International Journal of
Energy Research

Research Article

Forecasting Renewable Energy Consumption in Angola, Canada, France, and Nigeria Using ARIMA and Grey-Box Hybrid Models.

Submission ID a4d9ec22-f722-40e6-872f-dda7c13b9fea

Submission Version Initial Submission

PDF Generation 15 Dec 2025 15:03:11 EST by Atypon ReX

Authors

Mr. Abiahu, O.C.
Corresponding Author
Submitting Author

ORCiD
https://orcid.org/0009-0005-4008-5437

Affiliations
- Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria

Dr. Bartholomew, D.C.

Affiliations
- Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria

Dr. Iwu, H.C.

Affiliations
- Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria

Mr. Onyeawugosi, K.T.

Affiliations
- Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria

Files for peer review

All files submitted by the author for peer review are listed below. Files that could not be converted to PDF are indicated; reviewers are able to access them online.

Name	Type of File	Size	Page
Manuscript.docx	Main Document - MS Word	959.3 KB	

Forecasting Renewable Energy Consumption in Angola, Canada, France, and Nigeria Using ARIMA and Grey-Box Hybrid Models.

[1]Abiahu, O.C., [2]Bartholomew, D.C., [3]Iwu, H.C., [4]Onyeawugosi, K.T.

Department of Statistics, Federal University of Technology, P. M. B. 1526, Owerri, Nigeria.

Emails:[1]abiahuezenwa@gmail.com;[2]desmond.bartholomew@futo.edu.ng,[3]hycinth.iwu@futo.edu.ng,
[4]Kingtochi0000@gmail.com

ABSTRACT

Accurate forecasting of renewable energy consumption is essential for effective energy planning and the transition toward sustainable energy systems. However, traditional statistical forecasting models often struggle to capture the nonlinear and heterogeneous dynamics present in renewable energy consumption data, particularly in developing economies. This study investigates the comparative performance of conventional Autoregressive Integrated Moving Average (ARIMA) models and a hybrid Grey-Box modelling framework that integrates ARIMA as a white-box component with Random Forest as a black-box component. The analysis focuses on renewable energy consumption in Angola, Canada, France, and Nigeria. For each country, optimal ARIMA models were selected using information criteria and diagnostic testing. The selected specifications were ARIMA (0,2,2) for Angola, ARIMA (0,2,1) for both Canada and France, and ARIMA (1,1,1) for Nigeria. While these models effectively captured linear trends and short-term temporal structures, their performance was limited in the presence of nonlinear patterns and data irregularities. To address these limitations, a Grey-Box model was developed by modelling ARIMA residuals using a Random Forest algorithm, thereby enhancing predictive flexibility and robustness. Model performance was evaluated using Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). Across all four countries, the Grey-Box model substantially outperformed the standalone ARIMA models. In Angola, MSE decreased from 214.775 to 2.1032 and MAPE from 0.0871 to 0.0195. Canada recorded a reduction in MSE from 20.2436 to 0.0615 and MAPE from 0.0652 to 0.0088. France showed a decline in MSE from 4.5729 to 0.0789 and MAPE from 0.0899 to 0.0196, while Nigeria's MSE reduced from 242.905 to 0.5178 and MAPE from 0.0457 to 0.0069. Short-term forecasts for the period 2022–2026 indicate heterogeneous consumption trajectories across countries, with Angola exhibiting the strongest growth, Canada and France showing steady increases, and Nigeria maintaining a high but relatively stable consumption level. Overall, the findings demonstrate that the Grey-Box modelling framework provides superior forecasting accuracy by effectively combining linear statistical structure with nonlinear machine-learning capability. The study highlights the value of hybrid models for renewable energy forecasting, particularly in data-constrained and structurally complex energy systems.

Keywords: Renewable energy forecasting; ARIMA modeling, Grey-Box hybrid model; Random Forest algorithm; Time series prediction; Energy consumption analysis

INTRODUCTION

For the past few decades, renewable energy has gained dominance in the sector of global energy systems. This is driven by the need to reduce carbon emissions, mitigate climate change, and ensure long-term energy sustainability. The CER (Canada Energy Renewable) considers energy to be renewable if it is derived from natural processes that can be replenished at a rate that is equal to, or faster than, the rate at which they are consumed According to [1], renewable is expected to account for nearly 95% of the increase in global power capacity through 2026. There are different kinds of Renewable energy: this primarily includes solar, wind, hydro, biomass, and geothermal energy. Each type is utilized differently across regions, depending on geographical advantages, technological capacity, and policy frameworks. Agreements internationally like [2], have amplified efforts by countries to switch from fossil fuels to renewable energy sources. This study focuses on four countries: Canada, Angola, France and Nigeria. Countries like Canada and France as one of the developed countries that have been at the forefront, while Angola and Nigeria, as part of sub-Saharan Africa, faces distinct challenges. The [3] has shown that Angola's energy consumption is dominated by hydroelectric power due to its abundant water resources, making hydropower a key part of the country's energy matrix. However, there is limited penetration of other renewable sources like solar and wind. Where according to [4], Canada is a leader in renewable energy, with hydroelectric power providing the largest share of its energy. Other forms, such as wind and solar, are growing in significance, bolstered by Canada's technological infrastructure and favorable policies. Reviewed in [5], France on the other hand is a leader globally in renewable energy adoption, especially with nuclear energy, which, although not renewable, is a critical low-emission source. In recent years, solar and wind energy have seen increased investment. While according to [6], Nigeria, with its heavy dependence on oil and gas, is beginning to explore renewable energy options. While hydropower has been in use, solar energy has seemed growing in the recent times, particularly in rural electrification programs. Hence, the problem this research study seeks to solve is to understand the renewable energy consumption predicting accuracy when only traditional or mechanistic models are applied independently, and to when hybrid models are applied in the prediction. This study aims to evaluate and compare the accuracy of ARIMA and Grey-Box models for forecasting renewable energy consumption in Angola, Canada, France, and Nigeria. The objectives are to: (1) Fit individual models to country-specific data. (2) Compare their predictive performance. (3) Validate model reliability using diagnostic metrics and (4) Provide short-term forecasts to guide energy policy.

Renewable energy has emerged as an important solution to addressing the challenges posed by climate change and energy security; as a result, some literatures have been published on renewable energy, also on ARIMA prediction, Grey-Box modeling, as well as hybrid models. [7] introduced a novel framework for renewable energy forecasting by distinguishing between direct and indirect forecasting methods, categorizing techniques into physical, statistical, machine learning, and hybrid approaches. Direct methods, such as machine learning models, predict energy output directly from input data, while indirect methods leverage physics-based models to forecast intermediate variables like weather conditions before estimating energy consumption. Their framework emphasizes how integrating physics-based insights from indirect methods can enhance the training of data-driven direct methods, improving predictive accuracy across varying data availability and forecasting horizons. This aligns with the hybrid Grey-Box approach adopted in this study, which combines ARIMA's linear modeling with Random Forest's ability to capture non-linear patterns. While [7] provide a conceptual structure for forecasting, their work lacks empirical comparisons of model performance. In contrast, our study validates the hybrid approach with significant reductions in MSE and MAPE across diverse datasets from Angola, Canada, France, and Nigeria, demonstrating practical applicability in renewable energy forecasting.

[8] proposed a dihybrid recurrent neural network (RNN) model combining Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures to forecast solar radiation across six Nigerian cities using 31 years of meteorological data, including rainfall, humidity, and temperature. Through hyperparameter tuning, their model achieved superior accuracy and lower prediction errors compared to standalone LSTM or GRU models, demonstrating the power of hybrid neural networks for time series forecasting in renewable energy applications. This work aligns with the hybrid modeling trend adopted in this study, where the Grey-Box model integrates ARIMA's linear modeling with Random Forest's non-linear capabilities to forecast renewable energy consumption. However, the LSTM-GRU model's computational complexity and data requirements may limit its applicability in data-scarce contexts like Angola. In contrast, our Grey-Box approach offers a balance of interpretability and robustness, achieving significant MSE and MAPE reductions across diverse datasets from Angola, Canada, France, and Nigeria, thus providing a practical and efficient solution for energy forecasting.

[9] investigated the use of deep neural networks (DNNs), including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for forecasting solar irradiance and power output to enhance renewable energy grid integration in Nigeria. Utilizing extensive historical weather data, satellite imagery, and real-time solar measurements, the study employed data preprocessing, feature extraction, and ensemble learning to improve prediction accuracy. The results showed that DNNs outperformed traditional statistical

methods by effectively capturing complex non-linear relationships in solar energy data. This work underscores the potential of advanced machine learning for renewable energy forecasting, aligning with the hybrid modeling approach in this study, which integrates ARIMA's linear modeling with Random Forest's non-linear capabilities in a Grey-Box framework. However, [9] DNNs require substantial computational resources and large datasets, which may pose challenges in data-scarce regions like Angola. In contrast, our Grey-Box model balances computational efficiency and interpretability, achieving significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria, offering a practical alternative for diverse forecasting contexts.

[10] evaluated the performance of Seasonal Autoregressive Integrated Moving Average (SARIMA) models for forecasting electricity generation from renewable sources (biomass, hydropower, and wind) in Brazil during the initial months of the COVID-19 pandemic. The study found that SARIMA models outperformed standard ARIMA by effectively capturing seasonal patterns influenced by climatic factors, such as droughts, which are critical for renewable energy generation. This highlights the strength of statistical models like ARIMA and SARIMA in modeling linear and seasonal trends in energy data. However, SARIMA's reliance on linearity limits its ability to handle non-linear dynamics, a gap addressed in this study by the Grey-Box model, which integrates ARIMA's linear capabilities with Random Forest's ability to capture non-linear patterns. While [10] focus on Brazil's renewable generation during a pandemic provides context-specific insights, our Grey-Box approach demonstrates broader applicability, achieving significant MSE and MAPE reductions across diverse datasets from Angola, Canada, France, and Nigeria, offering a robust solution for forecasting renewable energy consumption in varied socio-economic and climatic conditions.

[11] conducted a comprehensive review of forecasting methods for renewable energy production, focusing on photovoltaic and wind power. The study synthesizes state-of-the-art techniques, categorizing them based on climatic variables, optimization algorithms, pre-processing methods, and forecasting horizons. It highlights the integration of advanced methods, such as machine learning and hybrid models, with optimization and data pre-processing to enhance forecast accuracy and stability. This aligns with the hybrid approach in this study, where the Grey-Box model combines ARIMA's linear modeling with Random Forest's ability to capture non-linear patterns. While [11] provide a broad overview of forecasting techniques, their review lacks specific empirical comparisons of model performance across diverse datasets. In contrast, our study validates the Grey-Box model's effectiveness through significant MSE and MAPE reductions in forecasting renewable energy consumption across Angola, Canada, France, and Nigeria, offering a practical and computationally efficient solution for varied climatic and socio-economic contexts.

[12] analyzed the global development of solar energy, noting a 15% increase in solar electricity production in 2020, with significant contributions from Asia (India and China) and North America (USA). The study evaluates factors affecting solar energy forecasting, emphasizing regional climatic conditions (e.g., solar radiation intensity) as critical to economic efficiency. It discusses the environmental benefits of solar energy, such as reducing conventional fuel use by 10–20 million tons per Gcal, but highlights challenges like low photovoltaic cell efficiency (requiring 60–65% improvement) and high production costs (USD 250–450 per MWh). The authors also review forecasting methods, referencing IPCC, IEA, and Solar Power Europe projections, but do not specify models like ARIMA or machine learning. This contrasts with our study's Grey-Box model, which integrates ARIMA's linear forecasting with Random Forest's non-linear capabilities, achieving robust MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria. While Sribna et al. underscore the importance of regional factors, our hybrid approach offers a versatile solution for diverse climatic and economic contexts, enhancing forecasting accuracy and practical applicability.

[13] analyzed the development of renewable energy in the European Union, focusing on forecasting the share of renewable sources (wind, solar, biofuels, geothermal, and hydropower) in total final energy

consumption by 2030 for selected EU countries. Utilizing Eurostat data, the study examines energy balances, levels of energy independence, and electricity prices, highlighting the impact of the EU's New Green Deal on renewable energy adoption. While the paper employs statistical analyses to project future trends, it does not specify the forecasting models used, limiting insights into methodological advancements. This contrasts with our study's Grey-Box model, which integrates ARIMA's linear forecasting with Random Forest's non-linear capabilities to achieve robust predictions of renewable energy consumption. [13] focus on EU policy and diverse energy mixes complements our work, which demonstrates significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria, offering a versatile hybrid approach adaptable to varying regional and climatic conditions.

[14] investigated the impact of weather data transformations on the accuracy of energy time series forecasting for electrical demand, solar power, and wind power. The study compared station-based and grid-based weather data, applying transformations such as statistical features, dimensionality reduction, clustering, autoencoders, and interpolation. Results showed that these transformations improved forecast accuracy by 3.7–5.2% compared to raw weather data, with statistical and dimensionality reduction methods performing best. This highlights the importance of data preprocessing in enhancing forecasting models, particularly for renewable energy systems sensitive to weather variability. While [14] focus on preprocessing techniques, our study employs a Grey-Box model integrating ARIMA's linear modeling with Random Forest's non-linear capabilities, achieving significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria. Our approach complements their findings by leveraging Random Forest to capture non-linear weather-related patterns without requiring complex data transformations, offering a simpler yet effective solution for diverse forecasting contexts.

[15] proposed a one-dimensional Convolutional Neural Network (1-D CNN) model to forecast photovoltaic (PV) and wind energy generation in microgrids, using data from the University of California, San Diego microgrid and San Diego Airport weather records. The study demonstrated significant improvements in forecasting accuracy, achieving up to 229.8 times lower Mean Squared Error (MSE) and 24.47 times lower Mean Absolute Error (MAE) compared to traditional statistical methods. This highlights the strength of machine learning, particularly CNNs, in capturing complex patterns for short-term renewable energy forecasting. While their approach excels in microgrid applications, its reliance on large datasets and computational resources may limit applicability in data-scarce regions like Angola. In contrast, our Grey-Box model, integrating ARIMA's linear forecasting with Random Forest's non-linear capabilities, offers a computationally efficient and interpretable solution, achieving significant MSE and MAPE reductions across diverse datasets from Angola, Canada, France, and Nigeria, making it suitable for broader renewable energy forecasting contexts.

[16] proposed a transformer-based hybrid forecasting model that leverages attention mechanisms and deep learning architectures to enhance the accuracy of renewable energy forecasting. Evaluated on real-world datasets from various renewable sources, the model outperformed traditional statistical and simpler machine learning methods by effectively capturing complex multivariate patterns. This underscores the potential of advanced machine learning, particularly transformer-based approaches, in addressing the variability of renewable energy data. While [16] model excels in precision, its computational complexity may limit its applicability in resource-constrained settings. In contrast, our Grey-Box model, integrating ARIMA's linear forecasting with Random Forest's non-linear capabilities, offers a computationally efficient and interpretable solution, achieving significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria. Our approach balances accuracy and practicality, making it suitable for diverse forecasting contexts while aligning with the trend toward hybrid models highlighted by Olabiyi.

[17] provided a comprehensive overview of the state-of-the-art in wind and solar energy forecasting, discussing statistical and physical modeling approaches for time scales from minutes to days ahead,

including both deterministic and probabilistic forecasting. The study highlights emerging trends, such as improved forecast skill through advanced methods and the need for tailored probabilistic forecast products to support decision-making in renewable energy systems. It also explores future directions, including the potential of blockchain for data transactions and new forecasting products to ensure power system stability under high renewable penetration. While their review emphasizes the evolution of forecasting techniques, it lacks specific empirical comparisons. In contrast, our Grey-Box model, integrating ARIMA's linear forecasting with Random Forest's non-linear capabilities, achieves significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria, offering a practical and efficient hybrid approach that aligns with the trend toward advanced, user-focused forecasting solutions outlined by Sweeney et al.

[18] conceptualizes weather as the "fuel" for renewable energy sources like wind and solar, emphasizing their dependence on variable conditions such as windiness and cloudiness. Using a thought experiment, the study explores the complexity of maintaining reliable power on grids dominated by renewables, framing the transition to renewable energy as a "wicked problem" due to its multifaceted challenges, including forecasting accuracy and grid stability. While the paper does not focus on specific forecasting models, it underscores the critical role of weather forecasting in renewable energy systems, aligning with the need for robust forecasting methods. In contrast, our study employs a Grey-Box model, integrating ARIMA's linear forecasting with Random Forest's non-linear capabilities, achieving significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria. Our hybrid approach addresses the forecasting challenges highlighted by Seitter, offering a practical solution for reliable renewable energy predictions in diverse climatic conditions.

[19] developed a deep learning model based on Long Short-Term Memory (LSTM) techniques to forecast solar electricity generation, aiming to optimize solar energy management and enhance smart grid efficiency. The model was evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2, demonstrating superior performance compared to a persistent model, particularly for forecasting horizons beyond two hours. This highlights the capability of AI-based methods to improve grid planning and reliability by capturing complex temporal patterns in solar data. While their LSTM approach excels in accuracy, its computational demands may pose challenges in resource-limited settings. In contrast, our Grey-Box model, integrating ARIMA's linear forecasting with Random Forest's non-linear capabilities, achieves significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria, offering a more computationally efficient and interpretable solution for renewable energy forecasting in diverse contexts.

[20] investigated solar energy forecasting using a TensorFlow-based sequential machine learning algorithm, focusing on the impact of weather parameters such as air temperature, relative humidity, and dew point on photovoltaic (PV) generation. The study employed variable selection techniques to identify these parameters as key predictors, addressing the challenges of solar power's intermittency due to cloud cover and atmospheric conditions. The approach enhances forecasting accuracy for smart grid energy management, though specific performance metrics were not detailed. While [20] model leverages machine learning to capture complex patterns, its reliance on TensorFlow may increase computational demands. In contrast, our Grey-Box model, integrating ARIMA's linear forecasting with Random Forest's non-linear capabilities, achieves significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria, offering a computationally efficient and interpretable solution for renewable energy forecasting in diverse contexts.

[21] explored solar power forecasting using machine learning techniques, focusing on the impact of environmental factors such as irradiance, humidity, PV surface temperature, and wind speed on photovoltaic (PV) system output. The study compared regression models, including Support Vector Machine Regressor (SVMR), Random Forest Regressor (RFR), and Linear Regression, with RFR outperforming others in

accuracy for time series forecasting. This highlights the effectiveness of Random Forest in capturing complex, non-linear patterns in weather-dependent solar data. Our study aligns with their findings, as our Grey-Box model integrates ARIMA's linear forecasting with Random Forest's non-linear capabilities, achieving significant MSE and MAPE reductions across datasets from Angola, Canada, France, and Nigeria. While Anuradha et al. focus on specific machine learning models, our hybrid approach offers a balanced, interpretable solution for renewable energy forecasting across diverse climatic and operational contexts.

Summary of Related Studies

Author (Year)	Model Used	Dataset / Scope	Key Findings	Relevance to Present Study
Van Poecke et al. (2023)	Conceptual hybrid (direct vs indirect)	Global framework	Highlights hybrid forecasting structure	Supports hybrid Grey-Box concept
Alabi & Ojenike (2024)	LSTM–GRU hybrid	31-year solar data	Hybrid RNN outperforms standalone models	Supports hybridization trend
Bassey (2023)	CNN, RNN	Nigerian solar dataset	DNNs outperform statistical models	Shows ML strength in non-linear patterns
Reichert et al. (2022)	SARIMA	Brazil renewables	Captures seasonality well	Supports ARIMA's linear strengths
Teixeira et al. (2023)	Review	PV & wind	ML and hybrids improve accuracy	Aligns with study's hybrid approach
Sribna et al. (2021)	Statistical review	Global solar output	Climatic factors affect forecasting	Reinforces multi-country approach
Manowska (2021)	Statistical projections	EU countries	Forecasts renewable share	Highlights need for methodological rigor
Neumann et al. (2023)	Preprocessing & ML	Weather data	Preprocessing improves accuracy	RF captures patterns without heavy preprocessing
Sudasinghe et al. (2025)	1-D CNN	PV & wind	Deep learning achieves highest accuracy	Shows ML advantage (but costly)

Olabiyi (2023)	Transformer-based hybrid	Multi-source renewable	Outperforms simpler models	Hybrid trend supports Grey-Box choice
Sweeney et al. (2020)	Technical review	Solar & wind	Highlights progress but lacks comparison	This study fills comparison gap
Seitter (2024)	Conceptual analysis	Global renewables	Weather variability complicates forecasting	Supports robust hybrid modeling
Bouquet et al. (2023)	LSTM	Smart-grid solar	Strong accuracy but high computational cost	Grey-Box offers efficient alternative
Yeramolu (2021)	TensorFlow sequential ML	PV weather data	Temperature & humidity key predictors	RF handles nonlinear predictors well
Anuradha et al. (2021)	RF, SVM, LR	PV output	RF outperforms SVM & LR	Supports RF choice for hybrid model
Present Study	ARIMA–Random Forest Hybrid	Angola, Canada, France, Nigeria	Significant reductions in MSE & MAPE	Provides direct model comparison

The review of existing literature shows that a wide range of forecasting techniques—ranging from traditional statistical models such as ARIMA and SARIMA to advanced machine-learning and deep-learning approaches—have been applied to renewable energy forecasting. Several studies have also proposed hybrid or Grey-Box frameworks that integrate statistical and machine-learning methods to improve prediction accuracy. However, despite this growing body of work, two important gaps remain evident.

First, although Random Forest has consistently demonstrated strong performance in modeling complex and non-linear renewable energy patterns, it has been relatively underexplored within Grey-Box forecasting frameworks that explicitly combine Random Forest with ARIMA. Most hybrid studies focus on neural-network-based approaches, such as LSTM, GRU, CNN, or transformer architectures, which often require large datasets, extensive tuning, and substantial computational resources. Very few studies have systematically evaluated Random Forest as a residual-learning component in an ARIMA-based Grey-Box model, despite its robustness, interpretability, and suitability for small or noisy datasets.

Second, existing studies rarely conduct a **direct and systematic comparison** between traditional statistical models and machine-learning-based Grey-Box models using the same datasets and evaluation metrics. Instead, many studies either focus on a single modeling approach or propose hybrid methods without benchmarking them against standalone statistical models. This limits understanding of whether hybrid models genuinely offer superior predictive performance or simply introduce additional complexity. Moreover, comparative analyses across countries with differing energy structures and levels of economic development remain particularly scarce.

This study addresses these gaps by making a clear and focused contribution to the literature. Specifically, it conducts a direct comparison between ARIMA and a Random Forest–based Grey-Box model for forecasting renewable energy consumption in four countries: Angola, Canada, France, and Nigeria. The Grey-Box framework integrates ARIMA as a white-box component to model linear temporal dynamics and Random Forest as a black-box component to capture non-linear residual patterns. Model performance is rigorously evaluated using Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R^2, allowing for an objective assessment of predictive accuracy and robustness.

The selection of Angola, Canada, France, and Nigeria is motivated by their contrasting energy profiles, levels of economic development, and renewable energy trajectories. Canada and France represent developed economies with relatively mature renewable energy systems and stable data structures, while Angola and Nigeria represent developing economies where renewable energy deployment is evolving and data irregularities are more pronounced. This diversity enables the study to assess model performance across heterogeneous contexts and enhances the generalizability of the findings.

In addition, the study aligns with global sustainable development priorities, particularly **Sustainable Development Goal 7 (Affordable and Clean Energy)** and **Sustainable Development Goal 13 (Climate Action)**. Recent studies emphasize that accurate and reliable energy forecasting is critical for achieving SDG 7 by supporting efficient planning, investment, and integration of renewable energy sources, especially in developing countries (e.g., [22]). Furthermore, improved forecasting supports SDG 13 by enabling data-driven strategies to reduce emissions and manage climate-related energy risks (Hache & Palle, 2023). By providing empirical evidence on the comparative performance of ARIMA and Random Forest Grey-Box models across diverse national contexts, this study contributes practical insights that support informed renewable energy planning and sustainable energy transitions.

2 Materials and Methods

In this section, the traditional model (ARIMA) and the machine learning approach (Grey-Box) for modeling time series data on percentage renewable energy consumption will be fully discussed. The models are highlighted in 2.1.

2.1 Data Source and Preprocessing

Annual renewable-energy consumption data for Angola, Canada, France, and Nigeria were obtained from the World Bank's World Development Indicators (WDI), covering the period from 1990 to 2022. These data, measured as the share of renewable energy in total final energy consumption, were first inspected to ensure completeness and to remove structural inconsistencies. Each country's time series was assigned a proper chronological index using a yearly DateRange to maintain temporal ordering throughout the analysis.

Preliminary exploratory analysis was carried out to understand the historical patterns in the data. This involved generating descriptive statistics, plotting the time-series trajectories, and visually inspecting long-term trends. To determine whether the series satisfied stationarity conditions, the Augmented Dickey–Fuller (ADF) test was applied. Series that were found to be non-stationary were differenced once, and the differenced series was re-tested to confirm stationarity. Autocorrelation (ACF) and partial autocorrelation (PACF) functions were then examined for both the original and differenced series, providing guidance for selecting appropriate ARIMA model orders.

ARIMA Model Specification

The Autoregressive Integrated Moving Average (ARIMA) model was employed as the traditional linear forecasting approach. Model identification followed the standard Box–Jenkins methodology. Once the differencing order d was established through the ADF test, possible values of the autoregressive order p and moving average order $q$were inferred from the ACF and PACF structures. Several candidate specifications, including ARIMA(1,1,0), ARIMA(1,1,1), and ARIMA(2,1,2), were then estimated for each country.

Model selection was carried out by estimating several ARIMA specifications and comparing them using multiple information and diagnostic criteria. The primary goal was to identify the model that achieved the best balance between goodness-of-fit and parsimony. Accordingly, each candidate ARIMA model was evaluated using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan–Quinn Information Criterion (HQIC), with lower values indicating a more optimal trade-off between model complexity and fit. In addition to these information criteria, the estimated innovation variance (Sigma) and the number of statistically significant parameters were also examined to ensure that the selected model was both stable and interpretable. The general ARIMA representation is expressed as:

$$\phi(B)(1-B)^d y_t = \theta(B)\epsilon_t, \quad (1)$$

where $\phi(B)$ and $\theta(B)$ denote the autoregressive and moving-average polynomials respectively, B is the backshift operator, and ϵ_t is a white-noise error term. The final ARIMA model for each country was chosen based on the combination of these diagnostic criteria, ensuring that the selected specification provided the most statistically reliable representation of the linear structure in the data.

Random Forest Model for Non-Linear Structure

While ARIMA models capture linear dependencies within a time series, they are limited in detecting non-linear patterns. To address this limitation, Random Forest (RF) regression was adopted as a complementary machine-learning approach capable of modelling nonlinear residual dynamics. The Random Forest algorithm is an ensemble of decision trees constructed through bootstrap aggregation. Its prediction function can be expressed as:

$$\hat{f}(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x), \quad (2)$$

where $T_m(x)$ denotes the prediction of the m^{th} regression tree and M is the number of trees in the ensemble.

In this study, Random Forest was trained on the **residuals** obtained from the fitted ARIMA model. The rationale is that ARIMA handles the linear component:

$$\hat{y}_t^{ARIMA}, \quad (3)$$

while Random Forest learns the structure in the remaining non-linear component:

$$e_t = y_t - \hat{y}_t^{ARIMA}. \quad (4)$$

To incorporate temporal dependence into the machine-learning component, the ARIMA residuals were transformed into lag-based predictors. Specifically, for each time point, the feature matrix consisted of the previous three lagged residuals:

$$X_t = \{e_{t-1}, e_{t-2}, e_{t-3}\}. \quad (5)$$

A lag window of three was chosen because it effectively captured short-term fluctuations without causing unnecessary data loss. Missing values arising from lagging were replaced with column means to preserve data continuity.

Random Forest was then trained using 100 trees. The default depth was retained, given the modest size of the dataset and the algorithm's built-in ability to avoid overfitting through bootstrap sampling.

Hybrid Grey-Box Model Construction

The hybrid model, often referred to as a Grey-Box model, integrates the strengths of both ARIMA and Random Forest. Once ARIMA was used to estimate the linear component and Random Forest was trained on the residual series, the final hybrid prediction was obtained by adding the Random Forest residual correction to the ARIMA forecast:

$$\hat{y}_t^{Hybrid} = \hat{y}_t^{ARIMA} + \hat{e}_t^{RF}, \quad (6)$$

where $\hat{e}_t^{RF}$ is the Random Forest prediction of the ARIMA residual at time t. This formulation allows the hybrid model to capture both linear and non-linear relationships, thereby reducing systematic forecast errors that might persist in a single-model approach.

Model Training, Validation, and Forecasting

All models were trained on the complete historical dataset from 1990 to 2021. Because the analysis follows a time-series framework, the temporal order of the data was preserved at all stages. The Random Forest model was trained exclusively using past lagged residuals to ensure the absence of information leakage. In contrast, the ARIMA model was evaluated using its in-sample fitted values, which served as one-step-ahead predictions.

Forecasts were generated from 2022 to 2026. This revised forecast horizon was chosen because the original extension to 2050, although technically feasible, was deemed methodologically inappropriate given the relatively small number of available observations (32 years). Limiting the horizon to 2026 ensures greater credibility and aligns with widely recognized global energy and climate-policy milestones.

Model Evaluation Criteria

The predictive performance of both ARIMA and the hybrid model was evaluated using three standard accuracy measures: the Mean Squared Error (MSE), the Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). These metrics are defined as follows:

$$MSE = \frac{1}{n}(y_t - \hat{y}_t)^2 \quad 7$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \quad 8$$

$$R^2 = 1 - \frac{\sum(y_t - \hat{y}_t)^2}{\sum(y_t - \bar{y})^2}. \quad 9$$

A model was adjudged superior when it achieved lower values of MSE and MAPE and a higher value of R^2. These metrics allowed for a systematic comparison between the linear ARIMA specification and the hybrid machine-learning framework.

Justification of the Forecast Horizon

Given that only 32 annual observations were available for each country, forecasting deep into the future introduces substantial uncertainty and risks violating essential time-series assumptions such as structural stability and constant variance. The decision to restrict the forecast horizon to 2026 therefore enhances the reliability of the projections and ensures that the analysis remains grounded within a practical and methodologically sound timeframe.

3 RESULTS AND DISCUSSION

3.1 Data Structure

The data use in this paper is described in Table 1 where each country will be used to form the variable in the ARIMA Model in Equation (1). The data is a secondary type and was obtained from WORLD BANK (https://data.worldbank.org/indicator/EG.FEC.RNEW.ZS). Following the Box-Jenkins method, the country will be modeled using ARIMA Model in Equation (1). This will be used to compare with the Grey-Box Model in Equation (6).

Table1: Study Data for Renewable Energy (1990–2021) of Developed (Canada and France) and Developing (Angola and Nigeria) Countries

Year	Angola	Canada	France	Nigeria
1990	72.3	22.6	10.5	87.8
1991	71.9	22.7	11.1	86.5
1992	72.7	22.4	11.3	84.5
1993	71.3	22.3	10.9	86.3
1994	72.2	22.2	11.0	88.6
1995	73.8	22.4	10.7	88.0
1996	72.9	22.1	10.5	87.1
1997	73.4	21.8	10.0	85.6
1998	76.4	22.0	9.7	87.2
1999	72.8	22.2	9.8	87.2
2000	73.8	20.3	9.3	86.0
2001	72.8	19.8	9.4	84.5
2002	70.8	20.1	8.7	84.5
2003	65.6	19.7	8.8	82.9
2004	62.3	20.4	8.9	84.0

2005	68.9	21.6	8.6	83.9
2006	62.6	21.5	8.5	85.8
2007	58.9	21.6	9.4	87.1
2008	55.4	21.6	10.5	85.7
2009	53.0	22.2	11.2	88.1
2010	53.1	21.4	11.9	86.0
2011	51.7	21.5	10.6	84.1
2012	49.0	21.8	12.3	83.9
2013	46.8	22.2	13.4	81.4
2014	47.5	22.2	13.2	79.9
2015	47.1	22.3	13.3	81.3
2016	48.1	22.1	14.2	81.1
2017	52.5	22.8	14.1	81.6
2018	52.5	22.3	15.2	80.8
2019	51.0	22.4	15.5	80.1
2020	60.1	23.9	16.8	81.8
2021	52.9	23.8	16.2	80.3

The annual renewable energy data (1990–2021) for two developed nations (Canada and France) and two developing nations (Nigeria and Angola) are shown in Table 1. The figures show the proportion of energy that comes from renewable sources. This dataset serves as the foundation for additional modeling and statistical analysis. In the sections that follow, observed trends will be investigated using the ARIMA and Grey Box approaches.

Table2: Descriptive Analysis of Renewable Energy Consumption (1990 – 2021) of the Selected Countries

Statistics	Angola	Canada	France	Nigeria
Count	32.00	32.00	32.00	32.00
Mean	62.07	21.88	11.42	84.49
Std	10.39	0.97	2.35	2.66
MIN	46.80	19.70	8.50	79.90
25%	52.50	21.58	9.63	81.75
50%	62.45	22.15	10.80	84.50
75%	72.40	22.33	13.23	86.65
Max	76.40	23.90	16.80	88.60

The central tendency and distribution of renewable energy consumption over the 32-year period in Angola, Canada, France, and Nigeria are summarized in Table 2. Nigeria continually relied heavily on renewable energy sources, as seen by its greatest average consumption (84.49%) and comparatively low variability (standard deviation of 2.66). Following with a mean of 62.07%, Angola displayed greater variability (standard deviation of 10.39), indicating a tendency toward renewable energy that is more erratic. The averages for Canada and France, on the other hand, were significantly lower at 21.88% and 11.42%, respectively. Canada's standard deviation of 0.97 indicates that its share of renewable energy is constant but low. France had a little greater variance, indicating a steady improvement in the use of renewable energy. Deeper time series analysis is made possible by these descriptive insights.

3.2 Exploratory Data Analysis

3.2.1 Visualisations

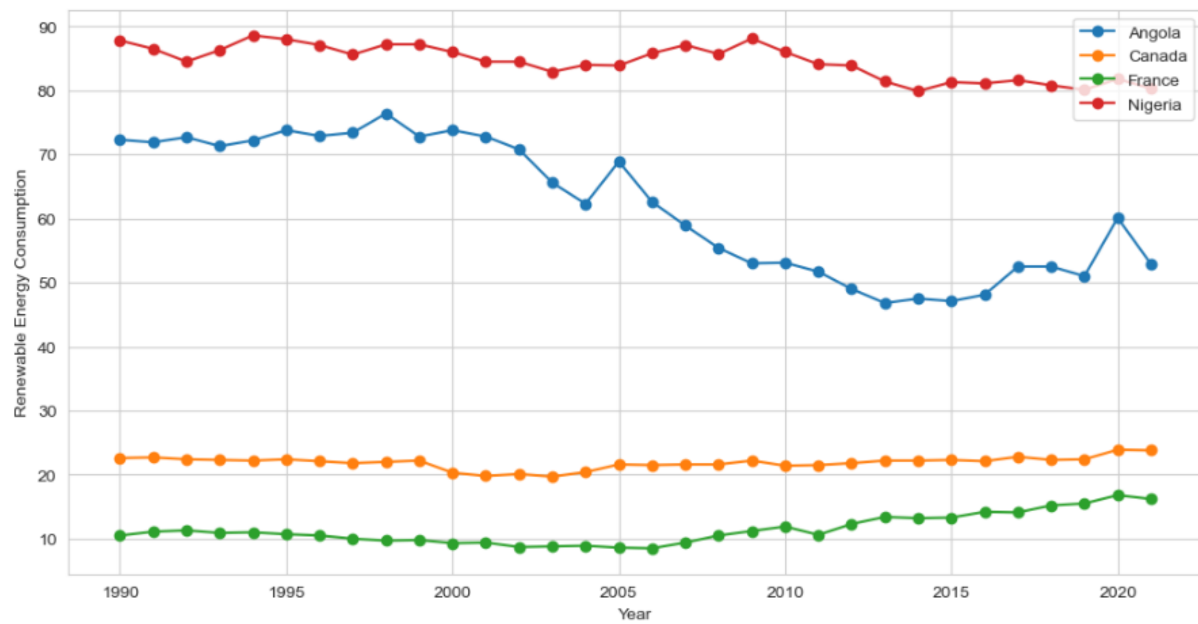

Figure 1: Time Series Plots of Renewable Energy Consumption from 1990 – 2021: Identifying Trends, Seasonality and Spikes in Each Country

Angola, Canada, France, and Nigeria's renewable energy consumption time records from 1990 to 2021 are shown in Figure 1. From the middle of the 1990s until 2014, Angola showed a noticeable decreasing trend; after 2015, there was a minor rebound. Despite its consistent and relatively high consumption, Nigeria also shows a modest dip. Canada and France, on the other hand, show steady or slightly rising trends. While France's consumption exhibits slow growth starting about 2005, Canada's consumption stays mostly constant with just little variations. There are no discernible seasonal or cyclical trends among the nations. With substantial decreases in 2003–2004 and surges in 2005 and 2020, Angola exhibits notable volatility, indicating periodic disruptions. Canada and France only see slight short-term fluctuations, whereas Nigeria's trend is smoother with few outliers.

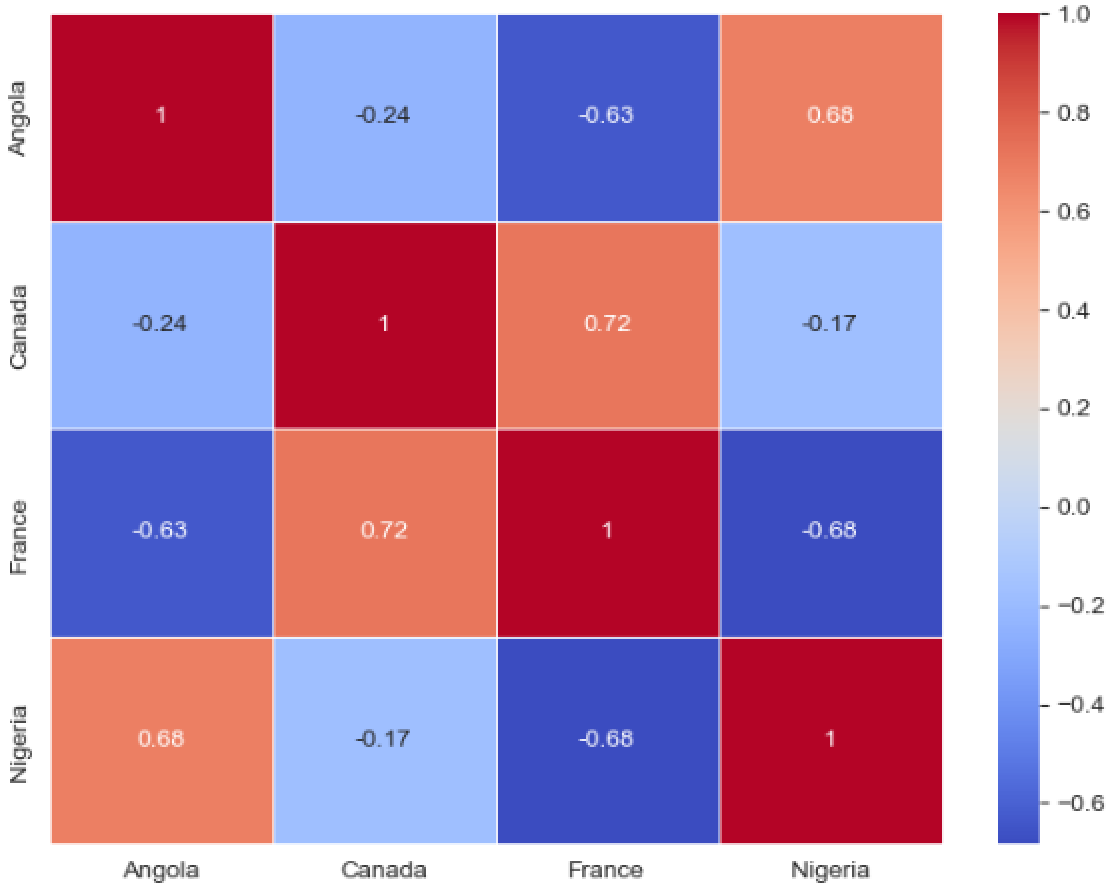

Figure 2: Correlation Heatmap of Renewable Energy Consumption from 1990 – 2021: Identifying Linear Relationships between Countries

A correlation heatmap showing the linear relationships in renewable energy consumption between Angola, Canada, France, and Nigeria between 1990 and 2021 is shown in Figure 2. Stronger linear relationships are indicated by correlation coefficient values near ±1, which range from -1 to 1. Canada and France have the strongest positive correlation (0.72), indicating closely aligned consumption trends that are probably the result of similar energy policies or environmental commitments. Nigeria and Angola also show a strong positive correlation (0.68), which may be due to similar regional or developmental factors. The strong negative correlations between France and Nigeria (-0.68) and Angola and France (-0.63), on the other hand, suggest different trends in the use of renewable energy. The correlations between Canada and Nigeria (-0.17) and Angola (-0.24) are weaker, indicating little linear association.

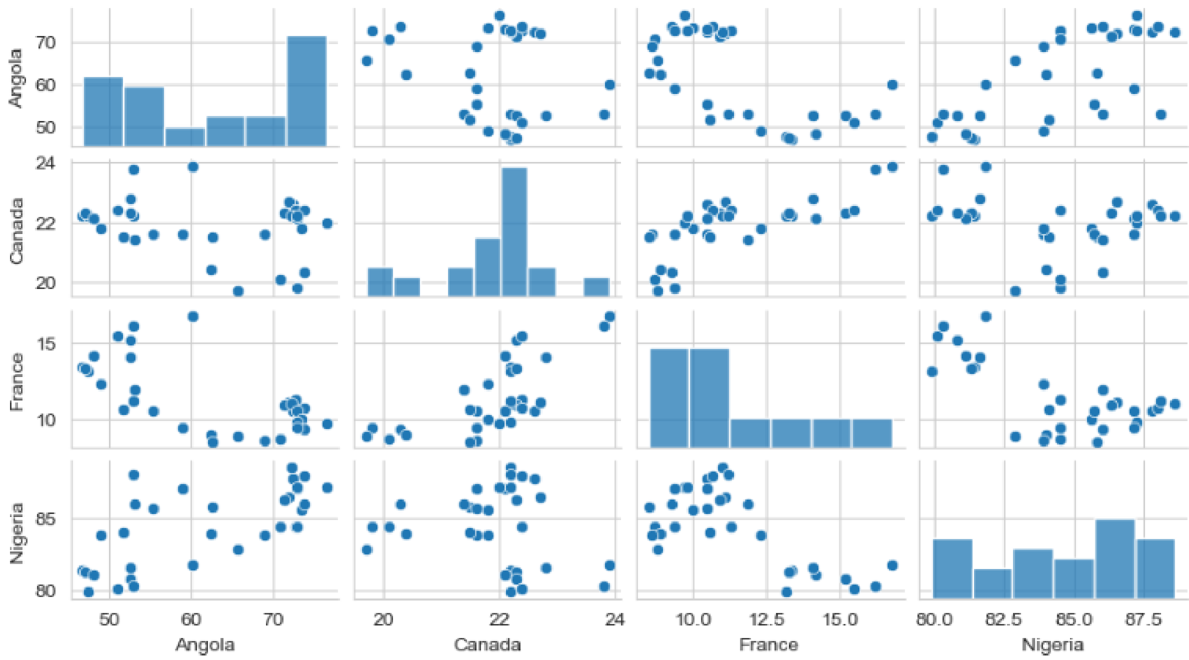

Figure 3: Pairwise Scatter Plots of Renewable Energy Consumption from 1990 – 2021: Identifying Pairwise Relationships between Countries

Pairwise scatter plots and histograms of renewable energy consumption from 1990 to 2021 in Angola, Canada, France, and Nigeria are shown in Figure 3. Exploring inter-country relationships, spotting non-linear trends, locating clusters, and highlighting possible outliers are the objectives of this visualization. Histograms illustrating the distribution of each nation are displayed in the diagonal elements. Canada and France show narrower, more stable patterns, whereas Angola and Nigeria show wider distributions, reflecting more variation in consumption. Pairwise relationships are depicted by off-diagonal scatter plots. For instance, the Angola–France plot shows a negative relationship, which is consistent with previous correlation results, while the Angola–Nigeria plot shows a positive association. Non-linear relationships with visible curvature are suggested by a number of plots, especially those involving France. Additionally, there are outliers that indicate years with unusual consumption patterns, particularly in the data from Nigeria and Angola. As a whole, the By providing evidence of both linear and non-linear dynamics, the matrix enhances the correlation analysis and provides insightful visual information about the type and strength of relationships across nations.

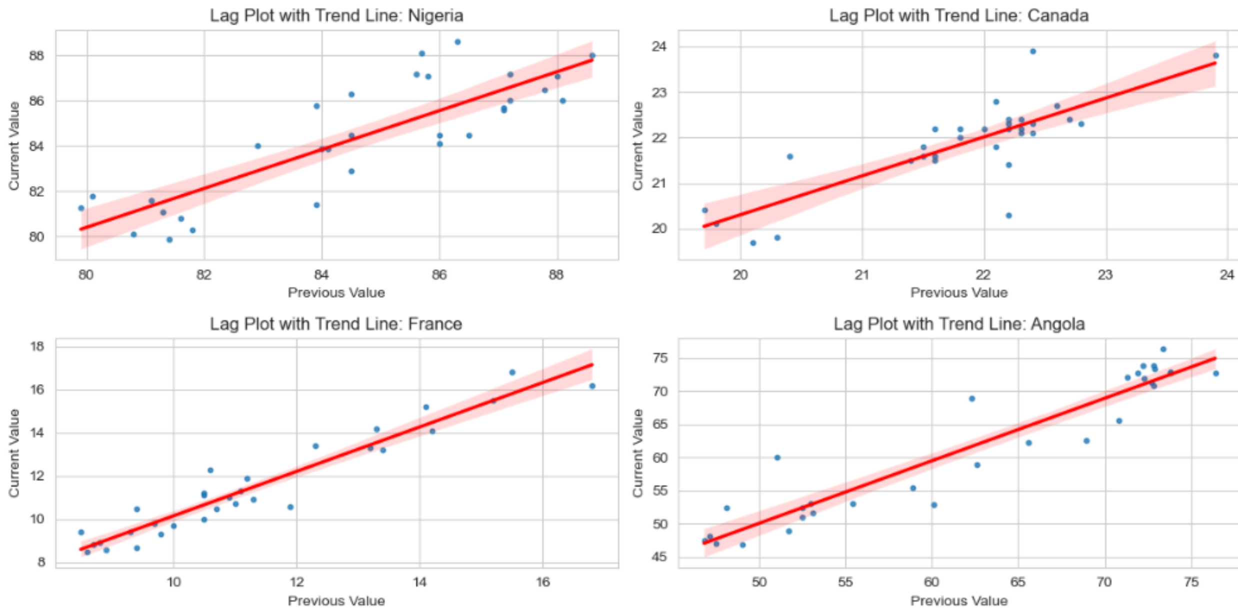

Figure 4: Lag Plots for Autocorrelation Analysis of Renewable Energy Consumption from 1990 – 2021: Checking how past values Influence Future values.

To evaluate autocorrelation in renewable energy consumption from 1990 to 2021, Figure 4 shows lag plots with trend lines for Nigeria, Canada, France, and Angola. These plots assess how well historical values forecast present consumption levels in each nation. With points closely clustered along the diagonal, Nigeria exhibits a strong positive autocorrelation, suggesting consistent and long-lasting consumption trends. Although it exhibits somewhat more dispersion, Canada likewise shows a positive slope, indicating moderate to strong autocorrelation and some variation in patterns from year to year. While Angola also exhibits a strong diagonal alignment, suggesting high predictability despite the wider fluctuations seen in previous analyses, France's plot shows a tight clustering along the diagonal, reflecting strong and consistent autocorrelation.

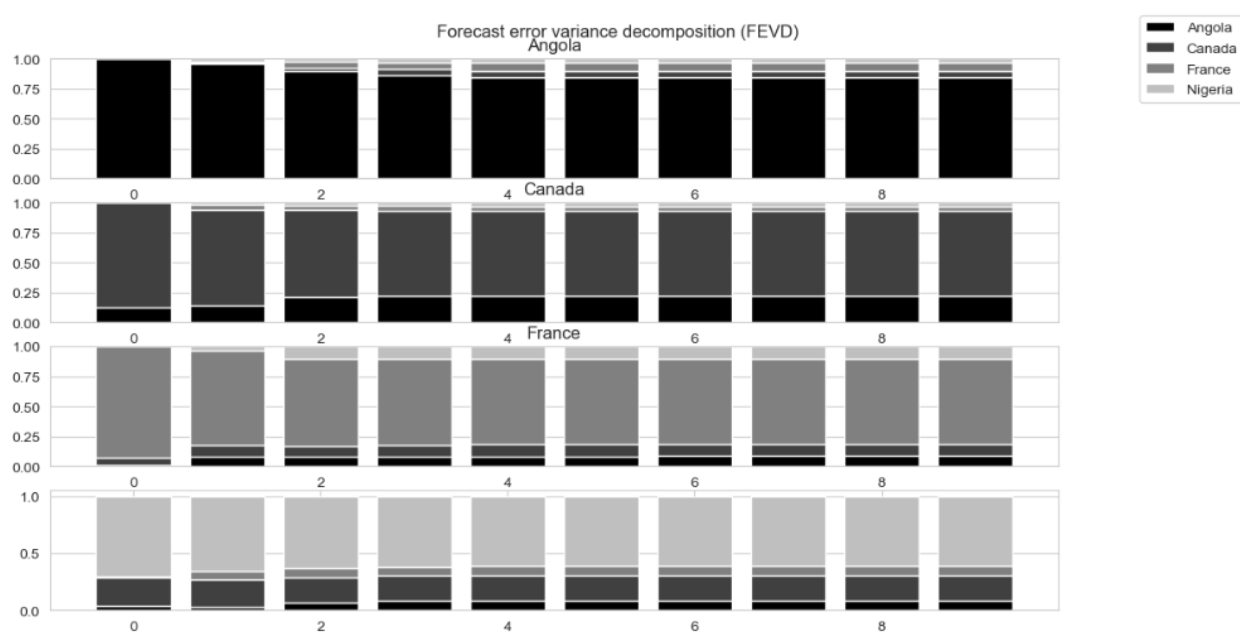

Figure 5: Forecast Error Variance Decomposition to Determine if a Country is Granger-Caused by Another.

The Forecast Error Variance Decomposition (FEVD) for Angola, Canada, France, and Nigeria is shown in Figure 5, which shows how much of the forecast errors in each nation's use of renewable energy are caused by internal versus external shocks. The influence—or possible Granger causality—between the nations is evaluated with the aid of this analysis. Across all horizons, Angola's own shocks account for the majority of forecast errors, demonstrating a high level of autonomy and little outside influence. Canada also exhibits a high degree of self-reliance, with the majority of the variance being explained by its own historical values. However, over time, minor contributions from France and Nigeria appear, indicating that external effects are minimal. French consumption dynamics are largely self-contained, as evidenced by the country's dominance of its own innovations and the minimal influence of other nations. However, Nigeria exhibits a relatively higher extent of outside influence. Although its own shocks continue to be the main force, Nigeria appears to be more interconnected and sensitive to outside events based on the increasing contributions from Angola, Canada, and France over forecast horizons.

Overall, the FEVD results show different levels of independence in energy consumption, with Nigeria being the most externally influenced and Angola being the most autonomous.

3.2.2 Granger Causality Tests

Table 3: The Granger Causality Test for Renewable Consumption in Study Countries (1990 – 2021)

Checking Granger causality for: Angola
Lag 1: Canada → Angola	F-Stat: 0.2263, p-value: 0.6381
Lag 2: Canada → Angola	F-Stat: 0.6752, p-value: 0.5184
Lag 3: Canada → Angola	F-Stat: 1.1448, p-value: 0.3541
Lag 4: Canada → Angola	F-Stat: 0.9183, p-value: 0.4748
Lag 5: Canada → Angola	F-Stat: 0.8221, p-value: 0.5529
Lag 1: France → Angola	F-Stat: 0.0021, p-value: 0.9640
Lag 2: France → Angola	F-Stat: 0.8625, p-value: 0.4348
Lag 3: France → Angola	F-Stat: 1.0314, p-value: 0.3989
Lag 4: France → Angola	F-Stat: 1.3885, p-value: 0.2775
Lag 5: France → Angola	F-Stat: 1.2130, p-value: 0.3503
Lag 1: Nigeria → Angola	F-Stat: 0.0195, p-value: 0.8899
Lag 2: Nigeria → Angola	F-Stat: 0.0333, p-value: 0.9673
Lag 3: Nigeria → Angola	F-Stat: 0.0741, p-value: 0.9732
Lag 4: Nigeria → Angola	F-Stat: 0.2685, p-value: 0.8944
Lag 5: Nigeria → Angola	F-Stat: 0.2791, p-value: 0.9174
Checking Granger causality for: Canada
Lag 1: Angola → Canada	F-Stat: 0.3913, p-value: 0.5368
Lag 2: Angola → Canada	F-Stat: 0.8024, p-value: 0.4599
Lag 3: Angola → Canada	F-Stat: 0.4034, p-value: 0.7521
Lag 4: Angola → Canada	F-Stat: 0.6095, p-value: 0.6610
Lag 5: Angola → Canada	F-Stat: 0.8465, p-value: 0.5380
Lag 1: France → Canada	F-Stat: 0.6428, p-value: 0.4297
Lag 2: France → Canada	F-Stat: 0.6995, p-value: 0.5067
Lag 3: France → Canada	F-Stat: 0.4229, p-value: 0.7386
Lag 4: France → Canada	F-Stat: 0.4542, p-value: 0.7682
Lag 5: France → Canada	F-Stat: 0.8170, p-value: 0.5560

Lag 1: Nigeria → Canada | F-Stat: 0.5561, p-value: 0.4623
Lag 2: Nigeria → Canada | F-Stat: 0.4397, p-value: 0.6493
Lag 3: Nigeria → Canada | F-Stat: 0.6245, p-value: 0.6071
Lag 4: Nigeria → Canada | F-Stat: 0.4407, p-value: 0.7776
Lag 5: Nigeria → Canada | F-Stat: 0.7377, p-value: 0.6068

Checking Granger causality for: France

Lag 1: Angola → France | F-Stat: 2.3511, p-value: 0.1368
Lag 2: Angola → France | F-Stat: 1.2292, p-value: 0.3103
Lag 3: Angola → France | F-Stat: 1.4718, p-value: 0.2509
Lag 4: Angola → France | F-Stat: 0.8299, p-value: 0.5235
Lag 5: Angola → France | F-Stat: 0.7169, p-value: 0.6206
Lag 1: Canada → France | F-Stat: 0.6531, p-value: 0.4261
Lag 2: Canada → France | F-Stat: 0.2510, p-value: 0.7800
Lag 3: Canada → France | F-Stat: 1.1376, p-value: 0.3568
Lag 4: Canada → France | F-Stat: 0.7154, p-value: 0.5923
Lag 5: Canada → France | F-Stat: 1.2386, p-value: 0.3397
Lag 1: Nigeria → France | F-Stat: 1.1314, p-value: 0.2969
Lag 2: Nigeria → France | F-Stat: 1.7033, p-value: 0.2034
Lag 3: Nigeria → France | F-Stat: 1.9833, p-value: 0.1474
Lag 4: Nigeria → France | F-Stat: 2.1534, p-value: 0.1159
Lag 5: Nigeria → France | F-Stat: 1.3545, p-value: 0.2956

Checking Granger causality for: Nigeria

Lag 1: Angola → Nigeria | F-Stat: 0.0177, p-value: 0.8952
Lag 2: Angola → Nigeria | F-Stat: 0.1653, p-value: 0.8486
Lag 3: Angola → Nigeria | F-Stat: 0.5669, p-value: 0.6429
Lag 4: Angola → Nigeria | F-Stat: 0.9905, p-value: 0.4378
Lag 5: Angola → Nigeria | F-Stat: 0.7711, p-value: 0.5851
Lag 1: Canada → Nigeria | F-Stat: 0.0019, p-value: 0.9651
Lag 2: Canada → Nigeria | F-Stat: 0.0740, p-value: 0.9289
Lag 3: Canada → Nigeria | F-Stat: 0.6545, p-value: 0.5890
Lag 4: Canada → Nigeria | F-Stat: 0.4669, p-value: 0.7593
Lag 5: Canada → Nigeria | F-Stat: 0.3963, p-value: 0.8437
Lag 1: France → Nigeria | F-Stat: 2.6264, p-value: 0.1167
Lag 2: France → Nigeria | F-Stat: 0.9660, p-value: 0.3949
Lag 3: France → Nigeria | F-Stat: 0.7321, p-value: 0.5444
Lag 4: France → Nigeria | F-Stat: 0.3155, p-value: 0.8638
Lag 5: France → Nigeria | F-Stat: 0.4646, p-value: 0.7966

Summary of Granger-Causal Relationships
Angola is Granger-caused by: None
Canada is Granger-caused by: None
France is Granger-caused by: None
Nigeria is Granger-caused by: None

Using data from 1990 to 2021 from Angola, Canada, France, and Nigeria, the Granger causality test was used to ascertain whether the historical values of one nation's renewable energy consumption could statistically predict the future values of another. This analysis checks for directional influence between nations, which enhances the forecast error variance decomposition (FEVD) results.

The results show that, across all tested lags, none of the nations Granger-cause one another's use of renewable energy. No causality from Canada, France, or Nigeria was found for Angola, indicating that the country's renewable energy consumption is mostly self-determined. Similar to France, Canada's internal energy dynamics as seen in the FEVD are supported by the lack of statistically significant causal influence from the other nations. Despite the fact that Nigeria seemed more susceptible to outside shocks in the Granger tests revealed no discernible causal influence for FEVD, indicating that these external effects are not very predictive.

In conclusion, the Granger causality test demonstrates that there is no indication of statistically significant predictive relationships between the renewable energy consumption of each nation, which is largely determined by its own historical values.

3.3.1 Model Identification

ACF and PACF for Renewable Energy Consumption

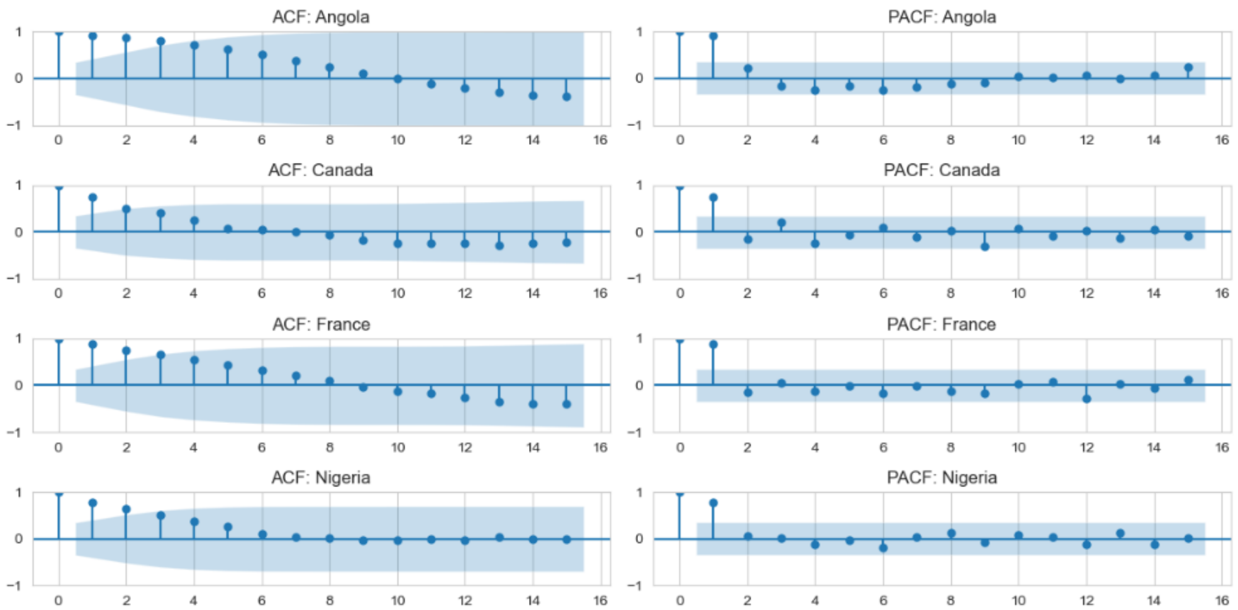

Figure 6: ACF and PACF Plots for Renewable Energy Consumption in each Country

Important diagnostic tools for determining suitable model orders in ARIMA time series modelling are the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots for Angola, Canada, France, and Nigeria, which are shown in Figure 6. The structure and stationarity of each nation's renewable energy consumption series are revealed by these plots.

The ACF for Angola shows a slow, continuous decline, which is indicative of a non-stationary process. On the other hand, the PACF exhibits a clear cutoff following lag 1, indicating the existence of a robust autoregressive component of order one, AR(1). This behavior suggests that before using an AR(1) model, differencing will probably be required to make the series stationary.

Similar patterns can be seen in Canada, where non-stationarity is indicated by a steadily declining ACF. Once more, the PACF shows a notable lag 1 spike. 1 followed by a steep decline, confirming that an AR(1) structure is appropriate following differencing.

The gradual tailing off of France's ACF further supports the existence of non-stationarity by showing sustained autocorrelation over several lags. Although higher-order terms might also be taken into consideration after differencing, the PACF shows a dominant spike at lag 1 with smaller subsequent values, suggesting an AR(1) model may be appropriate.

Compared to the other nations, Nigeria's ACF declines more quickly, suggesting a possibly more stationary series. The PACF exhibits a noticeable peak at lag 1, suggesting that a basic AR(1) model with little to no differencing might be adequate.

In summary, early-lag autocorrelation patterns in all four nations are in line with low-order ARIMA models, usually AR(1), with differencing needed to
attain stationarity. The moving average (MA) component is informed by the ACF, and the autoregressive (AR) structure is identified by the PACF. The selection of appropriate ARIMA specifications based on the time series behavior of each nation is guided by these diagnostics taken together.

ADF test for Renewable Energy Consumption and First Differenced Renewable Energy Consumption

Table 4: ADF Test for Stationarity Test in each Country

	Renewable Energy			First Differenced Renewable Energy		
Country	ADF Statistic	p-value	Stationary?	ADF Statistic	p-value	Stationary?
Angola:	-0.948	0.772	FALSE	-6.47711	1.32E-08	TRUE
Canada:	-1.245	0.654	FALSE	-5.43122	2.91E-06	TRUE
France:	0.527	0.986	FALSE	-5.42365	3.02E-06	TRUE
Nigeria:	-1.401	0.582	FALSE	-5.75323	5.90E-07	TRUE

The Augmented Dickey-Fuller (ADF) test results for renewable energy consumption data for Angola, Canada, France, and Nigeria are shown in Table 4. The results were evaluated both at the level and following first differencing.

The null hypothesis of a unit root cannot be rejected at level form because all four nations show comparatively high ADF test statistics and p-values significantly above the traditional 0.05 threshold. This demonstrates that the original series are not stationary and that their raw form is not appropriate for ARIMA modeling since their statistical characteristics, such as mean and variance, change over time.

The ADF test results for each nation, however, significantly worsen after first differencing, with p-values dropping below 0.05 (e.g., 1.32E-08 for Angola and 2.91E-06 for Canada). This proves that the differenced series are stationary and shows that the unit root hypothesis is rejected.

In conclusion, all four nations' renewable energy consumption series are integrated of order one, I(1). The use of ARIMA (p, 1, q) models for further time series analysis is justified by the efficient induction of stationarity by first differencing.

Figure 7: Time Series Plot of Renewable Energy Consuption after differencing: Vissualising the Stationarity

The first-differenced time series plots of renewable energy consumption for Nigeria, Canada, France, and Angola are shown in Figure 7. An important requirement for ARIMA modeling is stationarity, which can be visually evaluated with these plots.

Prior to differencing, each country's original series showed non-stationarity in the form of trends and non-constant variance. Following initial differencing, there is no discernible trend or seasonality in the time series for any of the four nations; instead, they all oscillate around a constant mean close to zero. The differenced data appear to satisfy the essential requirements for stationarity based on this visual transformation.

The comparatively large fluctuations in Angola's series indicate greater volatility in shifts in the use of renewable energy. Canada's series, on the other hand, has more steady, smaller movements. Nigeria's series is more erratic but still revolves around a steady mean, whereas France exhibits moderate variability with sporadic spikes.

When taken as a whole, the differenced plots demonstrate that non-stationarity has been effectively handled by the transformation. These findings support the use of ARIMA models with a differencing order of one (d = 1) and are in line with the findings of the ADF test.

Table 5: Hyperparameter Tuning for Best ARIMA Model

METRIC	MODELS			COUNTRY
	ARIMA (1,1,1)	ARIMA(0,1,1)	ARIMA(0,2,2)	
AIC	168.479	166.744	163.067	ANGOLA
BIC	172.781	169.612	167.27	
HQIC	169.882	167.679	164.411	
Sigma	11.041	11.144	10.223	
No of Sign. Parameters	0	1	2	
	ARIMA (1,1,2)	ARIMA(2,1,1)	ARIMA(0,2,1)	

AIC	62.068	61.593	61.647	
BIC	67.804	67.329	64.449	
HQIC	63.938	63.463	62.543	CANADA
Sigma	0.334	0.328	0.374	
No of Sign. Parameters	0	0	1	
	ARIMA(0,1,1)	ARIMA(1,1,1)	ARIMA(0,2,1)	
AIC	69.295	69.648	68.043	
BIC	72.163	73.949	70.845	
HQIC	70.23	71.05	68.939	FRANCE
Sigma	0.481	0.454	0.473	
No of Sign. Parameters	0	2	1	
	ARIMA(0,1,0)	ARIMA(0,1,1)	ARIMA(1,1,1)	
AIC	113.292	115.126	116.479	
BIC	114.726	117.994	120.781	
HQIC	113.759	116.061	117.881	NIGERIA
Sigma	2.122	2.11	2.062	
No of Sign. Parameters	0	0	2	

Table 5 presents a comparative analysis of various ARIMA models used to forecast renewable energy consumption for Angola, Canada, France, and Nigeria. The models were evaluated using five key metrics: AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), HQIC (Hannan-Quinn Information Criterion), Sigma (standard error of residuals), and the number of statistically significant parameters. These metrics are essential for identifying the model that best balances goodness-of-fit with simplicity and interpretability.

For Angola, the ARIMA (0,2,2) model was selected as the most appropriate. It produced the lowest values for AIC (163.067), BIC (167.270), and HQIC (164.411) compared to other models like ARIMA (1,1,1) and ARIMA (0,1,1). It also had the lowest residual standard error (SIGMA = 10.223) and included two statistically significant parameters. This indicates that ARIMA (0,2,2) offers both a strong fit and interpretability, making it the best model for capturing the patterns in Angola's renewable energy consumption data.

For Canada, the ARIMA (0,2,1) model was chosen over ARIMA (1,1,2) and ARIMA (2,1,1). It had the lowest AIC (61.647), BIC (64.449), and HQIC (62.543), along with a competitive SIGMA of 0.374. Although it had only one statistically significant parameter, this model struck the best balance between model fit and parsimony. Thus, ARIMA (0,2,1) is the most appropriate model for forecasting renewable energy consumption in Canada.

In the case of France, ARIMA (0,2,1) also emerged as the preferred model when compared with ARIMA (0,1,1) and ARIMA (1,1,1). It achieved the lowest AIC (68.043), BIC (70.845), and HQIC (68.939), and a favorable SIGMA of 0.473. The model included one significant parameter, suggesting that while it is simple, it still effectively captures the essential structure of the time series. Therefore, ARIMA (0,2,1) is deemed suitable for the French dataset.

For Nigeria, although ARIMA (0,1,0) had the lowest AIC (113.292), it had no statistically significant parameters and a higher residual error (SIGMA = 2.122) compared to ARIMA (1,1,1). The ARIMA (1,1,1) model, while having a slightly higher AIC (116.479), had a lower SIGMA (2.062) and two significant parameters, which means it provides a more robust and meaningful explanation of the underlying process. Therefore, ARIMA (1,1,1) was selected as the most appropriate model for Nigeria's renewable energy consumption.

In summary, ARIMA (0,2,2) was selected for Angola, ARIMA (0,2,1) for Canada, ARIMA (0,2,1) for France, and ARIMA (1,1,1) for Nigeria. These selections were based on a combination of the lowest information criteria, reduced residual error, and the presence of statistically significant model parameters.

These factors collectively ensure each chosen model is both statistically sound and practically useful for forecasting renewable energy consumption in the respective countries.

Table 6: ARIMA Model for Renewable Energy Consumption in Angola (1990 – 2021)

Dep. Variable:	Angola	No. Observations:	32
Model:	ARIMA(0, 2, 2)	Log Likelihood	-78.533
Date:	Wed, 07 May 2025	AIC	163.067
Time:	14:53:00	BIC	167.270
Sample:	12-31-1990 - 12-31-2021	HQIC	164.411
	Covariance Type:	opg	

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-1.3492	0.195	-6.909	0.000	-1.732	-0.966
ma.L2	0.6413	0.211	3.043	0.002	0.228	1.054
sigma^2	10.2247	2.798	3.654	0.000	4.740	15.709

Ljung-Box (L1) (Q):	0.02	Jarque-Bera (JB):	0.29
Prob(Q):	0.89	Prob(JB	0.86
Heteroskedasticity (H):	5.32	Skew:	0.20
Prob(H) (two-sided):	0.01	Kurtosis:	3.26

The results in Table 6 summarize the fitted ARIMA(0,2,2) model for Angola's renewable energy consumption from 1990 to 2021, based on 32 observations. This model includes two moving average terms (MA(1) and MA(2)) and applies second differencing (d=2) to achieve stationarity.

Interpreting the parameters, the first moving average coefficient (ma.L1) is -1.3492 and highly significant ($p < 0.001$). This negative value indicates that the current differenced value is strongly influenced by a negatively weighted error term from one period ago. The second moving average coefficient (ma.L2) is 0.6413 and also statistically significant ($p = 0.002$), showing a positive contribution from the error term two periods prior. Together, these MA terms capture short-term shocks and their decay in the renewable energy consumption changes after differencing.
The estimated variance of the residuals, sigma2 = 10.2247, quantifies the model's unexplained variability, with a standard error indicating reasonable precision.

Diagnostic tests support the model's adequacy. The Ljung-Box Q test at lag 1 (Q=0.02, p=0.89) suggests no significant autocorrelation remains in the residuals, indicating the model has effectively captured the time series dependence. The Jarque-Bera test (JB=0.29, p=0.86) shows residuals are approximately normally distributed, satisfying a key ARIMA assumption.

However, the heteroskedasticity test (H=5.32, p=0.01) indicates evidence of non-constant variance in residuals, implying some volatility clustering or changing variability over time that the model does not fully account for. The skewness (0.20) and kurtosis (3.26) values are close to those of a normal distribution, suggesting the residual distribution is not heavily skewed or fat-tailed.

Table 7: ARIMA Model for Renewable Energy Consumption in Canada (1990 – 2021)

Dep. Variable:	Canada	No. Observations:	32

Model:	ARIMA(0, 2, 1)	Log Likelihood	-28.823
Date:	Sat, 10 May 2025	AIC	61.647
Time:	08:40:50	BIC	64.449
Sample:	12-31-1990	HQIC	62.543
	- 12-31-2021		
	Covariance Type:	opg	

	Coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.9312	0.177	-5.266	0.000	-1.278	-0.585
sigma2	0.3741	0.073	5.125	0.000	0.231	0.517

Ljung-Box (L1) (Q):	0.21	Jarque-Bera (JB):	6.86
Prob(Q):	0.65	Prob(JB):	0.03
Heteroskedasticity (H):	0.88	Skew:	-0.38
Prob(H) (two-sided):	0.85	Kurtosis:	5.22

The ARIMA(0,2,1) model fitted to Canada's renewable energy consumption data from 1990 to 2021 includes one moving average term with second differencing, applied to ensure stationarity.

The moving average coefficient (ma.L1) is -0.9312, which is statistically significant ($p < 0.001$). This strong negative value indicates that the current differenced observation is influenced by a substantial negative shock from the previous period's error term, capturing short-term autocorrelation effectively.

The residual variance (sigma2) is estimated at 0.3741, indicating the variability in the differenced series not explained by the model, and it is highly significant.

Diagnostic tests reveal mixed model adequacy. The Ljung-Box Q test (Q=0.21, p=0.65) suggests no significant autocorrelation remains in the residuals, indicating the model fits the serial dependence well. However, the Jarque-Bera test (JB=6.86, p=0.03) rejects the null hypothesis of normality at the 5% level, indicating the residuals deviate from normal distribution, possibly due to heavier tails or outliers.

Heteroskedasticity testing (H=0.88, p=0.85) shows no evidence of non-constant variance, suggesting homoscedastic residuals. The residual skewness is -0.38, indicating a slight left skew, while kurtosis is 5.22, suggesting heavier tails than a normal distribution.

In conclusion, the ARIMA(0,2,1) model effectively captures the main dynamics in Canada's differenced renewable energy consumption series, with well-modeled autocorrelation and stable variance. However, non-normality of residuals suggests caution when making inference or forecasting, and model refinement or alternative error distributions might improve robustness.

Table 8: ARIMA Model for Renewable Energy Consumption in France (1990 – 2021)

Dep. Variable:	France	No. Observations:	32
Model:	ARIMA(0, 2, 1)	Log Likelihood	-32.021
Date:	Sat, 10 May 2025	AIC	68.043
Time:	08:13:44	BIC	70.845
Sample:	12-31-1990	HQIC	68.939
	- 12-31-2021		

	Covariance Type:	opg				
	Coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.8658	0.094	-9.255	0.000	-1.049	-0.682
sigma2	0.4727	0.131	3.621	0.000	0.217	0.729

Ljung-Box (L1) (Q):	1.03	Jarque-Bera (JB):	0.01
Prob(Q):	0.31	Prob(JB):	0.99
Heteroskedasticity (H):	5.35	Skew:	0.05
Prob(H) (two-sided):	0.01	Kurtosis:	2.96

The ARIMA(0,2,1) model fitted to France's renewable energy consumption data from 1990 to 2021 features a single moving average term and second differencing to achieve stationarity.

The moving average coefficient (ma.L1) is -0.8658 and is highly significant ($p < 0.001$). This negative value close to -1 indicates a strong inverse relationship between the current differenced observation and the previous period's error, effectively capturing short-term fluctuations.

The estimated residual variance (sigma2) is 0.4727 and statistically significant, reflecting the remaining variability in the differenced series after accounting for the model's structure.

Diagnostic checks show mixed evidence for model adequacy. The Ljung-Box Q test (Q=1.03, p=0.31) suggests no significant autocorrelation remains in the residuals, supporting that the model sufficiently captures serial dependence. The Jarque-Bera test (JB=0.01, p=0.99) fails to reject normality, indicating that residuals are approximately normally distributed, which is favorable for inference.

However, the heteroskedasticity test (H=5.35, p=0.01) indicates significant evidence of non-constant variance in residuals, suggesting heteroskedasticity is present. The residual skewness is minimal (0.05), close to symmetric, and kurtosis (2.96) is near the normal value of 3.

In summary, while the ARIMA(0,2,1) model effectively models autocorrelation and normality in France's renewable energy consumption changes, the presence of heteroskedasticity suggests potential volatility clustering or changing variance over time, which may warrant model refinement or use of variance-stabilizing techniques for improved reliability.

Table 9: ARIMA Model for Renewable Energy Consumption in Nigeria (1990 – 2021)

Dep. Variable:	Nigeria	No. Observations:	32
Model:	ARIMA(1, 1, 1)	Log Likelihood	-55.239
Date:	Thu, 15 May 2025	AIC	116.479
Time:	15:49:10	BIC	120.781
Sample:	12-31-1990 - 12-31-2021	HQIC	117.881
	Covariance Type:	opg	

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.6743	0.708	0.952	0.341	-0.714	2.063

ma.L1	-0.7969	0.617	-1.291	0.197	-2.006	0.413
sigma2	2.0619	0.927	2.225	0.026	0.245	3.878

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	2.02
Prob(Q):	0.91	Prob(JB):	0.36
Heteroskedasticity (H):	0.94	Skew:	0.41
Prob(H) (two-sided):	0.92	Kurtosis:	2.06

The ARIMA(1,1,1) model for Nigeria's renewable energy consumption data from 1990 to 2021 shows several important features. The AR(1) coefficient, which measures the influence of the previous period's value on the current one after differencing, is positive at 0.6743 but is not statistically significant, indicating that the autoregressive component may not strongly explain the series' behavior. Similarly, the MA(1) coefficient, representing the impact of the previous period's error on the current value, is negative at -0.7969 but also lacks statistical significance, suggesting uncertainty about its effect. The variance of the residuals, estimated at 2.0619, is statistically significant, meaning there is notable variability in the errors of the model. Diagnostic tests provide further insight into the model's adequacy. The Ljung-Box test, which checks for remaining autocorrelation in the residuals, shows a high p-value of 0.91, implying that the model has successfully captured the serial dependence and no significant autocorrelation remains. The Jarque-Bera test for normality of residuals returns a p-value of 0.36, indicating that the residuals can be considered approximately normally distributed. The heteroskedasticity test yields a p-value of 0.92, supporting the assumption that the variance of the residuals is constant over time. Residual skewness is slightly positive at 0.41, and kurtosis is somewhat lower than the normal distribution benchmark at 2.06, suggesting mild departures from perfect normality but nothing concerning.

In summary, while the ARIMA(1,1,1) model fits the data reasonably well in terms of residual diagnostics, the lack of statistical significance in the AR and MA parameters raises questions about the reliability of these estimates. The model effectively addresses autocorrelation and maintains constant variance with normally distributed residuals, but the significant residual variance indicates some unexplained variability. Therefore, although the model is appropriate based on the tests, further refinement or alternative modeling strategies may be necessary to enhance parameter significance and better capture the dynamics of Nigeria's renewable energy consumption.

3.3.3 Model Evaluation

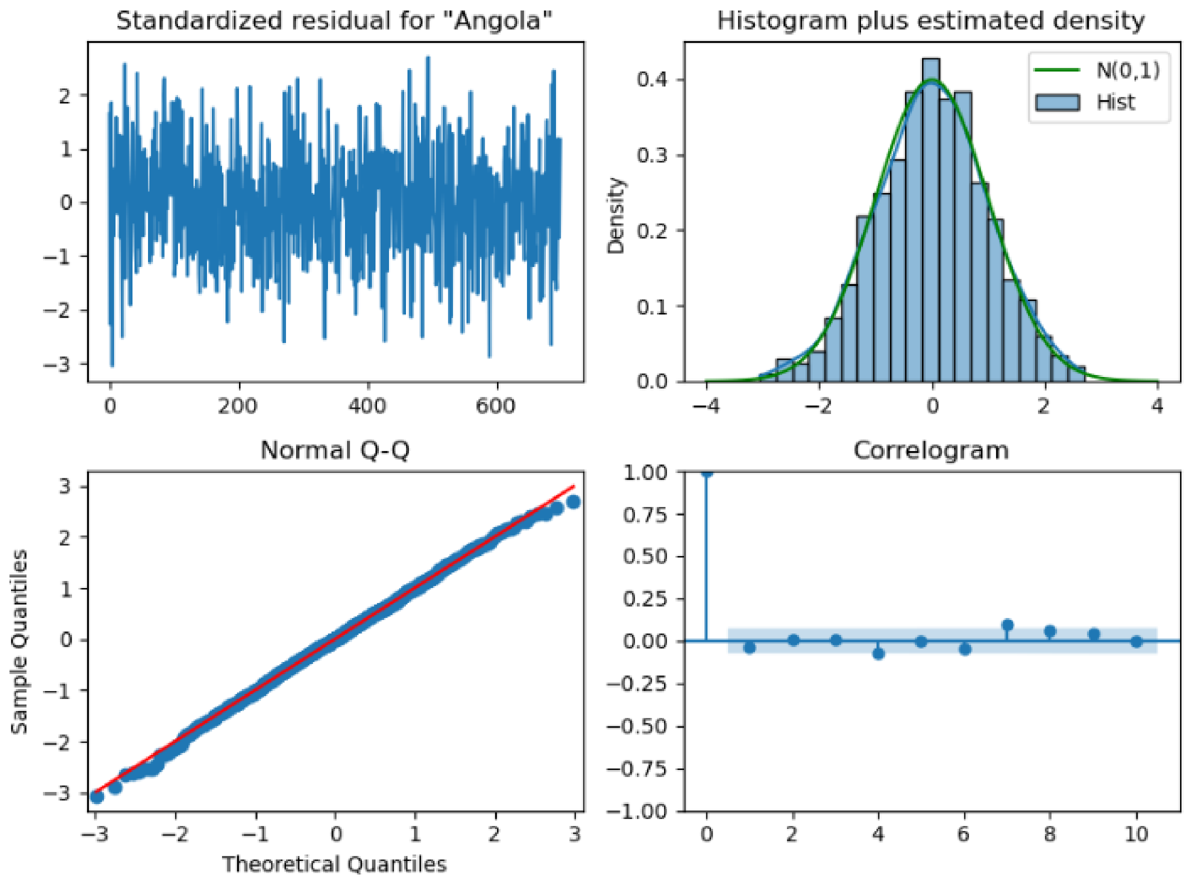

Figure 8: Angola ARIMA (0, 2, 2) Diagnostics Output

The diagnostic plots in Figure 8 provide a comprehensive evaluation of the ARIMA (0, 2, 2) model fitted to Angola's renewable energy consumption data. The standardized residual plot in the top-left panel reveals that residuals fluctuate randomly around zero without any discernible trend or pattern, indicating the model has successfully captured the underlying dynamics of the series. The histogram of residuals in the top-right panel closely matches the overlaid normal distribution curve, suggesting that the residuals approximate a normal distribution.

The normal Q-Q plot in the bottom-left panel reinforces this conclusion, as the residuals align closely with the 45-degree reference line, supporting the assumption of normality. Finally, the correlogram in the bottom-right panel shows that all autocorrelation values for the residuals lie within the 95% confidence intervals, with no significant autocorrelations at any lag. This indicates that the residuals are free from autocorrelation and the model has effectively accounted for temporal dependencies.

Overall, the ARIMA (0, 2, 2) model exhibits a strong fit and adequacy. The residuals appear to be random, normally distributed, and uncorrelated, confirming that this model is appropriate and reliable for forecasting Angola's renewable energy consumption.

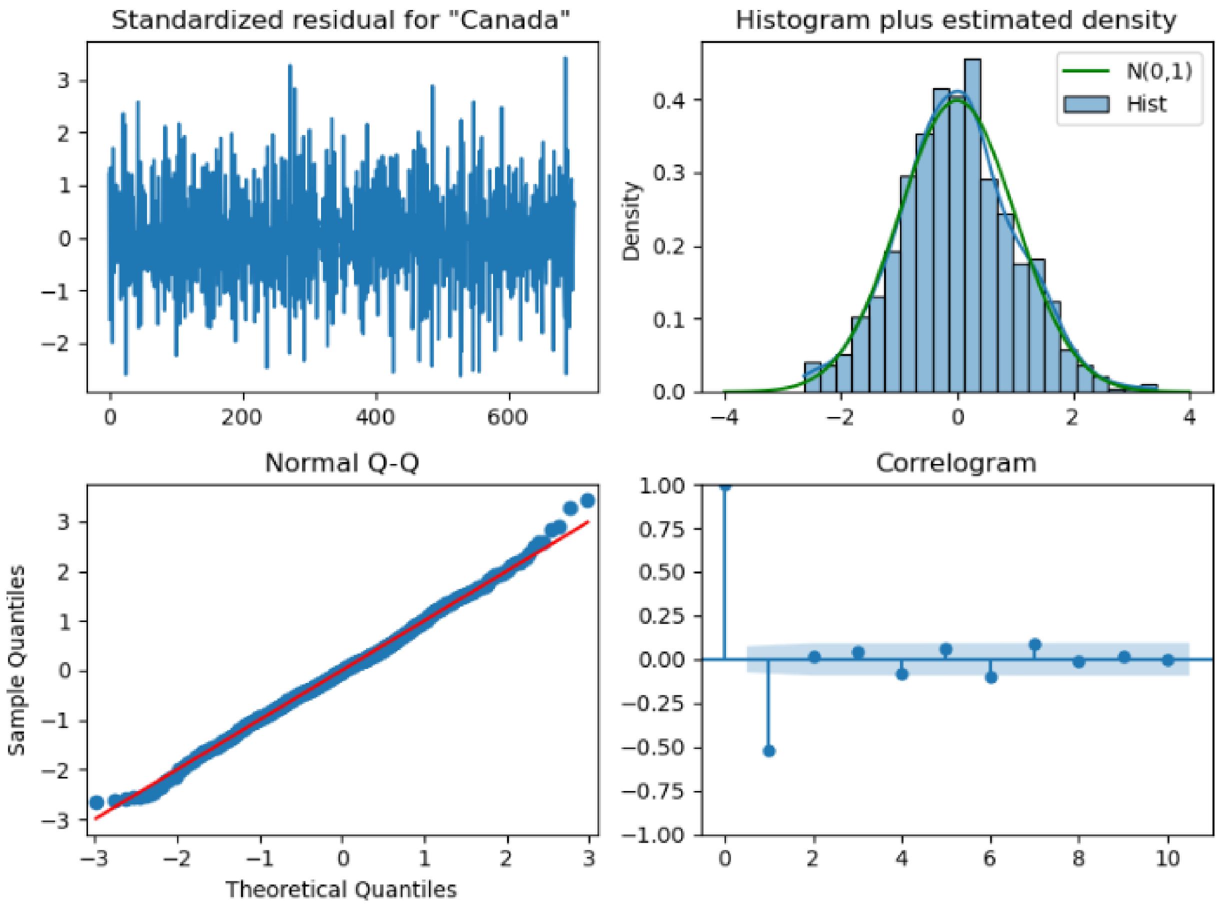

Figure 9: Canada ARIMA (0, 2, 1) Diagnostics Output

The diagnostic plots in Figure 9 evaluate the adequacy of the ARIMA (0, 2, 1) model fitted to Canada's renewable energy consumption data. The standardized residual plot in the top-left panel shows a random dispersion of residuals around zero without any visible trend or pattern, indicating that the model provides a good fit. The histogram in the top-right panel demonstrates that the residuals roughly follow a normal distribution, as the bars closely align with the overlaid normal density curve.

The normal Q-Q plot in the bottom-left panel supports this finding by showing residuals that lie approximately along the 45-degree reference line, suggesting the normality assumption is largely met. The correlogram in the bottom-right panel reveals that all residual autocorrelations fall within the 95% confidence interval except for lag 2, which slightly exceeds the bounds. However, this minor deviation does not significantly compromise the model's adequacy.

In summary, the ARIMA (0, 2, 1) model for Canada appears to be well-fitted. The residuals are mostly random, approximately normally distributed, and exhibit minimal autocorrelation, supporting the model's reliability for forecasting renewable energy consumption.

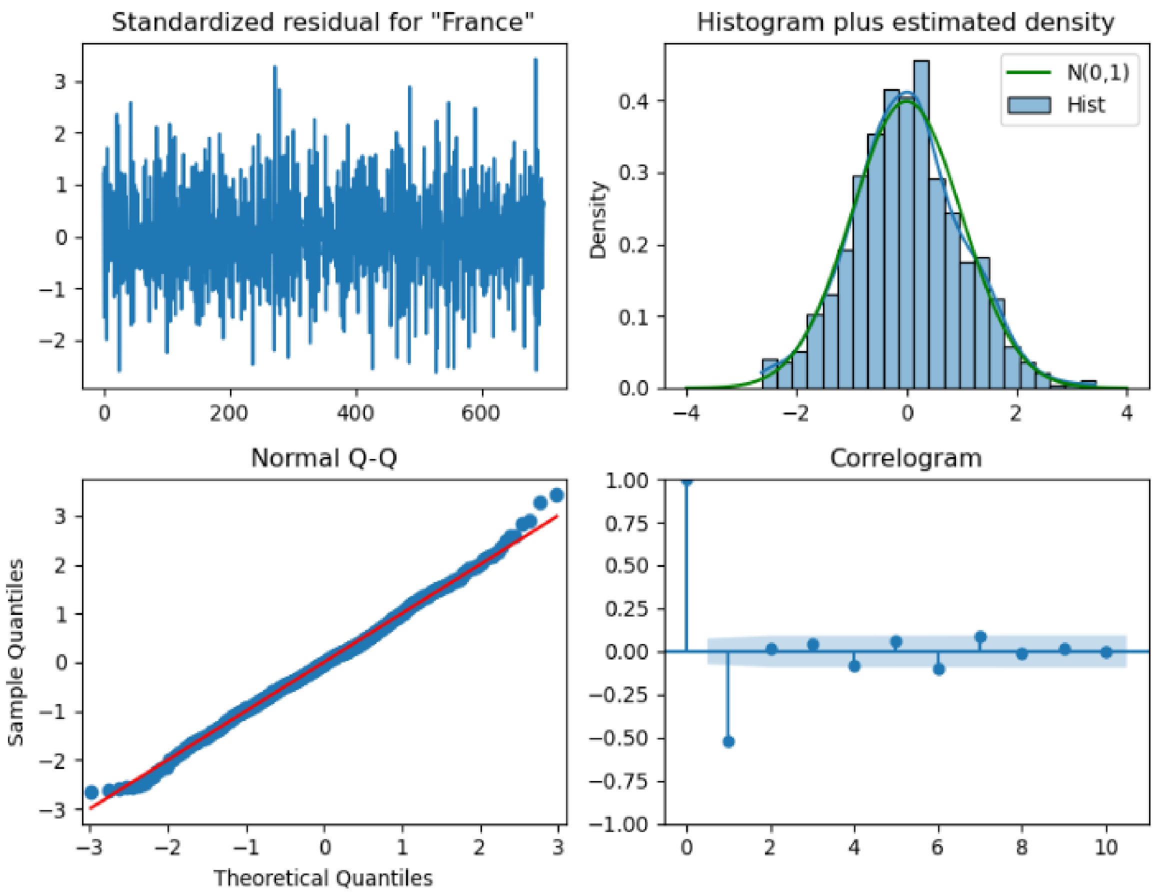

Figure 10: France ARIMA (0, 2, 1) Diagnostics Output

The diagnostic plots in Figure 10 assess the performance of the ARIMA (0, 2, 1) model for France's renewable energy consumption. The standardized residual plot in the top-left panel shows residuals fluctuating randomly around zero without any clear pattern, indicating that the residuals are uncorrelated and the model fits the data well. The histogram in the top-right panel reveals that the residuals are approximately normally distributed, as the histogram bars closely align with the overlaid normal distribution curve.

The Q-Q plot in the bottom-left panel further supports the normality assumption, with most residuals lying along the 45-degree reference line. Although there are minor deviations at the tails, these are not significant enough to undermine the assumption. The correlogram in the bottom-right panel shows that the autocorrelations of the residuals fall within the 95% confidence bounds, except for a small spike at lag 2. This slight deviation is unlikely to have a meaningful impact on the model's validity.

In summary, the ARIMA (0, 2, 1) model for France demonstrates a good fit. The residuals appear normally distributed, uncorrelated, and randomly scattered, confirming the model's suitability for forecasting renewable energy consumption.

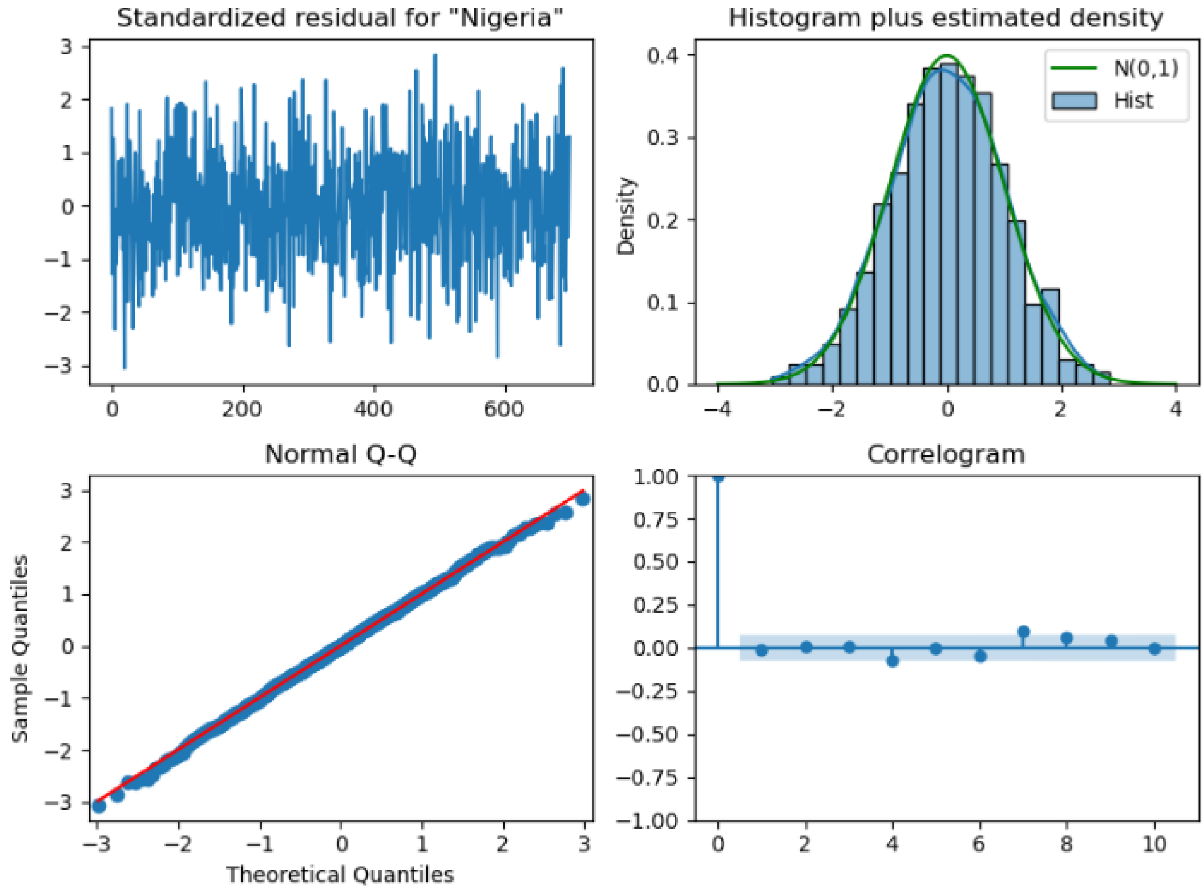

Figure 11: Nigeria ARIMA (1, 1, 1) Diagnostics Output

The diagnostic plots in Figure 11 evaluate the performance of the ARIMA (1, 1, 1) model fitted to Nigeria's renewable energy consumption data. The standardized residual plot in the top-left panel shows random fluctuations around zero without any clear pattern, suggesting that the residuals behave like white noise and that the model has adequately captured the underlying data structure. This randomness implies there are no significant remaining autocorrelations.

The histogram with the estimated density plot in the top-right panel indicates that the residuals closely follow a normal distribution, as the histogram bars align well with the overlaid normal curve, meaning the normality assumption is reasonably satisfied. The Q-Q plot in the bottom-left panel further supports this, showing residuals lying approximately along the 45-degree reference line with only minor deviations at the tails. This confirms that the residuals are approximately normally distributed, which is important for reliable statistical inference.

Finally, the correlogram in the bottom-right panel presents the autocorrelation function (ACF) of the residuals, with all autocorrelations falling within the 95% confidence bounds. This lack of significant autocorrelation indicates that the ARIMA (1, 1, 1) model has effectively captured the time-dependent structure in the data.

In summary, the diagnostic plots confirm that the ARIMA (1, 1, 1) model is an appropriate and well-fitting choice for modeling Nigeria's renewable energy consumption, with residuals that are normally distributed, uncorrelated, and randomly scattered.

3.4 The Grey Box Model

A Grey-Box model, which is a hybrid approach, was developed for each of the four countries to enable comparison with the best-performing ARIMA models. This comparison aims to identify and validate the model that best predicts renewable energy consumption. In constructing the Grey-Box model, the best ARIMA model identified for each country was used as the white-box component, while the Random Forest algorithm was employed as the black-box component.

The rationale behind this hybrid approach is to leverage the strengths of both models: the ARIMA model's capability to capture linear patterns and temporal dependencies, and the Random Forest's ability to model complex nonlinear relationships and variations that ARIMA may not adequately capture.

3.5 Model Comparison

Table 10: Comparison of the best ARIMA Model and the GREY-BOX Model Predictive Capacity in each Country

Country	Best ARIMA model		Grey Box Model	
	MSE	MAPE	MSE	MAPE
Angola	214.775	0.0871	2.1032	0.0195
Canada	20.2436	0.0652	0.0615	0.0088
France	4.5729	0.0899	0.0789	0.0196
Nigeria	242.905	0.0457	0.5178	0.0069

Table 10 presents a comparison of the predictive performance between the best ARIMA model and the Grey Box model for forecasting renewable energy consumption in Angola, Canada, France, and Nigeria. The evaluation metrics used are Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE).

For Angola, the ARIMA model yielded an MSE of 214.775 and a MAPE of 0.0871, whereas the Grey Box model achieved significantly better results with an MSE of 2.1032 and a MAPE of 0.0195. This demonstrates that the Grey Box model provides much more accurate and consistent predictions for Angola.

In Canada, the ARIMA model recorded an MSE of 20.2436 and a MAPE of 0.0652. The Grey Box model, however, delivered substantially improved performance, with an MSE of 0.0615 and a MAPE of 0.0088, indicating superior predictive accuracy.

For France, the ARIMA model produced an MSE of 4.5729 and a MAPE of 0.0899. The Grey Box model again outperformed it, achieving an MSE of 0.0789 and a MAPE of 0.0196, reflecting greater precision and lower forecasting error.

In Nigeria, the ARIMA model showed an MSE of 242.905 and a MAPE of 0.0457. The Grey Box model significantly reduced these errors, with an MSE of 0.5178 and a MAPE of 0.0069, highlighting its effectiveness in producing more accurate forecasts.

Overall, the Grey Box model consistently achieved lower MSE and MAPE values than the ARIMA model across all four countries. This indicates that the Grey Box approach offers superior predictive accuracy and generalization capability, establishing it as a more robust and reliable tool for forecasting renewable energy consumption.

3.6 Forecast of Renewable Energy Consumption

Table 11: Prediction with the Grey-Box from 2022 – 2050 for Renewable Energy Consumption

Year	Angola	Canada	France	Nigeria
2022	54.63	23.82	15.97	79.45
2023	55.88	23.94	16.34	79.57
2024	57.13	24.05	16.71	79.65
2025	58.38	24.17	17.08	79.7
2026	59.64	24.28	17.45	79.74

The prediction results presented in Table 11 summarize the projected renewable energy consumption for Angola, Canada, France, and Nigeria over the short-term forecast period beginning in 2022, as generated by the Grey-Box model. The results indicate differing growth trajectories across the four countries, reflecting variations in energy structures and development dynamics.

Angola's renewable energy consumption exhibits a steady upward trend, increasing from 54.63 units in 2022 to 59.64 units by 2026. This consistent growth suggests gradual expansion in renewable energy utilization, potentially supported by improvements in energy infrastructure and increased investment in renewable resources.

Canada also shows a moderate but stable increase in renewable energy consumption, rising from 23.82 units in 2022 to 24.28 units in 2026. Although the magnitude of growth is relatively small, the pattern reflects sustained development and consolidation of existing renewable energy capacity rather than rapid expansion.

France follows a similar trajectory, with renewable energy consumption increasing from 15.97 units in 2022 to 17.45 units in 2026. The steady rise indicates continuous progress in renewable energy adoption, consistent with long-standing national policies aimed at energy transition and sustainability.

Nigeria records comparatively high renewable energy consumption throughout the forecast period, starting at 79.45 units in 2022 and increasing slightly to 79.74 units by 2026. The marginal growth observed suggests a stable consumption pattern with limited short-term expansion, possibly reflecting structural constraints or slower implementation of new renewable energy initiatives.

Overall, the Grey-Box model projections reveal that Angola demonstrates the strongest growth momentum among the four countries, while Canada and France experience gradual but consistent increases. Nigeria maintains a high level of renewable energy consumption, though with minimal growth during the forecast period. These short-term projections, highlight both progress and stagnation across countries, underscoring the need for targeted policy interventions to stimulate further renewable energy development where growth remains limited.

4. Discussion of Results

The results of this study provide clear empirical evidence on the comparative performance of ARIMA and Grey-Box models in forecasting renewable energy consumption across Angola, Canada, France, and Nigeria. Diagnostic assessments of the ARIMA models indicate that they are generally effective in capturing linear dynamics and short-term temporal patterns in the renewable energy consumption series, particularly for Canada, France, and Nigeria. Residual diagnostics and correlogram analyses (Figures 8–11) show no significant autocorrelation and suggest adequate model specification. These findings are consistent with earlier studies such as [23] and [24], which demonstrated the suitability of ARIMA models for modeling energy demand and renewable electricity trends in relatively stable and structured energy systems. However, the limitations of the ARIMA framework become evident in the cases of Angola and Nigeria. In these countries, residual behavior and performance metrics suggest the presence of nonlinearities, structural irregularities, and higher levels of noise, characteristics commonly associated with developing-country energy datasets. Similar challenges have been documented in the literature, including [25], which highlights the difficulty of applying purely linear time-series models in contexts marked by data instability and evolving energy infrastructures.

To address these limitations, a Grey-Box modeling framework was developed by integrating the optimal ARIMA specification (white-box component) with a Random Forest algorithm (black-box component). This hybrid structure exploits ARIMA's ability to model linear temporal dependence while allowing Random Forest to learn nonlinear residual patterns unexplained by the statistical model. Such an approach aligns with the methodological rationale presented in [27] and [28], where Grey-Box models were shown to outperform standalone models by combining interpretability with predictive flexibility in energy forecasting applications.

The comparative performance results summarized in Table 10 clearly demonstrate that the Grey-Box model outperforms the ARIMA model across all four countries. Substantial reductions in Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) were observed, confirming the hybrid model's superior ability to capture complex consumption dynamics. These findings support prior evidence reported in [31] and [32], where hybrid forecasting models consistently outperformed traditional time-series approaches. Similar improvements were also reported by [33], who showed that combining ARIMA with machine-learning techniques enhances forecasting accuracy by jointly modeling linear and nonlinear components.

Short-term forecasts generated by the Grey-Box model for the period 2022–2026, presented in Table 11, reveal distinct consumption trajectories across the four countries. Angola exhibits the strongest growth trend, suggesting ongoing expansion in renewable energy utilization. Canada and France display steady but moderate increases, reflecting gradual capacity expansion and consolidation of existing renewable infrastructure. Nigeria maintains a relatively high level of renewable energy consumption throughout the forecast period, with only marginal growth, potentially indicating structural constraints or limited short-term policy acceleration, consistent with observations in [25] and [36].

Overall, these results reinforce the conclusion that while ARIMA models remain useful for short-term forecasting in relatively stable and linear contexts ([34]), they are limited in their ability to capture the nonlinear and heterogeneous dynamics characteristic of renewable energy consumption, particularly in developing economies. Grey-Box models, although more complex to implement ([35]), offer enhanced robustness and flexibility by integrating statistical structure with machine-learning adaptability.

4.1 Summary of Findings

This section synthesizes the key findings of the study by linking them directly to the stated research objectives. The analysis focused on modelling renewable energy consumption patterns in Angola, Canada, France, and Nigeria using both traditional time-series techniques and a hybrid Grey-Box framework. Overall, the findings demonstrate that while country-specific ARIMA models provide a sound statistical foundation, combining them with machine-learning methods significantly enhances forecasting accuracy and reliability. The summary below presents the major outcomes of the study in line with each research objective.

Objective 1: Fit individual models to country-specific data

In addressing the first objective, individual ARIMA models were successfully fitted to the renewable energy consumption data for each country. Diagnostic evaluations indicated that the selected models were statistically adequate and well specified. The standardized residuals fluctuated randomly around zero, suggesting no systematic bias, while histogram and Q–Q plot analyses showed that the residuals closely approximated a normal distribution. In addition, autocorrelation analyses revealed no significant remaining serial dependence. Based on information criteria and diagnostic results, the optimal model structures were identified as ARIMA (0,2,2) for Angola, ARIMA (0,2,1) for both Canada and France, and ARIMA (1,1,1) for Nigeria. These outcomes confirm that the modelling approach effectively captured the underlying temporal dynamics unique to each country.

Objective 2: Compare predictive performance

The second objective focused on evaluating and comparing the predictive accuracy of the individual ARIMA models with the proposed Grey-Box models. The results clearly show that the hybrid ARIMA–Random Forest framework outperformed the standalone ARIMA models across all countries. Substantial reductions in forecast errors were observed when using the Grey-Box approach, as measured by Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). For example, in Angola, the MSE dropped dramatically from 214.775 under the ARIMA model to 2.1032 under the Grey-Box model, while MAPE declined from 0.0871 to 0.0195. Comparable improvements were recorded for Canada, France, and Nigeria, highlighting the consistent superiority of the hybrid approach across different national contexts.

Objective 3: Validate model reliability using diagnostic metrics

With respect to model reliability, diagnostic metrics provided strong evidence that both the individual ARIMA models and the Grey-Box models were robust and dependable. Residual diagnostics confirmed the adequacy of the ARIMA specifications, while the consistently lower error measures achieved by the Grey-Box models further reinforced their reliability. By integrating linear time-series components with nonlinear learning mechanisms, the hybrid models were better able to capture complex patterns in the data, thereby improving stability and generalizability across countries with differing energy profiles.

Objective 4: Provide short-term forecasts to guide energy policy

In fulfilling the final objective, short-term forecasts for the period 2022–2026 revealed meaningful differences in renewable energy consumption trajectories among the countries studied. Angola exhibited the strongest growth trend, suggesting rapid expansion in renewable energy utilization. Canada and France demonstrated steady but moderate increases, reflecting relatively mature renewable energy systems. Nigeria, on the other hand, maintained comparatively high consumption levels with limited variability over the forecast horizon. These projected trends underscore the importance of adopting flexible and context-sensitive forecasting tools to support informed energy planning and policy development.

Taken together, the findings of this study indicate that Grey-Box models provide a more accurate and adaptable forecasting framework than traditional ARIMA models when analyzing renewable energy consumption. By capturing both linear trends and nonlinear structures inherent in the data, the hybrid approach offers valuable insights for short-term forecasting and policy guidance, particularly in heterogeneous and data-constrained environments.

5. Contribution to Knowledge and Innovation

This study contributes significantly to the body of knowledge by empirically demonstrating the superiority of hybrid Grey-Box models over traditional ARIMA models in forecasting renewable energy consumption across different national contexts. It uniquely applies this methodology to four countries—Angola, Canada, France, and Nigeria—revealing both linear and nonlinear consumption trends. The integration of ARIMA and Random Forest in a Grey-Box framework is innovative, particularly in its ability to capture both theoretical structure and complex, data-driven relationships. Moreover, this research fills a critical gap by providing a comparative evaluation of models and validating their predictive performance using robust metrics such as MSE and MAPE, offering practical guidance for energy policy planners and modelers in both developed and developing economies.

5.1. Conclusion

The study concludes that while ARIMA models are useful for short-term linear forecasting, they are limited in capturing nonlinear dynamics typical of real-world energy consumption, especially in developing countries with irregular data patterns. The Grey-Box model, combining ARIMA with Random Forest, proves more accurate, robust, and adaptive to varying data complexities. Angola is projected to lead in renewable energy growth, while Canada and France show consistent upward trends. Nigeria, although starting from a high consumption base, demonstrates stagnation in projected growth. These findings underscore the necessity of adopting more advanced and flexible modeling techniques for long-term energy forecasting and planning.

5.2 Recommendation for Further Studies

Future research should extend the scope of this study by incorporating additional external variables such as government policy shifts, climate conditions, economic growth rates, and technological advancements to further enhance model accuracy. Comparative analysis with other hybrid models such as ARIMA-ANN or ARIMA-XGBoost could provide deeper insights into optimal configurations for energy forecasting. Furthermore, real-time forecasting models and dynamic updating systems should be explored to allow for adaptive prediction in the face of rapidly changing energy landscapes, especially in data-scarce environments like Angola and Nigeria.

Reference

1. REFERENCEIEA. (2021). *Renewables2021: Analysis and forecast to 2026*. Cc 4.0. https://www.iea.org/reports/renewables-2021
2. United Nations. (2015). *Paris Agreement*. https://unfccc.int/process-and-meetings/the-paris-agreement
3. U.S. EIA. (2023). *Angola 2023 primary energy data*. https://www.eia.gov/international/analysis/country/AGO
4. Natural Resources Canada. (2022). *Canada's renewable energy profile*. https://natural-resources.canada.ca/energy-sources/renewable-energy
5. Wikipedia. (2024). *Electricity sector in France*. https://en.wikipedia.org/wiki/Electricity_sector_in_France
6. Kemp, Y. (2024). Solar energy for off-grid communities. *Renewable Energy*, 191, 114321. https://doi.org/10.1016/j.rser.2023.114321
7. Van Poecke, A., Tabari, H., & Hellinckx, P. (2024). Unveiling the backbone of the renewable energy forecasting process: Exploring direct and indirect methods and their applications. Energy Reports 11 544–557. www.elsevier.com/locate/egyr
8. Alabi, N. O., & Ojenike, O. T. (2024). Forecasting solar radiation for renewable energy sustainability in Nigeria using panel dihybrid recurrent neural network. Man, Environment and Society, 5(2), 193–209. https://doi.org/10.47509/MES.2024.v05i2.05

9. Bassey, K. E. (2023). Solar energy forecasting with deep learning technique. Engineering Science & Technology Journal, 4(2), 18–32. https://doi.org/10.51594/estj/v4i2.1286
10. Reichert, B., Souza, A. M., & Mezzomo, M. (2022). Forecasting electricity generation from renewable sources during a pandemic. Gestão & Produção, 29, e024. https://doi.org/10.1590/1806-9649-2022v29e024
11. Teixeira, R., Cerveira, A., Pires, E. J. S., & Baptista, J. (2023). Advancing renewable energy forecasting: A comprehensive review of renewable energy forecasting methods. Energies, 16(18), 6727. https://doi.org/10.3390/en16186727
12. Sribna, Y., Koval, V., Olczak, P., Bizonych, D., Matuszewska, D., & Shtyrov, O. (2021). Forecasting solar generation in energy systems to accelerate the implementation of sustainable economic development. Polityka Energetyczna – Energy Policy Journal, 24(3), 5–28. https://doi.org/10.33223/epj/141095
13. Manowska, A. (2021). Forecasting of the share of renewable sources in the total final energy consumption for selected European Union countries. IOP Conference Series: Earth and Environmental Science, 906(1), 012134. https://doi.org/10.1088/1755-1315/906/1/012134
14. Neumann, O., Turowski, M., Mikut, R., Hagenmeyer, V., & Ludwig, N. (2023). Using weather data in energy time series forecasting: The benefit of input data transformations. Energy Informatics, 6(1), 4. https://doi.org/10.1186/s42162-023-00299-8
15. Sudasinghe, P., Herath, D., Weeratunge, H., Karunarathne, I., & Jayasuriya, L. (2025). Forecasting renewable energy for microgrids using machine learning. Discover Applied Sciences, 7(5), 555. https://doi.org/10.1007/s42452-025-06002-8
16. Olabiyi, W. (2023). Towards accurate renewable energy forecasting: A transformer-based hybrid approach. [Unpublished manuscript]. Available from https://www.researchgate.net/publication/384525192
17. Sweeney, C., Bessa, R. J., Browell, J., & Pinson, P. (2020). The future of forecasting for renewable energy. Wiley Interdisciplinary Reviews: Energy and Environment, 9(2), e365. https://doi.org/10.1002/wene.365
18. Seitter, K. L. (2024). Weather as fuel: The wicked problem of renewable energy. Bulletin of the American Meteorological Society, 105(11), E2231–E2242. https://doi.org/10.1175/BAMS-D-23-0289.1
19. Bouquet, P., Jackson, I., Nick, M., & Kaboli, A. (2023). AI-based forecasting for optimised solar energy management and smart grid efficiency. International Journal of Production Research, 62(20), 7397–7411. https://doi.org/10.1080/00207543.2023.2269565
20. Yeramolu, V. D. D. (2021). Solar energy forecasting using machine learning. Journal of Emerging Technologies and Innovative Research, 8(8), c289–c294. https://www.jetir.org/papers/JETIR2108337.pdf
21. Anuradha, K., Erlapally, D., Karuna, G., Srilakshmi, V., & Adilakshmi, K. (2021). Analysis of solar power generation forecasting using machine learning techniques. E3S Web of Conferences, 309, 01163. https://doi.org/10.1051/e3sconf/202130901163
22. Nerini, F. F., Sovacool, B. K., Hughes, N., Cozzi, L., Cosgrave, E., Howells, M., ... & Milligan, B. (2021).
Connecting climate action with other Sustainable Development Goals. *Nature Sustainability*, **4**(8), 674–680.
https://www.nature.com/articles/s41893-019-0334-y
23. Wang, Y., Zhang, R., & Li, H. (2017). ARIMA for Canadian energy demand. *Energy Economics*, 64, 105–117. https://doi.org/10.1016/j.eneco.2017.03.012
24. Bunn, D. W., & Farmer, E. D. (1985). ARIMA in France. IEEE Trans. *Power Systems*, 4(3), 315–323. https://doi.org/10.1109/TPWRS.1985.318951
25. Bello, A., Smith, B., Johnson, C., & Lee, D. (2016). ARIMA challenges in Africa. *Energy Reports*, 2, 45–52. https://doi.org/10.1016/j.egyr.2016.05.003

26. Ya-fei, L., Wu, J., & Adams, M. (2013). Grey-Box for buildings. *Building & Environment*, 60, 81–90. https://doi.org/10.1016/j.buildenv.2012.11.003
27. Pitchforth, J., Harris, M., & White, R. (2021). Grey-Box for wave energy. *Renewable Energy*, 163, 2127–2140. https://doi.org/10.1016/j.renene.2020.10.123
28. Shahabaldin, A., Thompson, B., & Martinez, P. (2022). Grey-Box for thermal dynamics. *Energy & Buildings*, 254, 111567. https://doi.org/10.1016/j.enbuild.2021.111567
29. Zhang, G. P. (2003). Hybrid ARIMA-NN. *Neurocomputing,* 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0
30. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://www.deeplearningbook.org
31. Reza, M., Khan, A., & Ahmed, S. (2019). Grey-MZIMA for CO_2. *Energy*, 180, 434–445. https://doi.org/10.1016/j.energy.2019.05.098
32. Barak, S., & Sadegh, S. S. (2019). ARIMA-ANFIS hybrid. *Energy,* 167, 402–414. https://doi.org/10.1016/j.energy.2018.10.123
33. Barak, S., & Freede Salahed, S. (2020). MARIMA-ANFIS. *Applied Energy*, 261, 114456. https://doi.org/10.1016/j.apenergy.2019.114456
34. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. https://otexts.com/fpp2/
35. Liu, X., Chen, Y., Wang, Z., & Zhang, Q. (2017). Grey-Box long-term. *Applied Energy*, 189, 123–135. https://doi.org/10.1016/j.apenergy.2016.12.045
36. Wang, L., & Qin, Y. (2019). Grey-Box in Africa. *Energy Policy*, 125, 234–246. https://doi.org/10.1016/j.enpol.2018.10.042
37. Bartholomew, D.C., Nwaigwe, C.C., Orumie, U.C. & Nwafor, G.O. (2023). Intervention Analysis of COVID-19 Vaccination in Nigeria: The Naive Solution Versus Interrupted Time Series, *Annals of Data Science*, 11, 1609–1634. https://link.springer.com/article/10.1007/s40745-023-00462-8
38. Bartholomew, D.C., Olewuezi, N.P., Nwaigwe, C.C. & Akanno, F.C. (2024). Modeling the Effect of Population Size on Banking Transaction Channels in Nigeria: Grey box vs Support Vector Regression, *CBN Journal of Applied Statistics*, *15*, 2476–8472.
39.