Fueling the Future: A Comprehensive Analysis and Forecast of Fuel Consumption Trends in U.S. Electricity Generation

Md Monjur Hossain Bhuiyan * , Ahmed Nazmus Sakib , Syed Ishmam Alawee , Talayeh Razzaghi

Posted Date: 16 January 2024

doi: 10.20944/preprints202401.1244.v1

Keywords: Fuel Consumption; Forecasting; Time Series Analysis; Sustainable Energy Policies; RMSE.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 January 2024 doi:10.20944/preprints202401.1244.v1

Article

Fueling the Future: A Comprehensive Analysis and Forecast of Fuel Consumption Trends in U.S. Electricity Generation

Md Monjur Hossain Bhuiyan [1,*], Ahmed Nazmus Sakib [1], Syed Ishmam Alawee [2] and Talayeh Razzaghi [3]

[1] School of Aerospace and Mechanical Engineering, University of Oklahoma, USA
[2] School of Data Science and Analytics, University of Oklahoma, USA
[3] School of Industrial and Systems Engineering, University of Oklahoma, USA.
* Correspondence: mhbhuiyan@ou.edu

Abstract: The U.S. Energy Information Administration (EIA) provides crucial data on monthly and annual fuel consumption for electricity generation. This data covers significant fuels such as coal, petroleum liquids, petroleum coke, and natural gas. Fuel consumption patterns are highly dynamic, influenced by diverse factors. Understanding these fluctuations is essential for effective energy planning and decision-making. This study outlines a comprehensive analysis of fuel consumption trends in electricity generation. Utilizing advanced statistical methods, including time series analysis and autocorrelation, our objective is to uncover intricate patterns and dependencies within the data. This paper aims to forecast fuel consumption trend for electricity generation using data from 2015 to 2022. Several time-series forecasting models, including all four benchmark methods (Mean, Naïve, Drift, and seasonal Naïve), Seasonal and Trend Decomposition using Loess (STL), Exponential Smoothing (ETS), and Autoregressive Integrated Moving Average (ARIMA) methods, have been applied. The best-performing models are determined based on Root Mean Squared Error (RMSE) values. For Natural Gas (NG) consumption, the ETS model achieves the lowest RMSE of 20,687.46. STL demonstrates the best performance for coal consumption with an RMSE of 5,936.203. The seasonal Naïve (SNaïve) model outperforms others for petroleum coke forecasting, yielding an RMSE of 99.49. Surprisingly, the Mean method has the lowest RMSE of 287.34 for petroleum liquids, but the ARIMA model is reliable for its ability to capture complex patterns. Residual plots are analyzed to assess the models' performance against statistical parameters. Accurate fuel consumption forecasting is very important for effective energy planning and policymaking. The findings from the study help policymakers strategically allocate resources, plan infrastructure development and support economic growth.

Keywords: Fuel Consumption; Forecasting; Time Series Analysis; Sustainable Energy Policies; RMSE

1. Introduction

The United States relies on a diverse array of energy sources, broadly categorized into primary sources, including fossil fuels, nuclear, and renewables, and secondary sources, represented by electricity generated from primary sources [1]. Measurement units vary across energy types, with liquid fuels quantified in barrels or gallons, natural gas in cubic feet, coal in short tons, and electricity in kilowatts and kilowatt-hours [2]. British thermal units (Btu) serve as a standard for energy comparison, revealing that total U.S. primary energy consumption reached 100.41 quadrillion Btu in 2022 [3,4]. In the realm of electricity generation, the United States employs a spectrum of sources and technologies that have evolved over time. The three principal categories encompass fossil fuels (coal, natural gas, and petroleum), nuclear energy, and renewables. Predominantly, steam turbines, drawn from fossil fuels, nuclear, biomass, geothermal, and solar thermal energy, stand as the dominant force in electricity generation.

Other technologies contributing to this landscape include gas turbines, hydro turbines, wind turbines, and solar photovoltaics [5]. Notably, within the domain of fossil fuels, natural gas emerges as the predominant player, contributing approximately 40% to U.S. electricity generation in 2022 [6].

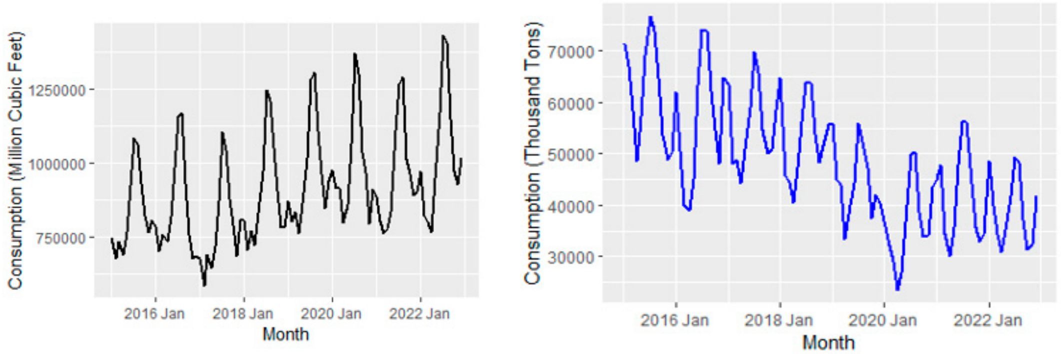
Coal, occupying the third-largest share at around 18%, predominantly fuels steam turbines, with certain facilities converting coal to gas for utilization in gas turbines. Petroleum's contribution is nominal, constituting less than 1%, where residual fuel oil and petroleum coke find application in steam turbines, and distillate fuel oil powers diesel-engine generators [6]. Nuclear energy commands a substantial portion, accounting for nearly one-fifth of U.S. electricity, generated through steam turbines and nuclear fission. Renewable energy sources, constituting approximately 22% of total U.S. electricity generation in 2022, have witnessed substantial growth since 1990, when their contribution to utility-scale electricity generation was approximately 12%. This discernible shift underscores the escalating significance of renewables in shaping the U.S. energy landscape [3,6]. In-depth analysis and insights from reputable sources further verify these trends and dynamics.

Navigating the undulating landscape of fuel consumption in electricity generation, as strictly narrated by the U.S. Energy Information Administration, presents a remarkable challenge. The inconsistent nature of these consumption patterns highlights a critical need for precise predictive models, as their absence impedes informed decision-making, leaving us grappling with the dynamic shifts in energy demands. This urgency compels a call to action for establishing research aimed at the development of advanced forecasting techniques, pivotal for strategic planning in the sphere of future electricity generation. Accurate predictions of fuel consumption not only facilitate efficient resource allocation and guide infrastructure development but also serve as the bedrock for informed energy policies [7]. Beyond the realms of numbers and projections, these forecasts wield a profound impact on investment decisions, environmental planning, grid management, cost optimization, and market strategies. They emerge as indispensable tools for emergency preparedness and play a pivotal role in orchestrating a seamless transition to renewable energy, ultimately forging a path toward a stable, sustainable, and resilient energy sector. As we investigate into the statistics of this fluctuation, the intricate jump of fuel consumption in electricity generation unveils itself, highlighting the pressing need for foresight in steering the future of our energy landscape [7].

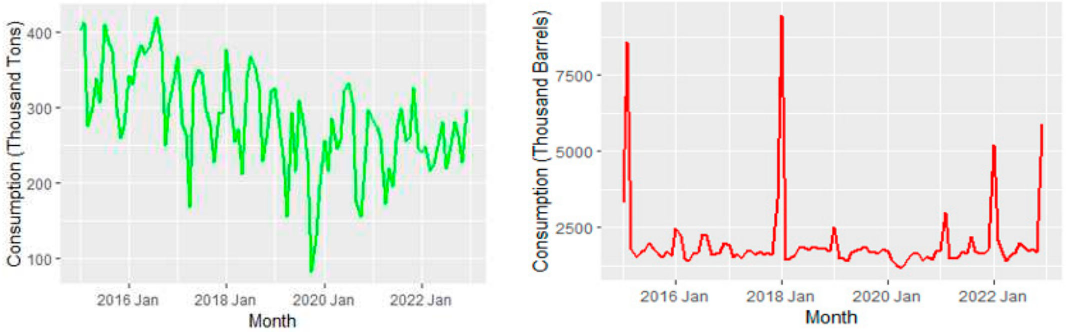
This study mainly focuses to construct a robust forecasting model for fuel consumption in U.S. electricity generation, utilizing comprehensive data spanning the years 2015 to 2022. The method involves systematic analysis of seasonal and trend patterns, rigorous autocorrelation analysis, and refinement by eliminating biases from historical trends. Leveraging actual consumption data, the approach includes model validation and fine-tuning through a detailed comparison of forecasted and real consumption figures. Employing data from the U.S. Energy Information Administration, this eight-year analysis aims to explain the primary trend of fuel consumption for electricity generation across the entire United States, mitigating potential uncertainties associated with regional variations. The forecasting process integrates diverse benchmark methods, including STL, ETS, and ARIMA, facilitating a robust comparison to identify the most accurate model. Evaluation metrics, such as Mean Error (ME), Mean Percentage Error (MPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), guide the selection of the optimal model. Once identified, this model will be deployed to predict future fuel consumption trends, offering valuable insights for strategic energy planning. The combination of a comprehensive approach, diverse forecasting models, specific fuel analysis, and a focus on policy impact distinguishes this work from previous studies in the field of fuel consumption forecasting for electricity generation. This research is crucial for informing stakeholders and decision-makers, enabling them to make well-informed decisions on resource allocation, energy planning, and sector-specific strategies, with the main goal of contributing to the formulation of sustainable energy policies for a stable and reliable national energy supply [8–10].

1.1. Overall Trend Analysis of Fuel Consumption

The trends observed in fuel consumption over time, as depicted in Figure 1 shows the dynamics of various fuel types utilized in U.S. electricity generation. Natural gas consumption exhibits a consistent upward trajectory from January 2016 to January 2022, with occasional fluctuations [6,11]. This notable increase, supported by statistical evidence [11] is attributed to several factors, including the growing preference for natural gas as a cleaner alternative, increased efficiency in gas-powered plants, and a shift towards renewables in the energy mix. In contrast, coal consumption displays a fluctuating trend with distinct periods of growth and decline, reaching a significant trough in January 2020.



(**a**) Natural gas consumption (**b**) Coal consumption



(**c**) Petroleum coke consumption (**d**) Petroleum coke consumption

Figure 1. Trend analysis of different fuel consumption for US electricity generation over time.

This decline aligns with a broader global trend of decreasing reliance on coal, driven by environmental concerns, regulatory shifts, and the pursuit of cleaner energy sources. Notably, the drop in January 2020 coincides with the onset of the COVID-19 pandemic, which induced a temporary reduction in industrial activities and energy demand, contributing to the observed drop in coal consumption. Similarly, petroleum coke usage exhibits variability, experiencing peaks at specific times. The visible drop in January 2020, concurrent with the pandemic's onset, can be attributed to a combination of reduced industrial activities, altered production patterns, and an overall decline in energy demand during the initial phases of the pandemic. Petroleum liquids demonstrate a diverse trend with notable spikes, particularly in early 2018. This increase can be linked to a confluence of factors, including economic conditions, geopolitical events, and shifts in energy policies. For instance, the spike in early 2018 may be associated with increased demand driven by economic growth and geopolitical factors affecting oil prices.

In summary, the observed trends emphasize the comprehensive nature of fuel consumption patterns, influenced by a complex interplay of economic, environmental, and technological factors. The distinct fluctuations and notable changes, especially during the COVID-19 pandemic, highlight the importance of considering external influences when analyzing fuel consumption dynamics for strategic energy planning and policy formulation.

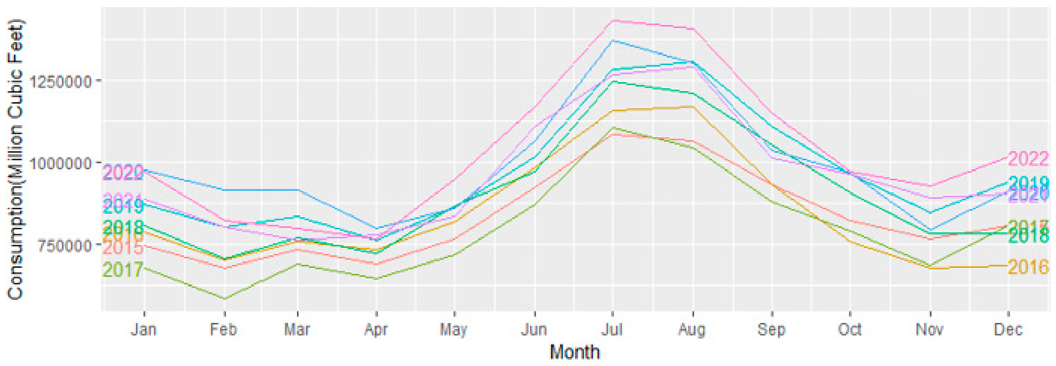
2. Methodology

2.1. Seasonality Analysis of Fuel Consumption

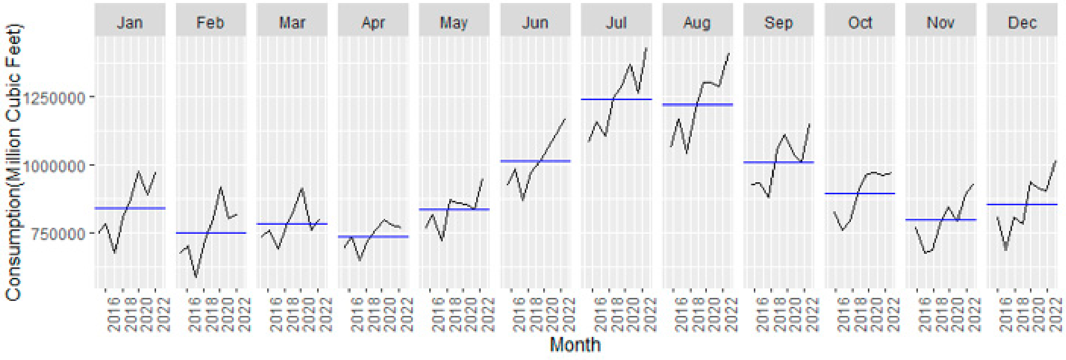
The seasonality analysis of fuel consumption is vital for understanding recurring patterns and cycles in fuel consumption over time. It enables accurate forecasting and prediction, aiding in operational planning and resource management. Seasonality analysis of different fuel consumption has been explained as follows.

2.1.1. Seasonality Analysis of NG Consumption

In Figure 2a, the upward trajectory of NG consumption from June to August each year reflects a consistent seasonal pattern, particularly during the summer months. This observed peak aligns with Figure 2b, illustrating the typical seasonal demand for electricity, notably driven by increased air conditioning needs in warmer weather. The surge in NG consumption during these summer periods can be attributed to its crucial role in meeting heightened electricity demand, with natural gas power plants playing a pivotal role. This seasonality underscores the importance of natural gas in addressing the specific requirements of peak periods, providing valuable insights into its usage dynamics. The ascending trend in NG consumption from 2016 to 2022 further emphasizes its increasing significance as an energy source for electricity generation, supported by advancements in natural gas technologies, its environmental advantages, and evolving energy policies.



(**a**) Monthly trend of NG consumption

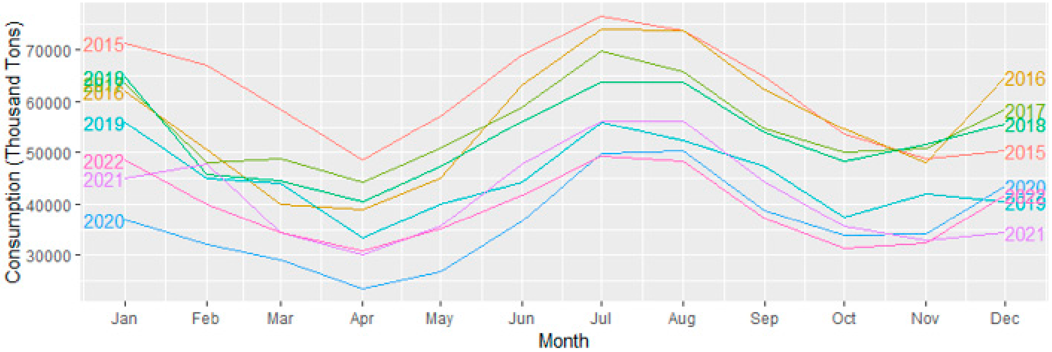


(**b**) Seasonality analysis of NG consumption

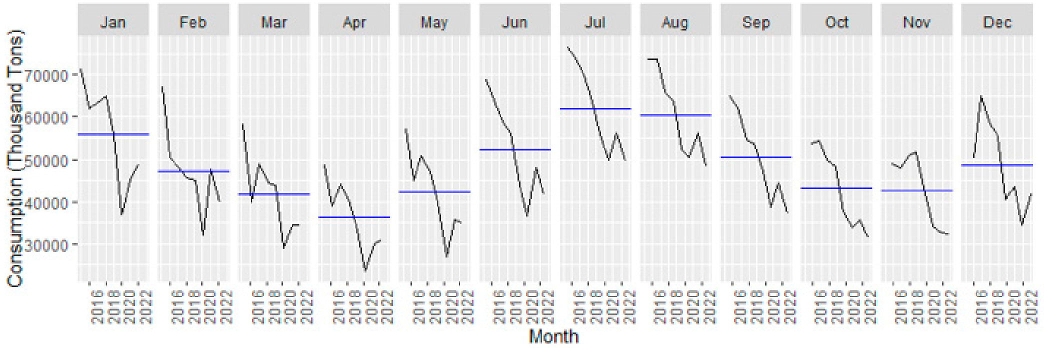
Figure 2. Monthly consumption and seasonality analysis of NG from the year 2015 to 2022.

2.1.2. Seasonality Analysis of Coal Consumption

The trend observed in coal consumption from 2015 to 2022, as shown in Figure 3, reveals a distinct seasonality pattern. The monthly trend, depicted in Figure 3a, consistently showcases a peak in consumption during the summer months of July and August, followed by a gradual decline through September and October. Subsequently, consumption reaches a relatively low in November and experiences an upturn again in December. This cyclicality suggests a pronounced seasonal influence on coal consumption, aligning with heightened energy demand during the summer for cooling purposes and a subsequent decrease as winter approaches.



(**a**) Monthly trend of coal consumption



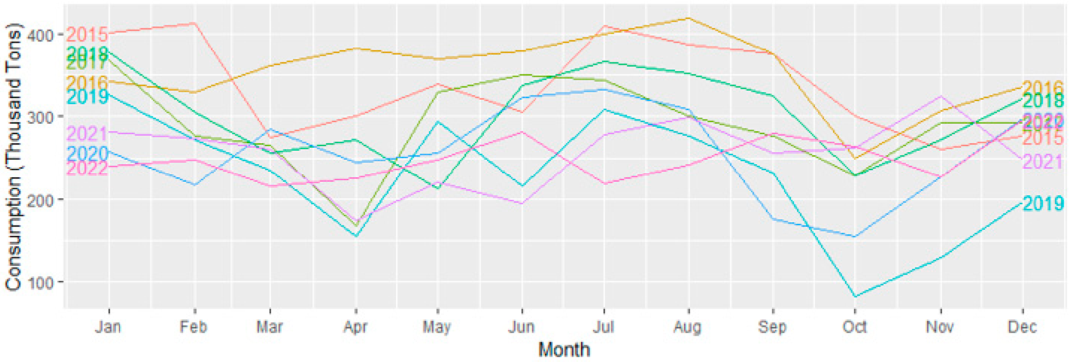
(**b**) Seasonality analysis of coal consumption

Figure 3. Monthly consumption and seasonality analysis of coal from the year 2015 to 2022.

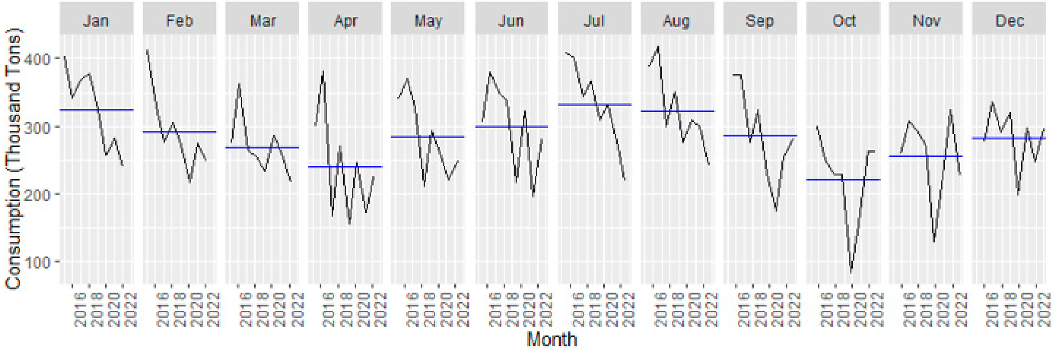
The statistics underlying this pattern demonstrate a consistent monthly fluctuation, with peak-to-trough differentials providing quantitative insights into the impact of seasonal variations on coal consumption [11]. The observed cyclicality aligns with established industry knowledge on the impact of weather-related demand fluctuations on energy consumption patterns [12,13]. This seasonality analysis, presented in Figure 3a and 3b, contributes valuable insights for energy planners and policymakers, enabling informed decisions to address the cyclical nature of coal consumption in the context of changing climate and energy demand dynamics.

2.1.3. Seasonality Analysis of Petroleum Coke Consumption

The monthly trend of petroleum coke consumption, depicted in Figure 4a from 2015 to 2022, exhibits noteworthy fluctuations, revealing distinctive patterns across different months. The seasonality analysis in Figure 4b further explains these trends, highlighting recurring patterns within the specified timeframe. Notably, petroleum coke consistently reaches its lowest consumption levels in October, with a recurring trend of reduced usage also observed in November. Conversely, the summer months of July and August consistently exhibit the highest consumption levels, indicative of a seasonal peak in fuel usage. This observed seasonality aligns with industry practices, where heightened energy demands during the summer, possibly attributed to increased industrial activities and higher temperatures, drive an uptick in petroleum coke consumption. The recurrent nature of these patterns emphasizes the importance of considering seasonal variations for accurate forecasting and strategic energy planning. These observations are grounded in data from reputable sources, providing a robust foundation for understanding the intricate seasonality dynamics in petroleum coke consumption.



(**a**) Monthly trend of petroleum coke consumption

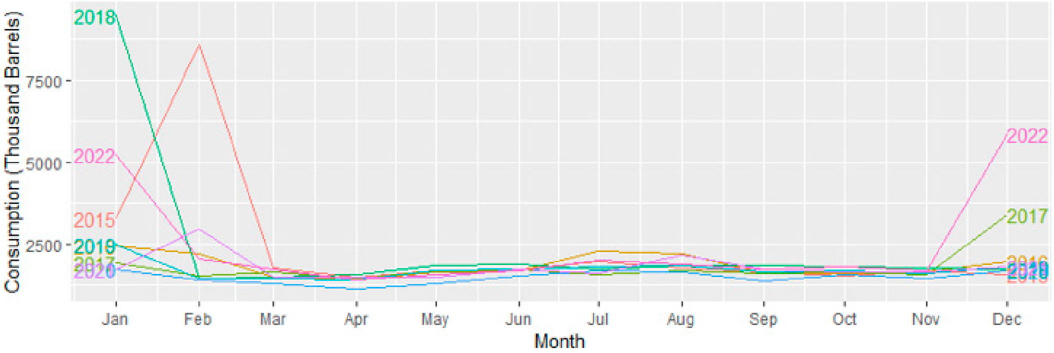


(**b**) Seasonality analysis of petroleum coke consumption

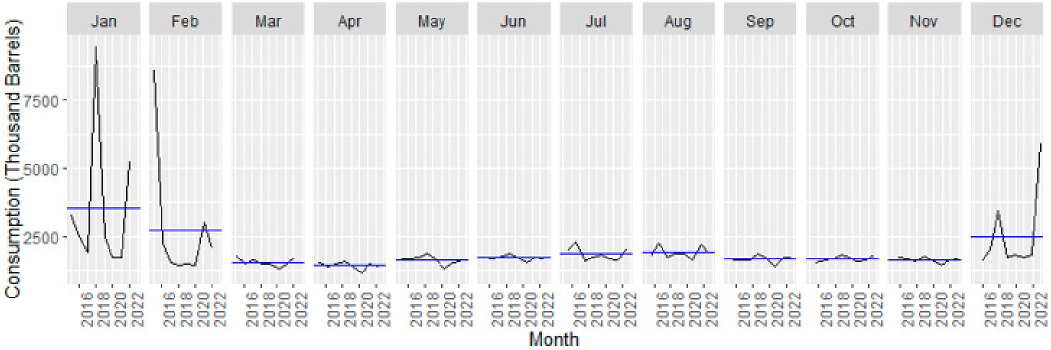
Figure 4. Monthly consumption and seasonality analysis of petroleum coke from the year 2015 to 2022.

2.1.4. Seasonality Analysis of Petroleum Liquids Consumption

The graphical representations in Figure 5a and Figure 5b offer a comprehensive insight into the monthly trends and seasonality analysis of petroleum liquids consumption in the United States from 2015 to 2022. In Figure 5a, the monthly trend reveals notable fluctuations, bringing attention to abnormal spikes in February 2015, January 2018, January 2022, and December 2022. These anomalies may be attributed to various factors, such as geopolitical events influencing global oil markets, economic shifts impacting demand, or specific regulatory changes affecting petroleum consumption. Figure 5b, the seasonality analysis, provides a deeper understanding of recurring patterns. The observed regular fluctuations suggest a potential influence of seasonal changes, economic cycles, or global events on petroleum liquids consumption. For instance, heightened demand during winter months or economic upturns may contribute to periodic spikes. These insights underscore the need for a nuanced understanding of the multifaceted factors influencing petroleum liquids consumption, integrating considerations beyond mere temporal patterns. Reference to observed data from the U.S. Energy Information Administration – EIA provides additional context for these observations [6,12,13].



(**a**) Monthly trend of petroleum liquids consumption



(**b**) Seasonality analysis of petroleum liquids consumption

Figure 5. Monthly consumption and seasonality analysis of petroleum liquids from the year 2015 to 2022.

2.2. ACF and PACF Analysis of Different Fuel Consumption

The analysis of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are a pivotal step in formulating an accurate forecasting model for fuel consumption in electricity generation spanning the years 2015 to 2022. These plots play a crucial role in detecting seasonality and identifying autoregressive (AR) and moving average (MA) components, vital for comprehending how fuel consumption changes over time. The ACF plot provides a comprehensive overview of autocorrelation at various lags, offering insights into longer-term patterns such as seasonality in fuel usage during specific calendar periods. Simultaneously, the PACF plot focuses on the direct relationship between the current observation and observations at individual lags, aiding in model selection by revealing correlations within the fuel consumption time series. This approach allows for a nuanced understanding of fuel consumption patterns, capturing both broader trends and specific lagged relationships. This relationship can be expressed mathematically as shown in Equation (1) [14,15].

$$PACF_k = \text{corr}\,(L_t, L_{t-k} \mid L_{t-1}, \ldots\ldots\ldots L_{t-k-1}) \quad (1)$$

Utilizing the ACF enables the identification of overarching temporal patterns, while the PACF emphasizes the immediate connections within the fuel consumption data. By employing Equation (1), the PACF quantifies the correlation at lag Lt, extending up to a defined number of lags, contributing to a more robust analysis [14,15]. The significance of ACF and PACF analyses lies in their ability to uncover temporal dependencies, aiding in the formulation of a forecasting model that accounts for seasonality and lagged relationships, ultimately enhancing the accuracy of predictions. This methodological approach is fundamental for gaining a comprehensive understanding of the intricate fuel consumption patterns crucial for effective energy planning and policy formulation.

doi:10.20944/preprints202401.1244.v1

2.2.1. Autocorrelation Analysis of NG Consumption

In the specific analysis focused on Natural Gas consumption using a lag of 36 in the ACF plot, the majority of lines are observed to cross the blue dashed line on the positive side suggest a significant positive autocorrelation at lag 36 (Figure 6). The positive crossings indicate a correlation between values at a given month and those from 36 months prior, revealing a prolonged cyclicality in consumption trends. As the line (blue dashed) typically represents the 95% confidence interval for the autocorrelation values, points outside this interval may indicate significant autocorrelation. In addition, in the PACF plot, lines crossing the blue dashed line up to lag 15 indicate statistically significant partial autocorrelations and partial autocorrelations beyond this point are not statistically significant.

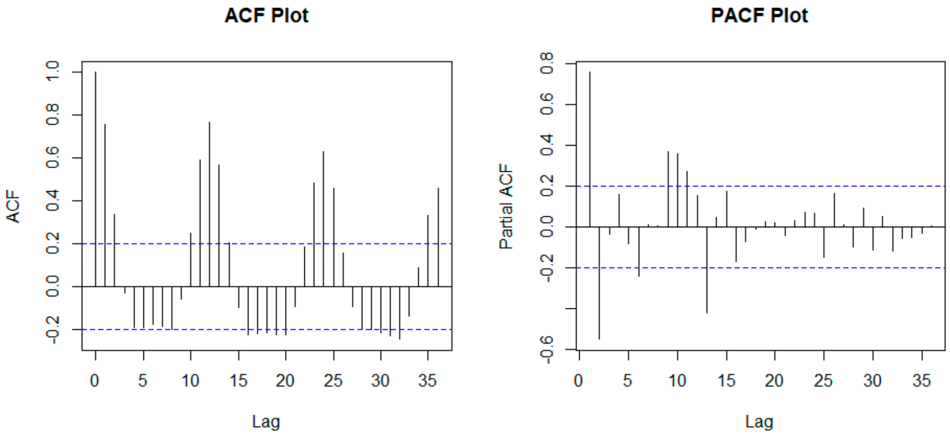


Figure 6. Autocorrelation analysis of NG consumption using ACF and PACF plot.

2.2.2. Autocorrelation Analysis of Coal Consumption

The ACF plot as shown in Figure 7 for Coal consumption with a lag of 36 provides insights into the autocorrelation structure, indicating the presence of seasonality within the first 27 months, followed by a change in the pattern. The positive crossing on the ACF plot up to lag 27 indicates significant positive autocorrelation at these lags. This suggests the presence of a repeating pattern or seasonality within the first 27 months of the Coal consumption data. The crossing to the negative side at lag 27 could indicate a reversal in the autocorrelation pattern. Moreover, in the PACF plot, the lines crossing the 95% confidence interval after lag 13 up to lag 36 indicate significant partial autocorrelation at these lags. The decrease in the number of lines crossing the confidence interval as the lag increases may signify a decay in the partial autocorrelation, indicating that the direct influence of observations on each other diminishes as the lag increases.

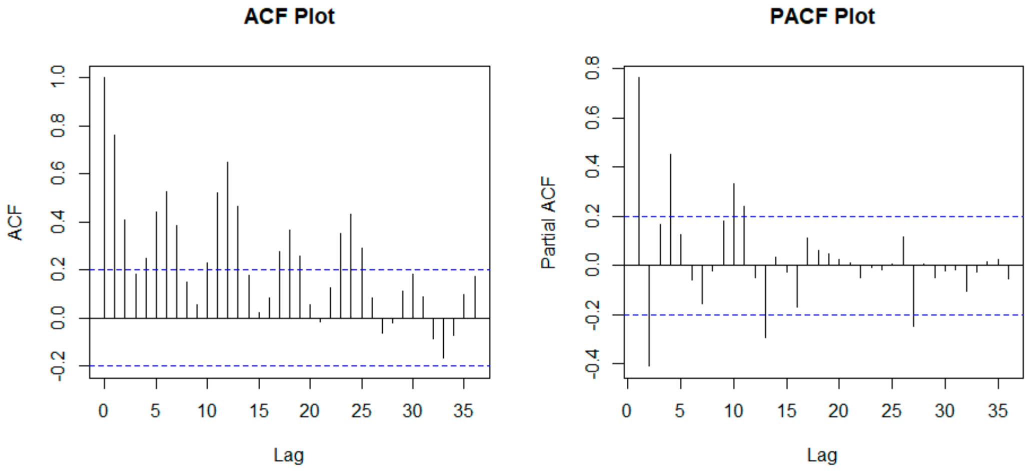


Figure 7. Autocorrelation analysis of coal consumption using ACF and PACF plot.

2.2.3. Autocorrelation Analysis of Petroleum Coke Consumption

The ACF plot for petroleum coke consumption, as depicted in Figure 8, indicates a positive correlation between observations at different lags, with the strength of correlation decreasing as the lag increases. The diminishing correlation suggests a changing influence of past observations on the current consumption pattern, possibly reflecting evolving trends or temporal dynamics. Conversely, the PACF plot shows no significant correlations, emphasizing the specific and direct influences captured by partial autocorrelations at each lag.

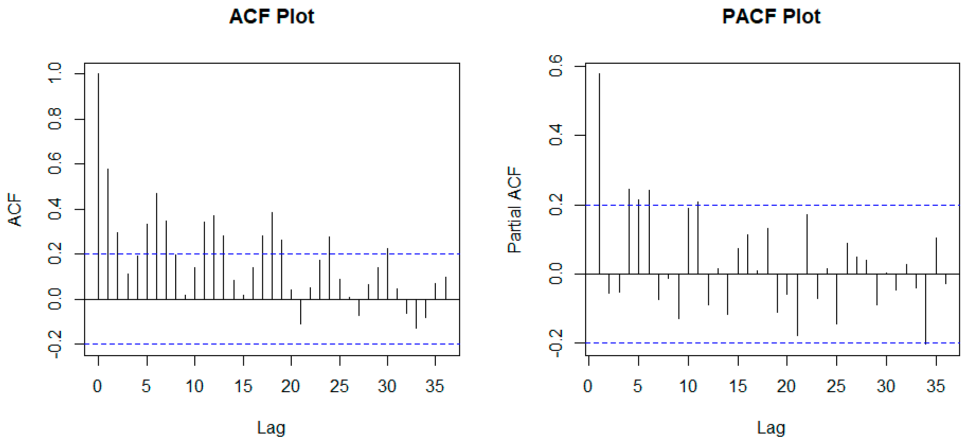


Figure 8. Autocorrelation analysis of petroleum coke consumption using ACF and PACF plot.

2.2.4. Autocorrelation Analysis of Petroleum Liquids Consumption

In the ACF plot of petroleum liquid consumption, the absence of lines crossing the dashed line, except at lag 0 and lag 35, indicates a lack of significant autocorrelation at most lags (Figure 9). At lag 0, the line crossing the blue dash line suggests a correlation with itself, which is expected. The PACF plot shows no significant correlations, emphasizing the specific and direct influences captured by partial autocorrelations at each lag.

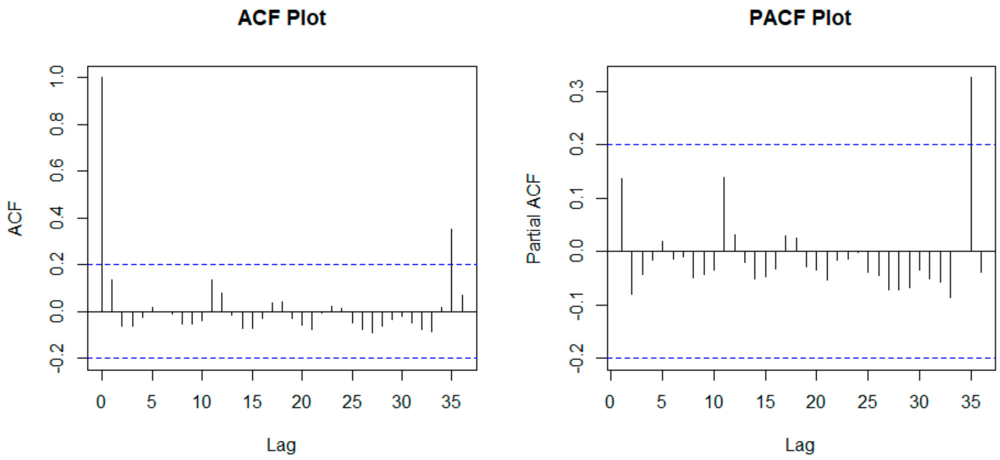


Figure 9. Autocorrelation analysis of NG consumption using ACF and PACF plot.

2.3. Forecasting Models

The development of forecasting models holds paramount importance in anticipating and adapting to the dynamic demand profiles inherent in fuel consumption, particularly for natural gas (NG) in electricity generation [16]. In this research endeavor, we deployed four distinct forecasting methodologies to project fuel consumption for electricity generation in the United States for the upcoming years 2023 and 2024. The models underwent a rigorous training phase spanning an 8-year period from January 2015 to December 2022, followed by meticulous validation using test data from January 2023 to August 2023. To comprehensively assess their performance, we conducted a detailed comparative analysis, evaluating the forecasting models based on their error metrics. The effectiveness of each model was further scrutinized through a meticulous comparison of forecasted data against actual consumption figures.

2.3.1. Forecasting Models for NG Consumption

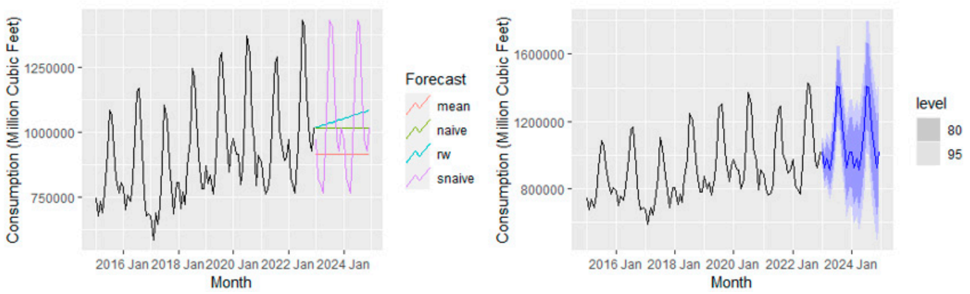
Figure 10 serves as a visual representation of the forecasted NG consumption for the years 2023 and 2024, employing diverse forecasting methodologies. The benchmark methods, STL decomposition, ETS method, and ARIMA method each contribute to a multifaceted analytical approach, offering a nuanced understanding of the varied strategies employed in forecasting. This rigorous methodology adheres to industry standards and contributes valuable insights to the realm of energy planning and policy formulation.

2.3.1.1. Forecasting using the Benchmark Methods

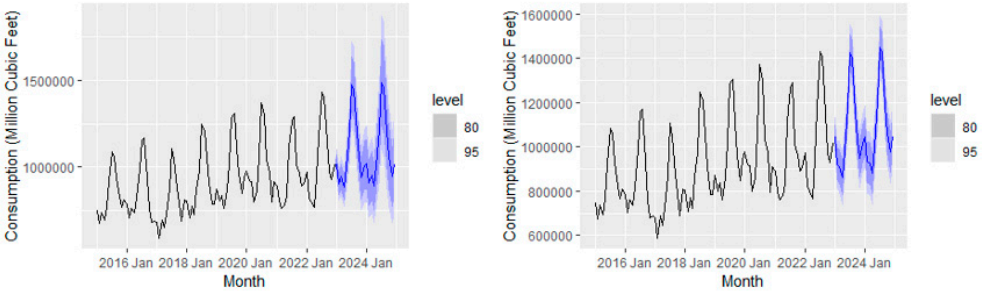
Benchmark approaches constitute foundational methodologies in time series forecasting, characterized by their simplicity and practicality. One such method is the Mean approach, which employs the average of historical observations to project future values. In contrast, the Naïve approach relies solely on the latest observation for forecasting. Extending the Naïve concept, the Drift method extrapolates future values by establishing a linear trend between the initial and final data points. Additionally, the Seasonal Naïve (SNaïve) method projects future values based on the last observed data point from the corresponding season in the preceding year. These benchmark methods serve as crucial reference points in fuel consumption forecasting, providing intuitive approaches for comparison with more sophisticated models [17]. Notably, among the illustrated benchmark methods (refer to Figure 10a), the Seasonal Naive (SNaïve) technique distinguishes itself by adeptly capturing the majority of fluctuations in fuel consumption.

2.3.1.2. Forecasting using the STL Decomposition Methods.

Decomposition techniques constitute a crucial element in the nuanced analysis of time series data, facilitating the discernment and isolation of pivotal components essential for unveiling trends, seasonal fluctuations, and cyclical patterns. One such sophisticated method is STL, denoting Seasonal and Trend decomposition using Loess. STL employs an intricate additive decomposition approach by iteratively applying the Loess smoother, which entails locally weighted polynomial regression at each data point.



(**a**) Benchmark methods (**b**) STL decomposition



(**c**) ETS method (**d**) ARIMA method

Figure 10. Forecasting NG consumption for the year 2023 and 2024 using different forecasting methods.

This method stands out for its robustness in handling outliers and its adaptability in addressing seasonal time series characterized by frequencies exceeding one. Unlike methodologies confined to specific temporal resolutions, such as monthly or quarterly intervals, STL demonstrates versatility by accommodating a broader spectrum of seasonal patterns, rendering it particularly well-suited for precise fuel consumption forecasting [18,19]. The forecasting outcome following STL decomposition is illustrated in Figure 10b, attesting to its efficacy in enhancing predictive accuracy.

2.3.1.3. Forecasting using the ETS Methods

The ETS (Error, Trend, Seasonality) forecasting methodology, renowned for its efficacy in time series prediction, is instrumental in modeling and anticipating fuel consumption patterns. This approach dissects time series data into three fundamental components: error, trend, and seasonality. The error term encapsulates stochastic fluctuations, the trend encapsulates long-term directional movements, and seasonality captures recurring patterns at fixed intervals. Diverse ETS model variations, such as ETS(AAA), ETS(AAN), and ETS(MAM), tailor to distinct time series characteristics by combining different error, trend, and seasonality components. Its adaptability to varied time series patterns and capacity to unveil insights into future trends make ETS models pervasive in forecasting applications.

Figure 10c illustrates the forecasted natural gas (NG) consumption data using the ETS method. The ETS decomposition plot in Figure 11, derived from the ETS (M, N, M) model—signifying multiplicative error, no trend, and multiplicative seasonality—was selected based on the minimal values of AIC, AICc, and BIC determined through the ETS () function in R. The estimated parameters for exponential smoothing include $\alpha = 0.583$ and $\Upsilon = 0.0001$, with a calculated $\sigma 2$ of 0.0024. Notably, α at 0.583 denotes a moderate emphasis on recent observations in forecasting, and the lower $\sigma 2$ value suggests heightened stability and predictability. These parameters assume a pivotal role in governing the rate of change for error, trend, and seasonality components, offering flexibility through α and Υ in adjusting level and trend, respectively.

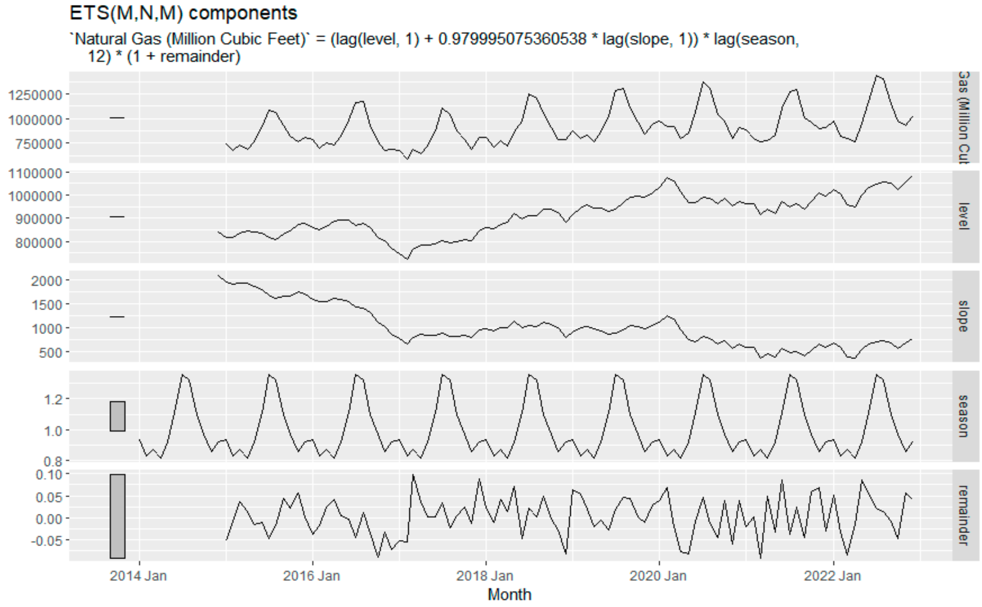


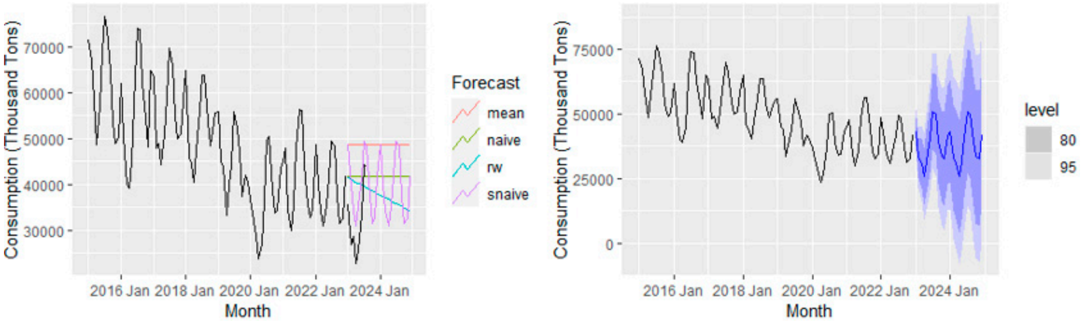
Figure 11. ETS (M, N, M) decomposition plot of NG consumption.

2.3.1.4. Forecasting using the ARIMA Methods

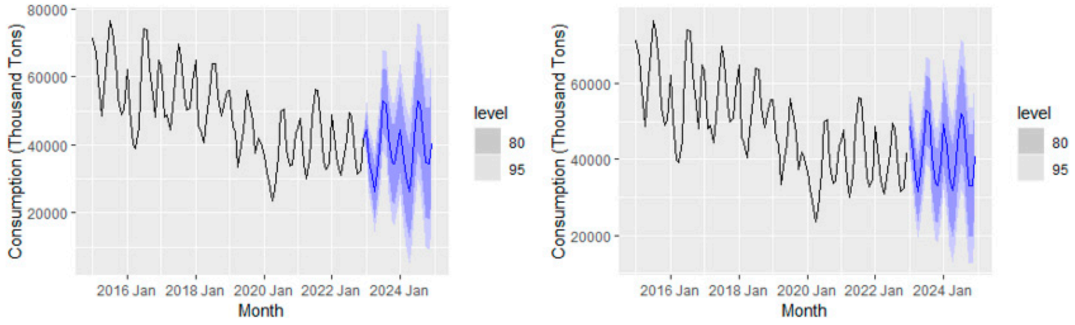
The ARIMA (AutoRegressive Integrated Moving Average) forecasting methodology stands as a widely embraced approach for time series prediction, notably effective in forecasting natural gas (NG) consumption for electricity generation. By seamlessly incorporating autoregressive (AR) and moving average (MA) components, along differencing to attain temporal stationarity, this model proves adept at capturing nuanced patterns inherent in time series data. The pivotal parameters of the ARIMA model encompass the autoregressive component (p), vital for discerning correlations with antecedent values, the integrated component (d) determining the order of differencing requisite for achieving stationarity, and the moving average component (q), effective in accounting for correlations with prior prediction errors. The optimal values for 'p,' 'd,' and 'q' are calculated through optimization techniques such as grid search. Post-training on historical data, the ARIMA model proficiently extrapolates future NG consumption values, adeptly accommodating both transient oscillations and enduring trends within the temporal dataset. This renders ARIMA an invaluable tool for precise and holistic forecasting within the realm of energy consumption. The forecasted natural gas (NG) consumption for electricity generation in 2023 and 2024, derived from the ARIMA forecasting method, is visually presented in Figure 10d, thereby showcasing the model's predictive prowess in delineating future trends.

2.3.2. Forecasting Models for Coal Consumption

In the prediction of coal consumption for electricity generation in the USA, we applied various forecasting methods, including the benchmark method, STL decomposition, ETS method, and ARIMA method, as outlined in Section 5. The coal consumption values from 2015 to 2022 were used as input data for predicting the values for the years 2023 and 2024. Figure 13 visually represents the forecasted coal consumption values obtained through different methods.



(**a**) Benchmark methods (**b**) STL decomposition

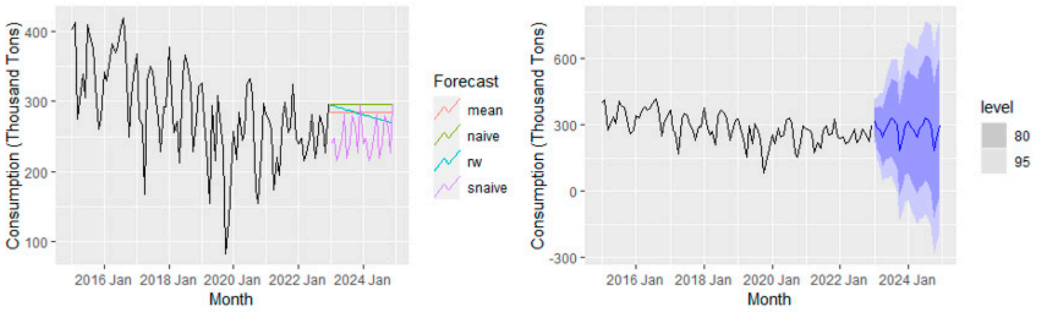


(**c**) ETS method (**d**) ARIMA method

Figure 13. Forecasting coal consumption for the year 2023 and 2024 using different forecasting methods.

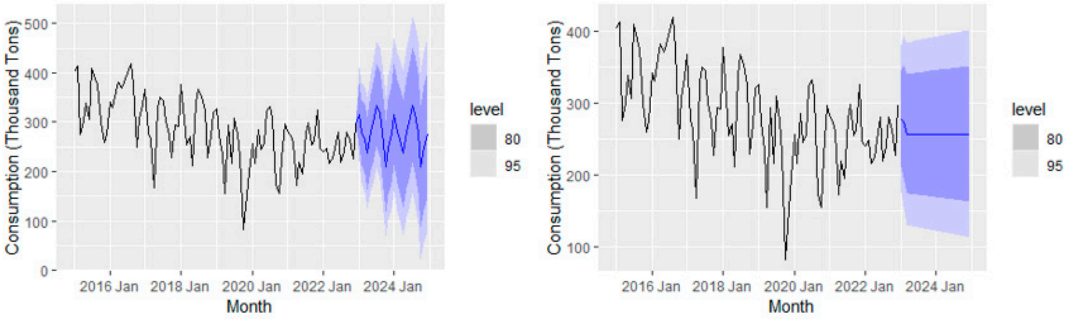
2.2.3. Forecasting Models for Petroleum Coke Consumption

The forecasting process for petroleum coke consumption in the years 2023 and 2024 followed the same methodology discussed in Section 5. Similar forecasting methods were applied to determine the most effective approach. Figure 15 visually presents the forecasted values for petroleum coke consumption obtained through the benchmark, STL decomposition, ETS, and ARIMA methods.



(**a**) Benchmark methods (**b**) STL decomposition

doi:10.20944/preprints202401.1244.v1

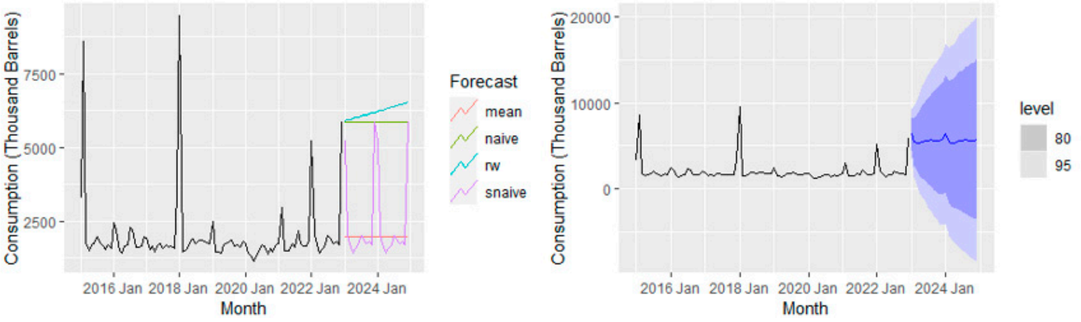


(**c**) ETS method (**d**) ARIMA method

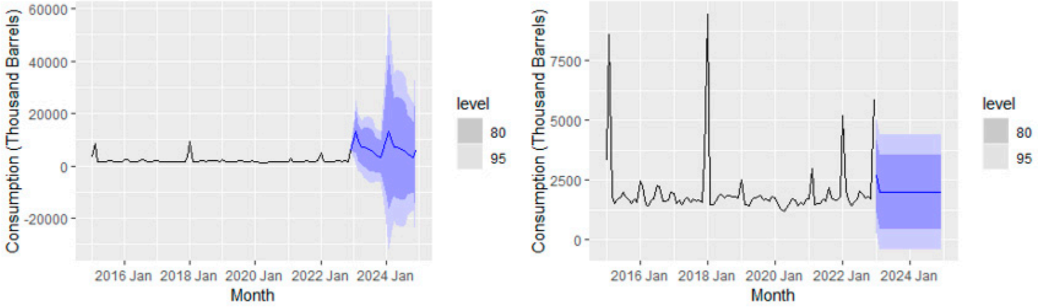
Figure 15. Forecasting petroleum coke consumption for the year 2023 and 2024.

2.3.4. Forecasting Models for Petroleum Liquids Consumption

Applying similar forecasting methods discussed in Section 5, Figure 17 visually shows the forecasted values for petroleum liquid consumption obtained through the benchmark, STL decomposition, ETS, and ARIMA methods.



(**a**) Benchmark methods (**b**) STL decomposition



(**c**) ETS method (**d**) ARIMA method

Figure 17. Forecasting petroleum liquid consumption for the year 2023 and 2024.

2.4. Model Comparison in Terms of Errors for the Energy Streams

2.4.1. Natural Gas

In assessing the accuracy of various forecasting methodologies for NG consumption, a comprehensive evaluation was conducted by partitioning the dataset into distinct training set (year 2015-2022) and test set (year 2023-2024). The performance of seven forecasting models was evaluated using six different accuracy measures, as demonstrated in Tables 1 and 2. The measures, designed to explain biases and precision across different models, encompass metrics such as the ME for bias estimation, MAE for precision measurement, and RMSE as an indicator of precision that penalizes larger errors. These three metrics operate on a scale-dependent basis. Conversely, the MPE and

doi:10.20944/preprints202401.1244.v1

MAPE are articulated in percentage terms, offering a more conducive platform for comparative analysis across diverse consumption levels. Remarkably, the ETS model emerged as the optimal performer, showing the lowest RMSE values in both the training (39237.5) and test (20687) datasets. The consistent excellence of the ETS model is evident in its ability to fit historical data and forecast future values. This makes it the chosen model for natural gas consumption forecasting in both training and test scenarios.

Table 1. Model comparison in terms of errors for the train dataset.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Train Data	STLF	2485.022	46311.86	37269.12	0.19733	4.054975	-0.31132
	ARIMA	-1003.74	42482.11	34075.3	-0.43476	3.772431	0.028523
	ETS	2776.487	**39237.5**	32580.19	0.124194	3.633987	0.029872
	MEAN	-3.88E-11	187831.5	150235	-3.91587	16.47926	0.758566
	NAÏVE	2849.021	129627.5	104644.3	-0.54117	11.12082	0.372243
	SNAÏVE	28184.8	85297.44	72281.51	2.45759	8.001223	0.716488
	RW-DRIFT	2.94E-11	129596.2	104914.3	-0.86377	11.16892	0.372243

Table 2. Model comparison in terms of Errors for the Test Data set.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Test Data	STLF	14741.56	50661.2	44064.48	0.566659	3.729218	0.648806
	ARIMA	25348.86	51424.27	46082.33	1.857718	4.030831	0.341779
	ETS	7600.542	**20687.46**	17204.31	0.513017	1.477106	0.117672
	MEAN	199798.9	308728.4	212774.1	14.57299	16.03116	0.650277
	NAÏVE	99858.38	255666.4	185054.6	5.251944	14.63274	0.650277
	SNAÏVE	75924.13	86594.7	75924.13	7.302419	7.302419	0.271527
	RW-DRIFT	87037.78	245731.8	181637.9	4.152805	14.53365	0.647994

Moreover, from the residual plot of the ETS model as shown in Figure 12, a normal distribution of residuals appears in the histogram plot, signifying a normal distribution of errors. The ACF plot, which identifies any remaining patterns in the residuals, illustrates that only a limited number of lines cross the blue dashed line, indicating statistical significance. This suggests that the model has successfully captured the majority of autocorrelation in the data. Moreover, random fluctuation in the residuals plot indicates that the model is capturing the underlying patterns in the data. In summary, the ETS model is a good fit for forecasting NG consumption as the residuals display randomness, normality, and lack of systematic patterns.

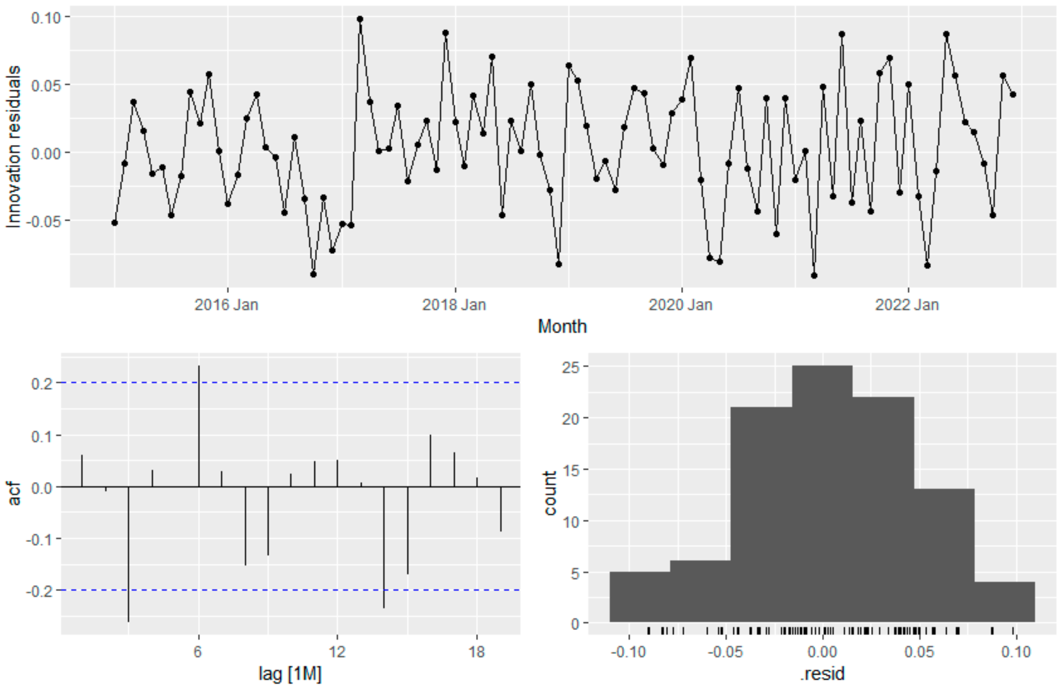


Figure 12. Residual analysis of ETS forecasting method.

2.4.2. Coal

Analyzing the performance on both training and test datasets reveals distinct characteristics of the models. While the ETS model boasts a lower RMSE of 3831.801 on the training data (Table 3), the STLF model demonstrates superior generalization on the test data with the lowest RMSE of 5936.203, compared to the ETS model's RMSE of 7288.344 (Table 4).

Table 3. Model comparison in terms of Errors for the Train Data set.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Train Data	STLF	-96.8973	4234.149	2884.178	-0.74241	6.481808	-0.32284
	ARIMA	-399.02	4385.426	3269.713	-1.15739	7.247969	0.012544
	ETS	-334.795	3831.801	2944.868	-1.1289	6.287089	0.036044
	MEAN	2.43E-12	12092.34	9731.963	-6.69494	21.85658	0.762233
	NAÏVE	-311.937	8018.467	6701.453	-1.98137	14.30493	0.269363
	SNAÏVE	-3190.68	7797.248	6340.655	-8.10497	14.95868	0.679204
	RW-DRIFT	9.19E-13	8012.397	6687.139	-1.29341	14.23191	0.269363

This difference in performance suggests that the STLF model outperforms the ETS model when forecasting coal consumption, particularly in terms of its ability to generalize well to unseen data. The choice to prioritize the STLF model is supported by its superior performance on the test data, highlighting its potential for accurate and reliable coal consumption predictions in real-world scenarios.

Table 4. Model comparison in terms of Errors for the Test Data set.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Test Data	STLF	-5630.45	5936.203	5630.449	-17.4302	17.4302	0.35842
	ARIMA	-10728.1	11186.03	10728.1	-35.0688	35.06884	0.030943
	ETS	-6864.8	7288.344	6864.795	-21.1329	21.13289	0.338486
	MEAN	-15923.6	17662.76	15923.57	-56.8521	56.85209	0.536813
	NAÏVE	-9111	11892.15	10296	-34.8482	37.53244	0.536813
	SNAÏVE	-8444.63	9014.585	8444.625	-28.1565	28.15649	0.300924
	RW-DRIFT	-7707.28	11166.73	10062.05	-30.5693	35.90604	0.560167

From Figure 14, The STLF model for coal consumption forecasting is robust, with residuals displaying normality, minimal significant autocorrelation in the ACF plot, and random fluctuation. Overall, the model effectively captures underlying data patterns.

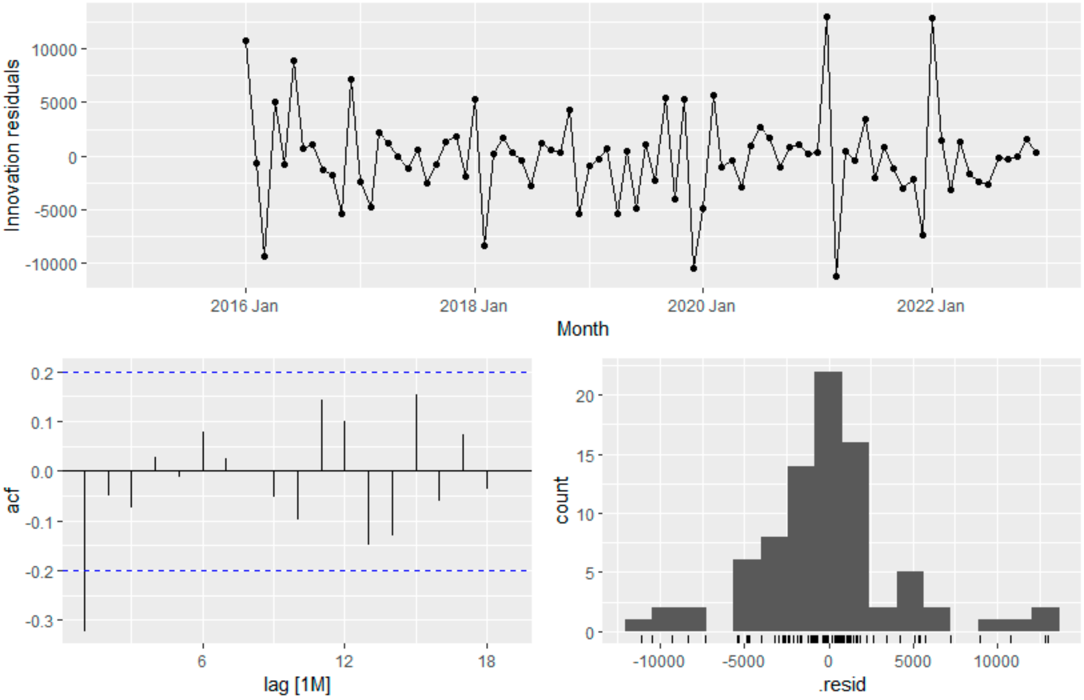


Figure 14. Residual analysis of STLF forecasting method.

2.4.3. Coak

From Table 5, the ETS model exhibits the lowest RMSE of 44.60 on the training data, indicating a good fit to the training set. However, the situation changes when assessing the test data, where the snaive model achieves the lowest RMSE of 99.49, outperforming the ETS model with an RMSE of 134.36 (Table 6).

Table 5. Model comparison in terms of Errors for the Train Data set.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Train Data	STLF	0.072135	50.30918	38.44801	-1.7798	15.68808	-0.31094
	ARIMA	-7.79429	50.97813	39.21442	-6.92597	16.82244	-0.04866
	ETS	-2.2064	44.60708	35.05145	-3.58674	14.38404	0.147046
	MEAN	0	66.17165	52.33333	-7.38547	21.85282	0.57797
	NAÏVE	-1.11579	59.88902	47.55789	-3.72212	19.53299	-0.17484
	SNAÏVE	-12.619	71.87241	60.09524	-9.37567	25.92356	0.360049
	RW-DRIFT	1.08E-14	59.87863	47.54859	-3.2987	19.48854	-0.17484

Table 6. Model comparison in terms of Errors for the Test Data set.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Test Data	STLF	-134.524	139.2002	134.5243	-97.8964	97.89637	0.392624
	ARIMA	-100.747	115.8635	100.7467	-79.5287	79.52873	0.571681
	ETS	-130.006	134.3647	130.0063	-94.2535	94.25347	0.375933
	MEAN	-122.75	134.7748	122.75	-94.6778	94.67779	0.558346
	NAÏVE	-134.75	145.7884	134.75	-102.904	102.9036	0.558346
	SNAIVE	-78.5	99.49749	90	-64.2448	68.79813	0.355008
	RW-DRIFT	-129.729	141.8381	129.7289	-99.7341	99.73414	0.565986

Based on the visual analysis as depicted in Figure 16, the snaïve model for petroleum coke consumption forecasting demonstrates robustness. The residuals exhibit a normal distribution in the histogram, minimal significant autocorrelation in the ACF plot, and random fluctuation.

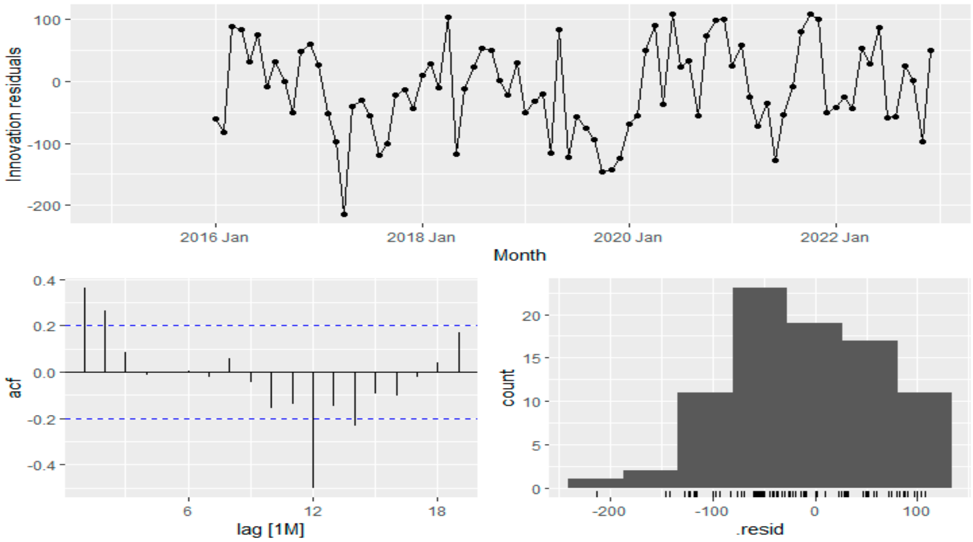


Figure 16. Residual analysis of SNAÏVE forecasting method.

2.4.4. Petroleum Liquid

From Table 7 and Table 8 ARIMA model exhibits the lowest RMSE on the training data at 1199.579, indicating a robust fit to historical observations. However, on the test data, the mean method surpasses the ARIMA model, achieving the lowest RMSE of 287.34, in contrast to ARIMA's RMSE of 424.909. Despite the mean method's superior performance on the test data, we prioritize the ARIMA model for its capability to capture underlying complex patterns, which is particularly advantageous in forecasting petroleum liquids consumption.

Table 7. Model comparison in terms of Errors for the Train Data set.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Train Data	STLF	51.27077	1206.734	469.8335	-5.79073	20.42126	-0.351
	ARIMA	-0.39939	1199.579	556.5674	-12.6417	22.80608	-0.00636
	ETS	-72.5277	1522.814	687.7297	-11.0457	34.25543	-0.26354
	MEAN	0	1214.503	563.3268	-12.923	23.11294	0.136773
	NAÏVE	27.18947	1548.012	604.3474	-9.65595	25.94808	-0.37293
	SNAÏVE	-1.96429	1494.727	578.4881	-6.82803	22.23511	0.142605
	RW-DRIFT	-1.82E-13	1547.773	604.41	-11.2077	26.08228	-0.37293

Table 8. Model comparison in terms of Errors for the Test Data set.

	Model	ME	RMSE	MAE	MPE	MAPE	ACF1
Test Data	STLF	-3879.82	3893.528	3879.82	-225.717	225.7167	-0.24153
	ARIMA	-353.011	424.909	354.831	-20.8028	20.8937	-0.39078
	ETS	-6402.64	6769.273	6402.639	-365.768	365.7678	0.550292
	MEAN	-260.063	287.3403	263.7344	-15.588	15.77136	0.232705
	NAÏVE	-4147.75	4149.55	4147.75	-241.594	241.5938	0.232705
	SNAÏVE	-475.75	1221.797	544.5	-26.5977	30.84144	-0.00337
	RW-DRIFT	-4270.1	4273.105	4270.103	-248.811	248.8107	0.476336

The visual analysis as in Figure 18 suggests the robustness of the ARIMA model for petroleum liquids consumption forecasting. Residuals exhibit a normal distribution in the histogram, minimal significant autocorrelation in the ACF plot, and random fluctuation.

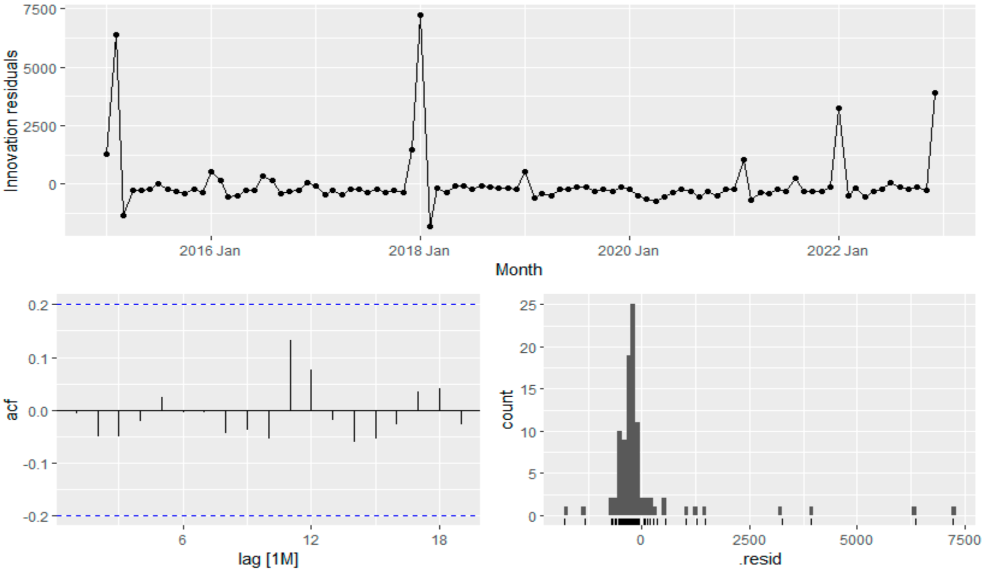


Figure 18. Residual analysis of ARIMA forecasting method.

3. Result and Discussion

3.1. Analysis of forecasting trends for different energy streams

3.1.1. Overall Trend of NG Consumption

As shown in Figure 19, the use of Natural Gas (NG) is still increasing, especially during the summer months. This is because natural gas is a versatile and efficient energy source, commonly used for generating electricity. It is considered better for the environment compared to other fuels, and its affordability and reliability make it a popular choice. During hot summer days, when people use more electricity for things like air conditioning, natural gas becomes a preferred option. Overall, because its environmentally friendly, cost-effective, and used in various ways, the demand for natural gas continues to rise.

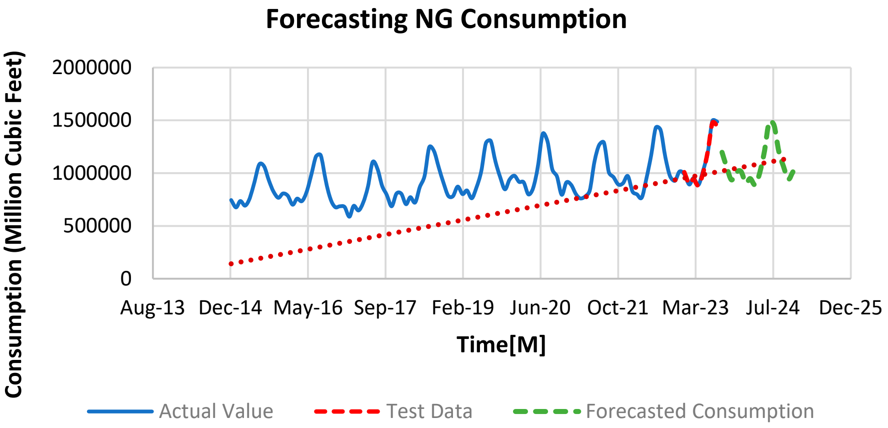


Figure 19. Overall trend analysis of NG consumption for US electricity generation.

3.1.2. Overall Trend of Coal Consumption

The forecast for coal consumption indicates a decreasing trend, raising questions about the underlying reasons. Analyzing the historical trend from 2015 to 2022, as shown in Figure 3, reveals a clear seasonality pattern. There is a noticeable peak in coal consumption during the summer months of July and August, followed by a gradual decline in September and October. Consumption reaches a low point in November and then starts to rise again in December. This cyclic pattern suggests a seasonal influence on coal consumption, with similar patterns observed in the forecasted coal consumption (Figure 20). The decreasing trend in coal consumption can be attributed to several factors. Firstly, there is a global shift towards cleaner and more sustainable energy sources, driven by environmental concerns and efforts to reduce carbon emissions. Governments and industries are increasingly adopting renewable energy sources and natural gas, which are considered more environmentally friendly alternatives to coal.

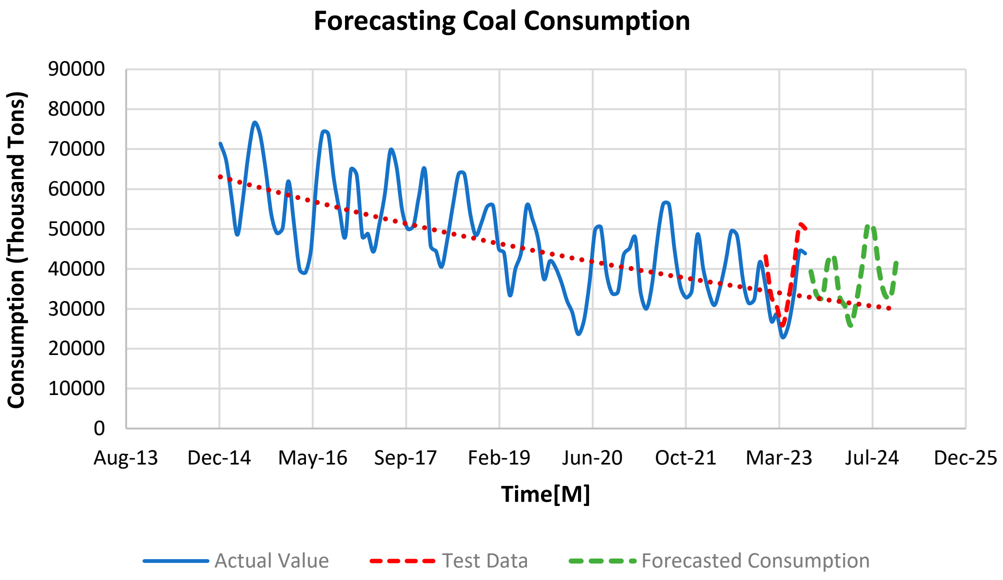


Figure 20. Overall Trend Analysis of Coal consumption for US electricity generation.

3.1.3. Overall Trend of Petroleum Coke Consumption

The data shows that petroleum coke use goes up a bit, especially in July and August, and then it goes down in October. This has happened not only in the past but is also expected in the future (Figure 21). Petroleum coke comes from refining oil and is used in different industries, like making cement and producing power. The small increase might be because some industries need more of it, or it could be happening where economies are growing. The ups and downs in different months might be connected to specific activities in those months. Additionally, changes in how oil is refined globally can affect how much petroleum coke is available and how much people want.

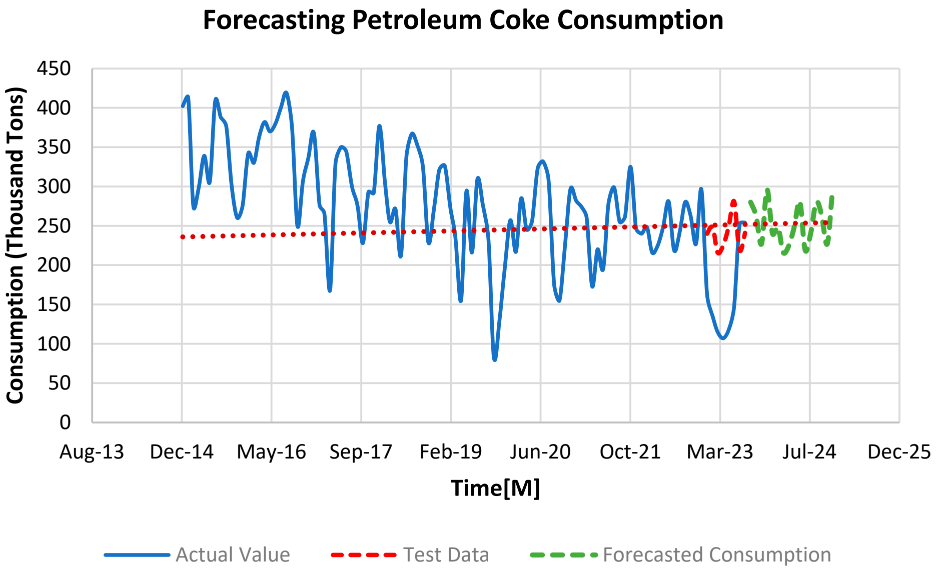


Figure 21. Overall Trend Analysis of Petroleum Coke consumption for US electricity generation.

3.1.4. Overall Trend of Petroleum Liquid Consumption

The monthly patterns and seasonality analysis of petroleum liquid consumption between 2015 and 2022 reveal spikes in February 2015, January 2018, January 2022, and December 2022. However, in the forecasted data, there are no unusual spikes. Instead, the forecast predicts consistent and steady consumption (Figure 22). This lack of anomalies in the forecast suggests that improved forecasting models have been employed. These models take into account historical irregularities but anticipate a more stable trend for the future.

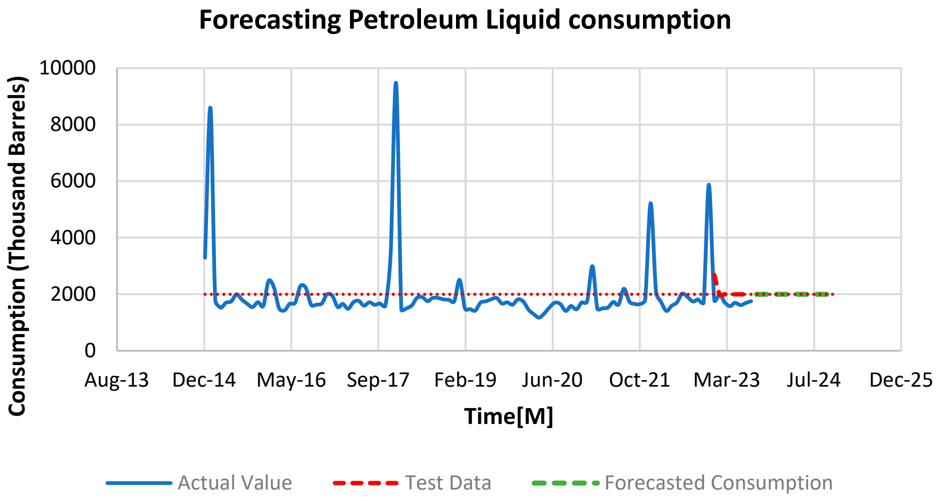


Figure 22. Overall Trend Analysis of Petroleum Liquid consumption for US electricity generation.

3.2. Comparing the forecasted trends of different energy streams in electricity generation

The graph as shown in Figure 23, explains the projected consumption patterns for four key energy streams used in electricity generation in the USA, spanning from September 2023 to December 2024. Over this period, natural gas consumption fluctuates, with notable peaks in July and August 2024 at approximately 1,542,143 MMBTUs and 1,508,448 MMBTUs, respectively. These surges are likely a response to heightened electricity demand for cooling during the hotter months, a recurring trend that emphasizes the influence of seasonal temperature variations on energy utilization.

Coal consumption, while demonstrating less volatility than natural gas, shows minor variations that could be linked to industrial demand cycles, regulatory impacts, and market pricing. For instance, coal consumption rises to 835,000 MMBTUs in December 2023, which may reflect an increase in energy needs during the winter season. Meanwhile, coke, used primarily in industrial processes like steel manufacturing, maintains a relatively flat demand, with a minor peak at 7,400 MMBTUs in December 2023, suggesting a consistent industrial requirement with little seasonal impact.

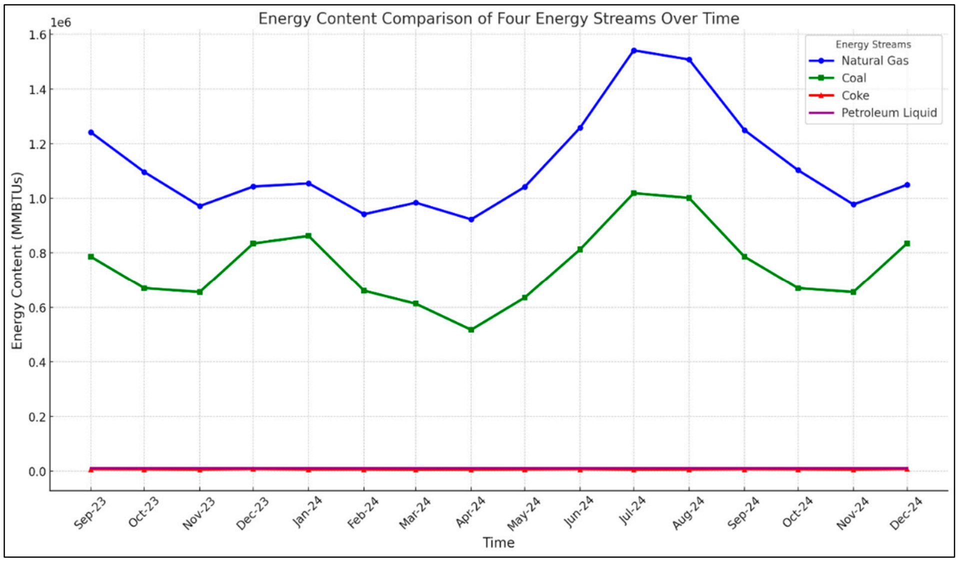


Figure 23. Forecasted trend of all four components for fuel consumption to generate electricity.

Petroleum liquids display an unwavering consumption value of approximately 11,575.176 MMBTUs throughout the forecasted period. This consistency indicates that petroleum liquids may have a set role in the energy mix for electricity generation, potentially due to established supply chains and a stable market for petroleum products.

The forecast result suggests that while natural gas and coal exhibit seasonality and potential sensitivity to external market and policy conditions, coke and petroleum liquids' demand appears more stable and possibly insulated from such factors. This stability could be due to long-term contracts, stockpiling strategies, or their specific usage domains within the industry, which might be less susceptible to short-term changes. Having said that, the natural gas will be still likely the dominating factor to generate electricity to near future as well for United States. To ensure the reliability of these projections and facilitate strategic energy planning, it is essential to corroborate this data with external literature, including market analyses, industry reports, and regulatory policy documents. This approach would provide a robust foundation for understanding the complex dynamics at play and for preparing the energy sector to meet future demand effectively and sustainably.

4. Conclusion

In this paper, our objective is to provide precise predictions for fuel consumption in electricity generation using time series forecasting models, making a substantial contribution to strategic planning and decision-making processes within the energy sector. Developing accurate predictive models is geared towards forecasting future fuel consumption, equipping stakeholders with valuable insights for well-informed decisions regarding resource allocation and sector-specific strategies.

Our analysis aims to identify sectors undergoing declining or increasing trends in fuel usage, enabling targeted interventions where necessary. This project seeks to deepen our understanding of energy market dynamics, facilitating proactive measures to efficiently meet future fuel demands. The insights derived from this work are poised to significantly impact the formulation of sustainable energy policies and strategies, ensuring a stable and reliable energy supply for the nation. The accurate forecasting of fuel consumption for electricity generation, as indicated by the forecasts for NG, coal, petroleum coke, and petroleum liquids, holds paramount importance for policymakers. These forecasted trends provide valuable insights that guide strategic decision-making in the ever-evolving landscape of the energy sector.

The expected rise in NG consumption suggests a need for increased infrastructure investment to support its growing demand, aligning with environmental goals. Conversely, the anticipated decline in coal usage allows policymakers to plan for a transition to cleaner alternatives, supporting affected communities. A slight increase in petroleum coke usage signals specific industrial needs, prompting policymakers to create supportive regulations. With a steady forecast for petroleum liquids, policymakers can plan infrastructure development to match stable energy needs. These forecasts empower policymakers to encourage sustainable energy practices, support economic growth, and ensure energy security, fostering a resilient and cleaner future.

Author Contributions: Conceptualization, M.M.H.B.; methodology, M.M.H.B.; software, M.M.H.B.; validation, M.M.H.B. and T.R.; formal analysis, M.M.H.B.; investigation, M.M.H.B.; resources, M.M.H.B.; writing—original draft preparation, M.M.H.B. and A.N.S.; writing—review and editing, T.R., S.I.A. and A.N.S.; visualization, supervision, T.R. and M.M.H.B.; project administration, A.N.S. and M.M.H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. W. Strielkowski, L. Civín, E. Tarkhanova, M. Tvaronavičienė and Y. Petrenko, "Renewable Energy in the Sustainable Development of Electrical Power Sector: A Review," Energies, vol. 14, no. 24, 07 December 2021.
2. "Units and calculators explained," U.S. Energy Information Administration-EIA, 2023. [Online]. Available: https://www.eia.gov/energyexplained/units-and-calculators/.
3. "U.S. energy facts explained," www.eia.gov, 2023. [Online]. Available: https://www.eia.gov/energyexplained/us-energy-facts/.
4. "Total primary energy consumption in the United States from 1995 to 2022," Statista, 2023. [Online]. Available: https://www.statista.com/statistics/183774/total-primary-energy-consumption-in-the-united-states-since-1995/.
5. J. Holechek, H. Geli, M. Sawalhah and R. Valdez, "A Global Assessment: Can Renewable Energy Replace Fossil Fuels by 2050?," Sustainability, vol. 14, no. 8, 2022.
6. "Electricity explained, Electricity in the United States," U.S. Energy Information Administration – EIA, 2023. [Online]. Available: https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php..
7. P. Manigandan, M. Alam, M. Alharthi, U. Khan, K. Alagirisamy and A. Rehman, "Forecasting Natural Gas Production and Consumption in United States-Evidence from SARIMA and SARIMAX Models," energies, vol. 14, 2021.
8. I. Karakurt, "Modelling and forecasting the oil consumptions of the BRICS-T countries.," Energies, vol. 220, no. 119720, 2021.
9. J. Liu, S. Wang, N. Wei, X. Chen, H. Xie and J. Wang, "Natural gas consumption forecasting: A discussion on forecasting history and future challenges.," J. Nat. Gas Sci. Eng., vol. 90, no. 103930, 2021.
10. L. Suganthi and A. Samuel, "Energy models for demand forecasting—A review. Renew.," Sustain. Energy Rev., vol. 16, pp. 1223-1240, 2012.
11. "Electric Power Monthly," U.S. Energy Information Administration – EIA, 2023. [Online]. Available: https://www.eia.gov/electricity/monthly/.
12. "Climate Impacts on Energy," United States Environmental Protection Agency-EPA, 2017. [Online]. Available: https://19january2017snapshot.epa.gov/climate-impacts/climate-impacts-energy_.html.
13. "Today in Energy," U.S. Energy Information Administration – EIA, 2020, 2020. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=42815#.
14. P. Brockwell and R. Davis, Introduction to Time Series and Forecasting, 2nd ed., New York: Springer: New York, 2002.
15. R. Ospina, J. Gondim, V. Leiva and C. Castro, "An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil.," Mathematics, vol. 11, 2023.
16. R. Park, K. Song and B. Kwon, "Short-Term Load Forecasting Algorithm Using a Similar Day Selection Method Based on Reinforcement Learning.," Energies, vol. 13, no. 2640, 2020.
17. R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed., OTexts, 2021.

18. S. Kang, C. Park, C. Lee and S. Lee, "The Stress-Induced Impact of COVID-19 on Tourism and Hospitality Workers.," Sustainability, vol. 13, no. 1327, 2021.
19. Z. Ouyang, P. Ravier and M. Jabloun, "STL Decomposition of Time Series Can Benefit Forecasting Done by Statistical Methods but Not by Machine Learning Ones.," Eng. Proc., vol. 5, no. 42, 2021.
20. P. Manigandan, M. Alam, M. Alharthi, U. Khan, K. Alagirisamy, D. Pachiyappan and A. Rehman, "Forecasting Natural Gas Production and Consumption in United States-Evidence from SARIMA and SARIMAX Models," Energies, vol. 14, no. 6021, 2021.