

# Prediction of Ozone Concentrations According the Box-Jenkins Methodology for Assekrem Area

Hicham Beldjillali<sup>1,\*</sup>, Nacef Lamri<sup>2</sup>, Nour El Islam Bachari<sup>2</sup>

<sup>1</sup>Department of Applied Statistics, Ecole Nationale des Statistiques et Economie Appliquée, 11, chemin Doudou Mokhtar Benaknoun-Alger

<sup>2</sup>Department of Ecology and Environment, Faculty of Biological Sciences, University of Science and Technology, Houari Boumedienne, USTHB, BP 32 El Alia, Bab Ezzouar Algiers

\*Corresponding author: [bachari10@yahoo.fr](mailto:bachari10@yahoo.fr)

**Abstract** The Box-Jenkins approach has been used to construct the forecast model of surface ozone ( $O_3$ ) concentrations. This forecast is important for monitoring  $O_3$  concentrations at a regional scale as well as at local level. We used the monthly average  $O_3$  concentrations covering the period going from January 2003 to December 2011. The accuracy of the models has been carried out with predicting and analyzing the average monthly  $O_3$  concentrations for 2012. By comparing the measured  $O_3$  concentrations values and the forecasted values, the AR(1) model is satisfactorily predicts monthly average  $O_3$  concentrations in the Assekrem area and the predictions of this model are loosely consistent with the measured values. The developed model can be used to forecast atmospheric tropospheric ozone concentrations in Assekrem area.

**Keywords:** Ozone, Assekrem (Tamanrasset), Box-Jenkins methodology, Autoregressive models

**Cite This Article:** Hicham Beldjillali, Nacef Lamri, and Nour El Islam Bachari, "Prediction of Ozone Concentrations According the Box-Jenkins Methodology for Assekrem Area." *Applied Ecology and Environmental Sciences*, vol. 4, no. 2 (2016): 48-52. doi: 10.12691/aees-4-2-3.

## 1. Introduction

Ozone ( $O_3$ ) has long been recognized as an important trace gas within the atmosphere influencing the climate, and the dynamics and chemistry of the stratosphere and troposphere [1]. The ozone concentration in any given area results from a combination of formation, transport, destruction and deposition [2]. Ozone ( $O_3$ ) is photochemically produced through a combination of chemical reactions involving a variety of volatile organic compounds (VOCs) and nitrogen oxides ( $NO_x$ ), which are emitted by motor vehicles, by large stationary sources, and by natural sources [3,4]. Ozone acts as an important infrared absorber (greenhouse gas), particularly in the upper troposphere, and also is an absorber of solar [5]. Its concentration in the atmosphere is under control of many worldwide meteorological services. Surface ozone (or tropospheric ozone) concentration measurements are performed in Algeria by sheltering in Assekrem (Tamanrasset) in the southern Sahara, one of the background monitoring stations of the Global Atmosphere Watch (GAW) program. Representativeness of this station privileged within the African continent, away from anthropogenic activities and biomass, is an ideal solution to perform measurements relating to the changes in the chemical composition of the atmosphere in the long term [6]. So, the variations in the  $O_3$  concentration can be detected easily and monitored precisely. Also, reliable models of monitoring  $O_3$  concentrations at a regional scale are needed for developing a prediction and helping to better control future atmospheric  $O_3$  levels around the globe. Moreover, numerous studies

have been made to develop models for predicting atmospheric  $O_3$  concentrations. Chaloulakou et al. [7] realized a comparison study with neural networks and multiple linear regression models to forecast the next day's maximum hourly ozone concentration in the Athens basin at four representative monitoring stations that show very different behavior. Heo and Kim [8] provided a study describing the method of forecasting daily maximum ozone concentrations at four monitoring sites in Seoul, Korea. Lengyel et al. [9] predicted the ozone concentration in ambient air using multivariate statistical methods including Principal Component Analysis (PCA), Multiple Linear Regression (MLR), Partial Least Squares (PLS), as well as Principal Component Regression (PCR) to evaluate the state of ambient air in Miskolc (second largest city in Hungary). Jasim et al. [10] combined the multiple regression method and PCA to obtain regression equations for total column of ozone with other measured ambient atmosphere parameters as predictor variables.

In this paper, we have developed a mathematical model able to forecast tropospheric ozone concentration in Assekrem area by considering the data available for nine years covering the period (2003 - 2011) observed at Assekrem station. For this, we use the Box-Jenkins time series technique to establish a monthly average  $O_3$  concentrations model.

## 2. Materials and Methods

### 2.1. Study Area and Data Collection

The study area of Assekrem is situated at a distance of about 50 km from Tamanrasset in southern Algeria, from

23°16' latitude north and 05°38' longitude east. It is located on the summit (plateau) of the second highest point of the Hoggar mountain range in the Saharan desert at an altitude of 2710 meters above sea level [11].

Measurements of tropospheric ozone concentrations are performed at Assekrem station since 1996. The measure is continuous, day and night, with a time step of one minute. In this study, we used the average monthly data covering the period from January 2003 to December 2011, containing 108 observations, taking into account the observations homogeneous, which do not have missing data. We have treated and structured the raw data before submitting them to statistical modeling, calculating the monthly average O<sub>3</sub> concentrations. The various kinds of software that were used during the process include Microsoft Excel (MS-Excel) and EvIEWS 5.

## 2.2. Method of ANALYSIS

In time series analysis, the Box–Jenkins methodology, named after the statisticians George Box and Gwilym Jenkins, applies Autoregressive Moving Average (ARMA) or Autoregressive Integrated Moving Average ARIMA models to find the best fit of a time series to past values of this time series, in order to make forecasts [12]. The Box–Jenkins approach for building a time-series models can be summarized in four steps: (a) Identification of preliminary specifications of the model; (b) Estimation of parameters of the model; (c) Diagnostic checking of model adequacy; and (d) Forecasting.

### (a) Identification step

The first step involves the selection of a general class of models using the autocorrelation function (ACF) and the partial autocorrelation function (PACF) [13]. Identification step is the most important and also the most difficult [14]. Using plots of the ACF and PACF of the time series to decide which autoregressive, moving average component or both together should be used to determine the order of the model required. The PACF plot helps to determine how many AR terms are necessary for the model while the ACF plot helps to determine how many MA terms are necessary. Before the identification step, the Box–Jenkins methodology begins by determining if the time series under consideration is stationary. Intuitively, a time series is stationary if the statistical properties (for example, the mean and the variance) of the time series are essentially constant through time. The Dickey-Fuller (DF) and augmented Dickey-Fuller (ADF) tests make it possible to test the null hypothesis that the time series contains a unit root against the alternative hypothesis which states that the series is stationary. We reject the null hypothesis of a unit root when the ADF test statistic value is smaller than the critical value and it can be concluded that the time series is stationary.

### (b) Estimation of parameters

After having identified a tentative model, the next step in Box-Jenkins methodology, consists of estimating the parameters identified in the model and discussing their quality and their aptitude to model the series given. We use the least squares procedure for finding the best possible estimates for the unknown parameters in the model.

### (c) Diagnostic checking:

Diagnostic checking is used to see whether or not the identified and estimated model is adequate. A good way of

finding the adequacy of an overall model is to analyze the residuals obtained from the model. In particular, the residuals should be independent of each other and constant in mean and variance over time [12]. To do this, we shall do some residual analysis. In particular we shall plot the histogram of the residuals and their correlogram. If the model is correct, the residuals would be uncorrelated and would follow a normal distribution with mean zero and constant variance. Assuming an adequate model, the autocorrelations of the residuals should therefore not be significantly different from zero [15].

## 3. Results and Discussion

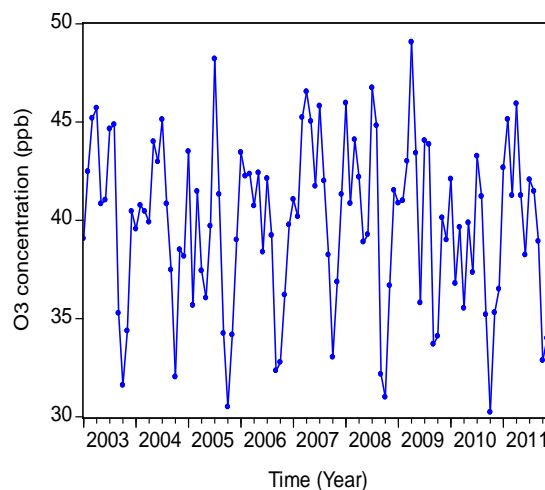
The Box-Jenkins time series technique has been used to establish a monthly average O<sub>3</sub> concentrations model.

### 3.1. Analyze Preliminary

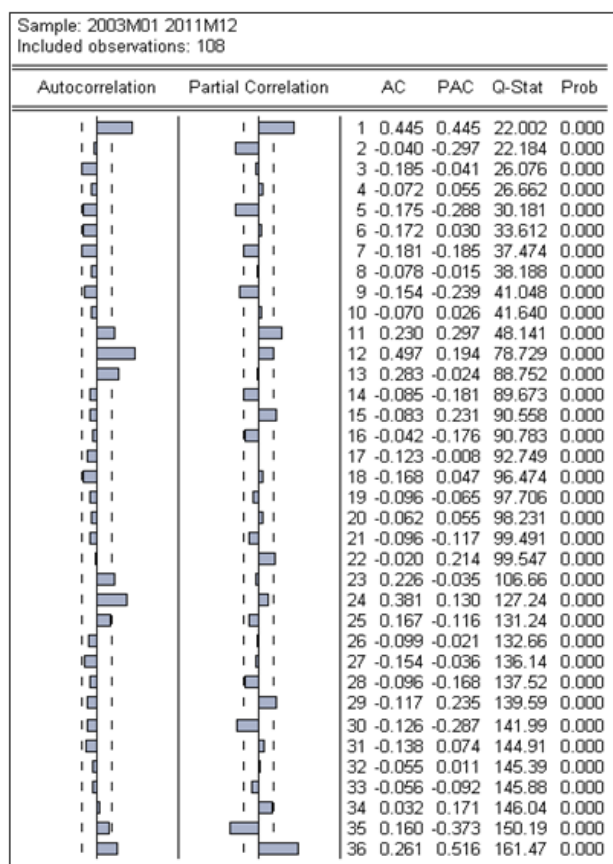
The preliminary step to take in any time series analysis is to plot the time series against time. The plot is often a valuable part of any data analysis, since qualitative features such as trend, seasonality, discontinuities and outliers will usually be present in the data [13]. In order to use the Box-Jenkins method, the time series should be stationary. If the time series presented to study is not stationary, it is necessary to transform it into a stationary one. To analyze the ozone data, the O<sub>3</sub> concentrations measured in Assekrem from January 2003 to December 2011 are plotted in Figure 1.

The graphical representation of the series O<sub>3</sub> presents a periodic movement characterized by upward and downward fluctuations throughout the study period. There is also a monthly effect represented by peaks on the level of each year what is called the seasonality. This is confirmed by the correlogram associated with the O<sub>3</sub> series, represented in Figure 2. Indeed, the correlogram of the raw series O<sub>3</sub> indicates several significant peaks that repeat. There is therefore a strong seasonality.

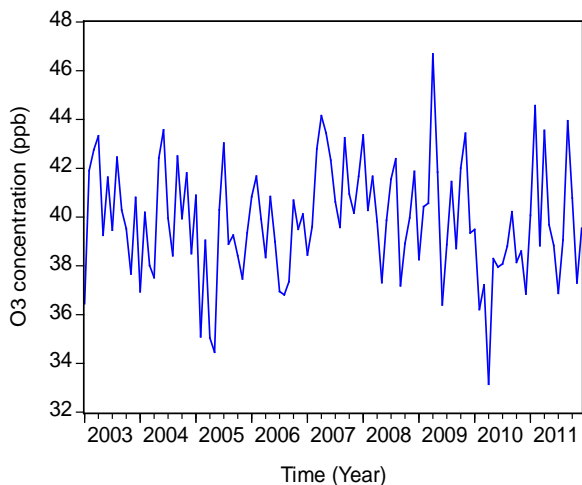
To deseasonalize the series, we have used the Moving Average method under EvIEWS which makes the series without seasonality. The seasonally adjusted series is noted "O3sa". After having deseasonalized the series, we have obtained the Figure 3.



**Figure 1.** Evolution of the O<sub>3</sub> concentrations during the period 2003-2011, observed at Assekrem



**Figure 2.** Correlogram of O<sub>3</sub> concentrations measured in Assekrem from January 2003 to December 2011



**Figure 3.** The seasonally adjusted concentration of O<sub>3</sub>

According to the graph of seasonally adjusted series, we find that the seasonal effect has disappeared. We will now study the stationarity of the series O<sub>3</sub>sa by applying the Dickey-Fuller test. The Dickey Fuller test is used to test for the presence of a unit root in the series and know whether the series is stationary or not. After having applied the Dickey-Fuller test for the model with constant and trend, we found that the latter is not significant. We therefore move to the model with constant. The results are reported in Table 1.

The test reveals that the empirical value of the t-statistic relating to the constant (C) which is equal to 7.895 exceeds the tabulated values (2.54) at the threshold of 5%. The critical probability associated with the constant is less

than 5 percent, which implies that the null hypothesis is rejected, so the constant is significantly different from zero.

We note also that statistics Dickey-Fuller (-7.901) is less than the critical values (-3.491, -2.888 and -2.581) relating to thresholds (1%, 5% and 10%, respectively). Thus, we can conclude that the series O<sub>3</sub>sa does not have a unit root and it is therefore stationary. So, parameter estimation will be performed using the model with constant.

**Table 1. The Augmented Dickey-Fuller test results**

Nul Hypothesis:O3SA has a unit root				
Exogenous: Constant				
Lag Lengh : O (Fixed)				
			t-Static	Prob.*
Augmented Dickey-Fuller test statistic			-7.901024	0.0000
Test critical values	1% level	-3.492523		
	5% level	-2.888669		
	10% level	-2.581313		
Mackinnon (1996) one-sided p-values				
Augmented Dickey-Fuller test statistic				
Dependent Variable : D(O3)				
Method least Square				
Sample (Ajusted) 2003M02 2011M12				
Included observation 107 after ajustement				
Variable	Coefficient	Std-error	t-Statistic	Prob
O3SA(-1)	-0.735812	0.093129	-7.901024	0.0000
C	29.35353	3.717962	7.895059	0.0000
R-squared	0.37285	Mean dependent var	0.028884	
Ajusted R-squared	0.36688	S.D dependent var	2.848627	
S.E Regression	2.26688	Akaike.info criterion	4.49296	
Sum Squared resid	539.439	Shwarz cretirion	4.54292	
Log likelihood	-238.373	F-statistic	62.4261	
Durbin Watson stat	1.94345	Prob (F-statistic)	0.0000	

### 3.2. Box-Jenkins Modeling

We have used the O<sub>3</sub>sa time series to identify the corresponding ARMA (p, q) process. We have examined the simple autocorrelation function (ACF) and partial autocorrelation function (PACF) of O<sub>3</sub>sa series in order to choose the appropriate order of the AR and MA terms of the model. The ACF and PACF plots are showed in Figure 4. We note that the simple autocorrelation function has a small peak at the shift 1, which allows us to suspect that an autoregressive candidate of order 1 would be adequate. The partial autocorrelation function also has a small peak at the shift 1. We can suspect that a moving average candidate of order 1 would be appropriate. Therefore, we identified three models: AR (1), MA (1), ARMA (1.1).

After identifying the models, we precede to the estimation of their parameters by the ordinary least squares, verifying their performance in order to choose the one that could better reflect the behavior of the ozone series. We have based on the Akaike information criterion (AIC) and Schwarz (SC) to select the appropriate model. Table 2 shows the different values of Akaike (AIC) and Schwarz (Sc) for each model, the statistic of Student and the probability.

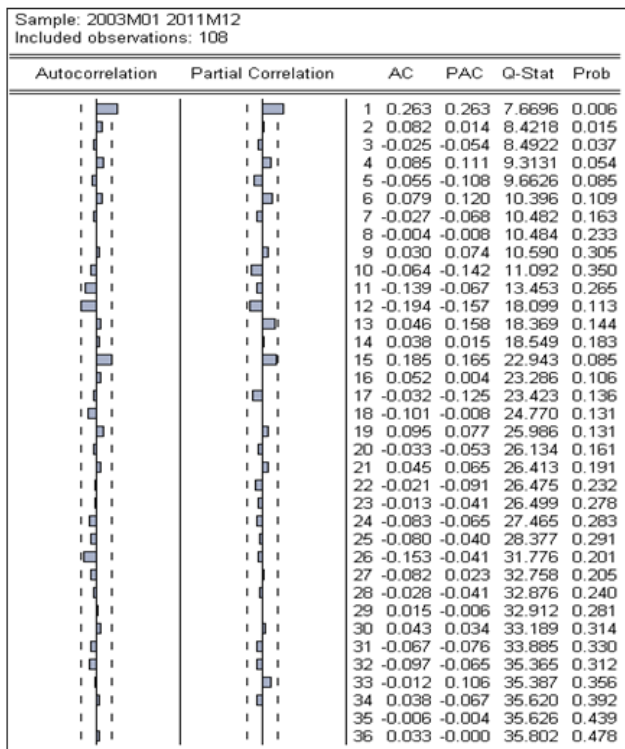
Figure 4. correlogram of the seasonally adjusted concentration of O<sub>3</sub>

Table 2. Different values of Akaike (AIC) and Schwarz (Sc).

Modèle	AIC	SCH	t-Statistic	Probabilité
AR(1)	4,492	4,542	2,836	0,0055
MA(1)	4,507	4,557	2,659	0,0090
ARMA(1,1)	4,508	4,583	AR(1) : 0,415	0,6787
			MA(1) : 0,459	0,6465

The analysis of the three processes shows that only the AR (1) process and MA (1) have coefficients significantly different of zero because the probabilities of Student's t-statistic are all less than 5%.

To decide on the right model, we have retained that which minimizes one of the Akaike Information Criterion (AIC) and the Schwarz Criterion (SC).

According to the table, the process AR (1) is the model chosen, because it minimizes both AIC and SC. The estimated coefficients of the model AR (1) by the ordinary least square method is given in the Table 3, of which the coefficients of the model are significantly different of zero.

Table 3. Estimated parameters of AR(1) model

Dependent Variable : D(O <sub>3</sub> )				
Method least Square				
Sample (Ajusted) 2003M02 2011M12				
Included observation 107 after ajustement				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std-error	t-Statistic	Prob
O <sub>3</sub> SA(-1)	39.89271	0.297837	133.94114	0.00000
AR(1)	0.264188	0.093129	2.836809	0.0055
R-squared	0.07118	Mean dependent var		39.88234
Ajusted R-squared	0.06234	S.D dependent var		2.340747
S.E Regression	2.26661	Akaike.info criterion		4.492963
Sum Squared resid	539.4399	Shwarz criterion		4.542922
Log likelihood	-238.3735	F-statistic		8.047488
Durbin Watson stat	1.94347	Prob (F-statistic)		0.005469

In order to make sure that this model is a representative of the data and could be used to forecast future O<sub>3</sub> concentrations at Assekrem, it is advisable to check whether the model is appropriate by applying tests to the residuals. The residual analysis consists of testing whether the residuals are white noise and normally distributed.

To check whether the residuals are white noise, we compute the sample ACF and sample PACF of the residuals to see whether they do not form any pattern and are all statistically insignificant. The correlogram of the residuals is shown in Figure 5. There isn't any term exterior to the confidence intervals and the Q-statistic has a critical probability greater than 5%. The residuals may be assimilating to a white noise process.

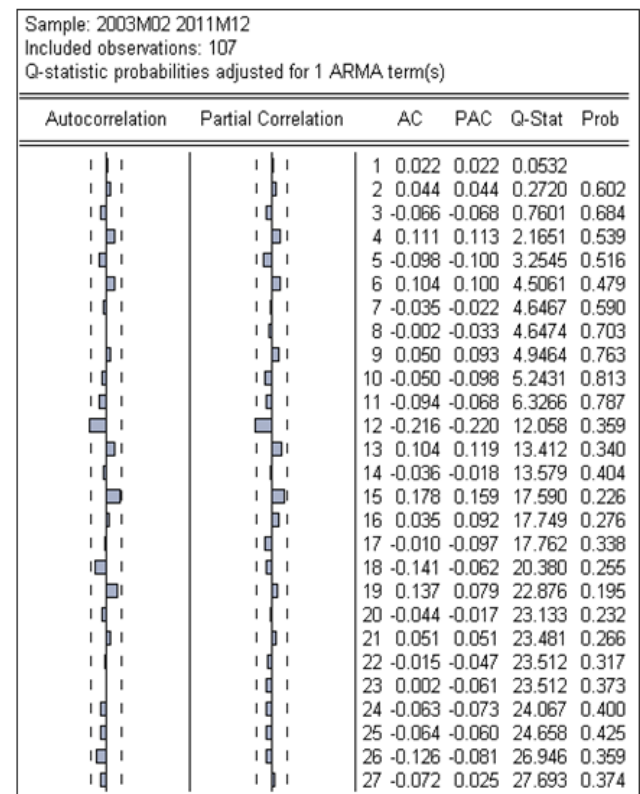


Figure 5. Correlogram of the residuals

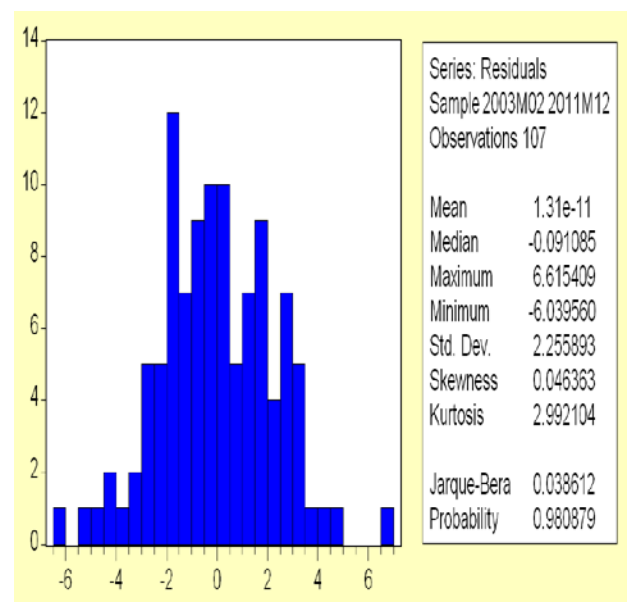


Figure 6. Histogram of residuals



To check for normality, the histogram of the residuals is shown in Figure 6. The Jarque-Bera test allows us to better assess the normality of residuals. The null hypothesis for this test is that the residuals are normally distributed. If the Jarque-Bera statistic is greater than the critical value of the chi-square with 2 degrees of freedom, we reject the null hypothesis. The Jarque-Bera test statistic of 0.038 is less than the critical value of the Chi-square with two degree of freedom of 5.99 at the five percent level of significance. Thus we do not reject the null hypothesis that the residuals follow a normal distribution.

Therefore, the estimation of AR(1) model is validated and the model could be used to represent the O<sub>3</sub> concentration data.

### 3.3. Forecasting

An AR (1) model of the monthly average O<sub>3</sub> concentrations was used to predict the monthly value of O<sub>3</sub> concentration for 2012. The forecast are shown in Table 4 with the error between observed and forecasted O<sub>3</sub> concentrations for the twelve months of the year 2012, as well as the Mean Forecast Error (MFE) and the Mean Absolute Error (MAE) for evaluating the accuracy of the forecast.

**Table 4. The O<sub>3</sub> concentration forecasts for 2012**

Month	Forecasted O <sub>3</sub>	Measured O <sub>3</sub>	Residuals
January	42.50	39.97	-2.53
February	40.46	43.52	3.05
March	42.32	43.87	1.55
April	42.27	45.00	2.72
May	41.47	42.10	0.63
June	39.29	39.85	0.56
July	45.07	46.29	1.22
August	42.31	48.58	6.27
September	34.87	36.41	1.53
October	31.97	33.87	1.89
November	36.59	36.73	0.14
December	39.54	41.80	2.26
Mean Forecast Error			1.606
Mean Absolute Error			2.028

As seen from the table, predictions of monthly average O<sub>3</sub> concentrations agree fairly well with measured values. The forecast produced errors are quite small compared to the forecasted and measured O<sub>3</sub>, as well as Mean Forecast Error and Mean Absolute Error are reasonably small, indicating that forecasts are on proper target and attesting to the goodness of the proposed model which can be used quite effectively in predicting future estimates of carbon dioxide concentration at Assekrem area. Although the AR(1) model should be tested with a new set of data to present it predicting power the future tropospheric ozone concentrations.

## 4. Conclusion

In this work, we apply the principles of George Box and Gwilym Jenkins to estimate the appropriate models that can be used for forecasting O<sub>3</sub> concentrations at Assekrem area. We used the monthly average O<sub>3</sub> concentrations covering the period (2003 - 2011). The proposed model has been applied successfully to forecast the average monthly O<sub>3</sub> concentrations for the year 2012

that are shown to be in good agreement with the measured values. The effectiveness of proposed forecast model was demonstrated through a comparison of the measured O<sub>3</sub> concentration data with forecasted values by calculating the residuals estimates. Thus, the AR (1) model can be used to make short-term predictions of tropospheric ozone concentrations to follow their evolution in the region of Assekrem (Tamanrasset).

## Acknowledgement

The authors gratefully acknowledge the Assekrem Global Atmospheric Watch (GAW) station for the provision of the data used in this paper.

## References

- [1] Henderson, G.S., McConnell, J.C., Templeton, E.M.J. & Evans, W.F.J. A numerical model for one-dimensional simulation of stratospheric chemistry, *Atmosphere-Ocean*, 25:4, 427-459, 1987.
- [2] Sicard, P., Dalstein-Richier, L. & Vas, N. Annual and seasonal trends of ambient ozone concentration and its impact on forest vegetation in Mercantour National Park (South-eastern France) over the 2000 - 2008 period. *Environmental Pollution* 159, 351-362, 2011.
- [3] Zheng, J., Swall, J. L., Cox, W. M. & Davis, J. M. Interannual variation in meteorologically adjusted ozone levels in the eastern United States: A comparison of two approaches. *Atmospheric Environment* 41, 705-716, 2007.
- [4] Abdul-Wahab, S.A., Bakheit, C.S., & Al-Alawi, S.M. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software* 20, 1263-1271, 2005.
- [5] Tarasick, D.W. & Slater, R. Ozone in the troposphere: Measurements, climatology, budget, and trends, *Atmosphere-Ocean*, 46:1, 93-115, 2008.
- [6] Malek, A., Drif, M., Chouder, A., & Chikh, M. Alimentation électrique par une installation photovoltaïque destinée pour des équipements de la Veille de l'Atmosphère Globale (Station de Recherche de l'Assekrem - Tamanrasset), *Bulletin des énergies renouvelables* - N° 2, 2002.
- [7] Chaloulakou, A., Saisana, M. & Spyrellis, N. Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. *The Science of the Total Environment* 313, 1-13, 2003.
- [8] Heo, J.S. & Kim, D.S. A new method of ozone forecasting using fuzzy expert and neural network systems. *Science of the Total Environment* 325 221-237, 2004.
- [9] Lengyel, A., Héberger, K., Paksy, L., Bánhidí, O. & Rajkó, R. Prediction of ozone concentration in ambient air using multivariate methods. *Chemosphere* 57, 889-896, 2004.
- [10] Jasim M., Rajab, M.Z. & MatJafri, H.S. Lim. Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia, *Atmospheric Environment* 71, 36-43, 2013.
- [11] Zellweger, C., Klausen, J. & Buchmann B. System and Performance Audit for Surface Ozone, Global GAW Station Tamanrasset / Assekrem, Algeria, WCC-Empa Report 03/1, 26 pp, Empa Dübendorf, Switzerland, 2003.
- [12] Adejumo, A. O., & Momo, A. A. Modeling Box-Jenkins Methodology on Retail Prices of Rice in Nigeria. *The International Journal of Engineering and Science (IJES)*, 2(9), 75-83, 2013.
- [13] Sharma, P., Chandra, A., & Kaushik, S. Forecasts using Box-Jenkins models for the ambient air quality data of Delhi City. *Environmental monitoring and assessment*, 157(1), 105-112, 2009.
- [14] Dobre, I. & Alexandru A. A. Modelling unemployment rate using Box-Jenkins procedure. *Journal of Applied Quantitative Methods*, 3(2), 156-166, 2008.
- [15] Etuk, E. H. Predicting inflation rates of Nigeria using a seasonal Box-Jenkins model. *Journal of Statistical and Econometric Methods*, 1(3), 27-37, 2012.