Forecasting of CO2 Emissions in Algeria Using Discrete Wavelet Transform – Based Autoregressive Integrated Moving Average Models

Abdelkader SAHED[1,*], Mohammed MEKIDICHE[2], Hacene KAHOUI[3]

[1] University Centre of Maghnia, Laboratory LEPPESE, (Algeria) (sahed14@yahoo.fr)
[2] University Centre of Maghnia, Laboratory LEPPESE, (Algeria) (mkidiche@yahoo.fr)
[3] University Centre of Maghnia, Laboratory MIFMA, (Algeria) (hacenkahwi@gmail.com)

Received: 28/05/2023; **Accepted:** 04/06/2023; **Publishing:** 10/06/2023

Summary: The increasing impact of climate change and rising temperatures has made the reduction of carbon dioxide emissions a top priority globally. Accurately forecasting these emissions is a crucial aspect of transitioning towards a clean energy economy. This paper introduces a new method for estimating CO2 emissions by combining the wavelet technique with both an autoregressive integrated moving average (DWT-ARIMA) and ARIMA model, applied to annual carbon dioxide emissions data in Algeria from 1970 to 2022. The study provides decision makers with crucial information to help find effective environmental protection solutions in Algeria. The results suggest that the wavelet-ARIMA model is more effective compared to the traditional ARIMA model.

Keywords: Forecasting; CO2 Emissions; Discrete Wavelet Transform; ARIMA.

Jel Classification Codes: Q543; C32; C323

.

* Corresponding author.

I- Introduction :

The global environment has faced numerous challenges in the past two decades, with global warming and climate change ranking among the most pressing. The increase in carbon dioxide (CO_2) levels, which are the primary contributor to the greenhouse effect, is exacerbating the situation (Ozturk, I., & Acaravci, A, 2010, p. 3220). CO_2 is considered one of the main drivers of global warming and climate change, as it is a vital component of greenhouse gases. One of its significant impacts is the rise in temperature, which is leading to a rise in sea levels due to thermal expansion of the water. According to the World Bank (2007), CO_2 is responsible for 58.8% of greenhouse gas (GHG) emissions and other pollutants contributing to climate change. Macroeconomic activities, such as manufacturing and energy usage, generate substantial amounts of CO_2, contributing to its increasing concentration in the atmosphere. Between 1750 and 2005, CO_2 was the primary contributor to climate change, according to statistics (Cosmas, N. C., Chitedze, I., & Mourad, K. A, 2019, p. 314).

The decline in carbon dioxide (CO_2) emissions in Algeria during the year 2020 was notable, registering a 7.45% decrease from the previous year to reach a total of 13,152 megatons. In terms of global rankings, Algeria placed 152nd out of 184 nations with a total CO_2 emission of 163,473 megatons in 2020. To gain a deeper understanding of the country's contribution to the atmospheric CO_2 levels, it is imperative to consider its per capita emissions. According to the data, Algeria's per capita CO_2 emissions in 2020 were estimated at 3.77 tons per person, which showed a decrease over the previous five years, despite an increase in overall CO_2 emissions.

The increasing concern surrounding global warming and the deterioration of the global climate has put the issue of CO_2 emissions at the forefront of public attention. Accurately predicting and evaluating the changing patterns of CO_2 emissions is crucial in finding solutions to the climate crisis. Although there has been considerable research on CO_2 emissions and the development of prediction models, there is still a need for a more quantified approach to selecting prediction indicators and studying CO_2 emissions from an economic perspective. (Lei Wen, Xiaoyu Yuan, 2020, p. 2).

Numerous studies have made predictions on the emissions of carbon dioxide (CO_2) in various regions and circumstances. Hosseini, et al. (2019) utilized both multiple linear regression (MLR) and multiple polynomial regression (MPR) analyses to forecast Iran's CO_2 emissions in the year 2030. The predictions were made under two scenarios: business as usual (BAU) and the Sixth Development Plan (SDP). The results of the study indicated that, under BAU assumptions, Iran is unlikely to meet its commitments under the Paris Agreement; however, the full implementation of the SDP might have achieved the objective by the end of 2018.

In a different study, Lei Wen and Xiaoyu Yuan (2020) utilized a novel BP neural network model that was based on random forest's index quantization capability and PSO's performance optimization capability. The PSO was improved in various ways to enhance the prediction accuracy. The validity of the model was verified using panel data from the Chinese business sector from the years 1997 to 2017. The results of the study showed that the newly built hybrid forecasting model could better estimate China's commercial department's CO_2 emissions than existing parallel models. The prediction indexes selected through quantification using the random forest helped increase the prediction accuracy. Additionally, the PSO modifications presented in the study significantly increased the prediction effect of the hybrid model.

Moreover, in a study conducted by Ning, Pei, and Li (2021), the authors first used the program Eviews to conduct a randomness analysis on carbon emission data from 1997 to 2017 in four typical provinces and cities: Beijing, Henan, Guangdong, and Zhejiang. The data preprocessing was executed using stationary processing of the difference, moving average, and substitution of strong impact points, based on the individual features of the data. Next, model identification, parameter estimation, and model testing were performed to build the ARIMA(p, d, q) model for predicting carbon emissions in the four areas. Finally, the model was used to forecast the data and examine their potential for carbon emissions over the following three years.

Mutascu (2022) proposes a study to forecast CO_2 emissions in the USA and analyze their components' contributions to the total volume using a mixed approach of artificial neural networks (ANN) and vector autoregressive (VAR) estimator. The study finds wind, solar, and total biomass energy consumption significantly impact CO_2 emissions compared to economic growth and net trade. The research emphasizes the importance of non-polluting energy capacities in controlling CO2 emissions and contributes to the literature by using ANN and VAR models to forecast CO2 emissions in the USA.

The problem being addressed by the topic of forecasting CO2 emissions is the need for a precise and trustworthy prediction of carbon dioxide (CO2) emissions in Algeria. CO2 emissions have a significant impact on global warming and climate change, making it crucial to understand and predict their future patterns in order to take effective action.

In Algeria, precise forecasting of CO2 emissions is especially critical as the country is undergoing rapid economic and industrial growth, which is likely to lead to a rise in energy consumption and, as a result, CO2 emissions. However, the data on CO2 emissions in Algeria is often contaminated with noise and has non-stationary and non-linear trends, making conventional forecasting methods unreliable.

The use of a Discrete Wavelet Transform (DWT) and Autoregressive Integrated Moving Average (ARIMA) model provides a solution to this problem by offering a robust and precise method for forecasting CO2 emissions. DWT breaks down the data into various frequency components, allowing for the discovery of trends and patterns that may go unnoticed by conventional methods, while the ARIMA model provides a flexible framework for modeling the fundamental structure of the data.

This research aims to provide a precise and trustworthy prediction of future CO2 emissions in Algeria by using various techniques on relevant data. The results of this study can guide decision-making and planning activities related to energy and environmental management in Algeria.

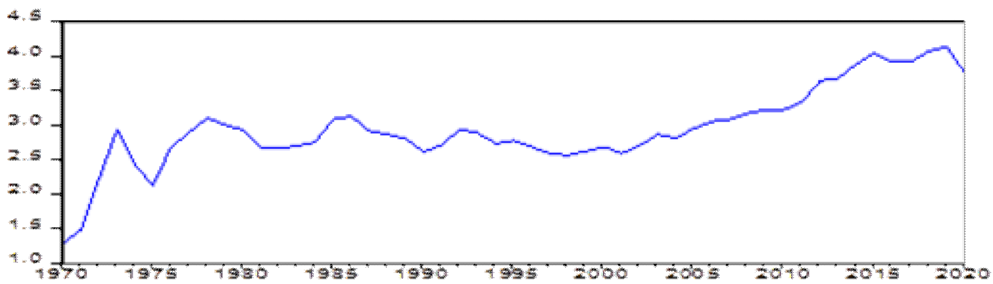
The paper is structured with three sections. Section II outlines methods and materials, Section III presents results and discussions, and Section IV concludes the paper. The clear organization makes the paper accessible and useful for scholars, practitioners, and policymakers.

II- Methods and Materials :

II.1.Data

The forecasting of CO2 emissions in Algeria using the Wavelet-ARIMA method is based on an extended sample of 50 observations covering the period from 1970 to 2020. The dependent variable targeted for this analysis is the CO2 emissions per capita. By using the Wavelet-ARIMA method, we aim to produce accurate predictions of future CO2 emissions in Algeria. Figure 1 shows that the highest amount of CO2 emissions per capita was 4.14 in 2019, while the lowest amount was 1.3 in 1970. This indicates that there has been a significant increase in carbon dioxide emissions in Algeria over the past few years.

Figure (1): CO2 Emissions in Algeria



The source: Author's own computation

II.2.Methodology

2.1.ARIMA model

ARIMA, a popular model for time series data, effectively combines two separate models: the Auto Regressive (AR) model and the Moving Average (MA) model. The AR model captures the auto regressive component, while the MA model deals with the moving average component. The integration element of the ARIMA model, represented by (I), connects these two models together.

In the ARIMA model, (p) represents autoregressive lags, (q) represents moving average lags, and (d) represents the order of differentiation. Using these parameters, we can estimate an autoregressive model and a composite moving average to describe the y process as follows:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

Overall, ARIMA is widely recognized as one of the best models for analyzing and forecasting time series data, as it takes into account both the auto regressive and moving average components of the data, as well as the order of differentiation.

Subsequently, to identify a stationary time series, the Dickey-Fuller test was employed followed by the difference method. The values of q and p were determined using the autocorrelation graph (ACF) and partial autocorrelation graph (PACF), respectively (Chyon, F.A., et al., 2021, p. 3). Using the auto.arima function, the ARIMA model parameters p and q were specified by minimizing AICc, and maximum likelihood estimates were used to estimate the coefficient values (Noah Aggeborn Leander, 2020, P4). The prediction process involved the use of both AR and MA models.Finally, the residual data was used to create a graph to ensure the estimated model's quality (Chyon, F.A., et al., 2021, p. 3).

2.2.Discrete Wavelet Transform (DWT)

The wavelet transform is an approach for signal analysis that simultaneously considers both the time and frequency of signals. It divides an input signal into two components, namely the low frequency information (approximation) and the high frequency information (detail), by using low and high frequency pass filters. The input of the higher decomposition level is an approximation of the preceding decomposition level, and this approach can be applied at multiple decomposition levels. There are two types of wavelet transformations, namely continuous and discrete wavelet transformations. In recent years, researchers in signal processing have increasingly used Daubechies wavelet transformations (Kumar, J., Kaur, A., & Manchanda, P., 2015, p. 433).

In our study, we used two decomposition levels and the Daubechies wavelet (db2) in MATLAB. The "db2" wavelet was used to decompose the original data into approximation and detail components at different levels (Yu, G., et al., 2021, p. 4). However, in practice, time series data are discrete, and the discrete wavelet transform (DWT) can be used to decompose them as shown below (Pannakkong, W., & Huynh, V. N., 2017, p. 162):

$$y_t = A_J(t) + \sum_{j=1}^{J} D_j(t) \quad (2)$$

$$y_t = \sum_{k=1}^{K} c_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{J} \sum_{k=1}^{K} d_{j,k}\psi_{j,k}(t) \quad (3)$$

Where:

y_t indicates the time series at period t.

$A_J(t)$ represents the approximation of the highest decomposition level (J).

$D_j(t)$ denotes the detail of decomposition level j.

$c_{j,k}$ and $d_{j,k}$ represent the coefficient of the approximation and detail respectively, at decomposition level j and period k.

$\phi_{j,k}(t)$ and $\psi_{j,k}(t)$ denote low (approximation) and high (detail) pass filters respectively, at decomposition level j and period k.

K indicates the total number of time series.

J represents the total level of decomposition.

2.3.Performance Criteria

To evaluate the predictive accuracy of the presented model, the researchers use the well-established criteria of the RMSE and MAPE. These two criteria allow for a quantitative assessment of the model's ability to predict outcomes. The RMSE measures the deviation between predicted and actual values, and the MAPE measures the average percentage difference between the two values. Both metrics are commonly used in a variety of fields to evaluate model performance and provide insights into potential areas of improvement.

The formulas for calculating the RMSE and MAPE are readily available in the literature, as provided by Shabri and Samsudin (2015, p. 3). These formulas enable researchers to compare the predictive ability of the presented model to other models and identify its strengths and weaknesses. The RMSE and MAPE can help pinpoint areas where the model needs improvement and inform decisions about refining the model. Obviously, the lower the RMSE and MAPE values, the greater the model's efficiency.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \qquad (4)$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{y_t - \hat{y}_t}{y_t} \times 100\right| \qquad (5)$$

where y_t denotes the actual data, $\hat{y}_t$ is the predicted value for period t, and n denotes the number of observations.

III- Results and discussion:

III.1.ARIMA estimations

To test for stationarity in the time series, an ADF test was conducted, which indicated nonstationarity with a P-value of 0.0723. However, the time series became stationary after one difference (ADF test P = 0.000). An ARIMA (0,1,2) model was selected as the best model with the lowest AICc of -0.273548, using the "auto.arima" function with d = 1. Figure 2 displays the residuals plot and accompanying ACF plot, which shows that the residuals are normally distributed and do not exhibit significant autocorrelation. Additionally, the Ljung-Box test of the residuals was performed and resulted in a P-value of 0.956, indicating no significant autocorrelation in the residuals.

Figure (2): Correlogram of residuals

Date: 01/25/22 Time: 05:44
Sample: 1970 2020
Included observations: 50
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.009	0.009	0.0039	
		2	0.079	0.079	0.3391	
		3	-0.060	-0.062	0.5387	0.463
		4	0.169	0.165	2.1447	0.342
		5	-0.005	-0.002	2.1464	0.543
		6	0.121	0.097	3.0103	0.556
		7	0.098	0.119	3.5853	0.611
		8	-0.055	-0.106	3.7699	0.708
		9	-0.041	-0.038	3.8774	0.794
		10	-0.033	-0.047	3.9489	0.862
		11	-0.133	-0.186	5.1237	0.823
		12	0.036	0.064	5.2121	0.877
		13	0.032	0.034	5.2835	0.917
		14	-0.055	-0.073	5.5024	0.939
		15	-0.040	0.059	5.6226	0.959

The source: Author's own computation

III.2.Discrete Wavelet Transform -based ARIMA estimations

Initially, we used MATLAB to decompose the original time series using wavelet analysis. The discrete wavelet transform was used for this purpose, as it has proven to be highly effective in handling time series data. The decomposition process segregated the data into two sections, coarser and finer, with the trend visible in the coarse sections and seasonal impacts and noise visible in the finer scales. Our technique employed Daubechies of order 2 and a decomposition level of 2. Thus, our wavelet decomposition technique resulted in the following output.

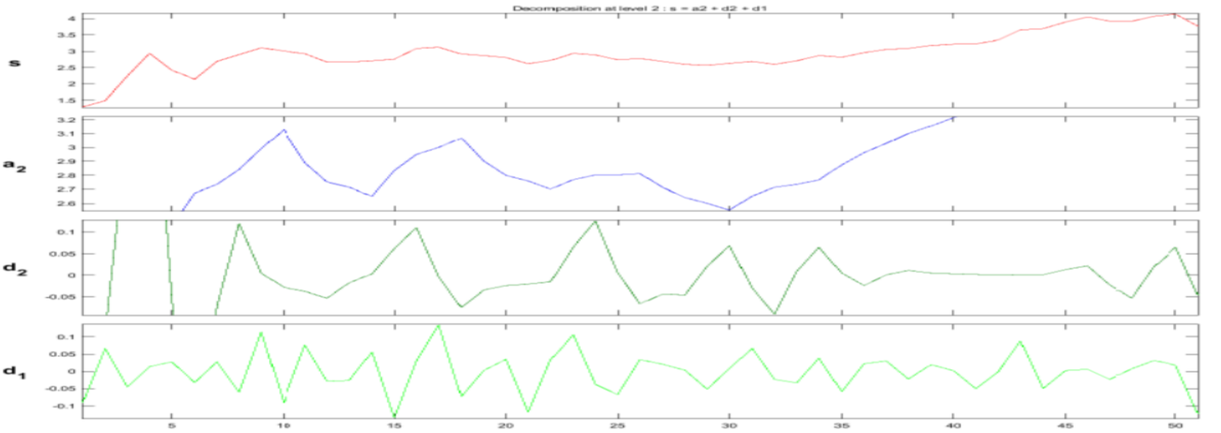
$$y_t = d_{1,t} + d_{2,t} + a_{2,t} \qquad (6)$$

Figure 3 depicts the results of decomposing the data using Daubechies 2 wavelet with level 2. This process generated three signals: wavelet signal d1 at the first level, wavelet signal d2 at the second level, and approximation signal a2 at level 2. We conducted stationary testing on each signal after wavelet and approximation signal generation.

The ADF Stationary Test showed that the series d1 and d2 were stationary at their respective levels, with a p-value less than 0.05. On the other hand, the approximation signal had a p-value greater than 0.05, specifically 0.8166, indicating that it was not stationary. Consequently, the approximation signal had to be differenced until it became stationary and could be modeled.

After analyzing the ADF Stationary Test results, we applied the Automatic ARIMA Forecasting algorithm to series d1. The ARMA(1,3) model achieved the lowest AIC value of -3.8853, while for series d2, the ARMA(4,4) model had the lowest AIC value of -2.8307. Additionally, the a2 series was best modeled using the ARIMA(0,1,4) model, which produced the lowest AIC value of -1.7497.

Figure (3): Decomposition at level 2



The source: Author's own computation

The DWT-ARMA model is constructed by applying equation (7) to combine the proposed models of all signals, including the detailed wavelet and approximation signals. This approach enables the modeling of the non-stationary components of the data via wavelet decomposition and modeling the stationary residuals with ARMA. The resulting DWT-ARMA model allows for a more accurate and robust modeling of time series data.

$$\hat{y}_t = \hat{d}_{1,t} + \hat{d}_{2,t} + \hat{a}_{2,t} \qquad (7)$$

Table 1 presents the root mean square error (RMSE) values of the DWT-ARIMA model and the ARIMA model alone for the CO2 Emissions data series. The RMSE of the DWT-ARIMA model is 0.151454668, which is significantly lower than the RMSE of the ARIMA model alone, which is 0.205168526. Additionally, the mean absolute percentage error (MAPE) of the DWT-ARIMA model is lower than that of the ARIMA model alone. These results demonstrate that the proposed DWT-ARIMA technique is more effective than using the ARIMA model directly for the given data set. Therefore, using the discrete wavelet transform as a preprocessor significantly improves the performance of the ARIMA model.

Table (1): Comparison of models

ERROR MEASURES	DWT-ARIMA	ARIMA
RMSE	0.151454668	0.205168526
MAPE	3.69902361	5.216105607

The source: Author's computation

The DWT-ARIMA model was employed to forecast carbon dioxide emissions in Algeria for the years 2021 and 2022, and the results indicate that the model is effective in predicting future emissions. The predicted emissions for 2021 and 2022 were 3.847282133 and 3.889891069, respectively. These forecasts can be used by policymakers to make informed decisions regarding carbon reduction initiatives and can also aid in assessing the effectiveness of existing measures. However, it is important to note that these are only forecasts, and actual emissions may differ due to various unforeseeable factors such as economic changes, political instability, or natural disasters. Nevertheless, the DWT-ARIMA model provides a useful tool for predicting future carbon dioxide emissions and can aid in developing strategies for mitigating the impact of climate change.

IV- Conclusion:

The present study proposes a hybrid approach for predicting CO2 emissions in Algeria by combining the ARIMA(0,1,2) model and WDT-ARIMA based on the annual data from 1970 to

2020. The results indicate that the WDT-ARIMA model outperforms the ARIMA(0,1,2) model in terms of accuracy. Therefore, the WDT-ARIMA model was employed to forecast CO2 emissions in Algeria from 2021 to 2022, and the results indicate that the estimated CO2 emissions will decrease to 3.847282133 tons per capita in 2021 and increase to 3.889891069 tons per capita in 2022. These findings can help the Algerian government to develop strategies and policies to reduce CO2 emissions.

Referrals and references:

1 Abdelaziz Naoum Aggeborn Leander, N. (2020). Forecasting the Daily Air Temperature in Uppsala Using Univariate Time Series.

2 Chyon, F. A., Suman, M. N. H., Fahim, M. R. I., & Ahmmed, M. S. (2021). Time series analysis and predicting COVID-19 affected patients by ARIMA model using machine learning. *Journal of Virological Methods*, 114433.

3 Cosmas, N. C., Chitedze, I., & Mourad, K. A. (2019). An econometric analysis of the macroeconomic determinants of carbon dioxide emissions in Nigeria. *Science of the Total Environment*, 675, 313-324.

4 Hosseini, S. M., Saifoddin, A., Shirmohammadi, R., & Aslani, A. (2019). Forecasting of CO2 emissions in Iran based on time series and regression analysis. *Energy Reports*, 5, 619-631.

5 Kumar, J., Kaur, A., & Manchanda, P. (2015). Forecasting the time series data using ARIMA with wavelet. *Journal of Computer and Mathematical Sciences*, 6(8), 430-438.

6 Mutascu, M. (2022). CO2 emissions in the USA: new insights based on ANN approach. *Environmental Science and Pollution Research*, 29(45), 68332-68356.

7 Ning, L., Pei, L., & Li, F. (2021). Forecast of China's Carbon Emissions Based on ARIMA Method. *Discrete Dynamics in Nature and Society,* 2021.

8 Ozturk, I., & Acaravci, A. (2010). CO2 emissions, energy consumption and economic growth in Turkey. *Renewable and Sustainable Energy Reviews*, 14(9), 3220-3225.

9 Pannakkong, W., & Huynh, V. N. (2017, October). A Hybrid Model of ARIMA and ANN with Discrete Wavelet Transform for Time Series Forecasting. *In International Conference on Modeling Decisions for Artificial Intelligence* (pp. 159-169). Springer, Cham.

10 Shabri, A., & Samsudin, R. (2015). Fishery landing forecasting using wavelet-based autoregressive integrated moving average models. *Mathematical Problems in Engineering*, 2015.

11 Wen, L., & Yuan, X. (2020). forecasting CO2 emissions in Chinas commercial department, through BP neural network based on random forest and PSO. *Science of The Total Environment*, 718, 137194.

12 Yu, G., Feng, H., Feng, S., Zhao, J., & Xu, J. (2021). Forecasting hand-foot-and-mouth disease cases using wavelet-based SARIMA–NNAR hybrid model. *Plos one*, 16(2), e0246673.

How to cite this article by the APA method:

Abdelkader SAHED & Mohammed MEKIDICHE & Hacene KAHOUI (2023), **Forecasting of CO2 Emissions in Algeria Using Discrete Wavelet Transform –Based Autoregressive Integrated Moving Average Models**, Journal of quantitative economics studies, Volume 09 (Number 01), Algeria: Kasdi Marbah University Ouargla, PP. 443-449.