International Journal of Environmental Monitoring and Analysis
2021; 9(6): 214-221
http://www.sciencepublishinggroup.com/j/ijema
doi: 10.11648/j.ijema.20210906.17
ISSN: 2328-7659 (Print); ISSN: 2328-7667 (Online)

Modeling and Forecasting Daily Temperature Time Series in the Memphis, Tennessee

Khayrun Nahar Mitu[1, *], **Khairul Hasan**[1, 2]

[1]Department of Civil and Environmental Engineering, Shahjalal University of Science and Technology, Sylhet, Bangladesh

[2]Department of Civil Engineering, University of Memphis, Memphis, USA

Email address:
mitu-cee@sust.edu (K. N. Mitu)

[*]Corresponding author

To cite this article:
Khayrun Nahar Mitu, Khairul Hasan. Modeling and Forecasting Daily Temperature Time Series in the Memphis, Tennessee. *International Journal of Environmental Monitoring and Analysis*. Vol. 9, No. 6, 2021, pp. 214-221. doi: 10.11648/j.ijema.20210906.17

Received: September 25, 2021; **Accepted**: November 2, 2021; **Published**: December 29, 2021

Abstract: Temperature is an essential weather component because of its tremendous impact on humans and the environment. As a result, one of the widely researched parts of global climate change study is temperature forecasting. This work analyzes trends and forecasts a temperature change to see the transient variations over time using daily temperature data from January 1, 2016 – November 3, 2019, collected from a weather station located at the Memphis International Airport. The Mann-Kendall (M-K) test is used to detect time series analysis patterns as a non-parametric technique. The result from the test revealed that the temperature time series increased by 0.0030 °F almost every day, implying that the location is becoming hotter. The other method of analysis is the autoregressive integrated moving average (ARIMA) model, which fits temperature time series using its three standard processes of identification, diagnosis, and forecasting. Considering the selection criteria, The seasonal autoregressive integrated moving average (SARIMA) (3, 0, 0) $(0, 1, 0)_{365}$ model is found as appropriate for the studied temperature data on a daily basis. Finally, the selected model is utilized to estimate the next 50 days; after November 3, 2019, the temperature forecast showed an increasing trend. This observed trend provides an understanding of daily temperature change in the studied area for that specific period.

Keywords: ARIMA, Daily Average Temperature Data, Mann–Kendall (M–K) Test, Trend, Memphis International Airport, SARIMA

1. Introduction

Temperature variation resulting from climate change has become a global concern as it is correlated to global warming. The fifth IPCC assessment report revealed that the mean temperature increased by 0.85°C through 1880 to 2012. [1]. Global warming significantly impacts the natural ecology, agricultural production, and human health [2]. Rising temperatures have already intensified drought, flooding, rising sea level, and weather extremes. [1]. Furthermore, temperature variations will delay the onset of the monsoon and cause water loss from the soil, reducing crop productivity and lowering water levels in surface and groundwater [3]. Because of ocean-atmosphere circulation, land cover use, and other linked characteristics, surface air temperature fluctuates more at the regional scale than at the global average scale [4]. However, because the temperature is affected by many climate elements, it is an ever-difficult endeavor to predict the changes in temperature for the projected duration [5]. Therefore, it is required to conduct quantitative analyses of temperature fluctuation to take the appropriate steps to mitigate adverse effects. In temperature forecasting research, time-series analysis is considered as an essential direction [6-8].

Since the Mann-Kendall (M-K) test incorporates the better treatment of outliers, It is frequently employed in weather and climate time series data to discover trends. [9, 10]. Various trend analysis studies have been carried out at various spatiotemporal scales, underscoring the importance of the M-K test. The M-K test was used in a research in Ethiopia's Woleka sub-basin to detect time series trends of precipitation and temperature. [11]. In a study the Mann-Kendall test, Sen's slope estimator, and linear regression were

used to examine yearly and seasonal temperature patterns, along with temperature extremes.[12]. One study used the M-K test to identify changes in environmental and meteorological features at a Kolkata station from 2002 to 2011, and the M-K test's performance was reliable at the verified significant level [13]. In addition, a study looked at the trend in rainfall time series at Fifteen sites in the Swat River basin from 1961 to 2011 using both non-parametric M-K and SR statistical tests, which provided a daily forecast of the parameters with precision [14]. Given this, it is reasonable to conclude that the standard-Kendall test is widely utilized to assess how parameters change over time.

In time series analysis, projecting values in the later phase are based on previous observations of the variable under examination. Numerous studies in hydrology and meteorology used the ARIMA method to achieve more accurate forecasts, and this method has essentially superseded older statistical techniques [15, 16]. Another study used a seasonal ARIMA model for agricultural irrigation and reported achievement of a significant level of model fitting in strategic planning [17]. Likewise, one study used a SARIMA model to assess temperature trend in Assam [18]. Moreover, an investigation used the ARIMA model to forecast monthly mean temperature and discovered a falling trend [19]. A substantial number of studies have successfully comprehended climate parameters and have provided a better understanding of the hydrological system using ARIMA and SARIMA models [4, 20-22].

This study is designed to investigate the temperature time series of daily average temperature data to discover trends using the non-parametric M–K test with the ARIMA model technique. The SARIMA model is fitted to daily temperature data (Jan 1, 2016 – Nov 3, 2019). The chosen model is used to forecast temperature for the next 50 days from Nov 4, 2019, to Jan 23, 2020, using Box-Jenkins's technique.

2. Materials and Methods

2.1. Data Collection and Study Area

The data for the temperature time series came from a weather station located at the Memphis International Airport, as seen in Figure 1. This data represented the local weather of Memphis, Tennessee, and offered the amount of data required to fit the SARIMA model. The data gathered from the station included temperature readings at the daily interval from Jan 1, 2016, to Nov 3, 2019.

Figure 1. *Memphis International Airport Weather Station Site Map.*

2.2. Trend Analysis

The Mann-Kendall test implies that data are not normally distributed, and it additionally considers the effect of outliers. As a result, trend analysis frequently employs the non-parametric M–K test and the Sen slope estimator [9, 12]. Using a two-tailed test with a 5% significance level [13], the alternative as well as null hypotheses are H_0 = There is no discernible trend in the time series. and H_1 = There is a rising or falling trend. [12]. Thus, the following equations (1) and (2) can be used to determine Mann–Kendall test statistics.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{sign}\,(x_i - x_k) \quad (1)$$

$$Var\,(S) = \frac{n\,(n-1)(2n+15) - \sum_{i=1}^{m} e_i(e_i-1)(2e_i+5)}{18} \quad (2)$$

where x_i and x_k are consecutive data in the series; n is the sample size; ei is the number of ties at the ith value, and m is the number of ties if the value is tied.

Z_C, the standard test statistic, was calculated as follows:

$$Z_C = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} \text{ when } S > 0 \\ 0 \text{ when } S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} \text{ when } S < 0 \end{cases} \quad (3)$$

The Z_C symbol indicates the trend's direction. A negative Z number indicates a downward trend, whereas a positive ZC value indicates an upward trend [13, 23]. The magnitude of the slope (change per day) was determined using Sen's estimator [9, 12].

2.3. ARIMA Model

ARIMA is the acronym for the autoregressive integrated moving average (ARIMA) model, widely known as the Box-Jenkins model (p, d, q). The order of the autoregressive (AR) is p, the degree of difference is d, and the order of the moving average (MA) is q. [24]. It is almost as if the independent variables in the regression model are the past values of the time series.

Equation 4 or 5 can be used to express the general equation [22].

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots\dots + \varphi_p Y_{t-p} + e_t + \theta_1 \theta_{t-1} + \theta_2 \theta_{t-2} + \cdots \dots \theta_q \theta_{t-q} \quad (4)$$

$$\varphi_p(L) * (1 - L)^d Y_t = c + \theta_q(L) * e_t \quad (5)$$

where φ_1 , $\varphi_2, \dots\dots\dots\dots \varphi_p$ and $\theta_1, \theta_2 \dots\dots\dots\dots\dots \theta_q$ are the regression coefficient, Y_t is the time series data (temperature), c is the intercept, φ_p indicates the AR part's order, $\theta_q(L)$ indicates the MA part's order, and d indicates the differencing, e_t is called the random error amount.

If seasonality is considered, then the ARIMA model will become a seasonal autoregressive integrated moving average (SARIMA) model and represented by ARIMA (p, d, q) (P, D, Q) S [19]. S stands for the number of seasons per year, P for the seasonal AR, D for the seasonal difference, and Q for the seasonal moving average.

The first stage in fitting the ARIMA model is to ensure that the time series is stationary. The Augmented Dickey and Fuller (ADF) unit root test are used to determine the stationarity of a time series data set [19]. The test's null and alternative hypotheses are H0: Series has a unit root and H1: Series has no unit root, respectively [19]. The ADF test statistics must be smaller than the crucial value to reject the null hypothesis. The transformation should be done using the differencing procedure if the time series is not stationary [13]. Following the discovery of the stationary time series, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to determine the appropriate order of AR (p) and MA (q) [18].

The model coefficients are estimated using the least square approach after the appropriate values of p, d, and q have been determined. The residuals are then examined using a set of criteria, assuming that they are not autocorrelated and normally distributed [24]. Within a 95 percent confidence interval, the residual's ACF should not differ from zero. Furthermore, the histogram of the residual will have a bell shape, indicating that it is normally distributed.

Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are used to select models [13]. Then the model with the least AIC and BIC values is selected as a best-fit model [21].

$$AIC = 2k - 2 * \ln(L) \quad (6)$$

$$BIC = k * \ln(n) - 2 * \ln(L) \quad (7)$$

where k is the number of model parameters, L is the likelihood function's maximum value, and n is the number of observations.

Finally, if the model has been evaluated using the root mean square error (RMSE), the mean absolute error (MAE) or mean absolute percentage error (MAPE) is employed for the predictive capability, as shown in equation 8-10. The minimum value of the RMSE, MAE, and MAPE is ideal for the model's adequacy. The RMSE, MAE, and MAPE equations are presented in the equation (8), (9), and (10).

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n} \frac{[Y_t(\text{obs}) - Y_t(\text{pred})]^2}{n}} \quad (8)$$

$$\text{MAE} = \sum_{t=1}^{n} \frac{[Y_t(\text{obs}) - Y_t(\text{pred})]}{n} \quad (9)$$

$$\text{MAPE} = \frac{1}{n} * \sum_{t=1}^{n} \left[\frac{[Y_t(\text{obs}) - Y_t(\text{pred})]}{n}\right] * 100 \quad (10)$$

The value obtained at time t is Y_t (obs), the predicted value is Yt (pred), and the number of observations is n.

3. Results and Discussion

3.1. Descriptive Statistics

Table 1 shows the descriptive statistics for the temperature time series. The skewness, in this case, is negative, indicating that the left-handed tail is longer than the right-handed tail. The first and third quartiles are 52.60 °F and 79.27 °F, respectively, according to the box plot in Figure 2.

Table 1. *Descriptive statistics of the temperature data.*

Temperature (degree F)	
Mean	64.83862698
Standard Error	0.434537702
Median	67.20285714
Mode	71.84
Standard Deviation	16.27632295
Sample Variance	264.9186888
Kurtosis	-0.681672829
Skewness	-0.529720862
Range	76.88296296
Minimum	13.03703704
Maximum	89.92
Count	1403

3.2. Mann–Kendall Trend Analysis

Table 2 shows the M–K test statistics for the time series data. Since the threshold value (p-value = 0.006) is less than 0.05, the M–K tests revealed patterns in temperature time series. Kendall's positive value implies a positive upward trend; hence, the temperature time series, which previously had a tendency, has been demonstrated to have a positive upward trend. According to Sen's slope, the temperature time series exhibits a trend of 0.003 ℉ every day, which is the slope of the trend.

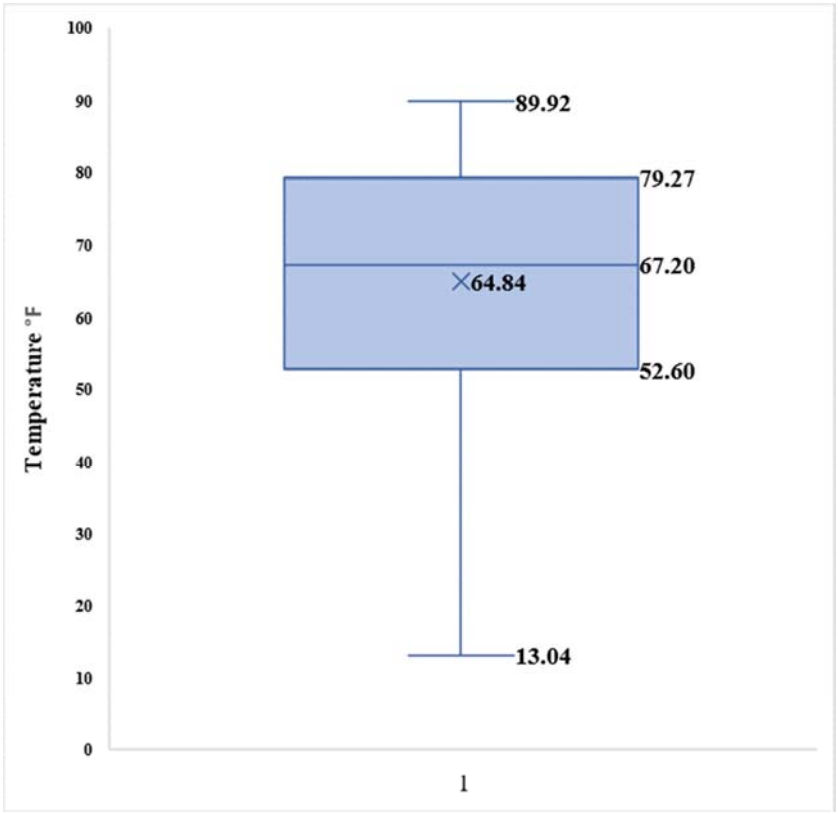

Figure 2. *Boxplot of the daily average temperature time-series data.*

Table 2. *MK Statistics at 5% significance level for Memphis International Airport.*

Station	Parameter	Sen's slope	Kendall's τ	p-value (Two-tailed test)	alpha, α	Test interpretation
Memphis International Airport	Temperature	0.003	0.049	0.006	0.05	Trend in series

3.3. ARIMA Model

Figure 3 illustrates a time series depiction of daily surface temperature. The data appear to be stationary in the graph. A consistent pattern in the data, on the other hand, suggests seasonality. As seen in Figure 4, this figure is further studied by deconstructing it using the additive method. Figure 4 shows that the data contains a seasonal component and a wavelike structure.

As a consequence, instead of the ARIMA model, the SARIMA model is investigated. Moreover, a trend in the time series data is also depicted in the figure. Furthermore, the inclusion of outliers in the data is indicated by the random component.

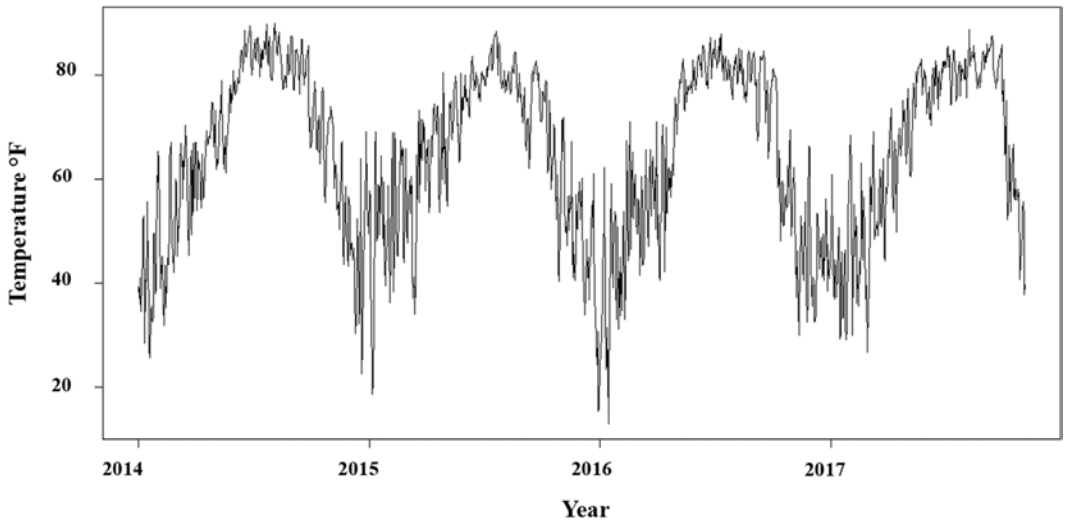

Figure 3. *Time Series plot for daily surface temperature.*

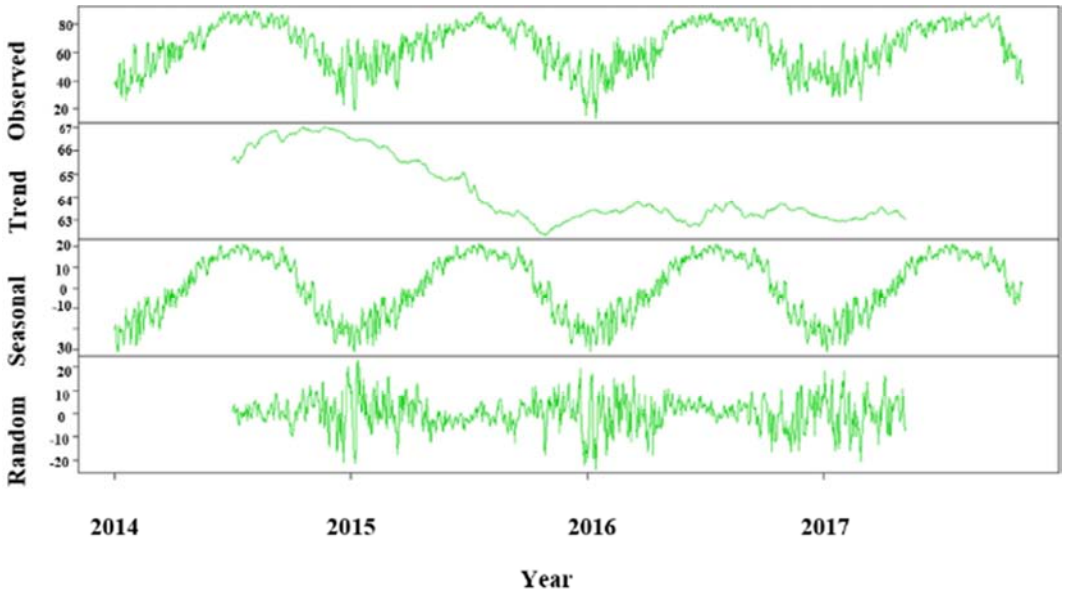

Figure 4. *Decomposition of daily time series surface temperature by the additive method.*

All The unit root in the daily temperature time series data is checked using the ADF test. The ADF data and accompanying p-value for the ADF test is shown in Table 3. The null hypothesis with a unit root is rejected because the p-value is less than 0.05. As an outcome, the daily temperature time series data is stationary, and there is no need for the difference.

Table 3. *ADF unit root test of the original daily temperature time series.*

	ADF statistics	p-value
Daily Temperature time series	-7.6689	0.01

All The ACF and PACF of the daily temperature time series are shown in Figures 5 and 6. The ACF plot in Figure 5 looks like a sine wave, indicating that the data has much seasonality. As a result, the seasonal difference should be considered to eliminate seasonality. PACF may be used to detect the order since ACF shows exponential series decaying to zero, suggesting the autoregressive model exclusively. As shown in figure 6, the PACF is significant at lags 1, 2, and 3, and after lag 3, the PACF shows an irregular pattern by being above and below the confidence limit. Furthermore, there is no discernible seasonal rise between 365 and 730. As a result, the non-seasonal AR term's order is possibly 3, whereas the seasonal AR term's order could be zero.

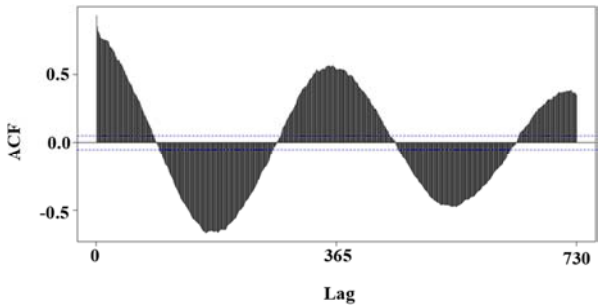

Figure 5. *ACF plot for daily surface temperature time series.*

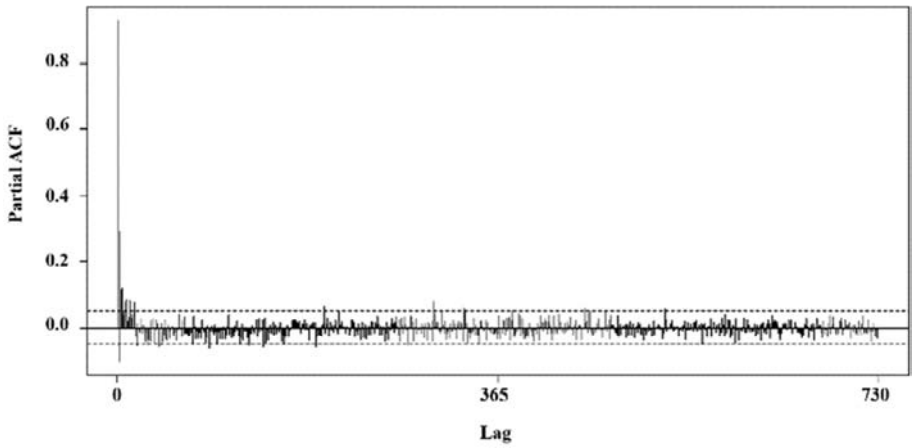

Figure 6. *PACF plot for daily surface temperature time series.*

The resulting model for this daily temperature data is SARIMA $(3,0,0)$ $(0,1,0)_{365}$, with three no seasonal

autoregressive parameters and one seasonal difference, considering the low AIC and BIC values. The estimated parameters for the selected model are shown in Table 4. The table shows that all coefficients are significant because the t-statistics are greater than 1.96 in all situations. Table 5 reveals that the highest RMSE, maximum MAE, and maximum MAPE for the selected model are 6.86, 4.28, and 8.23 percent, respectively. These numbers can be considered when determining whether or not the model is a good fit.

Table 4. *Parameter estimation for the daily temperature time series.*

Model	Variable	Coefficient	Std. error	t-statistics	AIC	BIC
$(3,0,0)\ (0,1,0)_{365}$	AR (1)	0.9043	0.0306	29.55229	7264.6	7284.38
	AR (2)	-0.3694	0.0401	-9.21197		
	AR (3)	0.1635	0.0307	5.325733		

Table 5. *Measures of the accuracy of the model fit.*

	RMSE	MAE	MAPE
Measures of accuracy	6.858348	4.285985	8.235479

The ACF of the residuals has no substantial autocorrelation, as seen in Figure 7. Furthermore, the histogram of the residuals is more or less normally distributed. As a result, the residuals are white noise, indicating that the chosen model can forecast.

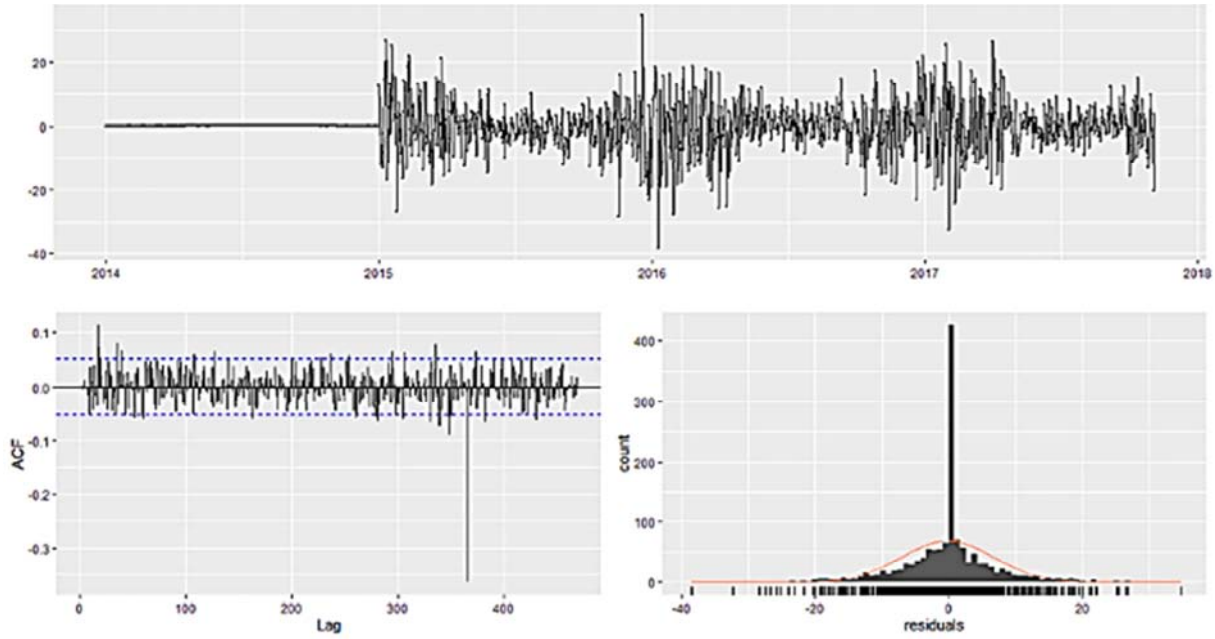

Figure 7. *Residual of the SARIMA (3,0,0) $(0,1,0)_{365}$ model.*

Figure 8 shows the time series forecast for the next 50 days. A blue line represents the forecasts. The light-colored area indicates a 95 percent prediction interval.

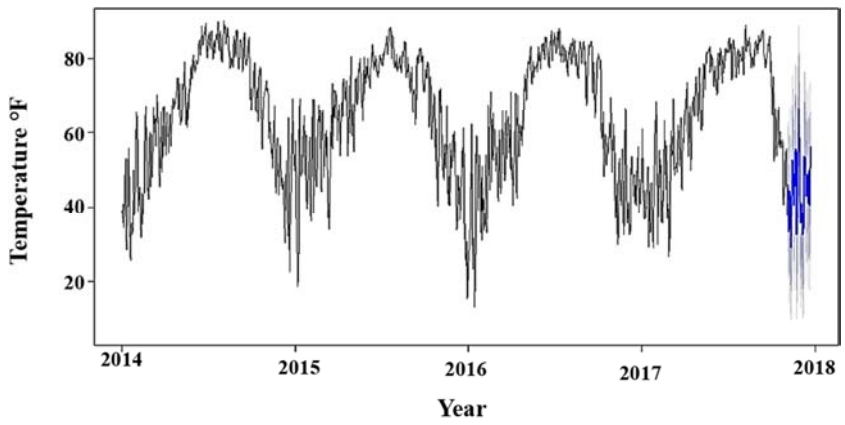

Figure 8. *Forecast using the SARIMA (3,0,0) $(0,1,0)_{365}$ model.*

4. Conclusions

The Mann–Kendall (M–K) test and the Box–Jenkins's method dubbed SARIMA were used in this work to determine daily average temperature variability and forecasting. For the Memphis international airport station, Mann–(M–K) Kendall's trend analysis showed a growing upward trend of 0.0030F each day. In addition, the identification and diagnosis for the SARIMA model reveal that the model fits well. The residuals analysis also shows that the model fits all assumptions. Moreover, the accuracy measures validate the model's predictive capacity. The following 50 days of data after November 3, 2019, has been projected using the $(3, 0, 0)\ (0, 1, 0)_{365}$ model. The analysis of this study will give policymakers insight into the rate of temperature change during that period and the scope and extent of possible temperature change.

Acknowledgements

The authors would like to thank Iowa State University for publishing daily air temperature data and giving open access to download.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

[1] IPCC, 2018: Summary for Policymakers. In: Global warming of 1.5 C. An IPCC Special Report on the impacts of global warming of 1.5 C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global. World Meteorological Organization, Geneva, Tech. Rep. 2018.

[2] H. Wang, J. Huang, H. Zhou, L. Zhao, and Y. Yuan, "An integrated variational mode decomposition and arima model to forecast air temperature," Sustainability 11 (2019) 4018. https://doi.org/10.3390/su11154018.

[3] K. Guan, B. Sultan, M. Biasutti, C. Baron, and D. B. Lobell, "What aspects of future rainfall changes matter for crop yields in West Africa?," Geophysical Research Letters, 42 (2015) 8001–8010. https://doi.org/10.1002/2015GL063877.

[4] E. S. El-Mallah and S. G. Elsharkawy, "Time-series modeling and short term prediction of annual temperature trend on Coast Libya using the box-Jenkins ARIMA Model," Advances in Research 6 (2016) 1–11. doi: 10.9734/AIR/2016/24175.

[5] S. Lee, Yung-Seop Lee, and Y. Son, "Forecasting Daily Temperatures with Different Time Interval Data Using Deep Neural Networks," Applied Sciences 10 (2020) 1609. https://doi.org/10.3390/app10051609.

[6] W. Wanishsakpong and B. E. Owusu, "Optimal time series model for forecasting monthly temperature in the southwestern region of Thailand," Modeling Earth Systems and Environment 6 (2019) 525–532. doi.org/10.1007/s40808-019-00698-5.

[7] T. Dimri, S. Ahmad, and M. Sharif, "Time series analysis of climate variables using seasonal ARIMA approach," Journal of Earth System Science volume 129 (2020) 149. doi: 10.1007/s12040-020-01408-x.

[8] M. BALIBEY and S. TÜRKYILMAZ, "A time series approach for precipitation in Turkey," Gazi University Journal of Science 28 (2015) 549–559.

[9] R. Ali, A. Kuriqi, S. Abubaker, and O. Kisi, "Long-term trends and seasonality detection of the observed flow in Yangtze River using Mann-Kendall and Sen's innovative trend method," Water 11 (2019) 1855. https://doi.org/10.3390/w11091855.

[10] F. Fathian, Z. Dehghan, M. H. Bazrkar, and S. Eslamian, "Trends in hydrological and climatic variables affected by four variations of the Mann-Kendall approach in Urmia Lake basin, Iran," Hydrological Sciences Journal 61 (2016) 892–904. doi: 10.1080/02626667.2014.932911.

[11] A. Asfaw, B. Simane, A. Hassen, and A. Bantider, "Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia: A case study in Woleka sub-basin," Weather and Climate Extremes 19 (2018) 29–41, doi: 10.1016/j.wace.2017.12.002.

[12] W. Wei, B. Wang, K. Zhang, Z. Shi, G. Ge, and X. Yang, "Temporal Trends and Spatial Patterns of Temperature and Its Extremes over the Beijing-Tianjin Sand Source Region (1960–2014), China," Advances in Meteorology 2018 (2018). doi.org/10.1155/2018/5473105.

[13] S. Chaudhuri and D. Dutta, "Mann–Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models," Environmental monitoring and assessment 186 (2014) 4719–4742. doi: 10.1007/s10661-014-3733-6.

[14] I. Ahmad, D. Tang, T. F. Wang, M. Wang, and B. Wagan, "Precipitation trends over time using Mann-Kendall and spearman's rho tests in swat river basin, Pakistan. Advances in Meteorology 2015 (2015) 431860. doi.org/10.1155/2015/431860.

[15] B. S. Kim, S. Z. Hossein, and G. Choi, "Evaluation of temporal-spatial precipitation variability and prediction using seasonal ARIMA model in Mongolia," KSCE Journal of civil Engineering 15 (2011) 917–925.doi.org/10.1155/2015/431860.

[16] M. Małgorzata, I. Malinowska, M. Gos, and J. Krzyszczak, "Forecasting daily meteorological time series using ARIMA and regression models," International agrophysics 32 (2018).

[17] S. Wang, J. Feng, and G. Liu, "Application of seasonal time series model in the precipitation forecast," Mathematical and Computer Modelling 58 (2013): 677–683. doi: 10.1016/j.mcm.2011.10.034.

[18] K. Goswami, J. Hazarika, and A. N. Patowary, "Monthly Temperature Prediction Based On Arima Model: A Case Study In Dibrugarh Station Of Assam, India.," International Journal of Advanced Research in Computer Science 8 (2017) doi.org/10.26483/ijarcs.v8i8.4590.

[19] E. Afrifa-Yamoah, "Application of ARIMA models in forecasting monthly average surface temperature of Brong Ahafo Region of Ghana," International Journal of Statistics and Applications 5 (2015) 237–246. doi: 10.5923/j.statistics.20150505.08.

[20] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, "Time series forecasting of temperatures using SARIMA: An example from Nanjing," IOP Conf. Series: Materials Science and Engineering 394 (2018) 052024 doi: 10.1088/1757-899X/394/5/052024.

[21] M. R. Karim, "Seasonal ARIMA for forecasting sea surface temperature of the north zone of the Bay of Bengal," Research & Reviews: Journal of Statistics 2 (2013) 23–31.

[22] A. Kumar, S. Puri, P. Selokar, Y. Talhar, R. Bhokre, and C. Chakole, "forecasting of temperature by using time series analysis," International Journal of Research and Analytical Reviews 7 (2020) 547–551.

[23] C. Zaiontz, "Real statistics using Excel. Effect Size for Chi-square test," 2019.

[24] B. George EP, G. M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. Time series analysis: forecasting and control. 2015.