

Research Article

# Predicting Ghana's Daily Natural Gas Consumption Using Time Series Models

Eric Broni-Bediako<sup>1,\*</sup> , Albert Buabeng<sup>2</sup> , Philip Allotey<sup>1</sup> 

<sup>1</sup>Department of Petroleum and Natural Gas Engineering, School of Petroleum Studies, University of Mines and Technology, Tarkwa, Ghana

<sup>2</sup>Department of Mathematical Sciences, Faculty of Computing and Mathematical Sciences, University of Mines and Technology, Tarkwa, Ghana

## Abstract

In recent years, natural gas utilisation has seen a considerable increase because, it presents an alternative energy source that is reliable, economical and environmentally friendly for consumers. In Ghana, natural gas consumption has over the years increased due to mainly the rise in industrial and residential demands. Accurate prediction of natural gas consumption will provide stakeholders with vital information needed for planning and making informed policy decisions. This paper explores the Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) to predict Ghana's daily natural gas consumption. The data employed for the study is daily natural gas consumption in Ghana from 2020 to 2022. The results show that both ARIMA and SARIMA models can predict the consumption of natural gas in Ghana with a good degree of accuracy. The SARIMA model slightly outperforms the ARIMA model, with a Root Mean Square Error (RMSE) of 22.25 and a Mean Absolute Percentage Error (MAPE) of 6.96%, compared to an RMSE of 23.27 and a MAPE of 7.29% for the ARIMA model. The model forecast suggests a steady natural gas consumption in Ghana but with some intermittent fluctuations.

## Keywords

ARIMA, SARIMA, Natural Gas Consumption, Mean Absolute Percentage Error, Root Mean Square Error, Time Series Analysis

## 1. Introduction

Natural gas plays a crucial part in Ghana's energy sector. Its utilization has experienced significant growth in recent years due to its environmental benefits, cost-effectiveness, and potential for reducing dependence on traditional fossil fuels [1]. In Ghana, natural gas is predominantly used for domestic power supply for industries, transport, and cooking [2-4]. This has increased natural gas consumption exponentially over the decades. Over two decades, Ghana's natural

gas consumption (NGC) has increased by 52.6% [5]. The Ghana Gas Company in 2020 revealed that natural gas makes up around 60% of Ghana's thermal energy generation. From just 20 million ft<sup>3</sup>/day in the early 2000s, total natural gas consumption increased significantly to around 161 million cubic feet per day in 2020 [6]. Ghana's NGC patterns might show seasonal variations, trends, and perhaps nonlinear correlations. Economic growth, population growth, urban-

\*Corresponding author: ebron-bediako@umat.edu.gh (Eric Broni-Bediako)

**Received:** 13 January 2024; **Accepted:** 25 January 2024; **Published:** 7 March 2024



Copyright: © The Author(s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

ization, growth in the manufacturing sector, the expansion of gas pipeline infrastructure, hydrology and droughts that affect hydropower generation, and government policies on energy pricing and subsidies are some of the major variables that affect consumption [7]. Accurate prediction of NGC in Ghana is vital for stakeholders such as the government, energy suppliers, and investors to make informed decisions regarding effective energy planning, resource allocation, supply chain management, and policy formulation to ensure a sustainable and reliable energy supply [8-11]. By anticipating future demand, the government can plan the expansion of gas infrastructure, optimise resource allocation, and ensure a stable supply to meet the growing energy requirements of various sectors.

In literature, several methods have been employed by researchers to predict NGC, including statistical methods [12-16], econometric methods [17-19], and machine learning methods [20-22]. However, these methods have certain limitations. Statistical methods such as simple linear regression may oversimplify the complex dynamics and non-linear patterns present in natural gas consumption time series data [23]. Econometric methods often require a strong theoretical foundation and may be sensitive to model assumptions [24]. Machine learning methods, although powerful, can be computationally expensive and may lack interpretability [25]. Notwithstanding the limitations of statistical models, this study explored statistics-based Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) to forecast natural gas consumption in Ghana due to their ability to account for various patterns such as linear and non-linear trends, varying or constant volatility, as well as seasonal or non-seasonal fluctuations [26, 27]. Also, these models are simple to implement as they require few parameters and assumptions which are based on statistical methods and theories. For these reasons, their application to various disciplines is numerous, including the petroleum industry. For instance, Ediger and Akar [26] estimated Turkey's projected primary energy consumption from 2005 to 2020 using ARIMA and SARIMA. Akpınar and Yumusak [28] employed the Holt-Winters exponential smoothing method and the ARIMA model to forecast natural gas demand. Erdogdu [29] used the ARIMA model to forecast the future growth of Turkey's gas demand. Akkurt et al. [30] employed SARIMA and linear regression models in a comparative study to forecast Turkey's natural gas consumption on a monthly and yearly base using time series data with a single seasonal pattern. These research works confirm the superiority of ARIMA and SARIMA models in modelling time series events when compared to other statistical techniques.

In continuance, the contribution of this study is to develop statistics-based models for predicting the consumption of natural gas in Ghana through a comparative study of ARIMA and SARIMA models. By incorporating historical gas consumption data, the ARIMA and SARIMA models identify and model the underlying patterns and dependencies, leading

to more accurate and reliable predictions. Also, the model's ability to handle non-linear and dynamic relationships within the data makes it a suitable choice for capturing the complexities of national gas consumption in Ghana. Hence, the research advances Ghana's energy landscape by addressing the need for predicting natural gas consumption patterns in Ghana, thus, allowing for effective decision-making and policy formulation.

## 2. Resources and Methods

### 2.1. Data Source and Description

In this study, secondary data on historical daily natural gas consumption in Ghana was obtained from the Ghana National Gas Company. The dataset was for three years spanning from 1<sup>st</sup> January 2020 to 31<sup>st</sup> December 2022. Figure 1 shows the plot of daily natural gas consumption in Ghana within the timeframe (1096 observations). As observed, there are upward and downward trends, and nonconstant variance, among others within the data, portraying the nonlinear aspects of daily natural gas consumption in Ghana.

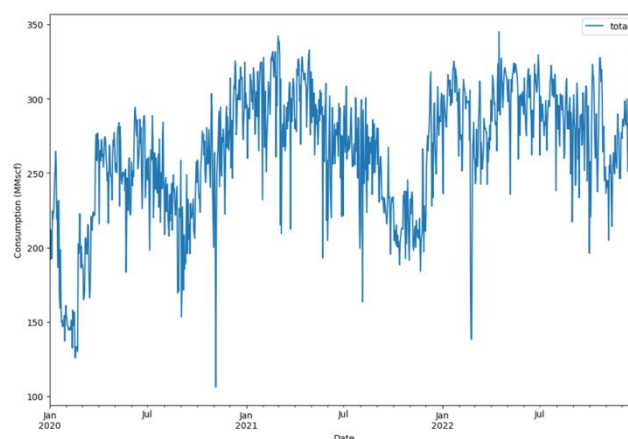


Figure 1. Plot of Daily Natural Gas Consumption in Ghana.

### 2.2. Testing for Stationarity

A crucial characteristic of time series data is stationarity, which shows that the data's statistical characteristics remain constant across time. It is necessary for many time series analytic approaches, such as modelling and forecasting. There are two primary methods for determining stationarity and they are visual inspection and statistical test. For visual inspection, time series data are plotted and trends, seasonality, or other patterns that change over time are looked for. Several statistical tests can be employed for testing stationarity, such as the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS).

The ADF test is commonly used to check the stationarity of

a time series by looking for the presence of a unit root in the data. If the ADF test rejects the null hypothesis of a unit root, then the time series is considered to be stationary [32]. In this study, the ADF test was used to find out whether the time series data used had unit root or were covariance stationary. This method was proposed by Dickey and Fuller [33] as an upgraded version of the Dickey-Fuller test. The unit root test is done by stating the null hypothesis as  $\gamma = 1$  (non-stationary) and the alternative hypothesis as  $\gamma < 1$  (covariance stationary). Where  $\gamma$  is the characteristic root of an AR polynomial. The ADF test statistic is given by Equation (1).

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + \sum_{j=1}^P \varphi_j \Delta y_{t-j} + \varepsilon_t \quad (1)$$

where  $D_t$  is a vector of deterministic terms (constant, trend *etc.*). The  $P$  lagged difference terms,  $\Delta y_{t-j}$ , are used to approximate the mean equation structure of the errors,  $\pi = \phi - 1$ , and the value of  $P$  is set so that the error,  $\varepsilon_t$  is serially uncorrelated.

Contrary to most unit root tests, like ADF, the absence of a unit root is not a proof of stationarity, but by design, of trend-stationarity. This was addressed in this work by using the KPSS test developed by Kwiatkowski *et al.* [34]. Upon testing the dataset any values above 0.05 meant the dataset was stationary. KPSS is defined by Equation (2).

$$KPSS = \frac{T^{-2} \sum_{t=1}^T \hat{S}_t^2}{\hat{\lambda}^2} \quad (2)$$

where  $\hat{S}_t = \sum_{j=1}^t \hat{u}_j$ ,  $\hat{u}_t$  is the residual of a regression of  $y_t$  on  $D_t$  and  $\hat{\lambda}^2$  is a consistent estimate of the long-run variance of  $u_t$  using  $\hat{u}_t$ .

### 2.3. Autoregressive Integrated Moving Average Model

The ARIMA is a statistical model that is used to predict future values of a time series [35]. The order of the ARIMA model is  $(p, d, q)$ . The ARIMA  $(p, d, q)$  model is made up of three components. The  $p$  or AR is the number of lag variables to be used as predictors. The  $q$  or MA shows that the present value has something to do with the past residuals. The “ $d$ ” is the order of differencing required to make the time series stationary. The ARIMA  $(p, d, q)$  process can be defined as expressed in Equation (3).

$$\phi_p(B)(B^d Y_t) = \phi_q(B)e_t \quad (3)$$

where  $Y_t$  is the number of TB cases recorded at time  $t$ ;  $\Delta^d$  is the order of differencing;  $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$  is the Autoregressive (AR) characteristic operator; and

$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$  is the Moving Average (MA).

The estimation of the model consists of three steps, namely: model identification, estimation of parameters and diagnostic checking:

#### 2.3.1. Model Identification

In identifying the appropriate model for predicting daily natural gas consumption, a visual assessment of the Auto-correlation Function (ACF) and the sample Partial Autocorrelation Function (PACF) together with two (2) information loss metrics were used. The two information criteria, the Akaike Information Criteria (AIC) and the Bayesian Information Criterion (BIC) are defined as shown in Equations (4) and (5).

$$AIC = \ln(\hat{\sigma}^2) + \frac{2K}{S} \quad (4)$$

$$BIC = \ln(\hat{\sigma}^2) + \frac{K}{S} \ln s \quad (5)$$

where  $\hat{\sigma}^2$  is the variance of the residuals,  $S$  is the sample size,  $K$  is the total number of parameters. The best model is the model that has the least AIC and BIC values [36].

#### 2.3.2. Estimation of Parameters

The second step is the estimation of the model parameters for the “best model” that has been selected.

#### 2.3.3. Diagnostic Checking

The estimated model is then checked to verify if it adequately represents the series. Diagnostic checks are performed to ascertain whether the residuals of the selected model are randomly and normally distributed. In this paper, the standardised residual, the ACF of residuals and Ljung-Box statistic plots were employed. Collectively, these tests establish the adequacy of the selected model.

### 2.4. Seasonal Autoregressive Integrated Moving Average Model

The SARIMA model is an extension of the ARIMA  $(p, d, q)$  model that incorporates seasonal patterns in the data [35, 37]. Seasonality is the tendency of a time series to repeat itself over regular intervals of time, such as daily, monthly, quarterly, or yearly. The order of the SARIMA model is represented as  $(p, d, q)(P, D, Q, s)$  and expressed as Equation (6).

$$\phi_p(B)\Phi_P(B^s)Y_t = \phi_q(B)\Theta_Q(B^s)e_t \quad (6)$$

where  $Y_t$  is the number of TB cases recorded at time  $t$ ;  $\Delta^d$  is the order of differencing;  $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$  is the non-seasonal AR operator;

$\Phi_P(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})$  is the seasonal AR operator;

$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$  is the non-seasonal MA; and

$\theta_Q(B^s) = (1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs})$  is the seasonal MA.

Similar to the ARIMA, SARIMA modelling also follows the three steps as discussed in Sections 2.4.1 - 2.4.3.

## 2.5. Performance Evaluation

The Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used to establish the forecasting performance of the ARIMA and SARIMA models. RMSE is calculated by taking the square root of the mean squared errors between the actual and predicted values. RMSE is expressed mathematically as presented in Equation (7) [38-40].

$$RMSE = \left[ \frac{\sum_{i=1}^N (y_{fi} - y_{oi})^2}{N} \right]^{1/2} \quad (7)$$

where  $(y_{fi} - y_{oi})^2$  is the difference squared of the actual and predicted value and  $N$  is the sample size.

MAPE is determined by taking the average of the absolute percentage errors between the actual and forecasted values. MAPE is expressed mathematically as shown in Equation (8) [36, 37].

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100\% \quad (8)$$

where  $n$  represents the number of observations,  $A_t$  represents the actual value and  $F_t$  is the forecasted or predicted value.

## 3. Results and Discussion

### 3.1. Preliminary Analysis

Table 1 shows the descriptive statistics of the daily natural gas consumption dataset shown in Figure 1. As seen, an average of 263.34 MMscf of natural gas was consumed daily. Within the timeframe, the minimum and maximum natural gas consumed daily was 106.19 and 345.13 MMscf respectively.

Table 1. Descriptive Statistics of the Data.

Variable	Unit	Mean	Standard Deviation	Minimum	Maximum	Observations
Consumption	MMscf	263.34	42.38	106.19	345.13	1096

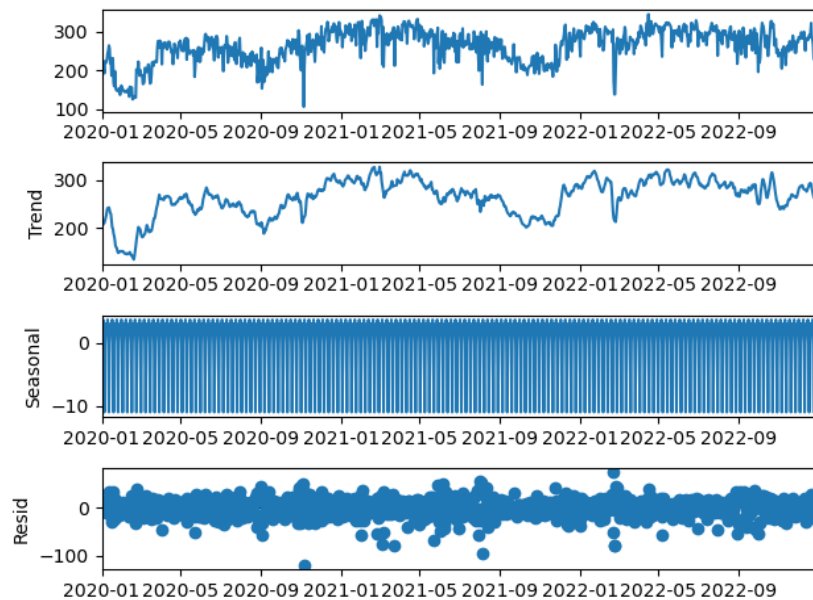


Figure 2. Decomposition Plot of the Daily Natural Gas Consumption Data.

Figure 2 shows the decomposed form of the daily natural gas consumption dataset (Figure 1) which reveals the various

time series components. In Figure 2, a trend existed but was not steady which suggests that the data was not stationary in

its initial form.

### 3.2. Data Processing

The dataset was split into training and testing. This was done to validate the performance of the selected models. The training set was from January 2020 to November 2022 and the testing set was December 2022. The models are developed using the training dataset and validated based on the reserved test data.

Before proceeding with the development of the ARIMA and SARIMA models, ACF and PACF plots (Figure 3) were used to assess the stationarity of the training data. From Figure 3, both the ACF and PACF plots of the dataset at initial order 0 were not stationary since there was a gradual drop in the spike level but showed a trend and hence confirmed that the data was not stationary. However, after differencing, the data became stationary since there is a significant drop in spike level with deflecting spikes in both negative and positive sides as indicated by the ACF and PACF plots in Figure 4.

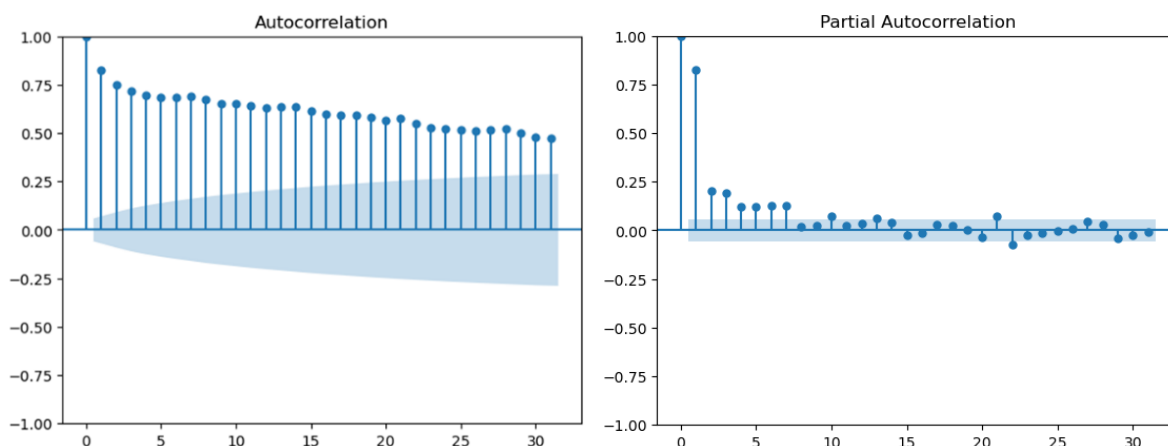


Figure 3. ACF and PACF Plot of Dataset (order 0).

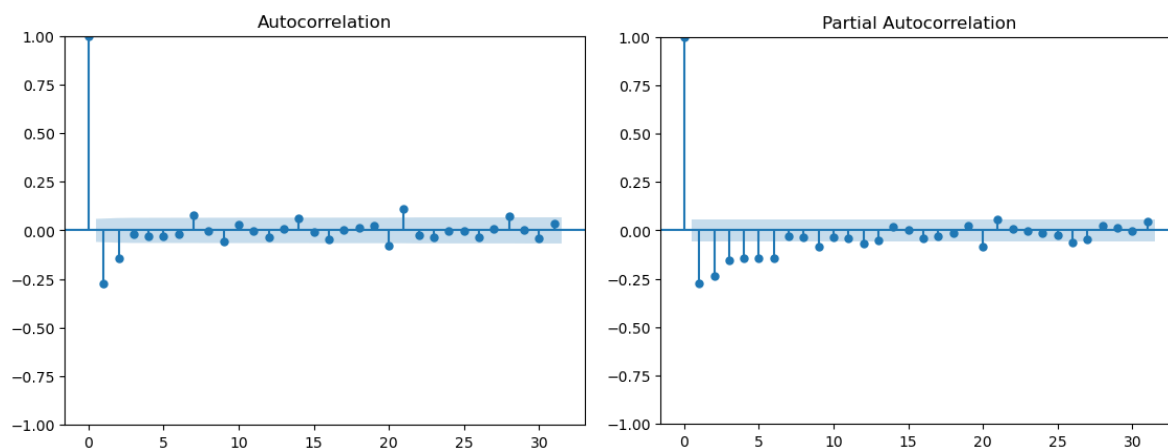


Figure 4. ACF and PACF Plot of Dataset (Order 1).

Table 2. Summary of Stationarity Test on Dataset.

Test	Hypothesis	Differencing (P-value)	
		Order 0	Order 1
ADF	H <sub>0</sub> : Not Stationary H <sub>1</sub> : Stationary	0.02	< 0.01
KPSS	H <sub>0</sub> : Stationary H <sub>1</sub> : Not Stationary	0.01	0.10



To confirm the results indicated by the ACF and PACF plots, a unit root test was conducted to check the non-stationary assumption using the ADF and KPSS tests. The unit root test was carried out on the training dataset and the results are presented in Table 2. KPSS test confirmed the trend assumption that the series was not stationary and the ADF confirmed that, the series had no unit root present. So, further tests for both ADF and KPSS were conducted after taking the first difference (Order 1), and both tests confirmed that the data was stationary (Table 2).

### 3.3. ARIMA Model Estimation

To find a suitable ARIMA model for forecasting the daily consumption of natural gas in Ghana, the ACF and PACF shown in Figure 4 were assessed. As shown in Figure 4, higher orders of AR and MA were suggested. However, the trial-and-error method of estimating the appropriate ARIMA model was adopted. Several models for different  $AR(p)$  and  $MA(q)$  orders (for  $p, q = 1, 2, 3, 4$ ) were generated and their respective AIC and BIC values were estimated. Four models with the least information loss were selected as being competitive as shown in Table 3. These competitive ARIMA models are further evaluated for the best model. The four competing models were used to predict the daily natural gas consumption for December 2022 (test data) and their prediction accuracy/error were computed in Table 3. From Table 3, the four competing models were evaluated using RMSE and MAPE metrics. It was observed that ARIMA (3, 1, 3) had the least error among the four competing models, and thus was selected as the best ARIMA model.

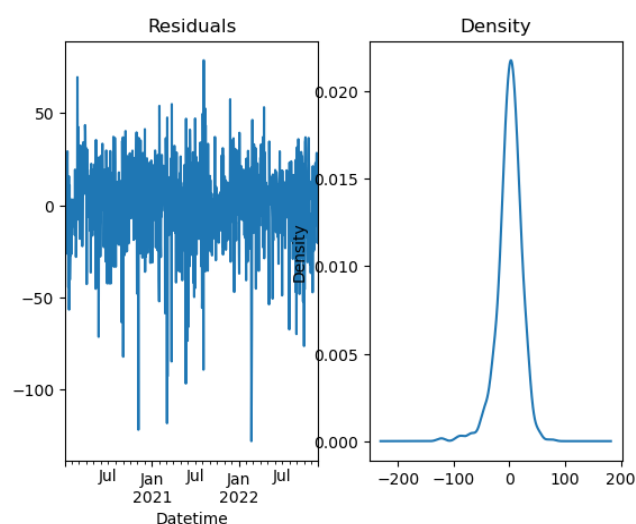
**Table 3.** Competing ARIMA Models and their Respective RMSE and MAPE.

Model	RMSE	MAPE (%)
ARIMA (2, 1, 1)	23.30	7.36
ARIMA (1, 1, 2)	23.31	7.36
ARIMA (0, 1, 2)	23.29	7.30
ARIMA (3, 1, 3)	23.27	7.29

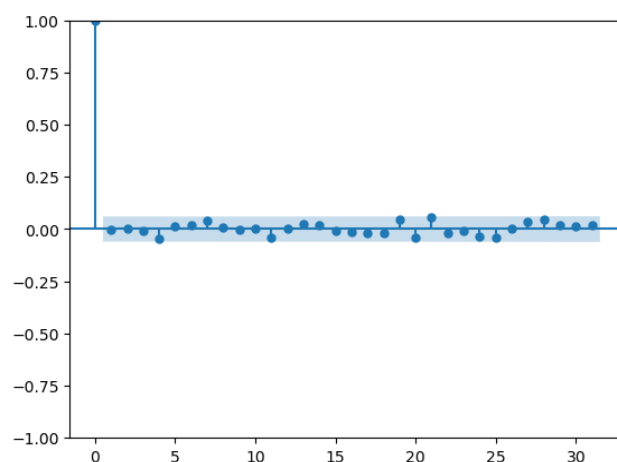
Table 4 shows the parameter estimates for ARIMA (3, 1, 3). As observed, apart from the first lag variable of the AR, the rest of the lag variables were all significant.

**Table 4.** Parameter Estimate for ARIMA (3, 1, 3).

Parameter	Coefficient	Std. Error	z value	$P >  z $
AR.L1	-0.0445	0.033	-1.346	0.178
AR.L2	-0.8114	0.015	-54.173	<0.0001
AR.L3	0.3980	0.032	12.528	<0.0001
MA.L1	-0.4100	0.027	-15.317	<0.0001
MA.L2	0.5891	0.018	32.517	<0.0001
MA.L3	-0.8426	0.025	-33.902	<0.0001
sigma2	454.5431	11.530	39.422	<0.0001



**Figure 5.** ARIMA (3,1,3) Standardized Residual and Density Plot.



**Figure 6.** ACF Plot for the ARIMA (3,1,3) Residuals.

Figures 5 and 6 show the diagnostic plot for the residuals of

ARIMA (3, 1, 3). As observed, the standardized residual and density plots (Figure 5) display normally distributed residuals as the points display zero trace of trend, no extreme outliers, and in general, no extreme change in variance across time. Also, the ACF of the residuals plot (Figure 6) show that only one lag out of the 31 lags of the series residuals exceeds the significant bounds. This is negligible since the probability of a spike being significant by chance is about one in 31. This simply indicates no significant autocorrelation, since it is expected that at most one out of 31 sample autocorrelations

exceed the 95% significance bounds. Collectively, these tests suggest that the model fits very well.

Figure 7 shows the observed and predicted daily natural gas consumed based on the selected ARIMA (3, 1, 3) model. As observed, the model produced an RMSE of 23.27 and a MAPE of 7.29% (refer to Table 3). As seen in Figure 7, the selected ARIMA (3, 1, 3) model relatively fitted the time series well despite the nonlinear nature of the data. The model was able to capture the trend component of the data to improve the prediction.

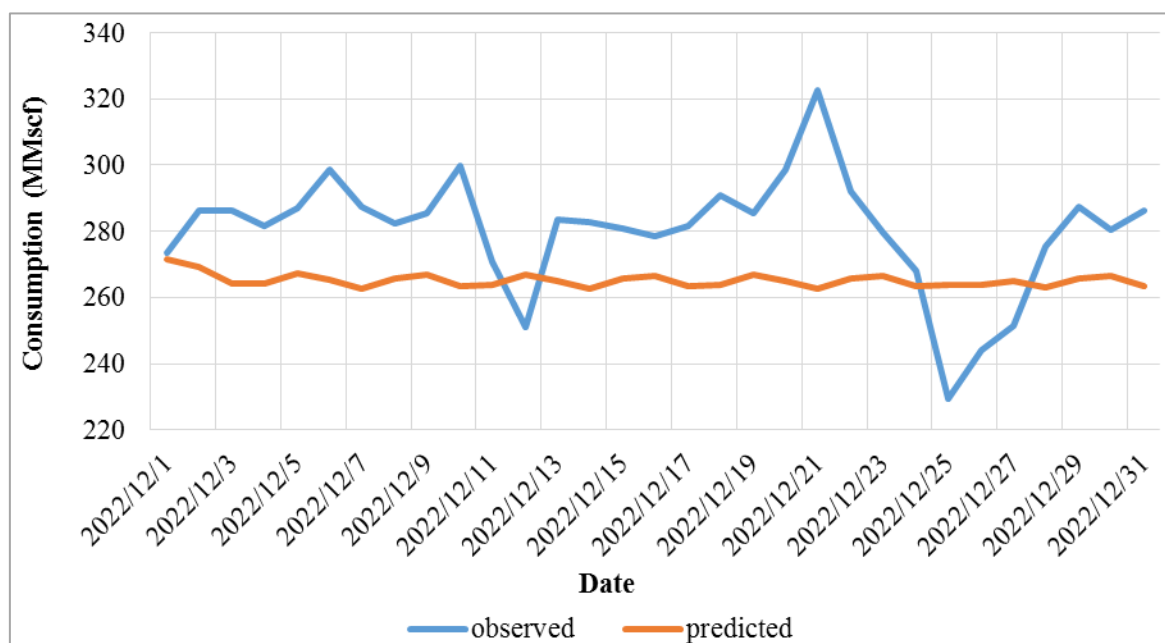


Figure 7. Plot of the Observed and Predicted Daily Natural Gas Consumption by ARIMA (3, 1, 3).

### 3.4. SARIMA Model Estimation

To find an appropriate SARIMA ( $p, 1, q$ )( $P, 0, Q, s$ ) 1,  $q$ ) model based on the ACF and PACF shown in Figure 4, the seasonal AR and MA which are represented by  $P$  and  $Q$  are evaluated. A careful inspection of the ACF plot in Figure 4 shows high correlations at lag 4, 14, 21 and so on. Hence, the period of seasonality  $s$  is represented as 7, where 7 represents weekly seasonality. Several models for different seasonal AR( $P$ ) and MA( $Q$ ) orders (for  $P, Q = 1, 2, 3, 4$ ) were generated, and their respective AIC and BIC values were used to select three best-competing models.

The four competing SARIMA models were used to predict the daily consumption of natural gas for December 2022 (test data) and their prediction accuracy/error was computed in Table 5. From Table 5, the four competing models were evaluated using RMSE and MAPE. It was observed that SARIMA (3,1,3)(1,1,1,7) had the least error among the four competing models and thus was adjudged the best SARIMA model.

Table 5. Competing SARIMA Models and their Respective RMSE and MAPE.

Model	RMSE	MAPE (%)
SARIMA (1, 1, 1) (1,0,1,7)	23.19	7.18
SARIMA (2, 1, 2) (1,0,1,7)	23.20	7.18
SARIMA (3, 1, 3) (1,1,1,7)	22.25	6.96

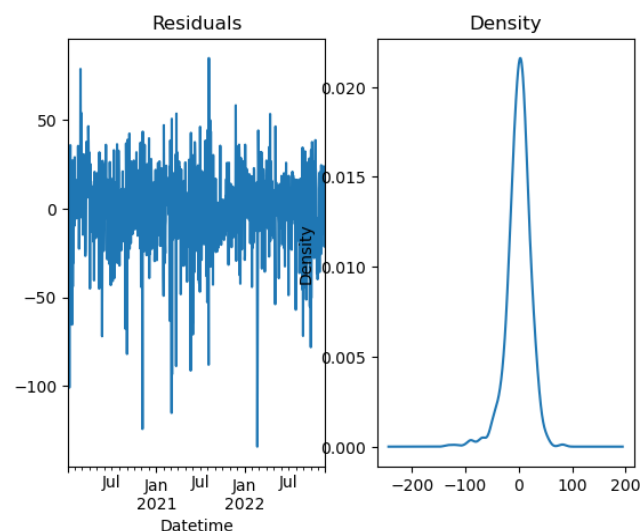
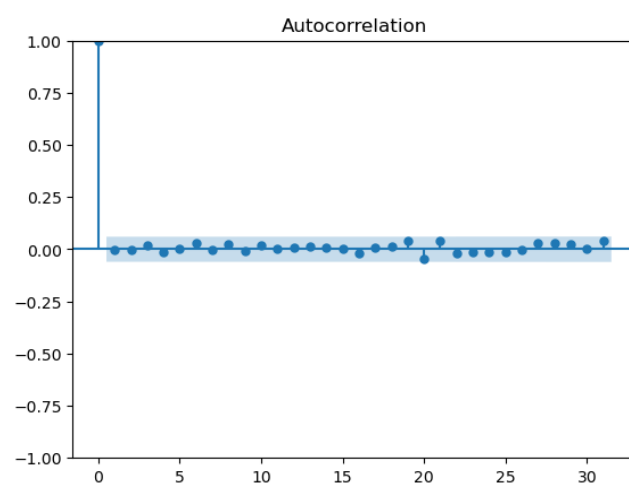
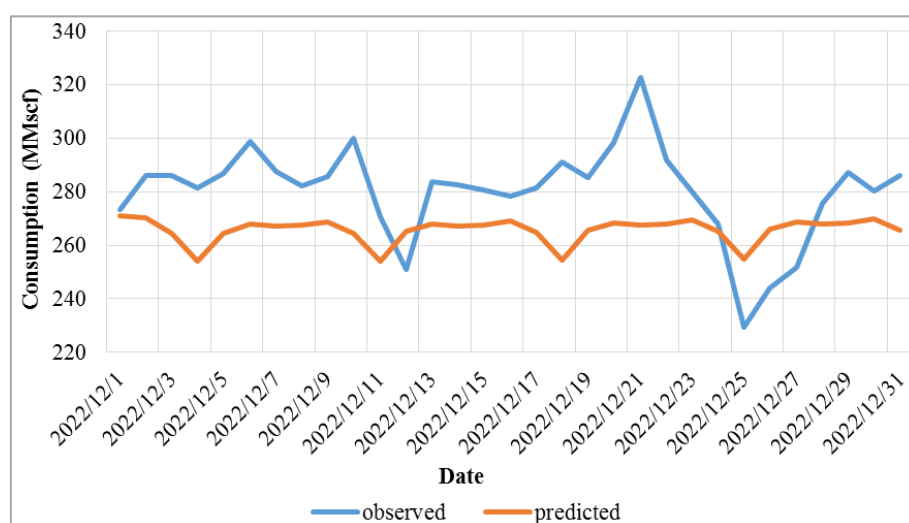
The parameter estimates of the selected SARIMA (3,1,3)(1,1,1,7) model are shown in Table 6. As observed, the only significant lag variable was the seasonal MA lag variable. This implies that there was a seasonal pattern in the dataset which occurred weekly.

**Table 6.** Parameter Estimate for SARIMA (3, 1, 3) (1, 1, 1, 7).

Parameter	Coefficient	Std. Error	z value	P >  z
AR.L1	-0.5308	18.798	-0.028	0.977
AR.L2	0.2668	7.213	0.037	0.970
AR.L3	0.0338	5.706	0.006	0.995
MA.L1	0.0742	18.799	0.004	0.997
MA.L2	-0.7050	1.572	-0.448	0.654
MA.L3	-0.0805	12.662	-0.006	0.995
AR.S.L7	0.0241	0.036	0.670	0.503
MA.S.L7	-0.9967	0.034	-29.731	<0.0001
sigma2	478.2706	17.335	27.590	<0.0001

Figures 8 and 9 show the diagnostic plot for the residuals of SARIMA (3, 1, 3) (1,1,1,7). As observed, the standardized residual and density plots (Figure 8) display normally distributed residuals as the points display zero trace of trend, no extreme outliers, and no significant change in variance across time. The ACF of the residuals plot (Figure 9) shows that only one lag out of the 31 lags of the series residuals exceeds the significant bounds. This is negligible since the probability of a spike being significant by chance is about one in 31. This simply indicates no significant autocorrelation, since it is expected that at most one out of 31 sample autocorrelations exceed the 95% significance bounds.

Figure 10 shows the plot of the observed as well as the predicted production values for December 2022 (test set). The model produced a RMSE of 22.25 and a MAPE of 6.96% (refer to Table 5). Relatively the selected SARIMA model fitted the time series even though the time series was nonlinear.

**Figure 8.** SARIMA (3,1,3) (1,1,1,7) Standardized Residual and Density Plot.**Figure 9.** ACF Plot of SARIMA (3,1,3) (1,1,1,7) Residuals.**Figure 10.** Plot of the Observed and Predicted Daily Natural Gas Consumption by SARIMA (3,1,3) (1,1,1,7).



### 3.5. Performance Comparison of ARIMA and SARMIA Models

The summary of the performance result of the developed ARIMA and SARIMA models is presented in Table 7. As seen, the model with the best performance was SARIMA (3,1,3) (1,1,1,7) even though the difference between it and ARIMA is close.

Both the ARIMA (3,1,3) and the SARIMA (3,1,3) (1,1,1,7) models were used to predict the daily consumption of natural gas for January 2023 (Figure 11). It was observed that the two models imitate the behaviour of the dataset by fluctuating. The forecast from the models suggests a steady consumption of natural

gas in Ghana but with some intermittent fluctuations.

Table 7. Summary of Performance Result.

Model	Performance Metrics	
	RMSE	MAPE
ARIMA (3,1,3)	23.27	7.29%
SARIMA (3,1,3) (1,1,1,7)	22.25	6.96%

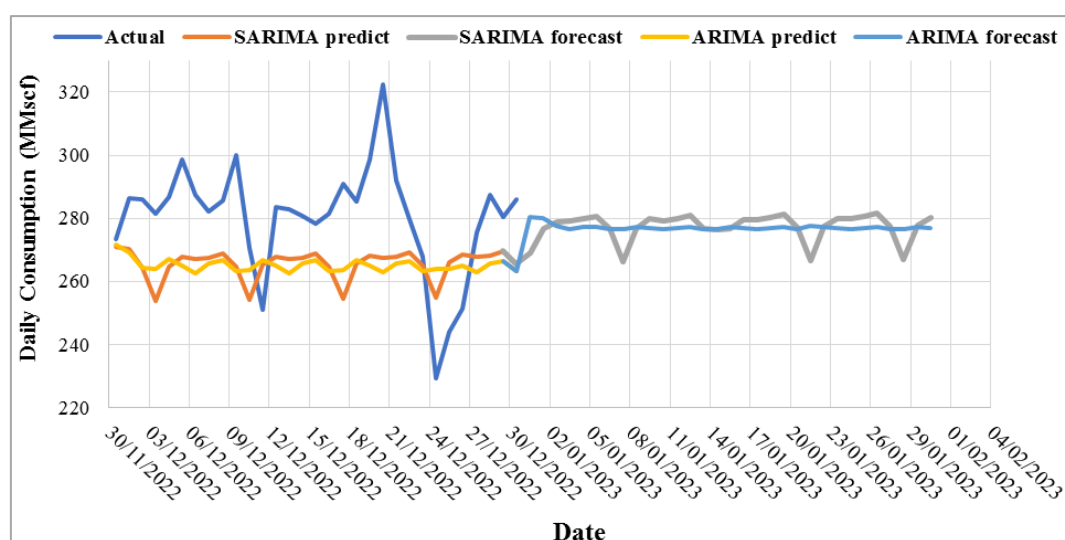


Figure 11. Plot of the ARIMA (3,1,3) and SARIMA (3,1,3) (1,1,1,7) Predicted and Forecasted Daily Consumption of Natural Gas.

## 4. Conclusions

The study utilises the concept of statistics-based ARIMA and SARIMA models for predicting daily natural gas consumption in Ghana as its accurate prediction helps in effective decision-making and policy formulation within the nation's energy landscape. From the preliminary analysis, it was seen that the daily consumption of natural gas in Ghana was non-stationary and exhibited seasonal features with a non-steady trend. Using performance indicators such as information loss (AIC and BIC) and error measures (RMSE and MAPE), ARIMA (3,1,3) and SARIMA (3,1,3) (1,1,1,7) models were noted to be suitable out of the several competing models, and thus, were selected for predicting the daily consumption of natural gas in Ghana. However, the SARIMA (3,1,3) (1,1,1,7) model comparatively showed better predictive accuracy than the ARIMA (3,1,3) model when tested on December 2022 consumption data, MAPE of 6.96% and 7.29% respectively. Hence, the SARIMA (3,1,3) (1,1,1,7) model

possesses the added advantage of being versatile in accounting for various seasonality and volatility patterns and, thus, could be implemented on a wide range of products in the extractive industry. A 30-day ahead forecast was then estimated which suggests a steady consumption of natural gas in Ghana but with some intermittent fluctuations. This study advances Ghana's energy landscape by addressing the need for the prediction of natural gas consumption patterns in Ghana, thus, allowing for effective decision-making and policy formulation. Despite the predictive performance of the ARIMA and SARIMA models, they are limited as they are well-suited for short-term predictions. Hence, future works will explore non-parametric intelligent techniques.

## Abbreviations

ADF: Augmented Dickey-Fuller  
 ARIMA: Autoregressive Integrated Moving Average  
 SARIMA: Seasonal Autoregressive Integrated Moving Average

RMSE: Root Mean Square Error  
 MAPE: Mean Absolute Percentage Error  
 ACF: Autocorrelation Function  
 PACF: Partial Autocorrelation Function  
 KPSS: Kwiatkowski-Phillips-Schmidt-Shin  
 AIC: Akaike Information Criteria  
 BIC: Bayesian Information Criterion

## Conflicts of Interest

The authors declare no competing interest.

## References

- [1] Boran FE (2015) Forecasting natural gas consumption in Turkey using grey prediction. *Energy Sources, Part B: Economics, Planning, and Policy* 10(2): 208-213. <https://doi.org/10.1080/15567249.2014.893040>
- [2] Broni-Bediako E, Dankwa OK (2013) Assessment of liquefied petroleum gas utilisation in Ghana – a study at Tarkwa. *International Journal of Scientific and Technology Research* 2(9): pp. 6-10.
- [3] Broni-Bediako E, Amarin R (2018) The Ghana liquefied petroleum gas promotion programme: opportunities, challenges and the way Forward. *Innov Ener Res* 7(2): 1-5.
- [4] Kumi EN (2017) The electricity situation in Ghana: challenges and opportunities. CGD Policy Paper.
- [5] Ayaburi J, Bazilian M (2020) Economic benefits of natural gas production: The case of Ghana's Sankofa Gas Project. *Energy for Growth Hub1-2*.
- [6] Anon (2023) Natural gas consumption. <https://www.ghanagas.com.gh/en/page/natural-gas-consumption-58>
- [7] Adom PK, Bekoe W, Akoena SKK (2012) Modelling aggregate domestic electricity demand in Ghana: An autoregressive distributed lag bounds co-integration approach. *Energy Policy* 42: 530-537. <https://doi.org/10.1016/j.enpol.2011.12.019>
- [8] Shin S, Cho M (2022) Green supply chain management implemented by suppliers as drivers for SMEs environmental growth with a focus on the restaurant industry. *Sustainability* 14(6): pp. 3515. <https://doi.org/10.3390/su14063515>
- [9] Gao F, Shao X (2021) Forecasting annual natural gas consumption via the application of a novel hybrid model. *Environ Sci Pollut Res* 21411–21424. <https://doi.org/10.1007/s11356-020-12275-w>
- [10] Owusu PA, Asumadu-Sarkodie S (2016) A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Engineering* 3(1). <https://doi.org/10.1080/23311916.2016.1167990>
- [11] Kemausuor F, Obeng GY, Brew-Hammond A, et al. (2011) A review of trends, policies and plans for increasing energy access in Ghana. *Renewable and Sustainable Energy Reviews* 15(9): 5143-5154. <https://doi.org/10.1016/j.rser.2011.07.041>
- [12] Scarpa F, Bianco V (2017) Assessing the quality of natural gas consumption forecasting: an application to the Italian residential sector. *Energies* 10: 13. <https://doi.org/10.3390/en10111879>
- [13] Cardoso CV, Cruz GL (2016) Forecasting natural gas consumption using ARIMA models and artificial neural networks. *IEEE Latin America Transactions* 14: 2233-2238. <https://doi.org/10.1109/TLA.2016.7530418>
- [14] Chkili W, Hammoudeh S, Nguyen DK (2014) Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics* 41: 1-18. <https://doi.org/10.1016/j.eneco.2013.10.011>
- [15] Feng X, Zhang J, Zou S, et al. (2014) Study on natural gas demand prediction model in China. *Advanced Materials Research* 962-965.
- [16] Vondracek J, Pelikan E, Konar O, et al. (2008) A statistical model for the estimation of natural gas consumption. *Applied Energy* 85(5): 362-370.
- [17] Khan MA (2015) Modeling and forecasting the demand for natural gas. *Renewable and Sustainable Energy Reviews* 49: 1145-1159.
- [18] Bianco V, Scarpa F, Tagliafico L (2014) Scenario analysis of nonresidential natural gas consumption in Italy. *Applied Energy* 113: 392-403. <https://doi.org/10.1016/j.apenergy.2013.07.054>
- [19] Kani AH, Abasspour M, Abedi Z (2014) Estimation of demand function for natural gas in Iran: Evidences based on smooth transition regression models. *Economic Modeling* 36: 341-347.
- [20] Beyca OF, Ervural BC, Tatoglu E, et al. (2019) Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics* 80: 937-949. <https://doi.org/10.1016/j.eneco.2019.03.006>
- [21] Behrouznia A, Saberi M, Azadeh A, et al. (2010) An adaptive network based fuzzy inference system-fuzzy data envelopment analysis for gas consumption forecasting and Analysis: The case of South America. *International Conference on Intelligent and Advanced System* 5716160. <https://doi.org/10.1109/ICIAS.2010.5716160>
- [22] Viet NH, Mandziuk J (2005) Neural and fuzzy neural networks in prediction of natural gas consumption. *Neural Parallel Scientific Computation* 13(3-4): 265-286. <https://doi.org/10.1109/NNSP.2003.1318075>
- [23] Liu T (2017) Statistical inference of partially linear panel data regression models with fixed individual and time effects. *Communications in Statistics - Theory and Methods* 46(15): 7267-7288. <https://doi.org/10.1080/03610926.2015.1116577>
- [24] Godfrey LG, Tremayne AR (1988). Checks of model adequacy for univariate time series models and their application to econometric relationships. *Econometric Reviews* 7(1): 1-42. <https://doi.org/10.1080/07474938808800138>

- [25] Carbó JM, Gorjon S (2022) Application of machine learning models and interpretability techniques to identify the determinants of the price of bitcoin. *SSRN Electronic Journal* 4: 24-56. <https://doi.org/10.2139/ssrn.4087481>
- [26] Ediger VŞ, Akar S (2007) ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* 35(3): 1701-1708. <https://doi.org/10.1016/j.enpol.2006.05.009>
- [27] Valipour M (2015) Long-term Runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications* 22(3): 592-598. <https://doi.org/10.1002/met.1491>
- [28] Akpınar M, Yumusak N (2016) Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies* 9(9): 1-17. <https://doi.org/10.3390/en9090727>
- [29] Erdogdu E (2010) Natural gas demand in Turkey. *Applied Energy* 87(1): 211-219.
- [30] Akkurt M, Demirel O, Zain S (2010) Forecasting Turkey's natural gas consumption by using time series methods. *European Journal of Economic and Political Studies* 3(2): 1-21.
- [31] Munandar D (2019) Multilayer Perceptron (MLP) and Autoregressive Integrated Moving Average (ARIMA) models in multivariate input time series data: solar irradiance forecasting. *International Journal on Advanced Science, Engineering and Information Technology* 9(1): 220.
- [32] Fossati S (2013) Unit root testing with stationary covariates and a structural break in the trend function. *Journal of Time Series Analysis* 34(3): 368-384. <https://doi.org/10.1111/jtsa.12020>
- [33] Dickey D, Fuller W (1979) Divolatility of the estimators for autoregressive time series with a unit root". *Journal of American Statistical Association* 74(366): 427-431.
- [34] Kwiatkowski D, Phillips PCB, Schmidt P., et al. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54(3): 159-178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- [35] Kofi A, Albert B, Joseph A (2019) Modelling the volatility of the price of bitcoin. *American Journal of Mathematics and Statistics* 9(4): 151-159.
- [36] Hyndman RJ, Athanasopoulos G (2018) *Forecasting: Principles and practice*", OTexts Publications, London.
- [37] Ismail MA, El-Metaal EMA (2020) Forecasting residential natural gas consumption in Egypt. *Journal of Humanities and Applied Social Sciences* 2(4): 297-308. <https://doi.org/10.1108/JHASS-03-2020-0046>
- [38] Zhang X, Pang Y, Cui M, et al. (2015) Forecasting mortality of road traffic injuries in China using seasonal autoregressive integrated moving average model. *Ann. Epidemiol.* 25: 101-106. <https://doi.org/10.1016/j.annepidem.2014.10.015>
- [39] Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. *Int. J. Forecast* 22: 679-688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- [40] Armstrong JS, Collopy F (1992) Error measures for generalizing about forecasting methods: Empirical comparisons. *Int. J. Forecast* 8: 69-80. [https://doi.org/10.1016/0169-2070\(92\)90008-W](https://doi.org/10.1016/0169-2070(92)90008-W)