Journal of Earth Sciences and Geotechnical Engineering, Vol.xx, No.xx, 2022, xx-xx
ISSN: 1792-9040 (print version), 1792-9660 (online)
https://doi.org/10.47260/jesge/
Scientific Press International Limited

Modeling and forecasting air temperature in Tetouan (Morocco) using SARIMA model.

Soufiane Ayad[1]

Abstract

The past decades have seen a growing concern to understand the impact of climate change at global and regional levels. In particular, air temperature has been considered as a key factor in climate impact studies on agricultural, ecological, environmental and economic sectors.
In this study, a seasonal ARIMA model is developed through the use of the Box and Jenkins method (1970) to predict the long-term air temperature in the city of Tetouan. Indeed, over the period of 1980 to 2022 from Sania Ramel station of the city of Tetouan, the monthly mean air temperature data are used to build and verify the model.
Four basic chronological steps, namely: Identification, Estimation, Validation and Prediction are established during the model development. The validity of the model is tested using the standardized residuals plots given by Box and Jenkins.
After carrying out the necessary checks, the ARIMA(1,0,0)(1,1,0)[12] model proved to be the most effective for predicting future air temperature.

Keywords: Air temperature, Time Series Forecasting, Box and Jenkins, Seasonal ARIMA.

[1] Natural Environments & Spatial Planning Research Group (University Abdelmalek Essaadi).
e-mail: s.ayad@uae.ac.ma

Article Info: *Received: ….. ..,* 2022. *Revised: ….. .., 2022.*
Published online: ….. .., 2022.

1. Introduction

Air temperature is a common weather variable that indicates how warm or cold the air is. It does not only affect the growth and reproduction of plants and animals, but also influences meteorological variables, such as relative humidity, evaporation, wind speed, and precipitation, as well as its direct impact on the wellbeing of the humans' live. Therefore, there is a need to forecast air temperature accurately, in order to prevent unexpected calamities caused by air temperature variation, such as drought, frost, snow melting, and wildfires which may cause financial and human losses. Furthermore, the accurate prediction of air temperature plays an important role in establishing a plan for energy policy, and business development.

In this article, we will analyze the monthly average air temperature of the city of Tetouan, and propose an adequate seasonal ARIMA model to describe the phenomenon.

2. Methodological Approach

2.1 Study Location and Data Collection

Tetouan's climate is Mediterranean, with short, sunny, warm, humid, and mostly clear summers and long, cool, rainy and partly cloudy winters. The city is located in the north of Morocco (Latitude: 35.58 | Longitude: -5.33 | Altitude: 10), near the coast of the sea (Figure 1).

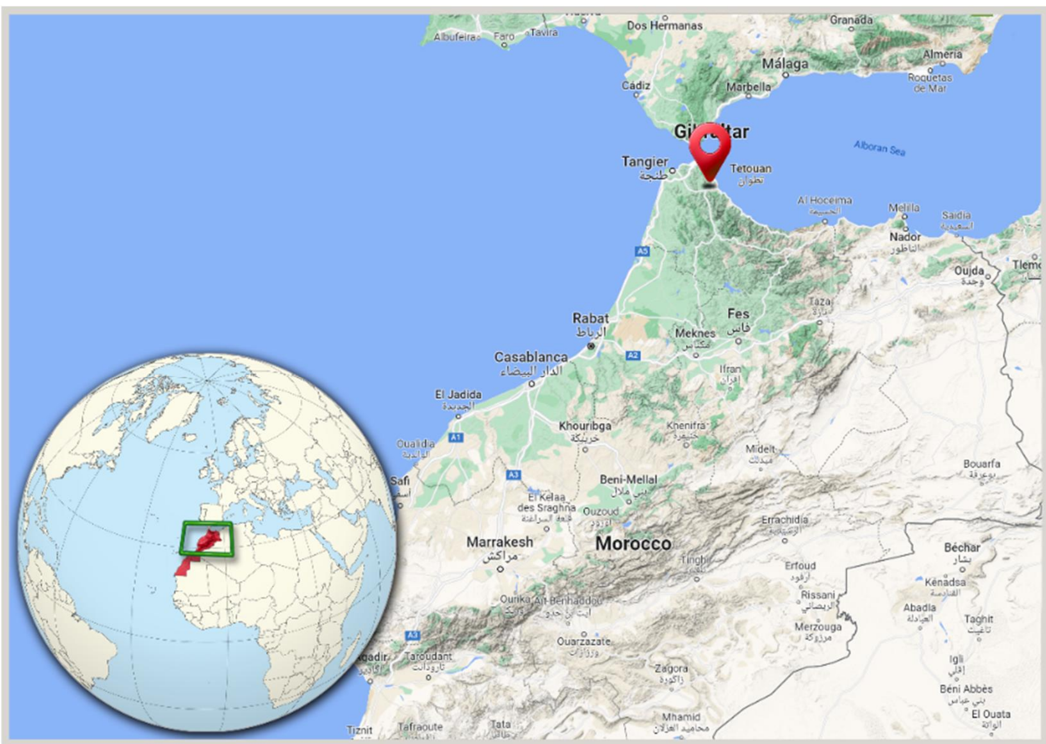

Figure 01: Location of the City of Tetouan ("*Google Maps*").

The data used in this study are those of the average monthly air temperature collected daily from the Sania Ramel station over the period from January 1980 to July 2022. According to our data, the average air temperature of the coldest month (January) is 13.3°C, and that of the hottest month (August) is 25.56°C, with an average monthly air temperature of 18.8°C and a standard deviation of 4.38 (Figure 2).

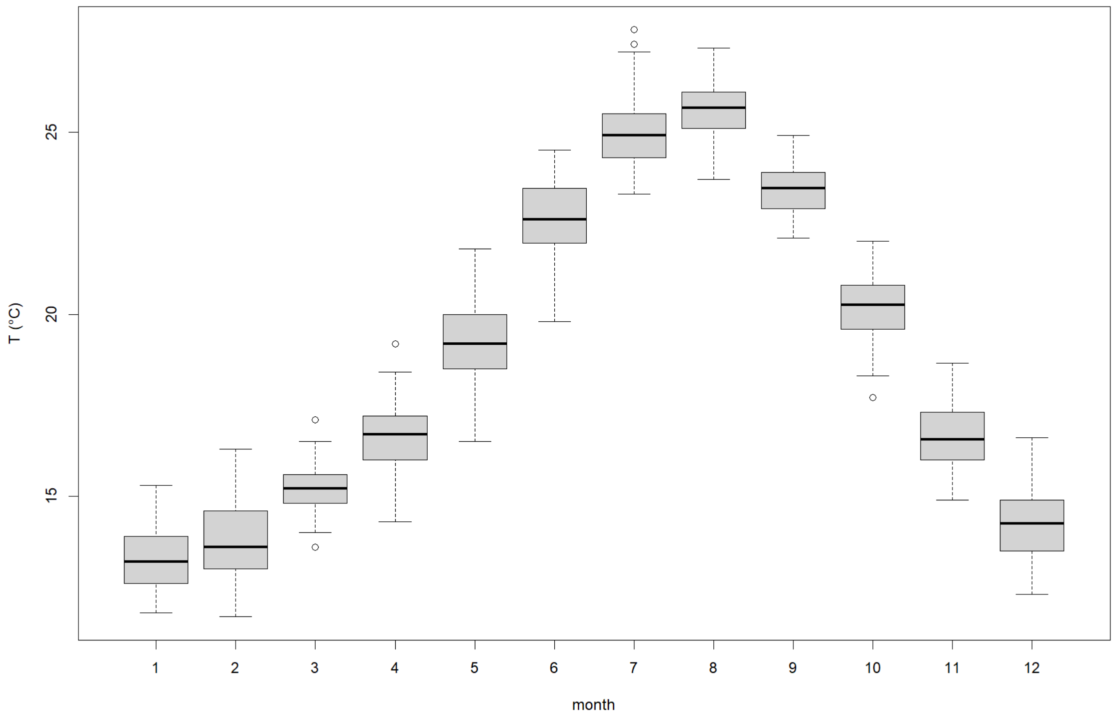

Figure 02: Box-Plots of the monthly average air temperature.

2.2 Methodology

To develop a seasonal ARIMA model, four basic steps must be followed: identification, estimation, validation (Diagnostic test) and prediction (Bari et al., 2015).

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors. A shorthand notation for the model is ARIMA(p, d, q) × (P, D, Q) [s], where “p” and “q” are non-negative integers that correspond to the order of autoregressive (AR) and moving average (MA) process respectively; whereas “d” stands for the order of the non-seasonal differencing (I), “s” stands for the period of repetition of the seasonal pattern and the parameters “P”, “D” and “Q” are the seasonal autoregressive parameter, the seasonal integrated parameter and the seasonal moving average parameter respectively. (Box et Al., 2015)

The transformed time series Y_t, seasonal ARIMA (p, d, q)(P, D, Q)[s], model may be written:

$$\phi_p(L)\ \Phi_P(L^s)\ (1-L)^d\ (1-L^s)^D Y_t = \theta_q(L)\ \Theta_Q(L^s)\ \varepsilon_t \qquad (1)$$

where,

- $\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$
- $\Phi_P(L^s) = 1 - \Phi_s L^s - \Phi_{2\times s} L^{2\times s} - \cdots - \Phi_{P\times s} L^{P\times s}$
- $\theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$
- $\Theta_Q(L^s) = 1 + \Theta_s L^s + \Theta_{2\times s} L^{2\times s} + \cdots + \Theta_{Q\times s} L^{Q\times s}$
- ε_t is an uncorrelated random variable with mean zero and constant variance.
- L is the BackShift or Lag operator.

The following algorithm illustrates the basic methodology of the seasonal ARIMA model development (Bisgaard and Kulahci, 2011):

1. Plot the series.
2. Is the variance stable?
   a. No, apply Box-Cox transformation, go to 1.
   b. Yes, continue.
3. Obtain ACF and PACF.
4. Is the series stationary?
   a. No, apply regular and/or seasonal differencing.
   b. Yes, continue.
5. Model selection.
6. Estimate parameter values.
7. Are the residuals uncorrelated?
   a. No, modify the model, go to 5.
   b. Yes, continue.
8. Are the parameters significant and uncorrelated?
   a. No, modify the model, go to 5.
   b. Yes, continue.
9. Forecasts.

3. Modeling

3.1 Model Identification

The first step toward modeling the air temperature is to check if there is seasonality in the observed data and if the data is stationary at constant variance.

The monthly mean air temperature time series plot (Figure 03) shows that there is a clear seasonality with a periodicity of one year (12 months) in the data set. This observation is supported by the ACF graph (Figure 04), which leads us to consider a seasonal differencing of order 1 (i.e., D=1). The ACF and PACF of the seasonal differenced data are shown in Figure 5.

For performing ARIMA modeling, time series should be stationary that bears a quasinormal distribution with zero mean and a constant variance.

The time series plot (Figure 3) doesn't show the presence of a trend in our data, and there is no evidence of changing variance.

The stationarity tests, in particular that of the Augmented Dickey-Fuller (ADF) (p-value < 0,0001) (Dickey and Fuller, 1981), and that of Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (p-value = 0,948) (Kwiatkowski et Al., 1992), make it possible to decide on the form of the stationarity of the series. Indeed, the tests clearly reject the hypothesis that the series is non stationary.

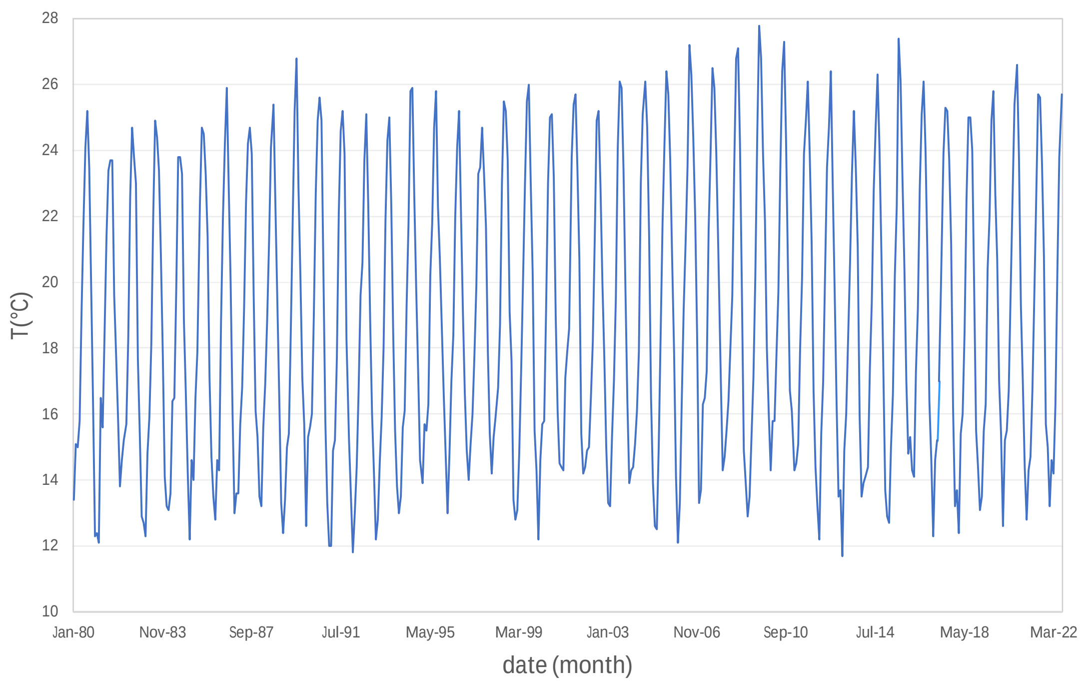

Figure 3: Time series plot of the observed data.

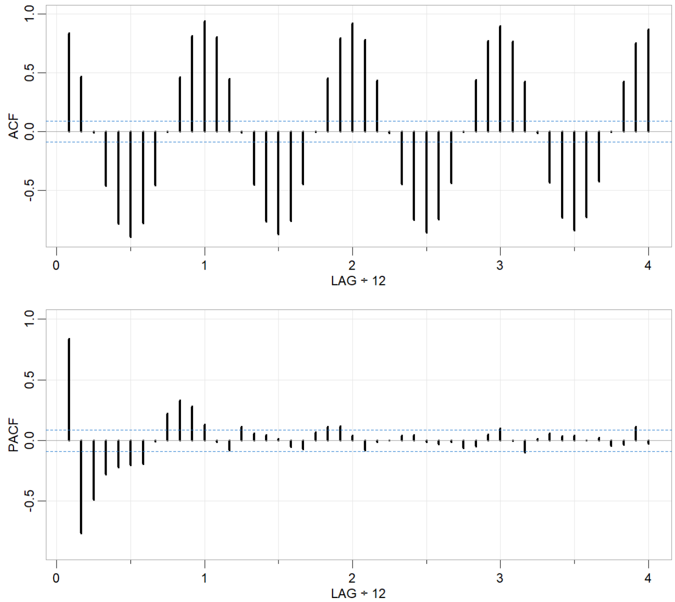

Figure 04: Time series ACF and PACF.

3.2 Model Estimation.

After seasonal differencing of order 1 (i.e., D=1), the ACF and the PACF of the new series (Figure 05) lead us to consider the model: ARIMA(1,0,0)(1,1,0)[12].
The general equation of the selected model can be expressed as (Hyndman, 2018):

$$(1-\phi_1 L)(1-\Phi_{12}L^{12})(1-L^{12})Y_t = \varepsilon_t$$

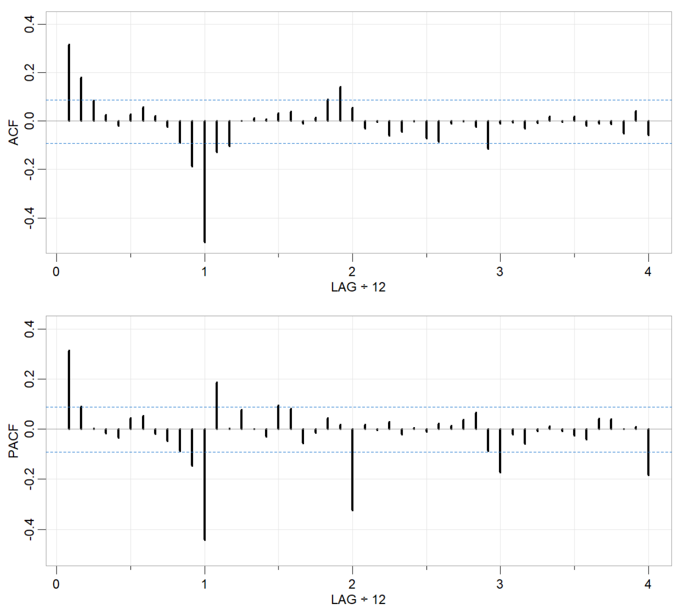

Figure 05: ACF and PACF of the series after seasonal differencing of order 1.

The coefficients of the considered model are as follows:

Table 01: Estimation Results.

Parameter	Value	Standard Error	t Statistic	P-Value
Constant	0	0		
AR{1}	0.317930	0.042375	7.5027	6.252e-14
SAR{12}	-0.518254	0.038747	-13.3752	2.3045e-51
Variance	0.9419	0.055696	16.4121	1.5658e-60

Table 02: Goodness of Fit.

AIC	AICc	BIC
1394.06	1394.11	1406.70

The estimated parameters of the selected model (Table 01) are significantly different from 0 (the Student's t-test is applied).
Therefore, the general equation of the selected model is:

$$(1 - 0{,}318\, L)\,(1 + 0{,}518\, L^{12})\,(1 - L^{12})\, Y_t = \varepsilon_t$$

where 'L', the BackShift operator, defined by:

$$L^m\, Y_t = Y_t - Y_{t-m}$$

and ε_t is white noise.

3.3 Model Validation

After estimating the parameters of this model, we assess their adequacy by analyzing the residuals.
Figure 06 suggests that the standardized residuals estimated from this model behave as an independent and identically distributed sequence with zero mean and constant variance.
The Q-Q plot (Figure 07) shows that the standardized residuals of the model approach a normal distribution. Moreover, the Shapiro-Wilk test gives no reason to reject the hypothesis that the distribution of the residuals is normal (p-value= 0.188).
From Figure 08, it is evident that autocorrelation and partial autocorrelation at different lags for the residuals lie within 95% confidence level, and, therefore, implying that the autocorrelation coefficients are statistically insignificant. From the above observation, we can deduce that the residuals are random (white noise) and not autocorrelated with each other.

Figure 09 shows the p-values for the Ljung-Box statistic (Ljung and Box, 1978). Given the high p-values associated with the statistics, we cannot reject the null hypothesis of independence in this residual series.
Therefore, we can say that the ARIMA(1,0,0)(1,1,0)[12] model fits the data well.
The predicted values taking into account the ARIMA(1,0,0)(1,1,0)[12] model are presented in Figure 10, where we compare these values with the observed values of air temperature in the city of Tetouan. The predicted values are relatively close to the observed values; this result indicates that the model provides an acceptable fit to predict the air temperature of the city of Tetouan.

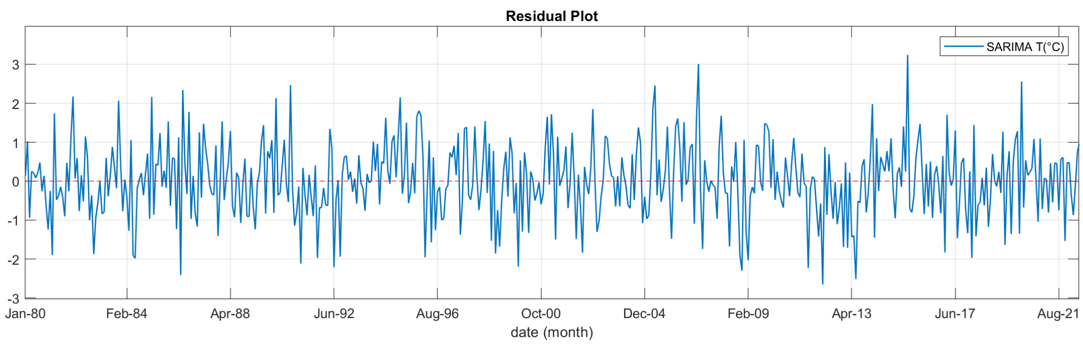

Figure 06: Residuals from ARIMA(1,0,0)(1,1,0)[12].

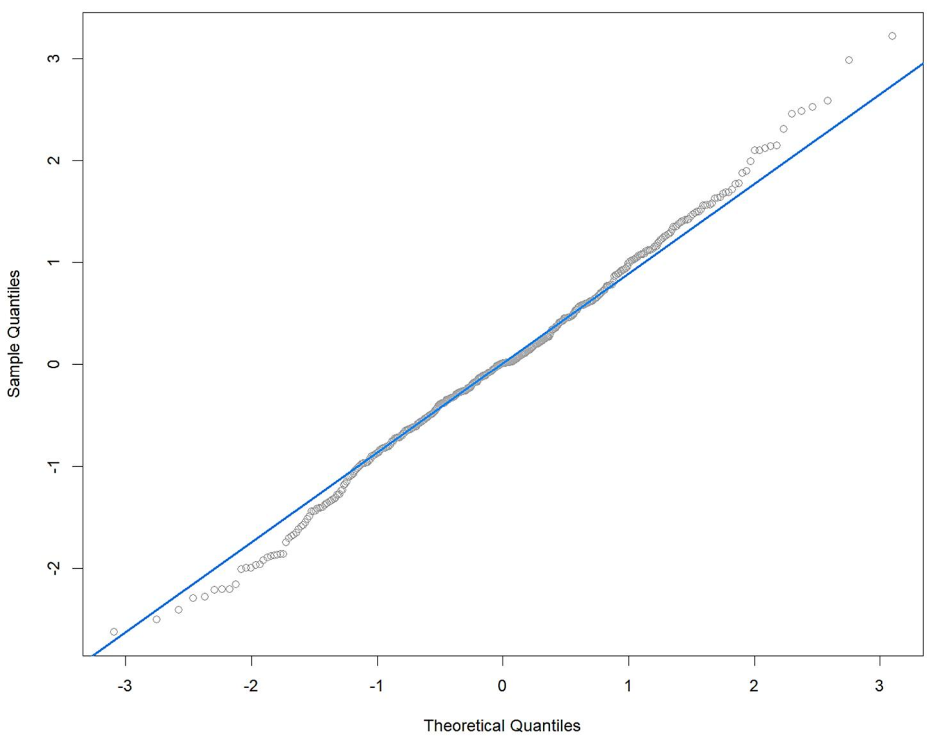

Figure 07: Normal Q-Q Plot, Residual.

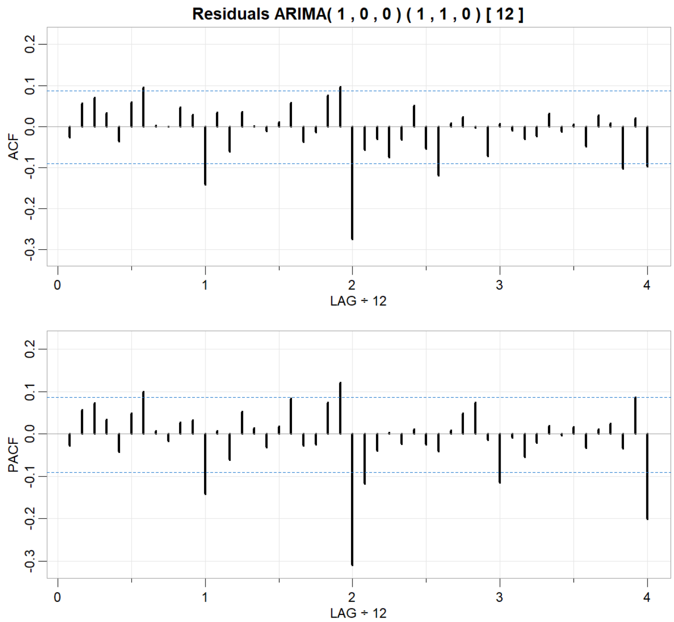

Figure 08: ACF and PACF residual plots.

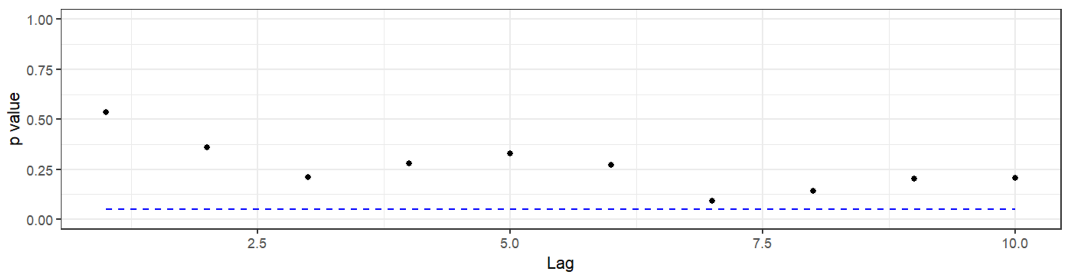

Figure 09: p-values for Ljung-Box statistic.

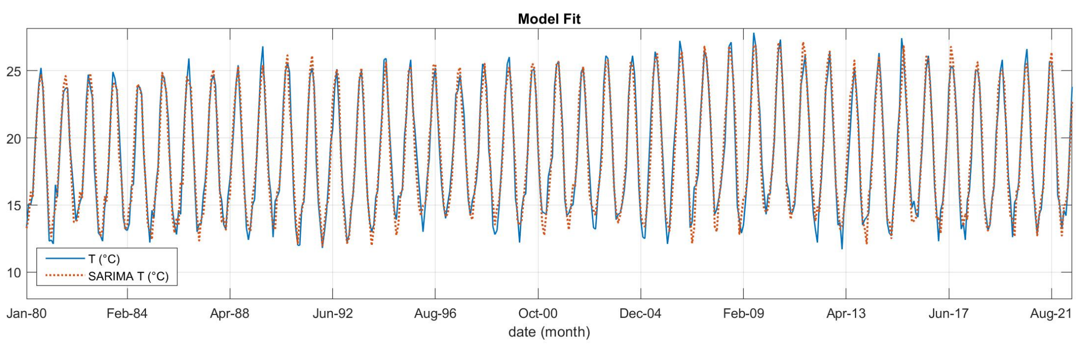

Figure 10: Observed Values vs Selected ARIMA Model Values.

3.4 Forecast

ARIMA(1,0,0)(1,1,0)[12] was applied to forecast monthly air temperature data from August 2022 to January 2024. The forecast time series and the observed time series with the error bound 80% and 95% confidence level are plotted (Figure 11). It is observed that the measured monthly values are within the error limit and that the predicted trajectory of the seasonal model fits reasonably well.

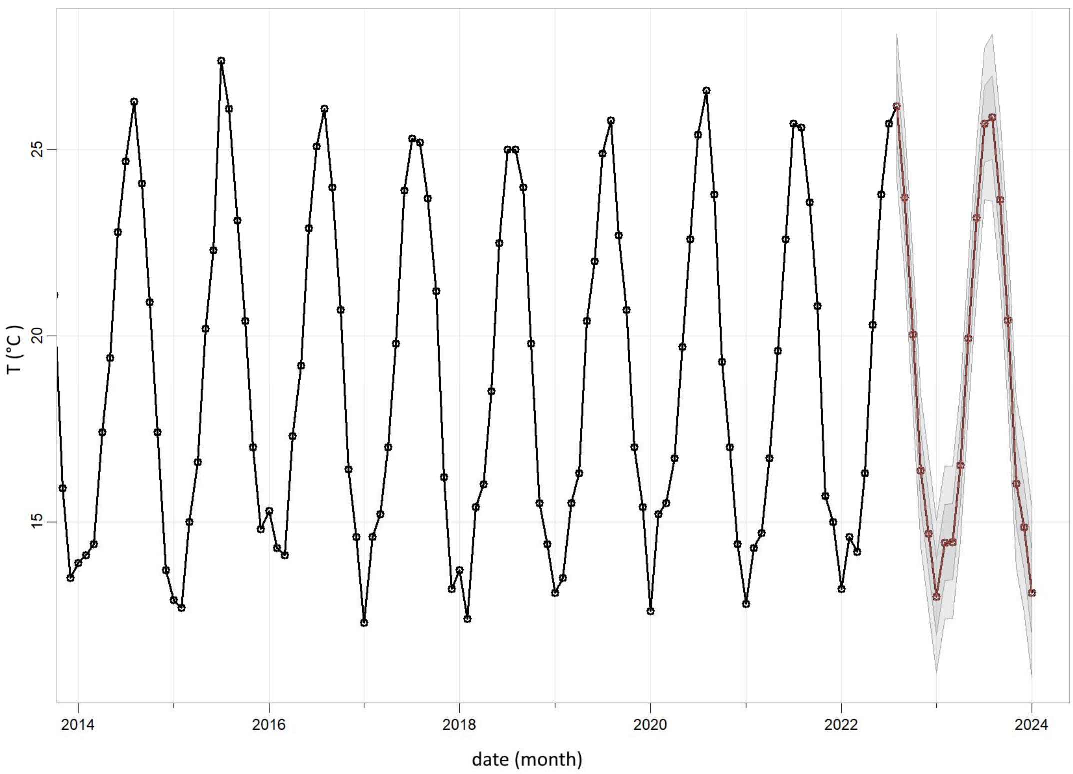

Figure 11: Observed and predicted data with 80% and 95% confidence limit.

Table 03: Predicted data values with 80% and 95% confidence limit.

date	Forecast	Lo.80	Hi.80	Lo.95	Hi.95
Aug 2022	26.17	24.92	27.41	24.27	28.07
Sep 2022	23.72	22.41	25.02	21.72	25.72
Oct 2022	20.03	18.72	21.34	18.02	22.03
Nov 2022	16.38	15.06	17.69	14.37	18.38
Dec 2022	14.69	13.38	16.00	12.68	16.70
Jan 2023	12.99	11.68	14.30	10.99	15.00
Feb 2023	14.44	13.13	15.76	12.44	16.45
Mar 2023	14.46	13.15	15.77	12.45	16.47
Apr 2023	16.51	15.20	17.82	14.50	18.51
May 2023	19.94	18.63	21.25	17.93	21.94
Jun 2023	23.18	21.87	24.49	21.17	25.18
Jul 2023	25.70	24.39	27.01	23.69	27.71
Aug 2023	25.87	24.43	27.32	23.67	28.08
Sep 2023	23.66	22.20	25.11	21.43	25.88
Oct 2023	20.43	18.97	21.88	18.20	22.65
Nov 2023	16.03	14.57	17.48	13.80	18.25
Dec 2023	14.85	13.39	16.31	12.62	17.08
Jan 2024	13.10	11.64	14.56	10.87	15.33

4. Conclusion

In this study, a seasonal ARIMA model for the air temperature in the city of Tetouan is developed. By comparing the observed and predicted values with a 95% confidence limit, the present model provides reasonable results. Therefore, the proposed model could help to determine a possible future strategy in the respective field for the city and its neighboring areas.

Références

[1] Bari, S.H., Rahman, M., Hussain, M. and Ray, S. (2015). Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model. Civil and Environmental Research, 7(1), pp.69 - 77.

[2] Bisgaard, S. and Kulahci, M. (2011). Time series analysis and forecasting by example. John Wiley & Sons Inc., New Jersey.

[3] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015). Time Series Analysis: Forecasting and Control. Fifth edition, Wiley.

[4] Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica: journal of the Econometric Society, 49(4), pp.1057-1072.

[5] Hyndman, R. J. and Athanasopoulos, G. (2018). Forecasting: Principles and Practice (3rd ed). OTexts. https://otexts.com/fpp3/

[6] Ljung G. M. and Box G. E. P. (1978). On a Measure of a Lack of Fit in Time Series Models. Biometrika. 65 (2), pp.297 - 303.

[7] Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of econometrics, 54, pp.159 - 178.