American Review of Mathematics and Statistics
June 2023, Vol. 11, No. 1, pp. 17-21
ISSN: 2374-2348 (Print), 2374-2356 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/arms.v11n1a2
URL: https://doi.org/10.15640/arms.v11n1a2

Application of Time Series to Daily Highest and Lowest Temperatures of Bangkok in Thailand

Sarisa Ruktametakul[1], Sittipong Ruktametakul[2], Pornpis Yimprayoon[2*]

Abstract

The purpose of this research is to create the time series models for predicting the maximum and minimum temperatures in each day of Bangkok in Thailand. The maximum and minimum temperature data sets were the number of 365 days between January 1, 2022 and December 31, 2022 from the Thai Meteorology Department. The results for these data sets show that the most suitable model for the highest temperature forecasting model is

$$\hat{Z}_t = 0.5937\hat{Z}_{t-1} - 0.0089\hat{Z}_{t-2} + 0.1374\hat{Z}_{t-3} + 0.008234105$$

and the lowest temperature forecasting model is

$$\hat{Z}_t = 0.4227\hat{Z}_{t-1} + 0.1978\hat{Z}_{t-2} + 0.014872327.$$

Keywords: Bangkok; forecasting; maximum temperature; minimum temperature; Thailand; time series models.

1. Introduction

Temperature is a comparative measure of any heat or cold source. The most popular unit of temperature scale is Celsius. At present the current climate is likely to change quickly, mainly caused by human behavior, such as deforestation, some types of farming, production of waste from industrial sector which make the concentrations of greenhouse gases in the atmosphere is increasing.

The amount of these gases is over an appropriate level. As a result, the global temperature rises to abnormal levels which contribute to climate change on Earth such as convection and its makes the ongoing phenomenon such as wind, clouds or rain. The difference of temperature in each area or time period may be caused by many factors such as sun rays, ground and surface water conditions, ocean currents, height of area, geographical location and the amount of cloud cause make the temperature varies.

From the problem of variability in temperature and climate bring to the attention of the researchers to find the perfect model for forecasting the weather. Also, the reason that we chose Bangkok is a case of our study because Bangkok is the capital and the most populous city in Thailand. The city's main economic of Thailand and the source of many cultural Gross domestic product by 25 percent from Bangkok. That is, it comes from the retail and wholesale of 24.31 percent, industry of 21.23 percent, transport and communications industry of 13.89 percent and hotels and restaurants of 9.00 percent by this research was to create the statistical models for predicting the daily maximum and minimum temperatures of Bangkok in Thailand by using simple time series analysis. To bring benefits for the plan in the future and allows residents to be prepared for prevent and solve problems or losses caused by meteorological phenomena. So be aware that the climate has changed in the future will be very useful.

The data in this analysis is the secondary data which was collected directly from the Thai Department of Meteorology such as the maximum and minimum temperature data sets were the number of 365 days between January 1, 2022 and December 31, 2022 by using the time series analysis for creating the forecasted statistical models. On this scale temperature is measured in degree celsius (°C). From the plots of time series for the lowest and highest temperature in each day of Bangkok, they indicate that the series are relatively jagged and are special cases of stationary time series in the mean. So the time series approach is proposed to forecast these data.

[1] Demonstration School of Nakhon Pathom Rajabhat University, Nakhonpathom 73000, Thailand
[2] Department of Computational Science and Digital Technology, Faculty of Liberal Arts and Science, Kasetsart University, Kamphaeng Saen Campus, Nakhonpathom 73140, Thailand
*E-mail: faasppy@ku.ac.th

In the context of time series analysis, many similar studies can be found in Yule (1927), Bartlett (1950), Whittle (1954), Brillinger and Rosenblatt (1967), Helmer and Johansson (1977), Rose (1977), Schwartz (1978), Tiao and Box (1981), Hughes et al. (2007), and others.

2. Methodology

In time series analysis, there are two useful representations to express a time series process. One is to write a process Z_t as a linear combination of a sequence of uncorrelated random variables, i.e.,

$$Z_t = \mu + \sum_{j=0}^{\infty} \varphi_j a_{t-j},$$

where $\varphi_0 = 1$, $\{a_t\}$ is a zero mean white noise process, and $\sum_{j=0}^{\infty} \varphi_j^2 < \infty$

Another useful is to write a process Z_t in an autoregressive (AR) representation, in which we regress the value of Zat time t on its own past values plus a random shock, i.e.,

$$\hat{Z}_t - \mu = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + a_t$$

Yule (1927) used an AR process to describe the phenomena of sunspot numbers and the behavior of a simple pendulum. Further, the AR representation is useful in understanding the mechanism of forecasting. In general, we choose a simpler model to describe the phenomenon. This is the principle of parsimony in model building recommended by Tukey (1967) and Box and Jenkins (1976). In the following topic sentence, we discuss some useful time series models and their properties.

2.1 *The General pth Order Autoregressive AR(p) Process*

The *p*th order autoregressive process AR(*p*) is

$$Z_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \cdots + \emptyset_p Z_{t-p} + \varepsilon_t$$

where

$$\varepsilon_t = a_t + \mu - \emptyset_1 \mu - \emptyset_2 \mu - \cdots - \emptyset_p \mu$$

$\{a_t\}$ is a zero mean white noise processwith constant variance σ^2

and μ is the mean of a stationary process $\{Z_t\}$.

Hence, we have the following recursive relationship for the autocorrelation function (ACF):

$$\rho_k = \emptyset_1 \rho_{k-1} + \emptyset_2 \rho_{k-2} + \cdots + \emptyset_p \rho_{k-p}, \qquad k > 0.$$

Then, we can easily see that when $k > p$ the last column of the matrix in the numerator of the partial autocorrelation function (PACF)

$$\emptyset_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

can be written as a linear combination of previous columns of the same matrix. So, the PACF $\emptyset_{kk}$ will vanish after lag *p*.

2.2 *The General qth Order Moving Average MA(q) Process*

The general *q*th order moving average process or model of order *q* is given by

$$Z_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} + \mu$$

where $\{a_t\}$ is a zero mean white noise process with constant variance σ^2

and μ is the mean of a stationary process $\{Z_t\}$.

Hence, the ACF is

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1,2, \ldots q, \\ 0, & k > q, \end{cases}$$

and the autocorrelation function of an MA(*q*) process cuts off after lag *q*. This important property enables us to identity whether a given time series is generated by a moving average process.

3. Results

To provide information become time series are stationary. We can choose the data transformation only one form only suitable for the analysis of the most. We calculated the following preliminary residual sum of squares $(Z_t^* - \bar{Z}_t)^2$ for various values of the power transformation parameter as shown in Tables 1, where $\bar{Z}_t$ is the corresponding sample mean of the transformed series. These calculations suggest that a $1/Z_t$ transformation be applied to the highest and lowest temperature data.

Table 1: Results of the power transformation on the lowest and highest temperature data

λ	Z_t^*	$(Z_t^* - \bar{Z}_t)^2$	
		Highest temperature	Lowest temperature
-1.0	$1/Z_t$	0.001631966	0.005277640
-0.5	$1/\sqrt{Z_t}$	0.013412261	0.031002087
0.0	$\ln Z_t$	1.771800122	2.943486286
0.5	$\sqrt{Z_t}$	14.694630984	17.639316232
1.0	Z_t	1958.498246575	1707.392219178

The sample ACF($\hat{\rho}_k$) and PACF($\hat{\phi}_k$) are calculated for $1/Z_t$ transformed data from a series of 365 real values as shown in Tables 2 and 3, respectively.

Table 2: Sample ACF for the highest and lowest temperature data

$\hat{\rho}_k$	Highest temperature	Lowest temperature
$\hat{\rho}_1$	0.687548294	0.626822909
$\hat{\rho}_2$	0.530810548	0.535650805
$\hat{\rho}_3$	0.476215215	0.452160369
$\hat{\rho}_4$	0.397916380	0.419384713
$\hat{\rho}_5$	0.433933352	0.410215720

Table 3: Sample PACF for the highest and lowest temperature data

$\hat{\phi}_{kk}$	Highest temperature	Lowest temperature
$\hat{\phi}_{11}$	0.687548294	0.626822909
$\hat{\phi}_{22}$	0.110165725	0.235126803
$\hat{\phi}_{33}$	0.145367850	0.083631276
$\hat{\phi}_{44}$	0.005837403	0.095873715
$\hat{\phi}_{55}$	0.211464277	0.105874976

To evaluate the significance of each parameter for the highest and lowest temperature data, we calculate the values of $\frac{\hat{\phi}_{kk}}{S_{\hat{\phi}_{kk}}}$ where $S_{\hat{\phi}_{kk}}$ is the standard error of the sample PACF, as shown in Table 4.

Table 4: Significance of each parameter of the highest and lowest temperature data

$t_k = \frac{\hat{\phi}_{kk}}{S_{\hat{\phi}_{kk}}}$	Highest temperature	Lowest temperature
t_1	13.135591720	11.975434870
t_2	2.104713228	4.492091263
t_3	2.777248878	1.597773281
t_4	0.111523432	1.831664747
t_5	4.040019348	2.022738585

From Table 4, the results for the highest temperature data indicate that the values of t_1 until t_3 are significant because they are greater than 1.96 at the significance level $\alpha = 0.05$, so the model of prediction for the series is AR3. While for the lowest temperature data, the results show that the values of t_1 and t_2 are significant, thus we get that the AR2 model fitting is adequate for the data. Then, we obtain the moment estimators for $\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ as

$$\begin{bmatrix} \widehat{\emptyset}_1 \\ \widehat{\emptyset}_2 \\ \vdots \\ \widehat{\emptyset}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{n-2} & \hat{\rho}_{n-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{n-3} & \hat{\rho}_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{\rho}_{n-1} & \hat{\rho}_{n-2} & \hat{\rho}_{n-3} & \cdots & \hat{\rho}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_n \end{bmatrix}.$$

Usually these estimators are called Yule-Walker estimators. This leads to the values of for $\widehat{\emptyset}_1, \widehat{\emptyset}_2, \ldots, \widehat{\emptyset}_p$ as shown in Table 5.

Table 5: Moment estimators of the highest and lowest temperature data

Yule-Walker estimator	Highest temperature	Lowest temperature
$\widehat{\emptyset}_1$	0.5937	0.4227
$\widehat{\emptyset}_2$	-0.0089	0.1978
$\widehat{\emptyset}_3$	0.1374	-

4. Conclusion

In summary, the AR(*p*) process has its autocorrelation tailing off and partial autocorrelations cutting off, but the MA(*q*) process has its autocorrelations cutting off and partial autocorrelations tailing off. Moreover, the AR processes are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock (a_t). So for μ is mean of data, the AR2 model is given by

$$a_t = (1 - \emptyset_1 B - \emptyset_2 B^2)(Z_t^* - \mu)$$

and

$$\hat{Z}_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \varepsilon_t$$

where

$$\varepsilon_t = a_t + \mu - \emptyset_1 \mu - \emptyset_2 \mu.$$

The AR3 model is defined as

$$a_t = (1 - \emptyset_1 B - \emptyset_2 B^2 - \emptyset_3 B^3)(Z_t^* - \mu)$$

and

$$\hat{Z}_t = \emptyset_1 Z_{t-1} + \emptyset_2 Z_{t-2} + \emptyset_3 Z_{t-3} + \varepsilon_t$$

with

$$\varepsilon_t = a_t + \mu - \emptyset_1 \mu - \emptyset_2 \mu - \emptyset_3 \mu.$$

Finally, based on the results of analyses, these imply that the fitted AR3 model could a good model for the original series of the highest temperature data. Parameter estimation gives

$$\hat{Z}_t = 0.5937 \hat{Z}_{t-1} - 0.0089 \hat{Z}_{t-2} + 0.1374 \hat{Z}_{t-3} + 0.008234105.$$

And for the lowest temperature data, we estimated the AR2 model identified for series and obtained the following estimation of this model:

$$\hat{Z}_t = 0.4227 \hat{Z}_{t-1} + 0.1978 \hat{Z}_{t-2} + 0.014872327.$$

Furthermore, we also calculate the mean square error (MSE) and mean absolute percentage error (MAPE) of the two models for the highest and lowest temperature data. The values of MSE and MAPE for these models are obtained by substituting the actual and forecast value into equations $\frac{1}{n}\sum_{t=1}^{n}\left(Z_t - \hat{Z}_t\right)^2$ and $\left(\frac{1}{n}\sum_{t=1}^{n}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right|\right) 100\%$, respectively. After we replace these values into this equation, the values of MSE for the highest and lowest temperature data are equal to 0.000002267 and 0.000004332, respectively. In addition, the values of MAPE for both data are equal to 3.59% and 3.71%, respectively, which are rather small values for the forecast accuracy.

Thus, they confirm that the fitted AR3 and AR2 model are adequate for the lowest and highest temperature in each day of Bangkok, respectively.

References

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika* 37:1–16.

Box, G. E. P. & Jenkins, G. M. (1976). *Time series analysis forecasting and control (2ndded.)*. San Francisco: Holden-Day.

Brillinger, D. R., Guckenheimer, J. & Guttorp, P. E., Oster, G. (1980). Empirical modeling of population time series data: The case of age and density dependent vital. *Lectures on Mathematics in the life Science (American Mathematical Society)* 13:65–90.

Helmer, R. M. & Johansson, J. K. (1977). An exposition of the Box-Jenkins transfer function analysis with an application to the advertising-sales relationship. *Journal of Marketing Research* 14:227–239.

Hughes, G. L.*, Rao, S. S.* & *Rao, T. S. (2007). Statistical analysis* and *time-series models* for *minimum/maximum temperatures* in the *Antarctic Peninsula. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 463:241–259.

Rangkakulnuwat, P. (2013). *Time series analysis for economics and business*. Bangkok: Chulalongkorn University Press.

Rose, D. E. (1977). Forecasting aggregates of independent ARIMA process. *Journal of Econometrics* 5:323–346.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6:461–464.

Tiao, G. C. & Box., G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association* 76:802–816.

Tukey, J. W. (1967). *An introduction to the calculations of numerical spectrum analysis*. New York: John Wiley.

William, W.S. Wei. (1989). *Time Series Analysis*. United Kingdom: Addison-Wesley.

Whittle, P. (1954). A statistical investigation of sunspot observations with special reference to H. Alfven's sunspot model. *Astrophysical Journal* 120:251–260.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot number. *Philosophical Transactions of the Royal Society of London (Series A)* 226:267–298.