Comparative study between real data and theoretical results calculated by ARIMA model to predict and calculate C02 emissions

DR. M.BELAIDI[1], DR. ABDERRAHMANE BELGUERNA[2], DR. OMAR ALAMI[3]

[1,3]Sciences institute, department of technology. University center of El Bayadh. Algeria
[2]Sciences and technologies institute, department of Mathematics and computer sciences. University center of Naama. Algeria

Abstract: In this article, we have predicted and calculated by a statistical process called ARIMA model C02 emissions in Algeria during the period from 1963 to 2019, the data used here is a time series of the evolution of pollution caused by emissions of CO2. Our approach is based on four essential steps; the first is the identification of the model, the estimation of the parameters, the validation step and finally the forecast of certain future values. in Algeria from 1963 to 2019. Generally this can be done through four essential steps, the first is the identification of the model, the estimation of the parameters, the validation step and finally the forecast of certain future values.

The R software enables us exploit the theoretical results. It is found that ARIMA(1, 1, 2) model appears suitable for predicting future values of CO2 emissions.

Keywords: Time Series Analysis, Modelling, Forecasting, Pollution.

1. Introduction

Every government, public and private corporations, and even an individual have usually been interested in the future. The ARIMA model takes an important place in predicting problems leading to decision making **(Bosq, 1998)**. Modeling and forecasting is the most powerful tool now, this can be done by one of the statistical approach referred to as a Time series analysis. **(Brockwell and Davis, 1996)**. The R software enables us exploit the theoretical outcomes obtained in this analysis. In this paper, we will focus on building an appropriate ARIMA model of the time series of the evolution of pollution caused by CO2 emission in Algeria. Then we discuss the validity of the fitted model and used it to forecast five future values.

In this study, we used the data collected by the World Bank **(World Bank, 2020),** on annual carbon dioxide emissions (in kilotons) in Algeria from 1963 to 2019. Everyone knows that the main culprit behind the greenhouse effect is CO2, due to the excessive emission of greenhouse gases by humans. The Earth traps too much gas in the atmosphere, which gradually warms it up. Carbon dioxide contributes to air pollution and traps radiation at ground level. This atmospheric layer prevents the earth from cooling down at night. The result is a warming of ocean waters. CO2 is considered a pollutant when associated with large amounts of different resources such as cars, airplanes, power plants and other human activities involving the combustion of fossil fuels such as gasoline and gas natural. They cause climate change and also contribute to respiratory diseases, extreme weather conditions and increased forest fires are other effects of climate change caused by CO2 emissions. We note in this study that the ARIMA model (1, 1, 2) appears suitable for predicting future values of CO2 emissions. We also found an upward trend for these shows.

The ARIMA modelling has been successfully applied in In China, **(Sun 2009)** studied CO2 emission. In Iran, **(Lotfalipour et al2013)** modeled and predicted CO2 emissions using Grey and ARIMA models over the period 1965 to 2010 . In Bangladesh, **(Rahman & Hasan 2017)**, uses annual ARIMA models, who is the optimal model for modeling on CO2 emissions from 1972 - 2015 based on. In Thailand using VARIMAX, **(Pruethsan 2017)[** test CO2 emissions the 2000 - 2015 and _nd that the VARIMAX (2, 1, 2) and VARIMAX (2, 1, 3) models are optimal models for modeling CO2 emissions in Thailand. In this work will make use of the ARIMA model for compared between real data and theoretical as a result forecasting CO2 emissions in

Algeria.2. ARIMA Process

The time series analysis theory and its applications have become more and more important in various domains. A model that explains the pattern or variation in an actual time series data is known as a time series model. The stationary process is a process where the statistical parameters do not change with time **(Challis and Kitney, 1991)**. An autocorrelation coefficient measures the correlation between successive observations of time series data at lag k denoted by ρk and defined by the equation (2.1) below:

$$\rho_k = \frac{\sum_{i=1}^{n-k}(Y_i - \overline{Y})(Y_{i+k} - \overline{Y})}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \quad \dots (2.1)$$

The ARIMA(p, d, q) given by the equation (2.2)

$$X_i = \mu + \frac{\theta(B)}{\varphi(B)}\varepsilon_i \quad \dots (2.2)$$

Where i is indexes time, μ is the mean term, B is the backward shift operator, φ(B) is the autoregressive operator and p is their order, θ(B) is the moving average operator q is their order, d is the number of necessary differentiation to make the process stationary, and ε_i is the independent error with mean 0 and variance σ_ε^2.

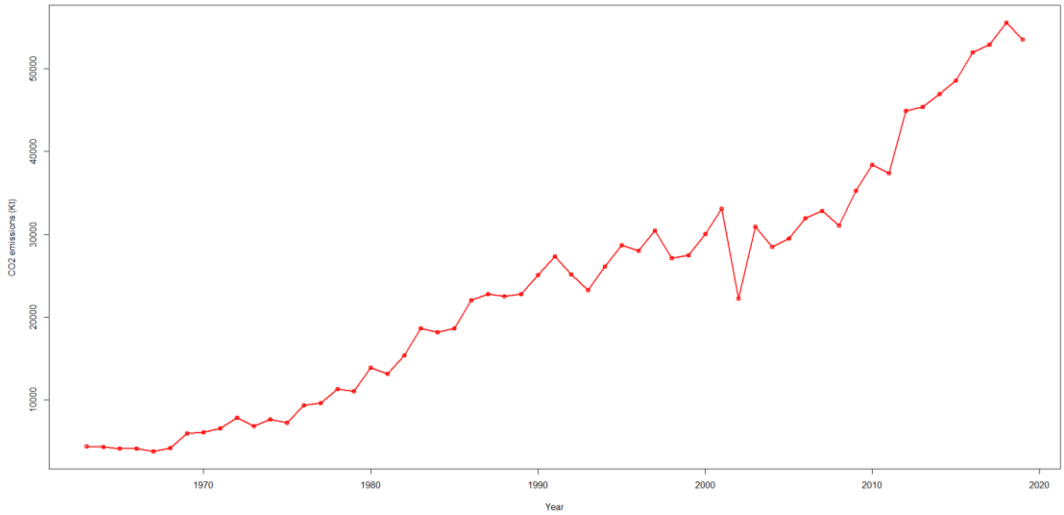
Us we said before, we focuses on the **(Box-Jenkins, 1976)** approach to identification, estimation, diagnostic and forecasting a univariate time series models. We start the analysis of time series by plotting and describing the data by using standard plots and summary statistics to see the behavior of the data. Draw the autocorrelogram and the partial autocorrelogram to see the stationary of the data, if not, stationary can be made by taking the d differences of data values.

A good model should have statistically significant coefficients and low AIC as compared to the other fitted model. Finally a **(Ljung and Box, 1978)** test can be used to analyze the residuals.

3. DATA Analysis

The data used in this study are the annual CO2 emissions in Algeria were collected from the World Bank. The data comprise 1 observation in each year from 1963 to 2019.

Fig.1. Figure CO2 emissions from 1963 to 2019 in Algeria



Source: World Bank

3.1 Summary Statistics:

The summary statistics of the CO_2 emissions time series are given in the table 1 by using the R software with function ”describe” from the ”psych” library.

Table 1. Summary statistics of the CO_2 emissions

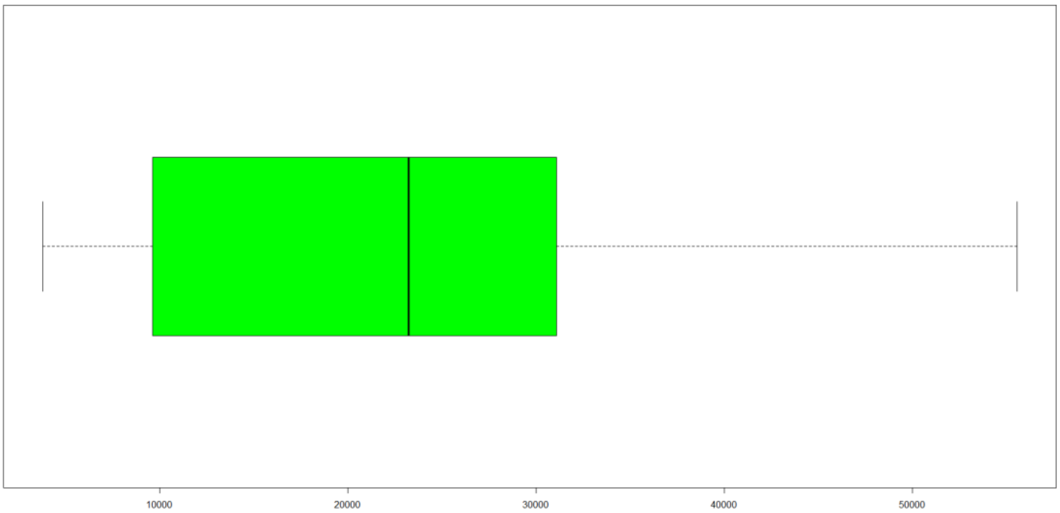
Size	Mean	S.D	Median	S.E
57	23721.24	14707.09	23223.11	1948
	Min	**Max**	**Skewness**	**Kurtosis**
	3784.34	55562.38	0.41	-0.75

3.2 The box plot

Visualization methods enhance our understanding on data and help us make comparisons. The box plot in figure 2 is a simple but powerful graphing tool that can be used in place of histograms to address both goals. This can be done by “box plot” function in R.

.

Fig.2. The box plot



4. ARIMA Model selection and estimation of parameters

The "auto.arima" function from the "forecast" library in R Returns best ARIMA model for our time series according to either AIC, AICc or BIC value. The function conducts a search over a possible model within the order constraints provided. The "auto.arima" function in R uses a variation of the Hyndman-Khandakar algorithm (Hyndman & Khandakar, 2008), which combines unit root tests, minimization of the AICc and MLE to obtain an ARIMA model. The arguments to "auto.arima" provide for many variations on the algorithm.

Running the "auto.arima" function in R, we get the best-selected model and parameters estimated, in the table below,

Table 2. ARIMA model selected and its parameters estimation

The model	φ_1	θ_1	θ_2	σ^2
ARIMA(1, 1, 2)	0.7335	-1.4645	0.7944	4905550
S.E	0.1284	0.1834	0.1624	
	AIC	**AICc**	**BIC**	**Log likelihood**
	1029.28	1030.48	1039.41	509.64

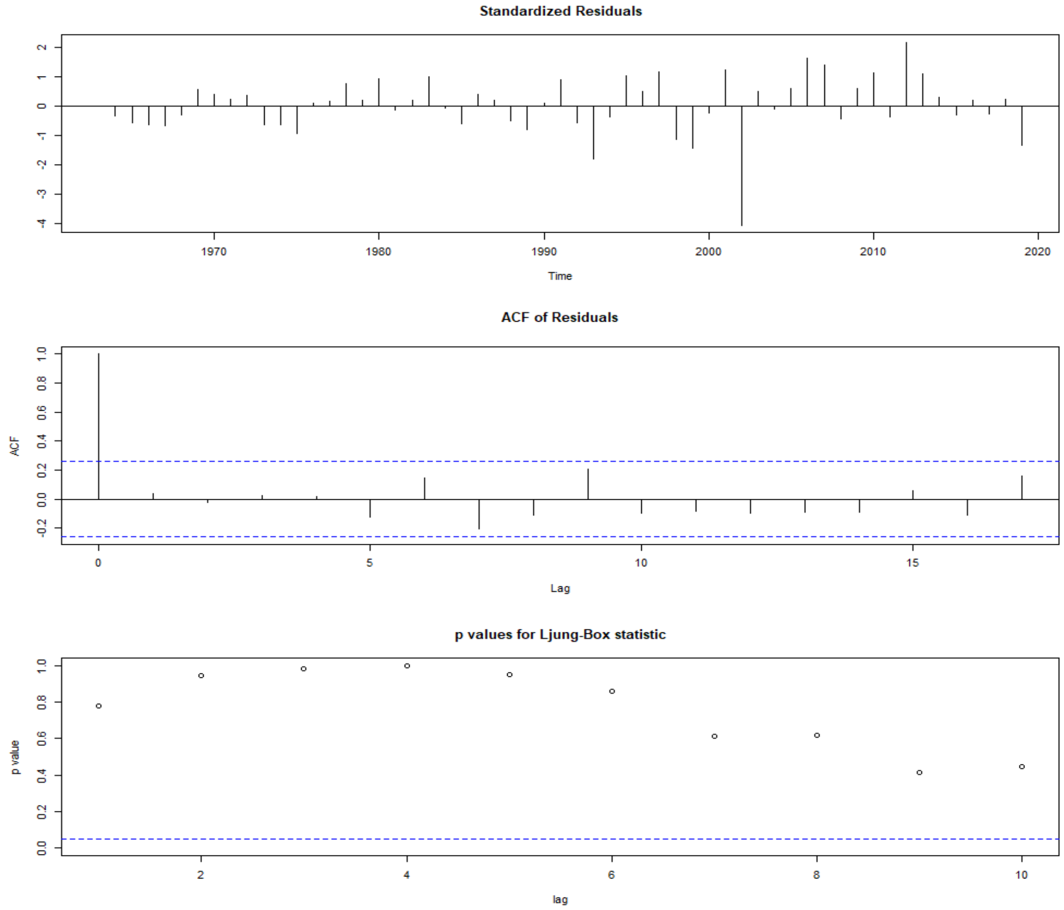
4.1 Diagnosis of residuals and validation of the model

R has the function "tsdiag" from "timeseries" library, which produces a diagnostic plot of a fitted time series model. The table 03 and the plots in figures 3 given below tell us the p-values for the Ljung-Box-Pierce statistics for each lag up to 8. These statistics consider the accumulated residual autocorrelation from lag 1 up to and including the lag on the horizontal axis. The dashed blue line is at 0.05. All p-values are above it. That’s a good result. We want non-significant values for this statistic when checking the residuals process.

Table 3. Box-Ljung test

X-squared	11.841
df	1
p-value	0.5794 e-3

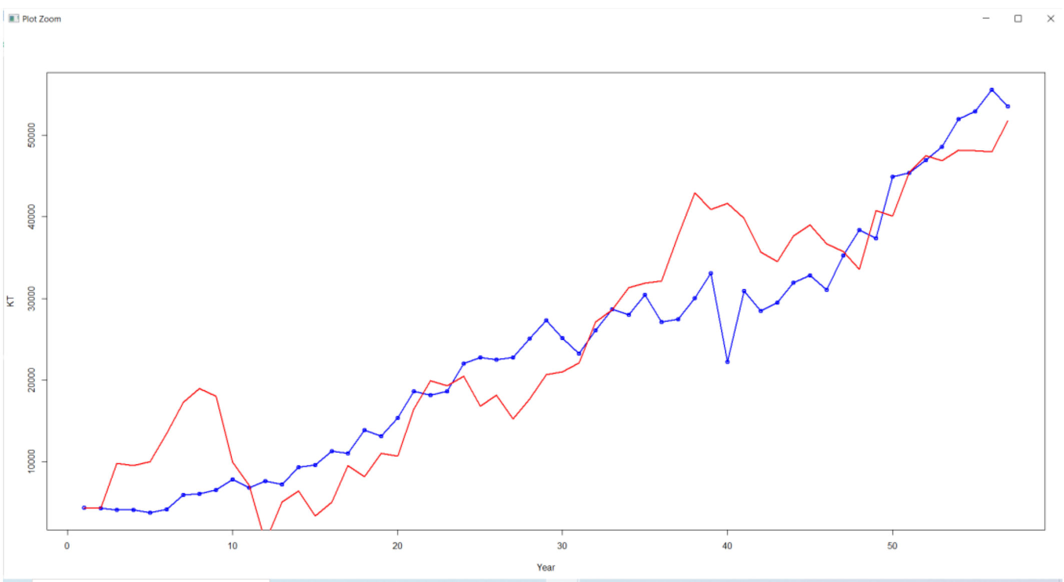
Fig.3. Residuals diagnosis



4.2 Forecasting

If we compare the data with a forecast last all the years, in other words the difference between the theoretical data of the deterministic in our model ARIMA with the parameters and the observed data. We find some correlation represented in the following figures.

Fig.3. Plot of observed values against theoretical values



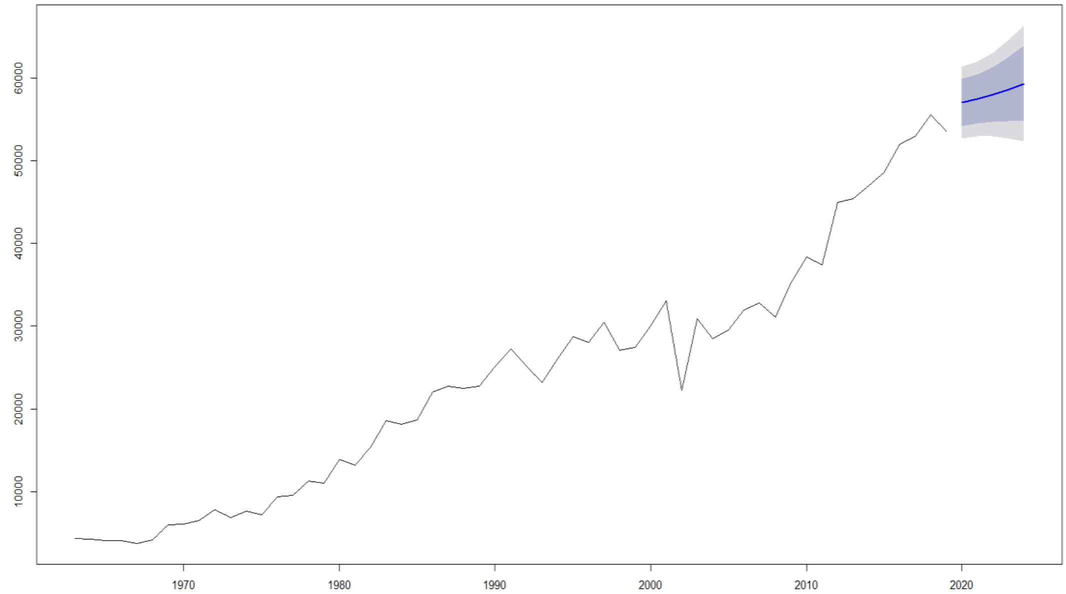
Once the model has been identified, its parameters estimated and diagnoses were checked, the aim task now is to forecast the future values of the time series. We can then use the fitted ARIMA model to make forecasts for 5 future values of the time series in the table 4 below, using the "forecast" function in the "forecast" library.

Table 4. Forecast 5 futures values

Year	2020	2021	2022	2023	2024
Forecasted values	57020	57455.76	57998.72	58620.31	59299.58

The plot in figure 4 given below, gives us the observed values from 1963 to 2019, as well as the next 5 year forecasted values using our ARIMA (1, 1, 2) model, by typing the code "plot" in R

Fig.4. Plot of our time series including forecasted 5 futures values



The results obtained in Figures 3 show that the statistical model used presents a correlation, between the parameters and the data obtained, reached at 0.9, which confirms that our ARIMA model is very useful for giving very appreciable forecasts and in good agreement perfect with experimental results.

5. CONCLUSION

We have shown in this study that our theoretical results are in very good agreement with the experimental data available by the World Bank on CO2 emissions, which allows us to predict CO2 emissions until 2040 in order to take the necessary measures in the useful time, these results confirm that our ARIMA model is very useful for giving very appreciable forecasts and in good agreement with experimental results. This study can be very useful in order to anticipate disasters that may be caused by an unexpected growth in CO2 emissions.

References

Box G. and Jenkins G. M. (1976), Time series analysis in forecasting and control Applied statistics, Holden-Day, San-Francisco;

Brockwell P. and Davis R. (1996), Introduction to Time series and Forecasting, Springer, New York.

Rob J. Hyndman, Yeasmin Khandakar (2008), Automatic Time Series Forecasting: The forecast Package for R, Journal of Statistical software, place, V27 Iss 3, page 1-22 ;

Ljung G. and Box G. (1978), a measure of lack of fit in time series models, Biometrika, 65: 297–303.

Granger C. and Newbold P. (1976), Forecasting transformed series, Journal of the Royal Statistical Society B, 38: 189–203.

Challis R. E. and Kitney R. I. (1991), Biomedical signal processing, part 1 time-domain methods, Medical and Biological Engineering and Computing, 28: 509–524.

World Bank (2020), the World Bank data by country, detailed web site: https://data.worldbank.org/country/DZ (consulted on 18/08/2020).

Mohammed Bouznita María del P.Pablo-Romerobc Energy Policy (2016),, CO2 emission and economic growth in Algeria, Energy Policy,Volume 96, Pages 93-104

Rébha Ghedamsia NoureddineSettoua AbderrahmaneGouarehb Adem Khamouli Nadia Saifia Bakhta Reciouib Boubekker Dokkara(2016), Modeling and forecasting energy consumption for residential buildings in Algeria using bottom-up approach, Energy and Buildings, Volume 121,, Pages 309-317

Miloud Lacheheb,A. S. Abdul Rahim , Abdalla Sirag (2015). Growth and Carbon Dioxide Emissions: Investigating the Environmental Kuznets Curve Hypothesis in Algeria, International Journal of Energy Economics and Policy, , 5(4), 1125-1132. Economic

Sidi Mohammed Chekouri, Abderrahim Chibi, Mohamed Benbouziane(2015), Examining the driving factors of CO2 emissions using the STIRPAT model: the case of Algeria Earth and Planetary Science Letters Volume 426, Pages 118-129

James P.Verdona Anna L.StorkaRob C.BissellbClare E.BondcMaximilian J.Wernera

(2014),Simulation of seismic events induced by CO2 injection at In Salah, Algeria,International Journal of Economics, Commerce and Research Vol. 4, Issue 4, 1-14

S.Ould-Amrouchea, D.Rekiouab, A.Hamidata(2010), Modelling photovoltaic water pumping systems and evaluation of their CO2 emissions mitigation potential, Applied Energy Volume 87, Issue 11, Pages 3451-3459

M.Allali, M.Tamali, M.Rahlia (2015),, The Impact of CO2 Emission on Output in Algeria, Energy Procedia, Volume 74, Pages 234-242