Journal of Earth Sciences and Geotechnical Engineering, Vol. 12, No. 1, 2022, 1-13
ISSN: 1792-9040 (print version), 1792-9660 (online)
https://doi.org/10.47260/jesge/1211
Scientific Press International Limited

Relative Humidity Forecasts in Tetouan (Morocco) with SARIMA model

Soufiane Ayad[1]

Abstract

In this study, a seasonal ARIMA model is built using the Box and Jenkins method (1970) to predict the long-term relative humidity in the city of Tetouan. To this end, the monthly average relative humidity data over the period of 1990 to 2022 from the Sania Ramel station are used to build and verify the model.
The methodology used in this work toward the development of the model includes five steps: exploratory analysis, model identification, parameter estimation, model validation and prediction.
The validity of the model is tested by using the standardized residual plots along with the appropriate statistical tests given by Box and Jenkins. In a second validation step, the predicted values of monthly relative humidity are verified through the use of real data series.
After carrying out the necessary checks, the *ARIMA(2,1,1)(2,0,0)[12]* model proved to be the most effective.

Keywords: Time Series, Seasonal ARIMA, Relative Humidity, Forecast.

[1] Natural Environments & Spatial Planning Research Group (University Abdelmalek Essaadi).

Article Info: *Received:* July 16, 2022. *Revised:* August 12, 2022.
Published online: August 22, 2022.

1. Introduction

The relative humidity, expressed as a percentage, is a common weather variable that indicates the amount of water vapor in the air that is defined as the ratio of the actual amount of water vapor present in a volume of air at a given temperature to the maximum amount that the air could hold at that temperature (Hanks, 2011). In other words, relative humidity is the ratio of how much water vapor is in the air and how much water vapor the air could potentially contain at a given temperature.

Relative humidity only considers the invisible water vapor, mists, clouds and fogs do not count towards the measure of relative humidity of the air. At 100% relative humidity, the air is saturated and, therefore, is at its dew point. The relative humidity can exceed 100%; a case in which the air is said to be supersaturated.

Relative humidity is an important metric that is used in weather forecasts and reports, as it is an indicator of the likelihood of precipitation, dew, or fog. It does not only affect the growth and reproduction of plants and animals, but also influences the human health and comfort.

In this study, we will analyze the monthly average relative humidity 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 mild rainy winter and hot sunny and humid summer. 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). Nearby are the Rif mountains, which can be covered in snow in winter. In the city, however, snow is extremely rare.

The data used in this study are those of the average monthly relative humidity collected daily from the Sania Ramel station over the period of January 1990 to May 2022. The missing values have been replaced by the arithmetic mean of the preceding and the following month of the missing data.

According to our data, the lowest relative humidity (48.8%) is recorded in the summer during the month of July and the highest (84.5%) is recorded in the winter during the month of October (Figure 2), with an annual average of 70.66% and a standard deviation of 6.28. The variation goes with the precipitation variation; it increases in the winter when high rainfall is recorded and decreases in the summer when the temperature increases.

2.2 Methodology

ARIMA (auto-regressive integrated moving average) forecasting methods were made popular by G. E. P. Box and G. M. Jenkins in the 1970s. These techniques, often referred to as Box-Jenkins forecasting methodology (Box & Jenkins, 1970), involve the following steps (Bari et al., 2015):

1. Identification and selection of the model,
2. Estimation of autoregressive (AR), integration or differentiation (I), and moving average (MA) parameters,
3. Verification and validation of the model,
4. Forecast.

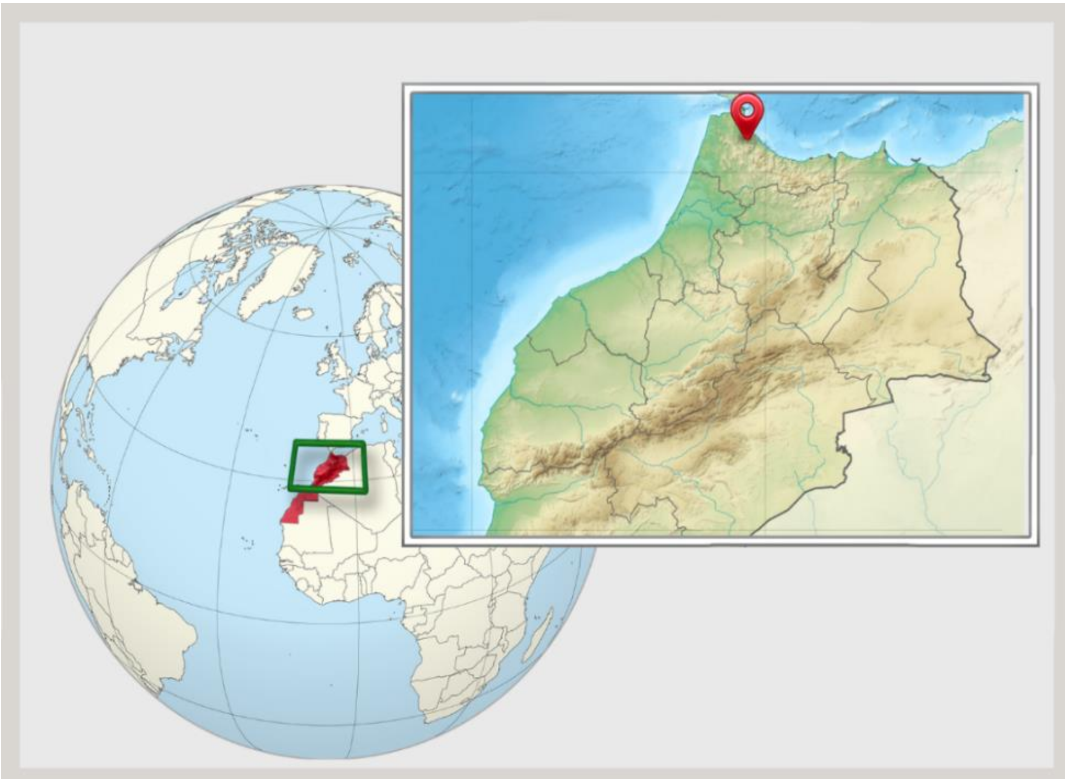

Figure 1: Location of the City of Tetouan.

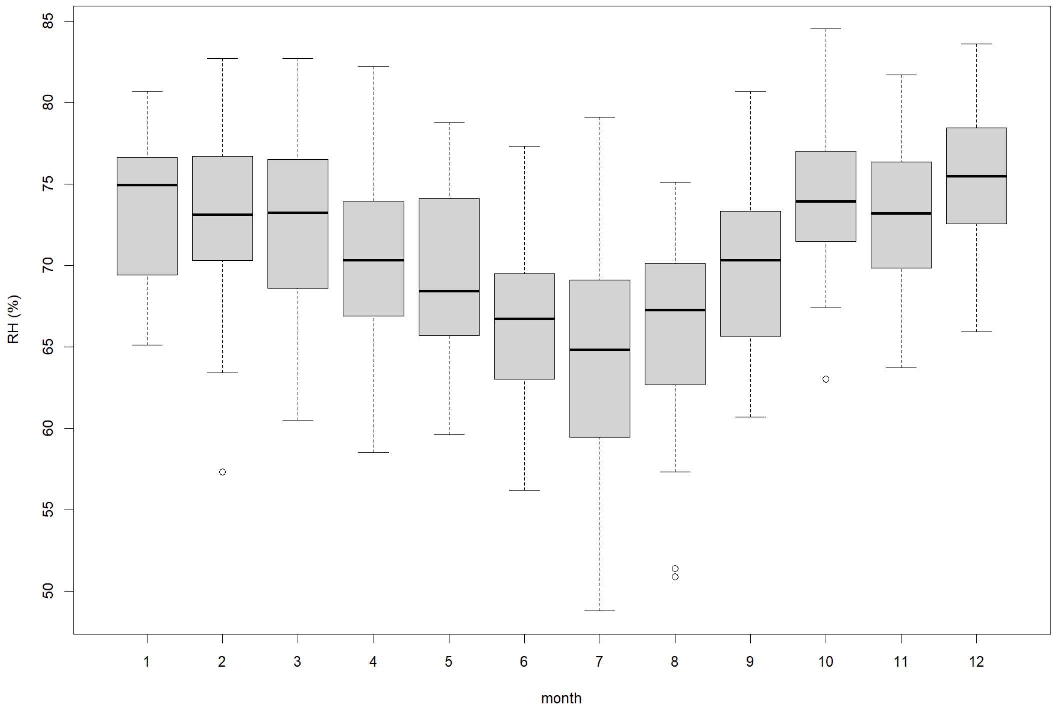

Figure 2: Box-Plots of the monthly average relative humidity.

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors (Box et al., 2015). A shorthand notation for the model is ARIMA(p, d, q)(P, D, Q)[s], with:

p : non-seasonal auto-regressive (AR) order,

d : non-seasonal differentiation,

q : non-seasonal moving average (MA) order,

P : seasonal autoregressive (AR) order,

D : seasonal differentiation,

Q : seasonal moving average (MA) order, and

s : period of repetition of the seasonal pattern.

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

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

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.

3. Modelisation

3.1 Model Identification

The monthly relative humidity time series plot (Figure 3) shows that there is seasonality with a periodicity of one year (12 months) in the data set. This observation is supported by the ACF graph (Figure 4).

For performing ARIMA modeling, time series should be stationary bearing a quasi-normal distribution with zero mean and a constant variance.

A decomposition of the series into its different components (Figure 5) shows the presence of a downward trend in our data, but there is no evidence of changing variance. The stationarity test, in particular that of the Augmented Dickey-Fuller (ADF), makes it possible to decide on the form of the non-stationarity of the series. Indeed, the test clearly rejects the hypothesis that the series is stationary (p-value = 0.183).

To address the non-stationarity (Bisgaard, 2011), we will take a first difference (d=1) of the data. The ACF and PACF of the differenced data are shown in Figure 6.

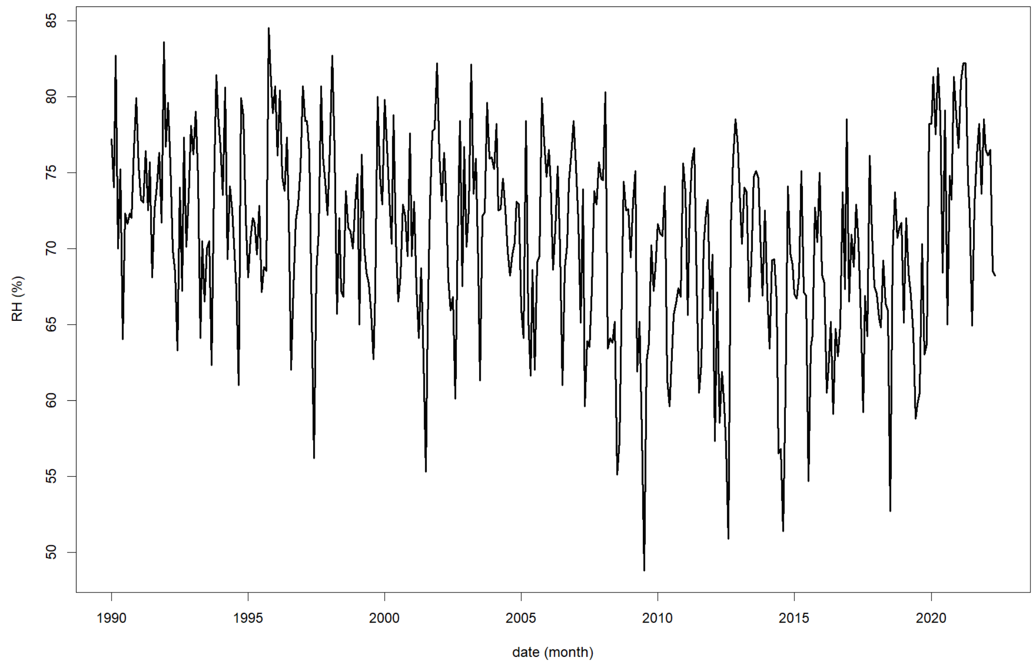

Figure 3: Time series plot of the observed data.

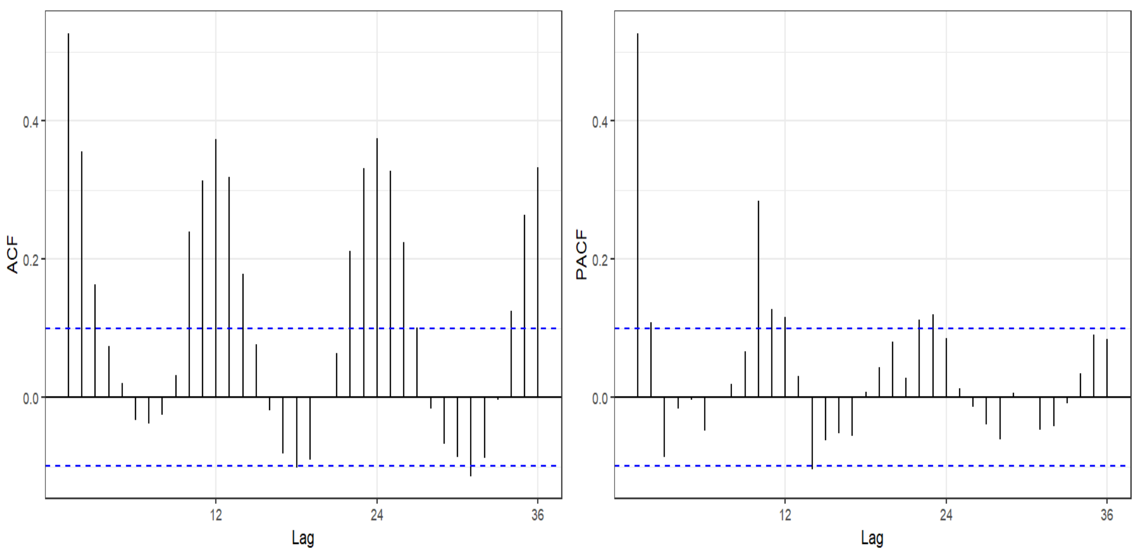

Figure 4: Time series ACF and PACF.

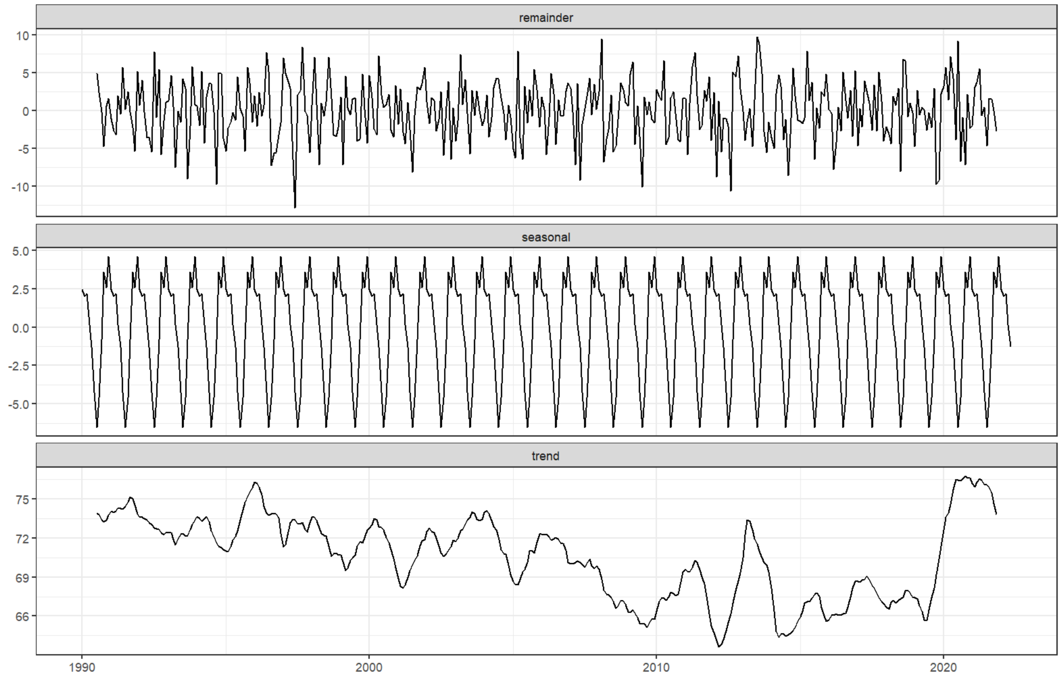

Figure 5: Graphical representation of the estimated states of the series over time.

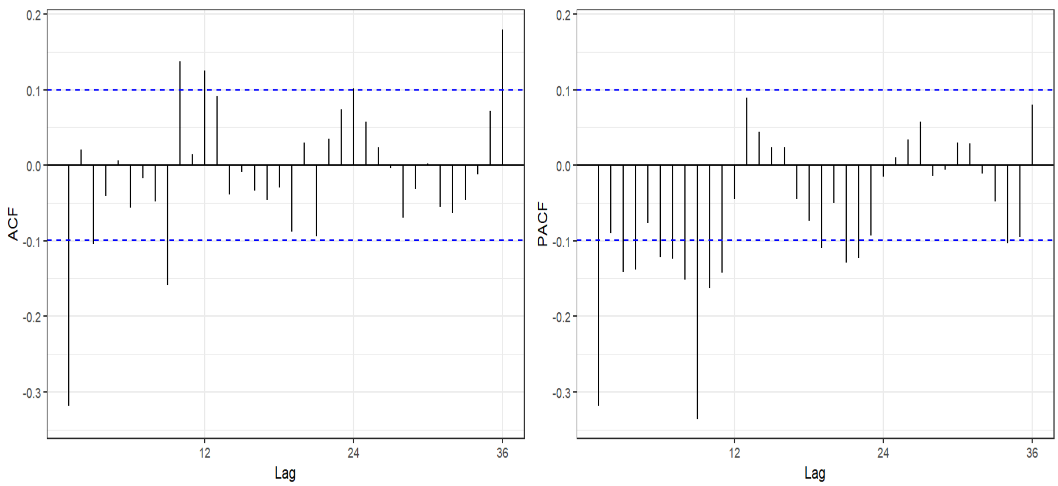

Figure 6: First differenced Time series ACF and PACF.

3.2 Model Estimation

Model parameters are approximated by incorporating the maximum likelihood procedure for which the best models were selected on the basis of statistical criterion namely: Akaike information criterion (AIC), corrected Akaike information criterion (AICc) and Bayesian information criterion (BIC). Table 1 illustrates six selected models amongst a number of combinations that did agree to the data with parameter estimations on the basis of the statistics criterion (Martinez et al., 2011); (Hyndman and Athanasopoulos, 2018).

Figure 7 shows autocorrelations and partial autocorrelations in terms of model residuals at different lags for the best six combinations selected on the basis of lowest AICc measures (Shadab et al., 2019).

From Figure 7, it is evident that autocorrelation and partial autocorrelation at different lags for the residuals lie within 95% confidence level, thus 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. Therefore, all the selected models can be considered as good fit.

ARIMA(2,1,1)(2,0,0)[12] with the lowest values of AICc (also lowest AIC and BIC) is selected as the best fit model that is used hereafter for predicting accurate time series forecasts.

The general equation of the ARIMA(2,1,1)(2,0,0)[12] model can be expressed as:

$$(1 - \phi_1 L - \phi_2 L^2)\,(1 - \Phi_{12} L^{12} - \Phi_{24} L^{24})\,(1 - L)\,Y_t = (1 + \theta_1 L)\,\varepsilon_t \qquad (2)$$

Table 1: Performance of selected ARIMA models.

ARIMA model	AICc
ARIMA(1,1,1)(2,0,0)[12]	2379.279
ARIMA(1,1,1)(0,0,2)[12]	2388.380
ARIMA(1,1,2)(2,0,0)[12]	2377.236
ARIMA(1,1,3)(2,0,0)[12]	2378.529
ARIMA(2,1,1)(2,0,0)[12]	*2376.540*
ARIMA(2,1,2)(2,0,0)[12]	2378.483

The estimated parameters of the selected model (Table 2) are significantly different from 0 (the Student's t-test is applied). If a coefficient is not significantly different from 0, consideration should be given to a new specification eliminating the invalid AR or MA model order.

Table 2: Model Parameters.

Parameter	Value	Standard Error	t-statistic	p-value
Constant	0	0		
AR{1}	0.32955	0.053076	6.209	5.3319e-10
AR{2}	0.12208	0.05159	2.3663	0.017968
SAR{12}	0.19029	0.049555	3.84	0.00012302
SAR{24}	0.18233	0.050828	3.5871	0.00033437
MA{1}	-0.96378	0.017413	-55.3492	0
Variance	26.1242	1.928	13.5499	7.9391e-42

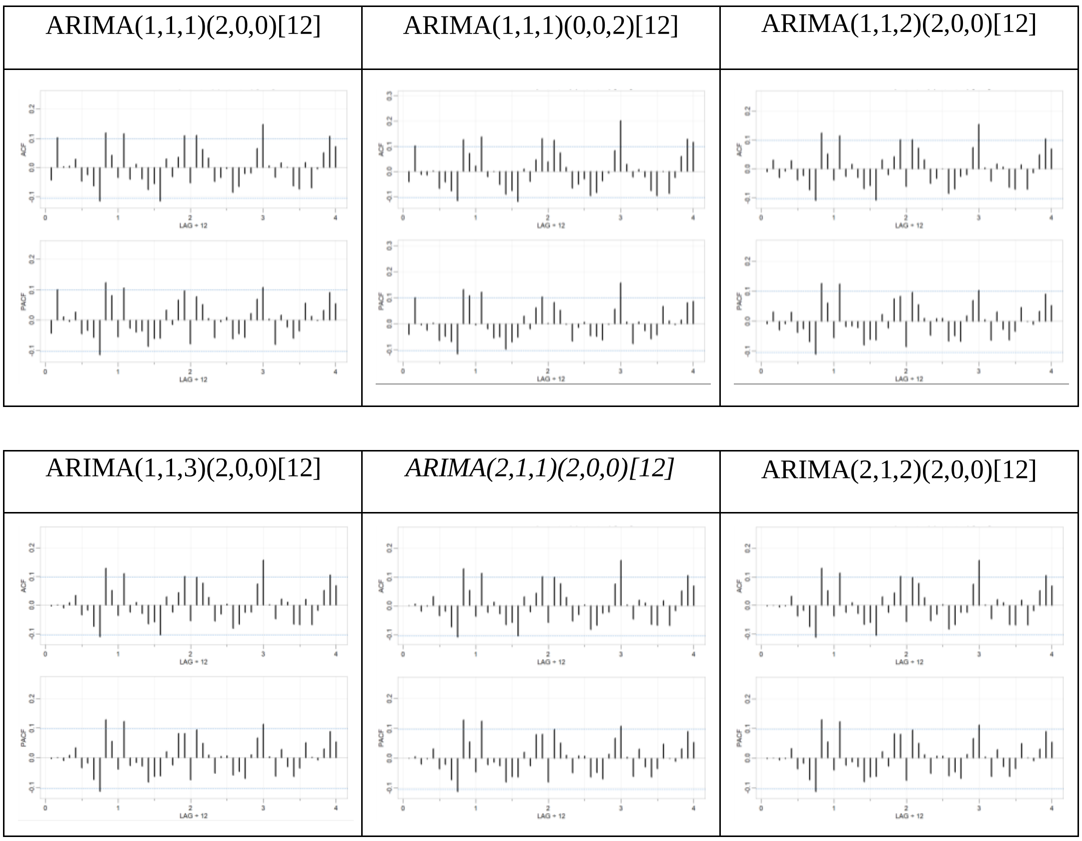

Figure 7: ACF and PACF residual plots for the selected models.

Table 3: Goodness of fit.

AIC	AICc	BIC
2376.32	2376.54	2400.09

Therefore, the general equation of the selected model is:

$$(1 - 0{,}33\ L - 0{,}122\ L^2)\ (1 - 0{,}19\ L^{12} - 0{,}182\ L^{24})\ (1 - L)\ Y_t = (1 - 0{,}963\ L)\ \varepsilon_t \quad (3)$$

3.3 Model Validation

The efficacy of the selected models can further be assessed by analyzing the residuals. Figure 8 shows the standardized residuals, their histogram, the respective ACF plot and the p-values for the Ljung-Box statistic.

Graph (a) suggests that the standardized residuals estimated from this model behave as an independent and identically distributed sequence with zero mean and constant variance.

The histogram (b) 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 = 0.1325).

The ACF of the residuals (c) suggests that the autocorrelations are close to zero. This result means that the residuals do not deviate significantly from a zero-mean white noise process.

Graph (d) shows the p-values for the Ljung-Box statistic. 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(2,1,1)(2,0,0)[12] model fits the data well.
The predicted values taking into account the ARIMA(2,1,1)(2,0,0)[12] model are presented in Figure 9, where we compare these values with the observed number of relative humidity 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 relative humidity of the city of Tetouan.

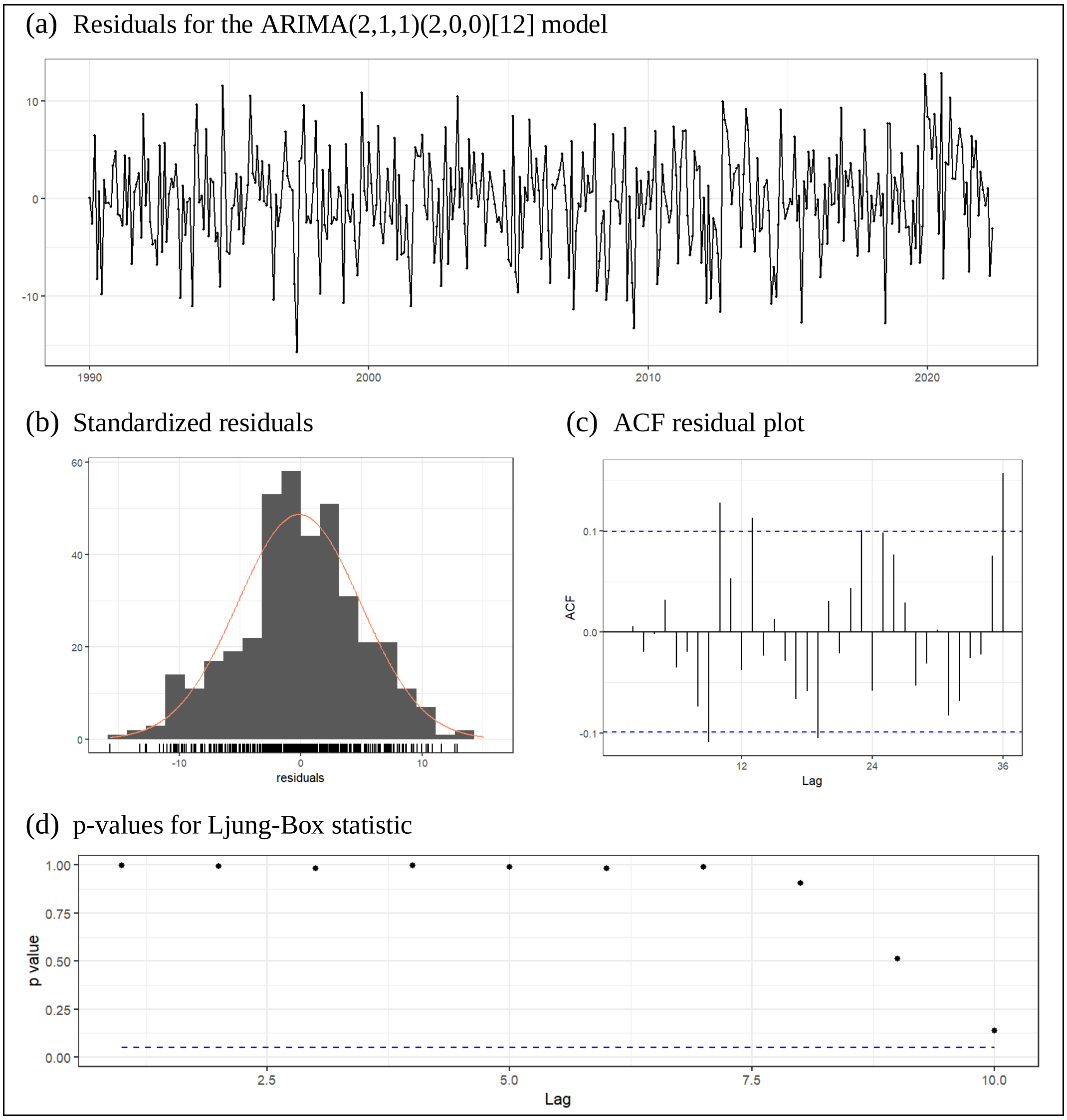

Figure 8: Graphical diagnostics to assess model fit.

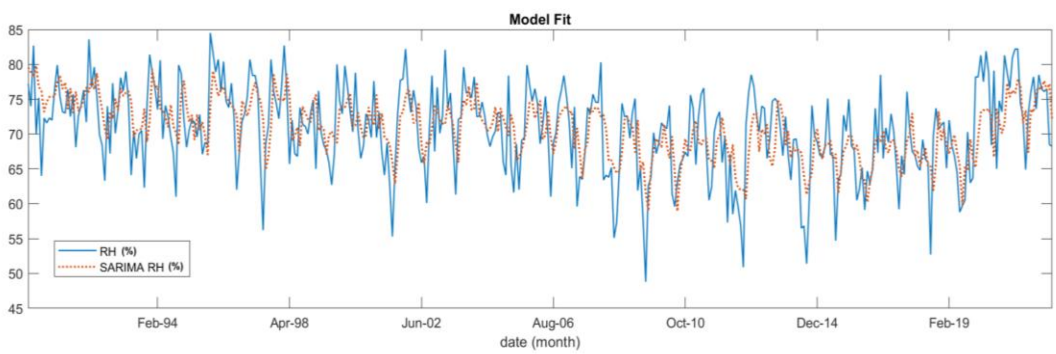

Figure 9: Observed Values vs Selected ARIMA Model Values.

3.4 Forecasts

ARIMA(2,1,1)(2,0,0)[12] was applied to forecast monthly relative humidity data from June 2022 to December 2023. The forecast time series and the observed time series with the error bound 80% and 95% confidence level are plotted (Figure 10).

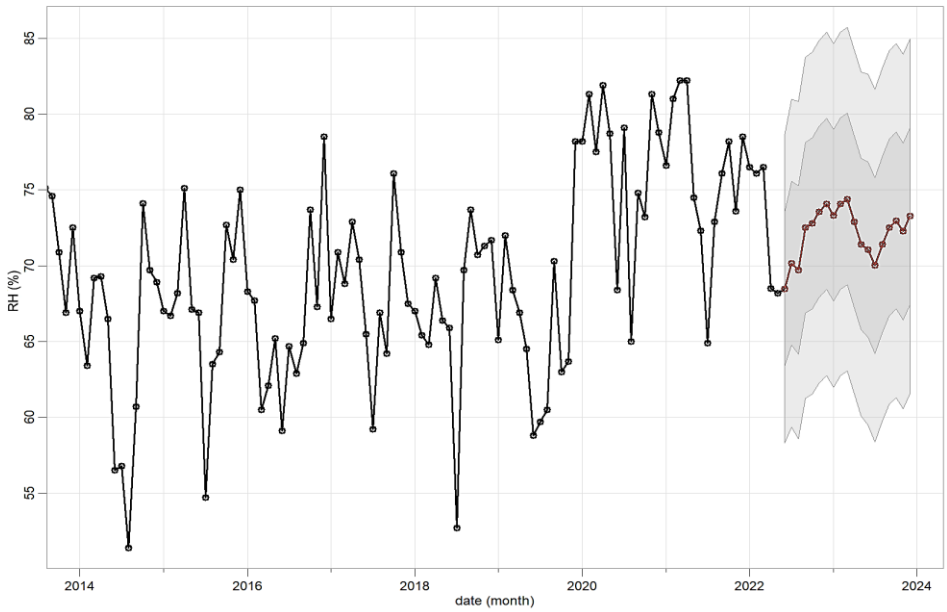

Figure 10: Observed and predicted data with 80% and 95% confidence limit.

Table 4: Predicted data values with 80% and 95% confidence limit.

Date	Forecast	Lo.80	Hi.80	Lo.95	Hi.95
Jun 2022	68.46	61.91	75.00	58.45	78.46
Jul 2022	70.15	63.19	77.12	59.50	80.81
Aug 2022	69.72	62.54	76.89	58.75	80.69
Sep 2022	72.50	65.26	79.74	61.42	83.58
Oct 2022	72.80	65.53	80.07	61.68	83.92
Nov 2022	73.56	66.27	80.84	62.41	84.70
Dec 2022	74.07	66.78	81.37	62.92	85.23
Jan 2023	73.33	66.02	80.63	62.16	84.49
Feb 2023	74.09	66.78	81.39	62.91	85.26
Mar 2023	74.39	67.08	81.70	63.21	85.57
Apr 2023	72.90	65.59	80.21	61.71	84.09
May 2023	71.42	64.10	78.74	60.23	82.61
Jun 2023	71.06	63.60	78.52	59.65	82.47
Jul 2023	70.01	62.52	77.50	58.55	81.46
Aug 2023	71.41	63.90	78.92	59.93	82.90
Sep 2023	72.53	65.01	80.05	61.02	84.03
Oct 2023	72.97	65.44	80.50	61.46	84.49
Nov 2023	72.26	64.73	79.80	60.74	83.79
Dec 2023	73.27	65.73	80.81	61.73	84.80

4. Conclusion

In this study, a Seasonal ARIMA model for the relative humidity in the city of Tetouan is developed. By comparing the observed and predicted values with a 95% confidence limit. The presented model produces 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.

References

[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. Wiley.

[3] Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

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

[5] Hanks, R.R. and Stadler, S.J. (2011). Encyclopedia of Geography Terms, Themes, and Concepts. ABC-CLIO.

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

[7] Martinez, E., Silva, E. and Fabbro, A. (2011). A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. Revista da Sociedade Brasileira de Medicina Tropical, 44, pp.436 - 40.

[8] Shadab, A., Said, S. and Ahmad, S. (2019). Box-Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study. International Journal of Energy and Water Resources, 3(4), pp.305 - 318.