

## **Forecasting of CO2 Emissions in Algeria (2000-2030): A Hybrid ARIMA-LSTM**

**التنبؤ بانبعاثات ثاني أكسيد الكربون في الجزائر (2000-2030): نهج هجين يجمع بين نماذج ARIMA و LSTM**

**TOUL Hamza<sup>1</sup>, SAHED Abdelkader<sup>2</sup>**

<sup>1</sup> University of Tlemcen, Laboratory LEPPSE, (Algeria), [hamzatoul@yahoo.com](mailto:hamzatoul@yahoo.com)

<sup>2</sup> University Centre of Maghnia, Laboratory LEPPSE, (Algeria),

[sahed14@yahoo.fr](mailto:sahed14@yahoo.fr); [a.sahed@cu-maghnia.dz](mailto:a.sahed@cu-maghnia.dz)

**Received on: 19/02/2025**

**Accepted on: 31/05/2025**

**published on: 06/06/2025**

**Abstract:** The study analyzed Algeria's CO<sub>2</sub> emissions from 2000 to 2030 by integrating the ARIMA model, which captures linear trends, with the LSTM network, known for detecting complex nonlinear patterns. This hybrid approach significantly outperformed traditional models by delivering highly accurate predictions, as evidenced by substantial reductions in error metrics—including a 58.2% decrease in MAE, a 58.7% drop in MAPE, a 71.4% reduction in MSE, and a 46.5% decline in RMSE. While emissions are projected to continue rising until 2030, the study anticipates a potential slight decline due to the adoption of renewable energy policies and the indirect effects of global crises. Overall, the findings underscore the effectiveness of combining artificial intelligence techniques with traditional statistical methods for enhancing environmental forecasting accuracy, particularly in fossil fuel-dependent nations, and emphasize the urgent need to transition towards cleaner energy sources amid evolving economic and political challenges.

**Keys words:** Carbon dioxide emissions, hybrid approach, ARIMA, LSTM, Algeria.

**JEL classification codes:** C53; Q54; Q25.

**ملخص:** قامت الدراسة بتحليل انبعاثات ثاني أكسيد الكربون في الجزائر من عام 2000 إلى 2030 من خلال دمج نموذج ARIMA الذي يلتقط الأنماط الخطية مع شبكة LSTM المعروفة بالكشف عن الأنماط غير الخطية المعقدة. وقد تفوق هذا النهج الهجين بشكل كبير على النماذج التقليدية من خلال تقديم تنبؤات دقيقة للغاية كما يتضح من الانخفاضات الكبيرة في مقاييس الخطأ، بما في ذلك انخفاض بنسبة 58.2% في MAE، و 58.7% في MAPE، و 71.4% في MSE، و 46.5% في RMSE. وعلى الرغم من أنه من المتوقع أن تستمر الانبعاثات في الارتفاع حتى عام 2030، فإن الدراسة تتوقع انخفاضاً طفيفاً بسبب تبني سياسات الطاقة المتجددة وتأثيرات الأزمات العالمية. في المجمل، تؤكد النتائج على فعالية دمج تقنيات الذكاء الاصطناعي مع الطرق الإحصائية التقليدية في تعزيز دقة التنبؤات البيئية، خاصة في الدول المعتمدة على الوقود الأحفوري، وتؤكد على الحاجة الملحة للتحويل نحو مصادر الطاقة النظيفة في ظل التحديات الاقتصادية والسياسية.

**الكلمات المفتاحية:** انبعاثات ثاني أكسيد الكربون، النهج الهجين، ARIMA، LSTM، الجزائر

**تصنيف JEL:** C53، Q54، Q25

**Corresponding author:** TOUL Hamza, e-mail: [hamzatoul@yahoo.com](mailto:hamzatoul@yahoo.com)

## **1-Introduction**

The Climate Change problem is one of the greatest global challenges. Emission of greenhouse gases such as carbon dioxide (CO<sub>2</sub>) is a crucial factor responsible for climate change (IPCC, 2021, p. 4). As countries aim to reduce their negative effects on the environment and manage sustainable growth, projecting CO<sub>2</sub> emissions has become a critical activity for governments, academics, and others interested in the areas (United Nations, 2015, p. 5). Algeria is among the top economies in Africa, and as one of the major producers of hydrocarbons, the country has to find a way to reconcile economic growth with environmental responsibility (OPEC, 2022, pp. 12-13). The dependence on fossil fuels is a double edged sword for the country since it drives substantial value to the economy but at the same time leads to high emissions of CO<sub>2</sub>. Thus, the development of advanced predictive models that are capable of resource planning and policy making becomes crucial (World Bank, 2020, p. 18).

Establishing plans to mitigate greenhouse emissions is crucial, but it requires careful monitoring and evaluation for CO<sub>2</sub> emissions forecasting (Al-Fawzan & Dincer, 2021, p. 7). Due to the effectiveness of traditional statistical methodologies like Autoregressive Integrated Moving Average (ARIMA) in addressing time-series predictions, these techniques have garnered significant attention (Box et al, 2015, p. 36). However, as with most traditional analytics, real world datasets are complex, contain non-linearities and these models struggle with such data (Hyndman & Athanasopoulos, 2018, p. 92). Lately, machine learning techniques have gained popularity, especially the Long Short-Term Memory (LSTM) networks, a form of recurrent neural network (RNN), because of their proficiency in capturing non-linear relations and intricate shifts in time (Hochreiter & Schmidhuber, 1997, p. 1736; Greff et al, 2017, p. 2223). The involvement of ARIMA and LSTM in a single framework is essential because, as is the nature of most techniques, one approach cannot solve everything and hybrid models are bound to emerge for this reason (Zhang et al, 2019, p. 3).

This study has developed a new hybrid ARIMA-LSTM model for the prediction of CO<sub>2</sub> emissions in Algeria for the years 2000 – 2030. This model seeks to outperform the accurate but ARIMA bound predictive model and the flexible, but LSTM proficient one as mentioned by (Siami-Namini et al. in 2018). Integrating ARIMA with LSTM in this manner is claimed to improve the overall prediction accuracy by splitting the time series into its linear and non-linear components and enabling each model to do what it does best (Dong et al. in 2020). This is particularly notable since the improvements in accuracy seem to come from infusing old statistical techniques with newer approaches to data science, which is a rapidly growing area of research (Kumar et al., 2021).

With regards to carbon dioxide emissions, Algeria depends significantly on oil and gas extraction, especially with regards to their share of total emissions

(BP Statistical Review of World Energy, 2022, p. 25). The country's industrialization and urbanization over the past twenty years came with an increase in energy consumption and high emission rates (International Energy Agency [IEA], 2020, p. 34). Even if some diversification has happened in the energy mix through renewable resources, the pace of change is slow and greatly insufficient for the Algerian economy (Ministry of Energy and Mines, Algeria, 2021, p. 15). CO<sub>2</sub> emissions need to be examined considering current policies to determine the ease with which targets set by international agreements such as the Paris Agreement could be achieved (UNFCCC, 2015, p. 2).

This hybrid ARIMA-LSTM model is more advantageous than other models in several ways. First, it does not suffer the drawbacks of neither fully statistical nor fully machine learning approaches, as it combines both of them (Zhang et al., 2019, p. 4). Second, it provides an adaptable framework that can easily extend beyond CO<sub>2</sub> emission forecasting to other time-series forecasting problems (Siami-Namini et al., 2018, p. 1396). Lastly, the model's use for Algerian data makes it relevant for developing countries with similar problems of achieving a balance between economic growth and environmental sustainability (World Bank, 2020, p. 20). Focusing on a particular region, this research aims to balance out the literature on climate change mitigation and provide useful information to local decision-makers.

Beyond its methodological contributions, this research places emphasis on the necessity of interdisciplinary integration in solving challenging environmental problems. The integration of environmental science, statistics, and artificial intelligence is one approach to devise new solutions for specific regions. (Mukherjee et al. 2021, p. 3). In addition, this study's results can serve the national and global level policy-makers to respond to climate change with the urgency it deserves (United Nations, 2015, p. 6).

Last but not least, this study is divided into main sections, five in total. Section 2 discusses previously conducted works on CO<sub>2</sub> emission modeling and forecasting, and identifies and explains the challenges of traditional models. Section 3 sketches out the basics of the methodology, which includes the theoretical background of the ARIMA and LSTM models and details the implementation of the hybrid ARIMA-LSTM model. The findings of the study are presented and the results are analyzed in Section 4, where the accuracy and forecasting efficiency of the individual models are compared to the hybrid model. Finally, in Section 5, the study makes its conclusions and, while revisiting the findings, puts forward policies that can enhance emission reduction efforts.

## **2-Literature Review**

The issue of carbon dioxide emissions (CO<sub>2</sub>) is of global concern and requires sophisticated predictive models to assist in climate change mitigation. The emergence of hybrid models such as ARIMA-LSTM has made it possible to combine statistical methods for dealing with linear trends and

artificial neural network methods for interpreting complex non-linear relationships. However, Algeria's emission patterns are considerably under-researched compared to more developed economies, and this is largely due to the exceptional context that Algeria presents with its heavy natural gas dependency, rapid urbanization, and lack of investments into advanced modeling mechanisms.

Al-lami and Török (2025) examined CO<sub>2</sub> emissions from transportation in Central Europe through the Kaya Identity, the Logarithmic Mean Divisia Index approach (LMDI) and ARIMA forecasting. The research involved five countries – Hungary, Czech Republic, Poland, Slovakia, and Austria - for the years 2001 to 2021, with projections going as far as 2050. The emissions were driven by GDP intensity, population emissions intensity, energy intensity, and carbon intensity. The findings revealed that Austria was able to make significant strides in emissions reduction as a result of energy efficiency measures and is expected to reduce emissions by 7.6% in 2050. Poland, on the other hand, had problems stemming from strong dependency on fossil fuel sources with energy intensity expected to increase by 10.2%. Austrian data showed that ARIMA performs very well in quite stable contexts, for example R<sup>2</sup> was approximately 0.9999 while in more volatile situations, i.e. Poland, it did not do as well, having MAPE as high as 196.24%.

Rahman et al. (2024) focused on estimating Bangladesh CO<sub>2</sub> emissions from a natural gas power plant using the ELM model and data from January 2015 to December 2022. They used an IPCC (2006) approach for the estimation, and the preliminary ELM fuel usage forecasts in the study provided projections with decreasing emissions trends. Accuracy of the ELM model was acceptable as it achieved RMSE of 3494.46, MAE of 2013.42, and MASE of 0.93. The model also showed the highest emissions value was in January of 2023 and the lowest value of emissions was in December of 2026. There is a noted gap in this study due to the focus of the model around a single plant which does not account for other factors such as fuel quality and plant running hours, but this study does provide an insight to other gaspowered plants in Bangladesh.

Wen et al. (2023) used a hybrid ARIMA-LSTM to study the CO<sub>2</sub> emissions in China in terms of emissions analysis and forecasting. To enhance its feature, the ESM model incorporated some transportation variables such as freight volume and road mileage because the 14 factors identified on their own were not sufficient. MAPE for the entire country in this model was zero point zero one eight six, which indicates the eclipsed all other models like ARIMA, LSTM, and other models such as linear regression and BPNN. This model was able to capture and provide insights for carbon reduction policies for the regions utilizing both linear and nonlinear patterns.

In developing a method for improving predictions on CO<sub>2</sub> emissions from fuel used in roads vehicles in Cameroon, Sapnken, Noume and Tamba (2023) created the OWTGGM(1,N) model. This model features a wavelet transform, a

noise filtration mechanism, a Hausdorff derivative-based time response function and algorithmic parameter optimization using Rao's method. These estimations performed the OWTGGM(1,N) model on data spanning from 1995 to 2020 and found astounding results, MAPE = 1.27%, and FT score = 99%. While the benchmarks were less favorable, conventional models were greatly outperformed. The estimates clearly stated that emissions would significantly rise from 3,779 kt in 2020 to 4,600 kt by 2030. This advanced CO<sub>2</sub> projection model emphasizes data cleansing and parameter optimization necessary for frameworks for other countries such as Algeria.

Nevertheless, significant barriers remain in predicting Algeria's CO<sub>2</sub> emissions: 1. Geographical Context: Available research concentrates on Central European regions, Asia and countries situated in Sub Saharan Africa. There is little recognition for Algeria, which is heavily reliant on fossil fuel, and has emerging renewable energy strategies. 2. Methodological Limitations: Previous work focuses on linear models like ARIMA or grey models such as GM(1,N), which tend to perform poorly on volatile data, supports their premise. Wen et al.'s ARIMA-LSTM hybrid has, to my knowledge, not been evaluated in any African or oil-dependent countries. 3. Integration of Decomposition and Forecasting: It appears that no one has managed to address two approaches simultaneously to combine decomposition techniques like LMDI with hybrid machine learning which are used to single out important factors before undertaking forecasting. 4. Temporal Scope: The emission trajectory of Algeria from 2000 – 2030, specifically during the critical decades of energy transition (2020s), is still to be holistic.

This study recommends using the ARIMA-LSTM hybrid model for Algeria's CO<sub>2</sub> emission forecast in contexts where the gap exists. The research intends to assist Algeria in meeting its climate obligations under the Paris Agreement by providing the country with the first national specific emission projection that is tailored over the long term. The study also attempts to fill the methodological and contextual gaps in the decarbonization approach of fossil fuel reliant economies with a powerful device for policy makers. This not only aids in superseding voids in emission understanding, it also contains vital information necessary for crafting sustainable energy policies in Algeria.

### **3-Methodology**

#### **3.1 ARIMA Model: Analysis and Forecasting**

The ARIMA model (short for Autoregressive Integrated Moving Average) is a popular statistical method for analyzing and forecasting time series data. Developed by Box and Jenkins in 1970, it brings together four main components (Box & Jenkins, 1970):

##### **1. Autoregressive (AR) Process**

This part of the model suggests that the current value of a time series can be explained by its past values. Mathematically, this is written as (Yasmin&Moniruzzaman, 2024, p. 3):

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + e_t$$

where:

- $Y_t$  is the variable being studied (such as area, production, or yield) at time  $t$ .
- $\alpha_1, \alpha_2, \dots, \alpha_p$  are the coefficients for the past values.
- $e_t$  is the random error term.

## 2. Integrated (I) Component

Sometimes, the data we analyze isn't stationary—it might have trends or other patterns that change over time. To fix this, the ARIMA model uses a process called 'differencing,' which essentially means looking at the differences between consecutive values. The number of times this differencing is applied is indicated by the order  $d$ .

## 3. Moving Average (MA) Process

Instead of using past values directly, this component uses past errors (or 'shocks') to predict the current value. The idea is that the errors from previous predictions can provide valuable information for the current prediction. Mathematically, this is written as:

$$Y_t = e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \dots + \beta_q e_{t-q}$$

where:

- $Y_t$  is the observed value at time  $t$ .
- $e_t$  is the error term at time  $t$ .
- $\beta_1, \beta_2, \dots, \beta_q$  are the coefficients for the past errors.

## 4. ARIMA Model Representation

When you put these components together, the full ARIMA model is typically denoted as  $ARIMA(p, d, q)$ . Its general mathematical form is (Yasmin & Moniruzzaman, 2024, p. 3):

$$(1 - \varphi^1 B - \varphi^2 B^2 - \dots - \varphi_p B^p)(1 - B)^d Y_t = (1 + \theta^1 B + \theta^2 B^2 + \dots + \theta_q B^q) \varepsilon_t$$

where:

- $B$  is the lag operator (a tool for shifting data points in time).
- $\varphi$  and  $\theta$  represent the coefficients for the AR and MA parts, respectively.
- $\varepsilon_t$  is the white noise error term.

The Box-Jenkins methodology involves four iterative stages (Angelo et al., 2023, p. 494):

## 1. Model Identification

Stationarity Check: Use the Augmented Dickey-Fuller (ADF) test to detect unit roots. If non-stationary, apply differencing until stationarity is achieved.

Parameter Selection: The appropriate parameter values for the ARIMA model are determined using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the data. PACF plots help identify the optimal order of the model (Tarmanini et al., 2023, p. 553).

## 2. Model Estimation

Estimate coefficients using  $(\phi, \theta)$  maximum likelihood or non-linear least squares.

Validate models using AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) to balance fit and complexity.

## 3. Diagnostic Checking

Residual analysis is a critical step in validating the model's performance, ensuring that the residuals  $r_t$  approximate white noise. It involves examining the residuals to confirm that their autocorrelations  $\rho_k$  for all lags  $k > 0$  are nearly zero (i.e.,  $(\rho_k \approx 0)$ ). Mathematically, this implies that the residuals follow a normal distribution,  $r_t \sim \mathcal{N}(0, \sigma^2)$ , where  $\sigma^2$  is the error variance. To rigorously assess this property, the Ljung-Box test is used, with the test statistic defined as

$$Q = n(n+2) \sum_{k=1}^h \frac{\widehat{\rho_k^2}}{n-k},$$

where  $n$  is the sample size and  $h$  is the number of lags considered. Ensuring that  $Q$  remains below the critical chi-square value for  $h$  degrees of freedom boosts confidence in the model's forecasts and overall reliability (Tolesh & Biloshchytska, 2024, p. 178).

## 4. Forecasting

In the forecasting stage, which is the final step in ARIMA modeling, we rely on past time series data to predict future values. This is done by combining errors from past forecasts with differentiating observations from historical data, which helps improve the accuracy of forecasts. This process can be expressed by the following equation:

$$\Delta^d Y_t = \phi_0 + \phi_1 \Delta^d Y_{t-1} + \phi_2 \Delta^d Y_{t-2} + \dots + \phi_p \Delta^d Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

To determine the accuracy of forecasts, we use metrics such as RMSE and MAPE, which help measure the quality of forecasts and ensure their reliability to make more accurate decisions. (Siddique et al., 2024, p. 8).

## 3.2. Long short term memory networks (LSTM)

LSTM is an advanced type of recurrent neural network (RNN), designed to solve problems faced by traditional models when processing long temporal data, such as gradient vanishing or exploding. What distinguishes it is its ability to remember important information and forget unnecessary information, thanks to the memory cell and three control mechanisms that help it understand long-term relationships efficiently. This technique first appeared in 1997 thanks to Hochreiter and Schmidhuber, who presented the LSTM model as a solution to the problems that traditional recurrent networks suffered from, which helped greatly improve the processing of temporal data (Hochreiter&Schmidhuber, 1997; H. Xue et al., 2025, p. 4).

### 1. Forget Gate

This gate determines which parts of the previous memory state should be discarded. It is computed as:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

### 2. Input Gate

This gate decides which new information should be added to the cell state. It is computed using:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$C_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

### 3. Cell State Update

The cell state is updated by combining the previous state (after applying the forget gate) with the new candidate values (after the input gate):

$$C_t = f_t C_{t-1} + i_t C_t$$

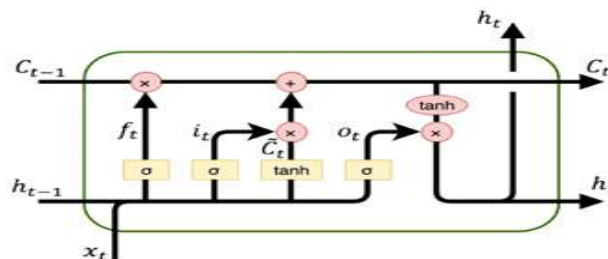
### 4. Output Gate

This gate determines what information is output from the cell. It is computed as:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \tanh(C_t)$$

**Fig.1.**Schematic Representation of an LSTM Network



**Source:**(Qiao et al., 2024, p. 5)

Figure 1 shows a simple diagram of the LSTM cell architecture, clearly illustrating the role of the input, forget, and output gates and their activation

functions, which work together to control the flow of information. This diagram shows how the LSTM cell processes sequential data by determining what information to keep or discard, which helps it effectively understand short- and long-term relationships, which is very important for ensuring accurate time series prediction (Jin et al., 2025, p. 4).

### **3.3.A Hybrid ARIMA +LSTM model**

The time series analysis process using the hybrid ARIMA-LSTM model begins with reading and examining data through the Pandas library, where data is extracted from files like Excel or CSV and displayed in graphical forms to understand its initial patterns, such as seasonal fluctuations or general trends.

Then, the time series is decomposed into its basic components: Trend which shows long-term data movements, Seasonality which reflects cyclical repetitions, and Residuals. Here, it's determined whether these components interact additively (Additive) — where changes are constant — or multiplicatively (Multiplicative) — where fluctuations change as the trend increases. A crucial stage follows to ensure model quality: testing data stationarity through tools like ADF test or Phillips-Perron. If the data is non-stationary (like having a clear upward or downward trend), mathematical transformations (like taking differences) are applied until it becomes stationary, which determines the value of parameter  $d$  in the ARIMA model.

As for parameters  $p$  (number of previous time periods affecting the current value) and  $q$  (number of previous periods for the random element), they are determined through analyzing PACF and ACF plots, or using the `auto_arma` tool which saves considerable time in automatically searching for the best parameters. After training the ARIMA model on training data, its performance is evaluated through metrics like RMSE, MAPE, and MAE, which show how close its predictions are to actual values. But even accurate models leave behind Residuals, and this is where LSTM comes in. These errors are used as inputs for the LSTM network, which processes complex non-linear patterns that ARIMA failed to capture. The optimal number of previous time periods (Lag) affecting the prediction is determined through the ACF plot, then the neural network is built using libraries like Keras and TensorFlow, adjusting parameters like number of layers, batch size, activation function (ReLU), and optimization algorithm (Adam).

Finally, ARIMA predictions (covering linear patterns) are merged with LSTM predictions (handling residual errors) to produce more accurate final predictions. The same error metrics are recalculated to ensure accuracy has improved, confirming the effectiveness of this hybrid approach in dealing with time series data complexities (Lama et al. 2024, p. 3).

#### **4. Results:**

Statistics published on the International Energy Agency (IEA) website through its database on carbon dioxide emissions in Algeria during the period 2000-2023 indicate a significant increase in emissions rates due to economic growth and expansion in the energy and industrial sectors. They recorded their highest levels in 2019 at about 150 million tons of carbon dioxide equivalent, before witnessing a slight decline in 2020 and 2021 as a result of the repercussions of the Covid-19 pandemic on economic activity. However, emissions rose again in 2022 and 2023 with the recovery of the economy and increased consumption of fossil fuels. Energy sector emissions constitute the largest share of total emissions, followed by the industrial and transport sectors. As part of Algeria's commitments under the Paris Climate Agreement, the government seeks to reduce its emissions by diversifying energy sources and expanding the use of renewable energies and low-carbon technologies, according to the database ([www.iea.org/countries/algeria/emissions](http://www.iea.org/countries/algeria/emissions)).

From 2000 to 2023, the data presented in Table 1 shows that the average CO<sub>2</sub> emissions from fuel combustion in Algeria were approximately 106.80 million tons, with a median value close by at 107.69 million tons. This indicates a balanced distribution without significant outliers. The table also reveals a variance of 902.02 and a standard deviation of 30.03 million tons, with a range of 91.49 million tons, indicating noticeable fluctuations in emissions over the years. In terms of distribution shape, the skewness of -0.01 indicates near symmetry, while the kurtosis of -1.41 suggests a flatter distribution compared to the normal one. The Shapiro-Wilk test (with a p-value of 0.0986) further confirms that the data is approximately normally distributed.

**Table 1:** Descriptive Statistics of CO<sub>2</sub> Emissions from Fuel Combustion in Algeria (2000-2023)

| STATISTIC          | VALUE                      |
|--------------------|----------------------------|
| MEAN               | 106.80 million tons        |
| MEDIAN             | 107.69 million tons        |
| VARIANCE           | 864.43                     |
| STANDARD DEVIATION | 29.40                      |
| RANGE              | 91.49                      |
| SKEWNESS           | -0.01                      |
| KURTOSIS           | -1.41                      |
| SHAPIRO-WILK TEST  | Statistic=0.9302, p=0.0986 |

Source: Compiled by the authors

**4.1. Prediction of CO2 emissions by using ARIMA model**

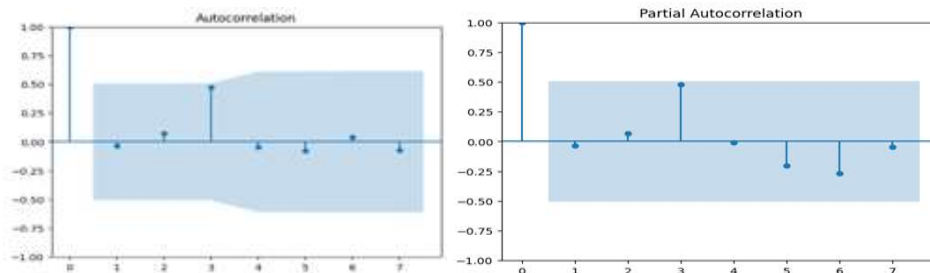
As shown in Table 2, the results of the stationarity tests present mixed conclusions. The Augmented Dickey-Fuller (ADF) test for the original time series indicates non-stationarity, with a test statistic (2.058) greater than the critical values at all significance levels and a p-value (0.9987) far above 0.05. Similarly, the KPSS test result for the original series (0.6335) exceeds the critical values at the 5% and 10% levels, further confirming non-stationarity. After first differencing, the ADF test indicates stationarity with a test statistic of -3.672, which is lower than the critical value at 5% (-3.104) and a p-value of 0.0045. However, the KPSS test still suggests non-stationarity, as the test statistic (0.5346) remains above the 5% critical value (0.463). These results imply that the series may require further differencing or model adjustments to achieve full stationarity.

**Table 2:** Results of Stationarity Tests (ADF and KPSS) Before and After Differencing

| Test                      | Statistic | p-value | Lags Used | Critical Value (1%) | Critical Value (5%) | Critical Value (10%) | Result         |
|---------------------------|-----------|---------|-----------|---------------------|---------------------|----------------------|----------------|
| <b>ADF (Original)</b>     | 2.058     | 0.9987  | 4         | -4.223              | -3.189              | -2.729               | Non-stationary |
| <b>KPSS (Original)</b>    | 0.6335    | 0.0196  | 2         | 0.739               | 0.463               | 0.347                | Non-stationary |
| <b>ADF (Differenced)</b>  | -3.672    | 0.0045  | 0         | -4.012              | -3.104              | -2.691               | Stationary     |
| <b>KPSS (Differenced)</b> | 0.5346    | 0.0339  | 0         | 0.739               | 0.463               | 0.347                | Non-Stationary |

**Source:**Compiled by the authors

From Figure 2, the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots support the selection of an ARIMA(1,1,1) model. The ACF shows a sharp drop after lag 1, indicating a moving average component ( $q=1$ ). Meanwhile, the PACF exhibits a significant spike at lag 1 before cutting off, suggesting the presence of an autoregressive component ( $p=1$ ). The series required first differencing ( $d=1$ ) to achieve stationarity, as previously determined. Therefore, the ARIMA(1,1,1) model is appropriate as it captures both short-term dependencies through AR and MA terms while accounting for the non-stationarity through differencing.

**Fig.2.** Autocorrelation (ACF) and Partial Autocorrelation (PACF) Plots for the Differenced Time Serie

From the ARIMA(1,1,1) model output, shown in Table 3, the estimated coefficients demonstrate statistical significance, indicating a well-fitted model. The autoregressive coefficient (AR1) is 0.9904 with a p-value of 0.000, confirming a strong lag-1 dependence, while the moving average coefficient (MA1) is -0.7276 with a p-value of 0.010, highlighting short-term noise correction. The Ljung-Box Q-statistic (2.70,  $p=0.10$ ) suggests that residuals are uncorrelated, indicating a good fit. Additionally, the Jarque-Bera test (0.26,  $p=0.88$ ) confirms that the residuals are normally distributed. The heteroskedasticity test ( $H=3.34$ ,  $p=0.21$ ) indicates homoscedasticity, further validating the model's suitability for CO2 emissions forecasting.

**Table 3:**ARIMA(1,1,1) Model Estimation Output

| PARAMETER               | COEFFICIENT | STD. ERROR | Z-STATISTIC | P-VALUE | 95% CONFIDENCE INTERVAL |
|-------------------------|-------------|------------|-------------|---------|-------------------------|
| AR(1)                   | 0.9904      | 0.034      | 28.858      | 0.000   | [0.923, 1.058]          |
| MA(1)                   | -0.7276     | 0.283      | -2.567      | 0.010   | [-1.283, -0.172]        |
| VARIANCE ( $\Sigma^2$ ) | 6.1442      | 3.046      | 2.017       | 0.044   | [0.175, 12.113]         |
| MODEL FIT               |             |            |             |         |                         |
| LOG-LIKELIHOOD          | -35.979     |            |             |         |                         |
| AIC                     | 77.959      |            |             |         |                         |
| BIC                     | 80.083      |            |             |         |                         |
| HQIC                    | 77.936      |            |             |         |                         |

|                               |      |      |
|-------------------------------|------|------|
| <b>DIAGNOSTIC TESTS</b>       |      |      |
| <b>LJUNG-BOX Q (L1)</b>       | 2.70 | 0.10 |
| <b>JARQUE-BERA</b>            | 0.26 | 0.88 |
| <b>HETEROSKEDASTICITY (H)</b> | 3.34 | 0.21 |

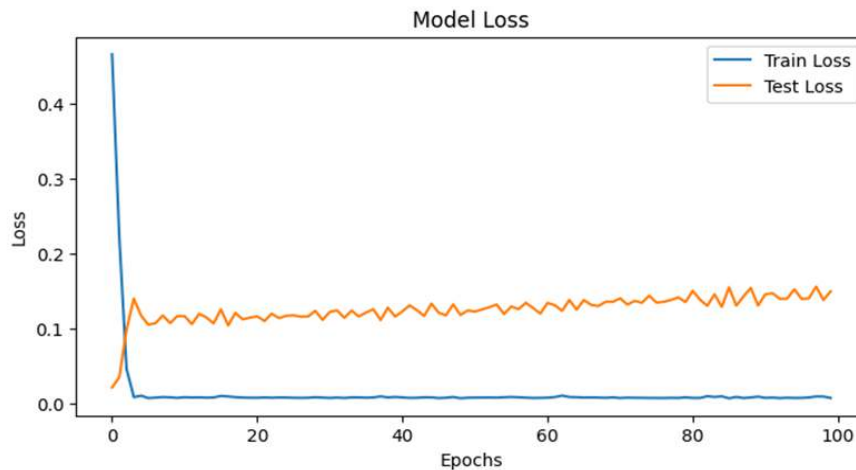
**Source:**Compiled by the authors

#### **4.2. Prediction of CO2 emissions by using LSTM**

The data was divided into two groups: training and testing with a 70/30 split, where the training data consisted of 16 samples and the testing data consisted of 8 samples. A model was adopted to analyze the residual series extracted from the ARIMA model to enhance the model's performance and address the challenges associated with time-series data. The data structure showed that each sample contained 3 time steps and a single feature, reflecting the suitability of this structure for handling multi-step time-series models efficiently. The input shapes were Xtrain(13,3,1) and Xtest(5,3,1), while ytrain and ytest represented the target values corresponding to both the training and testing data.

The RMSE results revealed a clear disparity between the model's performance on training and testing data, with a training loss of 2.45 and a testing loss of 11.58, indicating the presence of overfitting. This discrepancy highlights that the model learned the training data very well but failed to generalize to the testing data, which is a common issue when building time-series models in the absence of proper regularization techniques or when using limited training data.

Figure 3 shows the loss progression over 100 epochs, highlighting a significant gap between the model's performance on training and testing data. In the early stages of training, the training loss dropped sharply from over 0.45 to below 0.05 after the first epoch, indicating the model's rapid learning of the training data. As training continued, the training loss gradually decreased and stabilized at a very low level of approximately 0.01, demonstrating the model's excellent learning of the training data. However, this perfection came at the expense of generalization, as the testing loss remained relatively high throughout the training period.

**Fig.3.** Training and Testing Loss Curve Over 100 Epochs

### 4.3. Prediction of CO2 emissions by using ARIMA +LSTM

In the hybridization phase, which is the stage of combining predictions, the ARIMA predictions (the linear component) are added to the LSTM predictions (the non-linear component) to produce the overall forecast. The main goal of this phase is to obtain a final, comprehensive prediction of CO2 emissions in Algeria. This is achieved by combining the linear predictions from ARIMA with the non-linear predictions from LSTM. This process is based on the assumption that emissions are influenced by both linear and non-linear factors. By merging the two models, we can capture both types of influences. The end result is a more accurate and comprehensive forecast. This forecast takes into account both general trends (linear) and complex fluctuations (non-linear). Consequently, the overall predictive performance is improved. Finally, this combined forecast represents the expected value of future CO2 emissions.

### 4.4. Comparison of prediction results

The table (4) clearly demonstrates the superior performance of the ARIMA+LSTM model compared to the standard ARIMA model in forecasting CO2 emissions. Across all four-evaluation metrics—MAE, MAPE, MSE, and RMSE—the ARIMA+LSTM model exhibits significantly lower values. This indicates a substantial improvement in prediction accuracy, with the hybrid model producing smaller errors in both absolute and percentage terms. The reduction in MAPE to 4.704% for the ARIMA+LSTM model is particularly noteworthy, suggesting a high level of accuracy in capturing the relative changes in CO2 emissions. The lower MSE and RMSE values further confirm that the ARIMA+LSTM model is less prone to large errors and provides a better fit to the actual data.

The consistent improvement across all metrics strongly suggests that the addition of the LSTM component to the ARIMA model enhances its ability to capture complex patterns and long-term dependencies within the CO2 emissions

data. While ARIMA is a robust baseline model, it appears to be less effective at modeling the intricacies of this specific dataset. The LSTM's strength in handling sequential data with long-range correlations likely explains the performance gain. Therefore, based on this analysis, the ARIMA+LSTM model is the preferred choice for forecasting CO2 emissions, offering a more precise and reliable prediction compared to the traditional ARIMA approach.

**Table 4:** Comparison of ARIMA and ARIMA+LSTM Model Performance on CO2 Emissions Forecasting

| Metric      | ARIMA   | ARIMA+LSTM |
|-------------|---------|------------|
| <b>MAE</b>  | 16.071  | 6.720      |
| <b>MAPE</b> | 11.383  | 4.704      |
| <b>MSE</b>  | 292.811 | 83.748     |
| <b>RMSE</b> | 17.112  | 9.151      |

**Source:**Compiled by the authors

**Fig.4.** Actual vs. Predicted CO2 Emissions (ARIMA+LSTM)

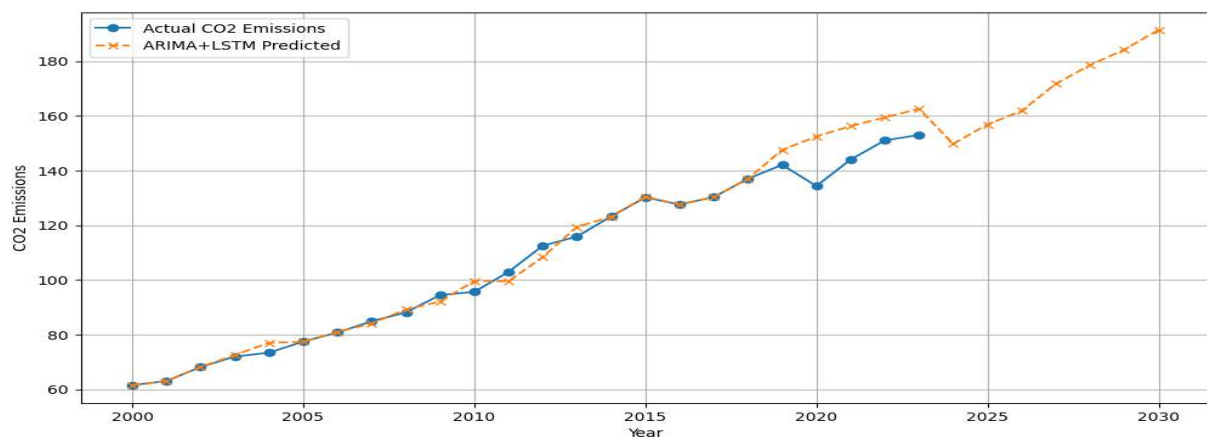


**Table (5):** ARIMA+LSTM Model Forecasts for CO2 Emissions in Algeria (2024-2030)

| Year | CO2 emissions<br>(ARIMA+LSTM) |
|------|-------------------------------|
| 2024 | 149.728                       |
| 2025 | 156.8078                      |

|      |          |
|------|----------|
| 2026 | 161.8126 |
| 2027 | 171.7038 |
| 2028 | 178.5405 |
| 2029 | 184.211  |
| 2030 | 191.3733 |

**Source:**Compiled by the authors

The table5and Figure 5presents the forecast for carbon dioxide emissions in Algeria from 2024 to 2030 using the ARIMA+LSTM model. Emissions start at 149.728 million tons in 2024 and gradually increase to 191.3733 million tons by 2030, reflecting a total rise of approximately 41.645 million tons (about 27.8%), with an annual growth rate of around 4.1%. The yearly increases, ranging between 5.00 and 9.89 million tons, suggest a continuing reliance on traditional energy sources or a rise in industrial activity. This highlights the need for measures to mitigate potential environmental impacts and promote sustainable development.

## **6. Conclusion**

In conclusion, the ARIMA+LSTM model presents a concerning outlook for Algeria's CO<sub>2</sub> emissions, highlighting not only an environmental challenge but also the shortcomings of current policies. Urgent action is needed to reduce carbon emissions before it's too late. While ARIMA+LSTM is a powerful tool for analyzing trends, it is inherently tied to past data. Algeria's present reality—its reliance on fossil fuels, pursuit of economic growth, and vulnerability to climate risks like drought and desertification—must guide every future step.

This is not just about transitioning to renewable energy or improving efficiency, though both are essential. It's about finding solutions tailored to Algeria's unique challenges, such as protecting water resources and curbing desert expansion. To remain relevant and reliable, the model must be continuously updated with local data and a deeper understanding of Algeria's social and economic landscape, ensuring that future policies are both practical and effective for the nation and its people.

## **5.Citations and bibliography list:**

- Al-Fawzan, K. A., & Dincer, I. (2021). Forecasting CO<sub>2</sub> emissions using hybrid models: A review.Renewable and Sustainable Energy Reviews, 143, 110853. <https://doi.org/10.1016/j.rser.2021.110853>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., &Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.

- BP Statistical Review of World Energy. (2022). BP Statistical Review of World Energy 2022. Retrieved from <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2022-full-report.pdf>
- Dong, Z., Wang, J., Zhang, X., & Xu, Y. (2020). A hybrid ARIMA-LSTM model for short-term wind speed prediction. *Energy Conversion and Management*, 221, 113077. <https://doi.org/10.1016/j.enconman.2020.113077>
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2605843>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>
- IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- Kumar, V., Jain, A., & Singh, P. (2021). Hybrid models for time series forecasting: A systematic literature review. *Journal of Ambient Intelligence and Humanized Computing*, 12(7), 6677–6700. <https://doi.org/10.1007/s12652-020-02544-y>
- Ministry of Energy and Mines, Algeria. (2021). *National Renewable Energy Development Strategy*. Retrieved from <http://www.mine.gov.dz>
- Mukherjee, S., Chakraborty, S., & Dutta, P. (2021). Interdisciplinary approaches to climate change mitigation: A review. *Environmental Research Letters*, 16(5), 053001. <https://doi.org/10.1088/1748-9326/abf0d5>
- OPEC. (2022). *OPEC Annual Statistical Bulletin 2022*. Retrieved from [https://www.opec.org/opec\\_web/static\\_files\\_project/media/downloads/publications/ASB2022.pdf](https://www.opec.org/opec_web/static_files_project/media/downloads/publications/ASB2022.pdf)
- Siامي-ناميني, S., Tavakoli, N., & Namini, A. S. (2018). A comparison of ARIMA and LSTM in forecasting time series. *2018 IEEE 17th International*

Conference on Machine Learning and Applications (ICMLA), 1394–1401.  
<https://doi.org/10.1109/ICMLA.2018.00227>

- United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. Retrieved from <https://sdgs.un.org/2030agenda>
- UNFCCC. (2015). Paris Agreement. Retrieved from [https://unfccc.int/sites/default/files/english\\_paris\\_agreement.pdf](https://unfccc.int/sites/default/files/english_paris_agreement.pdf)
- World Bank. (2020). Algeria Economic Update: Toward a Resilient Recovery. Retrieved from <https://openknowledge.worldbank.org/handle/10986/34174>
- Zhang, W., Zhang, H., & Huang, Q. (2019). A hybrid ARIMA-LSTM model for wind power prediction. *Applied Energy*, 254, 113628.  
<https://doi.org/10.1016/j.apenergy.2019.113628>
- International Energy Agency [IEA]. (2020). World Energy Outlook 2020. Retrieved from <https://www.iea.org/reports/world-energy-outlook-2020>
- Al-lami, A., & Török, Á. (2025). Regional forecasting of driving forces of CO<sub>2</sub> emissions of transportation in Central Europe: An ARIMA-based approach. *Energy Reports*, 13, 1215–1224. <https://doi.org/10.1016/j.egyr.2025.01.004>
- Rahman, M., Rashid, F., Roy, S. K., & Habib, M. A. (2024). Application of extreme learning machine (ELM) forecasting model on CO<sub>2</sub> emission dataset of a natural gas-fired power plant in Dhaka, Bangladesh. *Data in Brief*, 54, 110491.  
<https://doi.org/10.1016/j.dib.2024.110491>
- Qiao, Q., Eskandari, H., Saadatmand, H., & Sahraei, M. A. (2024). An interpretable multi-stage forecasting framework for energy consumption and CO<sub>2</sub> emissions for the transportation sector. *Energy*, 286, 129499.  
<https://doi.org/10.1016/j.energy.2023.129499>
- Wen, T., Liu, Y., Bai, Y. H., & Liu, H. (2023). Modeling and forecasting CO<sub>2</sub> emissions in China and its regions using a novel ARIMA-LSTM model. *Heliyon*, 9 (e21241). <https://doi.org/10.1016/j.heliyon.2023.e21241>
- Sapnken, F. E., Noume, H. C., & Tamba, J. G. (2023). Forecasting CO<sub>2</sub> emissions from road fuel combustion using grey prediction models: A novel approach. *MethodsX*, 11, 102271. <https://doi.org/10.1016/j.mex.2023.102271>
- Angelo, M. D., Fadhilrahman, I., & Purnama, Y. (2023). Comparative analysis of ARIMA and Prophet algorithms in Bitcoin price forecasting. *Procedia Computer Science*, 227, 490–499. <https://doi.org/10.1016/j.procs.2023.10.550>

- G.E.P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden –Day, San Francisco, 1970.
- Tolesh, F., & Biloshchytska, S. (2024). Forecasting international migration in Kazakhstan using ARIMA models. *Procedia Computer Science*, 231, 176–183. <https://doi.org/10.1016/j.procs.2023.12.190>
- Tarmanini, C., Sarma, N., Gezegin, C., & Ozgonenel, O. (2023). *Short term load forecasting based on ARIMA and ANN approaches*. *Energy Reports*, 9, 550–557. <https://doi.org/10.1016/j.egy.2023.01.060>
- Yasmin, S., & Moniruzzaman, M. (2024). *Forecasting of area, production, and yield of jute in Bangladesh using Box-Jenkins ARIMA model*. *Journal of Agriculture and Food Research*, 16, 101203. <https://doi.org/10.1016/j.jafr.2024.101203>
- Siddique, M. A. B., Mahalder, B., Haque, M. M., Shohan, M. H., Biswas, J. C., Akhtar, S., & Ahammad, A. K. S. (2024). Forecasting of tilapia (*Oreochromis niloticus*) production in Bangladesh using ARIMA model. *Heliyon*, 10, e27111. <https://doi.org/10.1016/j.heliyon.2024.e27111>
- Jin, X., Wei, F., Kandala, S. S., Umesh, T., Steele, K., Galgiani, J. N., & Laubichler, M. D. (2025). Time series forecasting of Valley fever infection in Maricopa County, AZ using LSTM [PDF]. *The Lancet Regional Health - Americas*, 43, 101010. <https://doi.org/10.1016/j.lana.2025.101010>
- Lama, A., Ray, S., Biswas, T., Narsimhaiah, L., Raghav, Y. S., Kapoor, P., Singh, K. N., Mishra, P., & Gurung, B. (2024). *Stepwise procedure*. In *Python code for modeling ARIMA-LSTM architecture with random forest algorithm* (p. 3). *Software Impacts*, 20, 100650. <https://doi.org/10.1016/j.simpa.2024.100650>