for simple model with order 1 or larger. For seasonal model, the ACF curve had one value cutoff and a sine wave for PACF which mean there is moving average model with one order or larger for seasonality Figures 5 to 10 indicated the suggestion models with different simple and seasonal differencing . The parameter of these eight models is shown in Table 1.
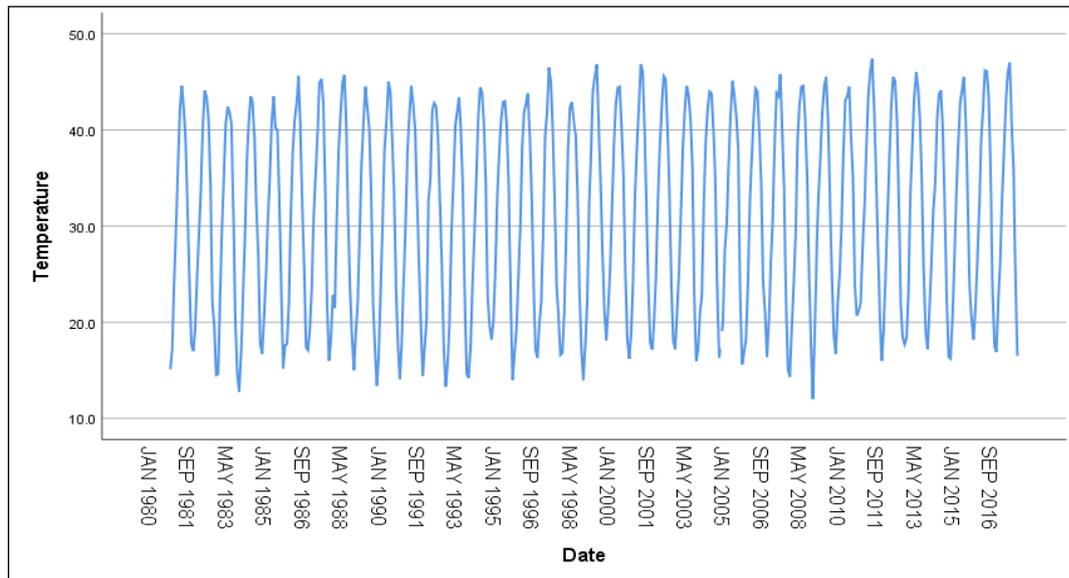
Fig.2: Monthly Max. Temperature of Kerbala (Jan.1980 - Dec.2016)
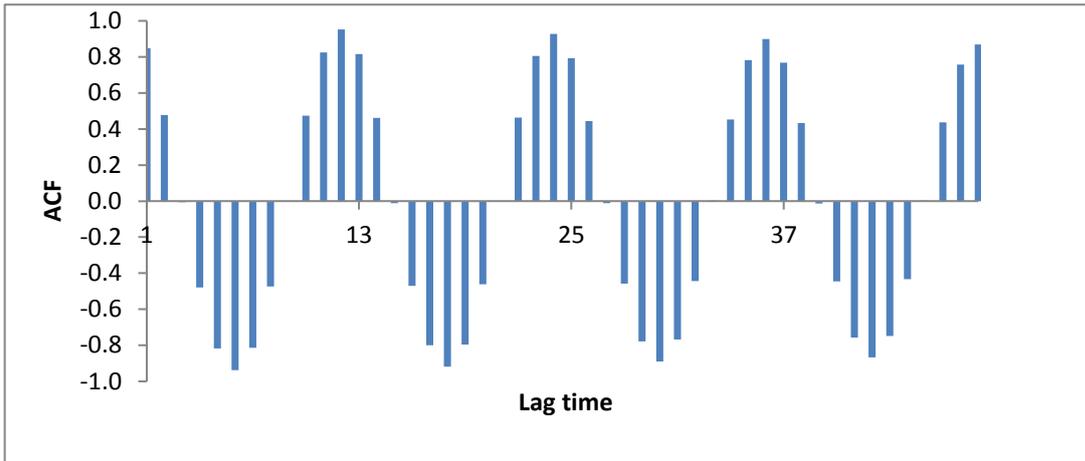
Fig.3: ACF for Max Temperature with d=0 & D=0
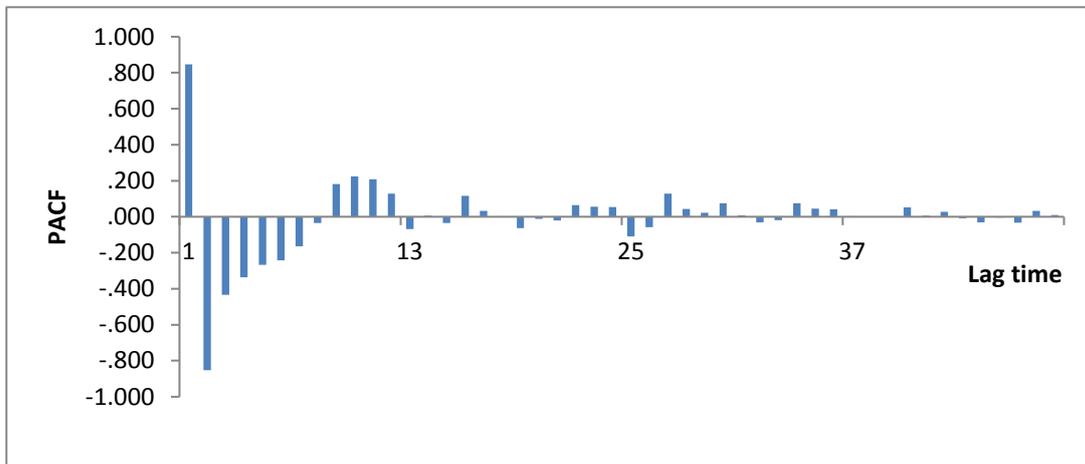
Fig.2: PACF for Max Temperature with d=0 & D=0
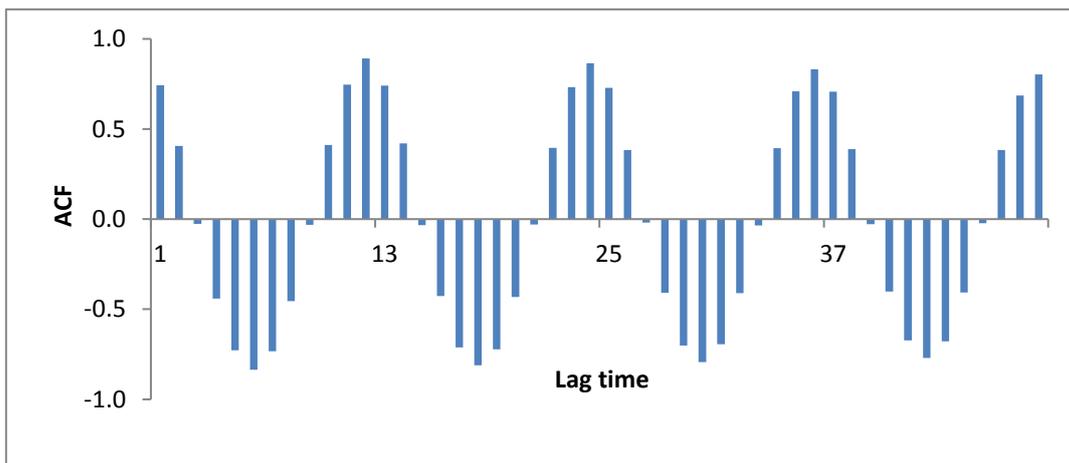
Figure 5: ACF for Max Temperature with d=1 & D=0
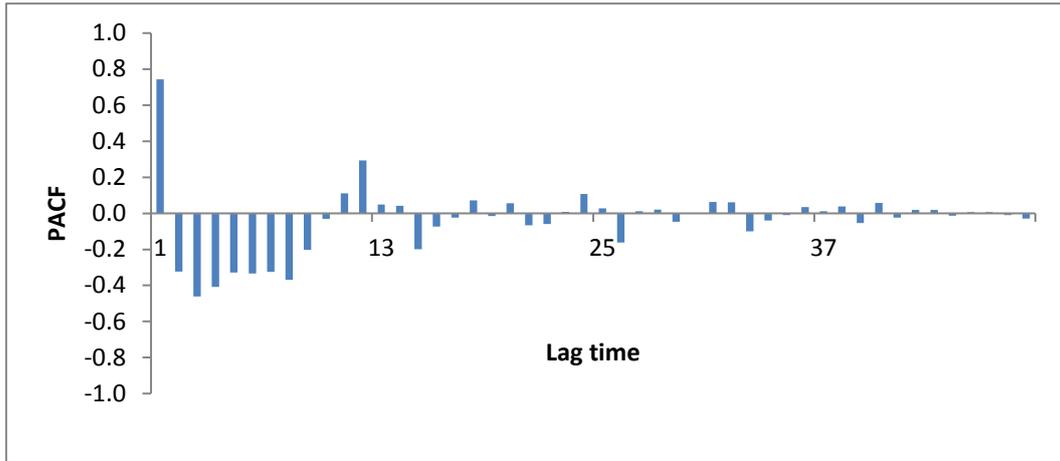
Fig.6: PACF for Max Temperature with d=1 & D=0
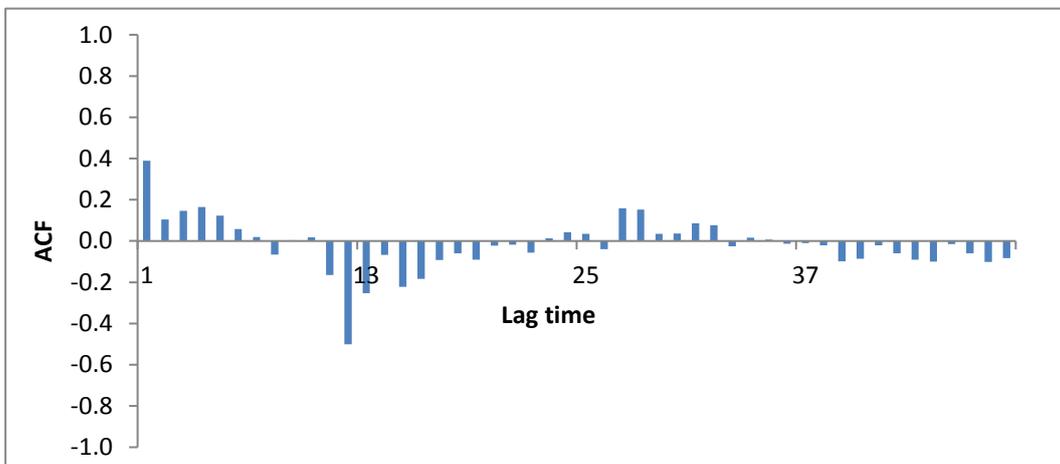
Fig.3: ACF for Max Temperature with d=0 & D= 1
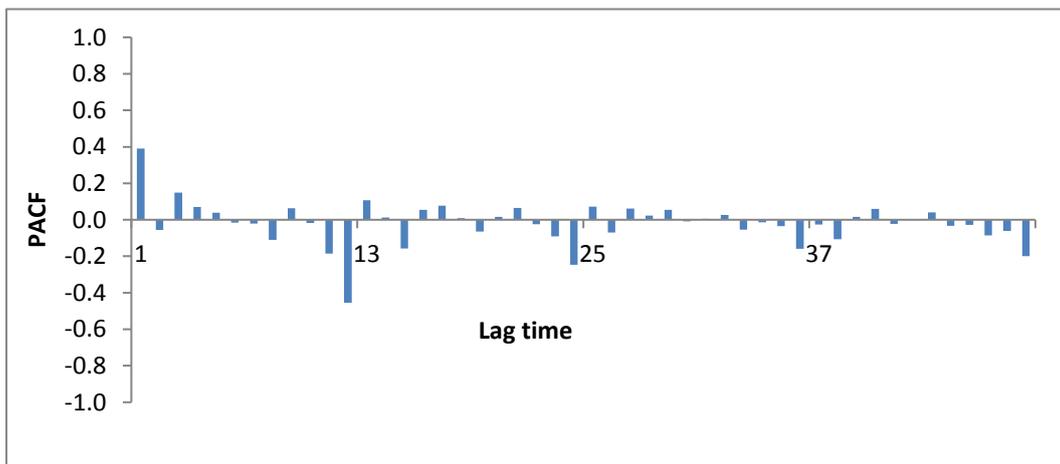
Fig.8: PACF for monthly max temperature with d=0 & D=1
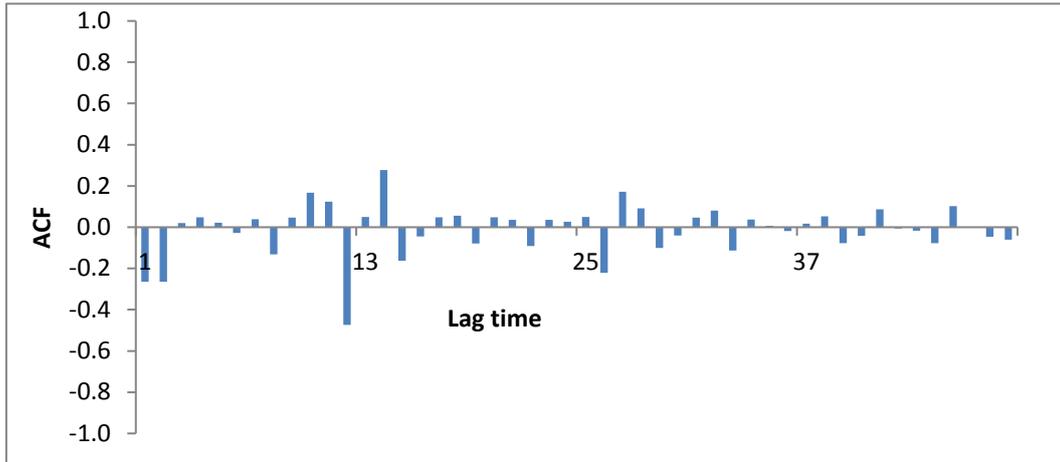
Fig.9: ACF for monthly max temperature with d=1 & D=1
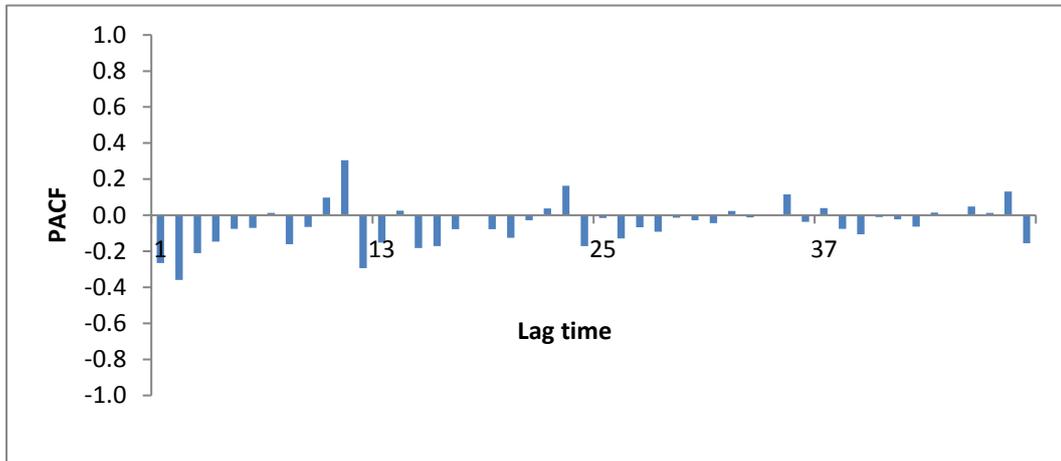
Fig.4: PACF for monthly max temperature with d=1 & D=1

Table 1: The predicted parameters of proposed ARIMA models

No.	Seasonal ARIMA Model	Simple AR			Simple MA		Seasonal AR	Seasonal MA
		ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	ϕ_{S1}	θ_{S1}
1	$(1,0,0)X(1,1,0)_{12}$	0.373					-0.49	
2	$(2,0,0)X(1,1,0)_{12}$	0.371	0.005				-0.49	
3	$(3,0,0)X(1,1,0)_{12}$	0.371	-0.01	0.054			-0.49	
4	$(1,1,0)X(1,1,0)_{12}$	-0.29					-0.50	
5	$(1,1,0)X(0,1,1)_{12}$	-0.31						0.987
6	$(1,1,1)X(0,1,1)_{12}$	0.372			1.00			1.00
7	$(2,1,2)X(1,1,1)_{12}$	-0.408	0.231		0.215	0.785	0.045	0.998
8	$(1,0,2)X(0,1,1)_{12}$	0.837			0.48	0.20		0.991

As illustrated in the table (1), the models (2 & 3) have parameters which approximately approach to zero so can be neglected [14].

To test the forecasting skill of the remaining models, they were applied to forecast the max. monthly temperature for the year 2016. The most significant criterion for selecting a forecast model is its accuracy i.e. how close it to forecast the actual data. In this research, some statistical criteria such are root mean square error RMSE, mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R2) were used to compare between models. An evaluation criterion is explained in the table (2). It can be inferred that all models gave good results but the forecasting capability of the ARIMA $(2, 1, 2) \times (1, 1, 1)12$ model was stronger as it has min. errors and high R2. Figure 11 presents the observed and forecast temperature for the best model.

The selected model was used to forecast the max. monthly temperature of the Karbala city during the period from 2017- 2021 as shown in Figure 12. As it is clear from the series, the data are nearly close because the forecasting process was performed without updating the data with a lag time equal 60 month.

Table 2: statistic parameters of ARIMA models

No.	ARIMA Model	RMSE	MAE	MAPE	R^2
1	$(1,0,0)X(1,1,0)_{12}$	1.188	0.972	0.036	0.9878
2	$(1,1,0)X(1,1,0)_{12}$	1.556	1.399	0.048	0.9875
3	$(1,1,0)X(0,1,1)_{12}$	1.188	0.972	0.036	0.9878
4	$(1,1,1)X(0,1,1)_{12}$	1.222	0.981	0.038	0.9871
5	$(2,1,2)X(1,1,1)_{12}$	1.121	0.962	0.035	0.9878
6	$(1,0,2)X(0,1,1)_{12}$	1.252	0.986	0.037	0.9875
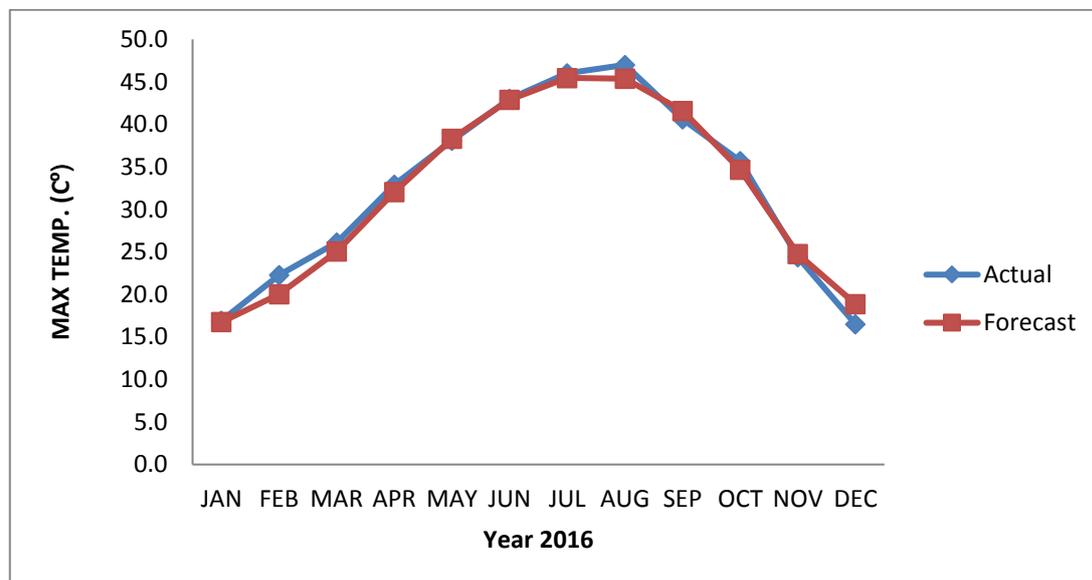
Fig.5: The actual and forecasting monthly max.temparture for the year 2016 using ARIMA $(2, 1, 2) \times (1, 1, 1)_{12}$ model.
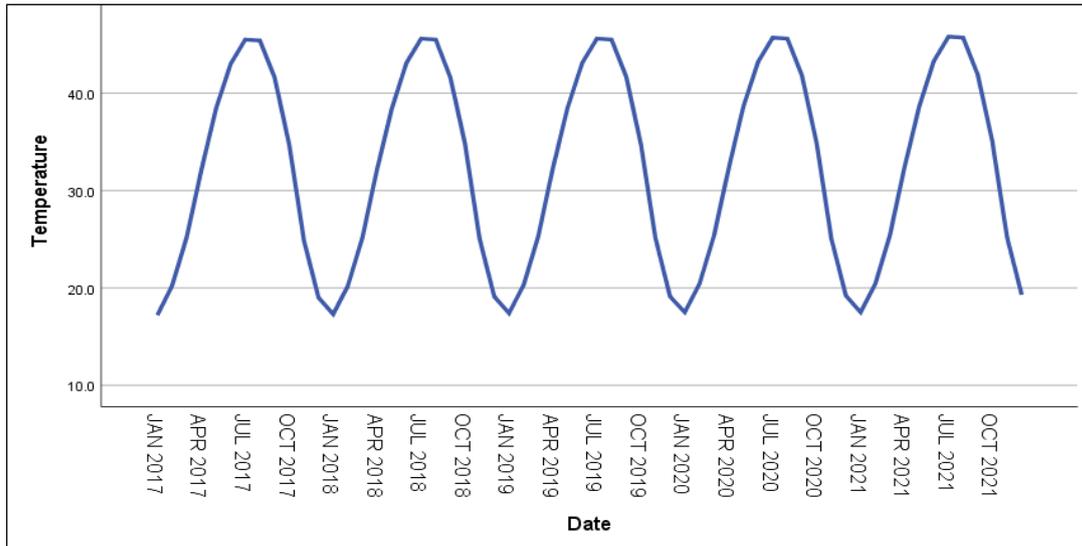
Fig.6: The forecasted time series of max. Monthly temperatures for the period (2017-2021).

4. Conclusion

In this paper, the Box- Jenkins technique has been used to develop monthly max. temperature with the aid of SPSS.23 software. A number of ARIMA models were proposed based on the auto correlation and partial autocorrelation functions of the time series. Among these models, seasonal ARIMA $(2, 1, 2) \times (1, 1, 1)12$ model gave the best results, so it can be adapted to forecast the monthly temperature of Karbala city. A Nomo- graph chart can be used instead of the forecasting equation to account the future temperature directly.

CONFLICT OF INTERESTS.

- There are no conflicts of interest.

Reference

[1] T. Raicharoen, C. Lursinsap, and P. Sanguanbhokai, "Application of critical support vector machine to time series prediction", Proceedings of the 2003 International Symposium on Circuits and Systems, vol. 5, pp. 741–744, 2003.

[2] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing Journal, vol. 50, pp. 159–175, 2003.

[3] A. K. Shathir and L. A. M. Saleh, "Best Arima Models for Forecasting Inflow of Hit Station", Basrah Journal for Engineering Sciences, vol. 16, no. 1, pp. 62–71, 2016.

[4] E. M. Lungu, "Detection of changes in rainfall and runoff patterns", Journal of Hydrology, vol. 147, pp. 153–167, 1993.

[5] P. Momani, "Time Series Analysis Model for Rainfall Data in Jordan : Case Study for Using Time Series Analysis", American Journal of Environmental Sciences, vol. 5, no. 5, pp. 599–604, 2009.

[6] A. H. Nury, M. Koch, and M. J. B. Alam, "Time Series Analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh", 4th Int. Conf. Environ. Asp. Bangladesh, pp. 24–26, 2013.

[7] A. Sarraf, S. F. Vahdat, and A. Behbahaninia, "Relative Humidity and Mean Monthly Temperature Forecasts in Ahwaz Station with ARIMA Model in time Series Analysis", 2011 Int. Conf. Environ. Ind. Innov. IPCBEE, vol. 12, pp. 149–153, 2011.

[8] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, "Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing", IOP Conf. Ser. Mater. Sci. Eng., vol. 394, no. 5, pp. 1–7, 2018.

[9] M. Tektaş, "Weather forecasting using ANFIS and ARIMA models", Environmental Research, Engineering and Management, vol. 1, no. 51, pp. 5–10, 2010.

[10] G. M. Box, G.E.P. and Jenkins, "Time Series Analysis, Forecasting and Control", Holden-Day, San Francisco., 1976.

[11] S. Shamsnia, N. Shahidi, and A. Liaghat, "Modeling of Weather Parameters Using Stochastic Methods, ARIMA Model, Case Study: Abadeh Region, Iran", Int. Conf. Environ. Ind. Innov., vol. 12, pp. 282–285, 2011.

[12] S. L. Ho, M. Xie, and T. N. Goh, "A comparative study of NN and BoxJenkins ARIMA modeling in time series prediction", Computers & Industrial Engineering Journal, vol. 42, pp. 371–375, 2002.

[13] A. H. Nury, K. Hasan, K. M. Erfan, and D. C. Dey, "Analysis of spatially and temporally varying precipitation in Bangladesh", Asian Journal of Water, Environment and Pollution, vol. 13, no. 3, pp. 15–27, 2016.

[14] H. Akaike, "On the Likelihood of a Time Series Model", Journal of the Royal Statistical Society, vol. 24, no. 3/4, pp. 217–235, 1978.

التنبؤ بدرجات الحرارة الشهرية العظمى لمدينة كربلاء باستخدام نماذج ARIMA الفصلية

عدنان خلف شذر   ليلى علي محمد صالح   سمية امال الدين مجيد

قسم الهندسة المدنية، جامعة كربلاء، كربلاء، العراق

eng.sumayah@gmail.com   Lazu200919 @yahool.com   Adnanshathir@gmail.com

الخلاصة:

يعتبر التنبؤ بالعوامل الجوية مسالة مهمة في مجال الأرصاد الجوي والبحث العلمي. في هذا البحث، تم اعتماد نموذج الانحدار الذاتي المتكامل والمتوسط المتحرك الفصلي (ARIMA) والذي يستند على نظرية.Box –Jenkins استخدمت البيانات الشهرية لمعدل درجة الحرارة العظمى لمدينة كربلاء للفترة (من يناير 1980 إلى ديسمبر 2016). تم اقتراح عدة نماذج للتنبؤ اعتمادا على دالة الارتباط الذاتي ودالة الارتباط الذاتي الجزئي لبيانات السلسلة الزمنية للسنوات من 1980 الى 2015. تم اختبار النماذج المقترحة باستخدام درجات الحرارة الشهرية العظمى لسنة 2016. من اجل اختبار دقة النماذج والمقارنة بينها استخدمت المعلمات الاحصائية مثل معدل الخطأ المطلق MAE، الجذر التربيعي لمتوسط مربعات الخطأ RMSE، المتوسط المطلق للخطأ النسبيMAPE ومعامل التحديد R^2. بينت النتائج ان النموذج $(1, 1, 1)(2, 1, 2)_{12}$ كان الاكثر دقة واستخدم للتنبؤ بمعدل درجات الحرارة الشهرية العظمى لمنطقة الدراسة للفترة من 2017 الى 2021.

الكلمات الداله: الانحدار الذاتي، بوكس وجينكنز، التنبؤ، المتوسط المتحرك، درجة الحرارة.