Jon Taylor

Forecasting Electricity Generated by Fossil Fuel in the United States

Forecasting and Time Series Models

Abstract

Due to global warming and increasing temperatures, the urgency to transition from fossil fuels to sustainable energy has become critical. This paper displays a statistical analysis and forecast of fossil fuel electricity generation in the United States, using time series data from *Our World in Data*. The study utilizes time series models, including benchmark and ARIMA methods, to predict fossil fuel use over the next decade.

Initial analysis reveals a steady decline in fossil fuel electricity generation over the past decade, with significant autocorrelation in the data. After accounting for heteroskedasticity and decomposing the series, ARIMA (2,1,0)(0,1,0) was identified as the most accurate model. Forecasts suggest that fossil fuel-generated electricity will stay consistent in the future, with fossil fuel use in 2033 being only slightly below 2023's numbers. This stagnation highlights limited progress toward reducing fossil fuel dependency and underscores the need for targeted energy policies.

The results provide a critical benchmark for policymakers to design interventions such as caps on fossil fuel use, subsidies for renewable energy, and tax credits to encourage green energy investments. While the forecast suggests a static trend, it emphasizes the potential for disruption through legislative action and societal efforts. The paper concludes that immediate and sustained policy initiatives are essential to shift the trend downward and accelerate the transition to sustainable energy.

Introduction

In the current political climate, global warming is a huge area of focus. Given that the world is experiencing record high temperatures, the need to utilize different means of energy is urgent as ever. Policy makers are under more pressure than ever to create legislation that would limit the use of fossil fuels and pursue different avenues of generating electricity. In order to make practical legislation, or legislation that would result in slowly becoming independent from fossil fuels, policy makers need to know how much fossil fuels are being used, and whether there is a trend in that usage. A good estimation of dependency would help legislators to allocate the proper resources towards making the energy companies that still rely heavily on fossil fuels to use them in a cleaner way as well as a provide a good idea of the timeframe of transitioning away from oil, gas, etc. The country could then practically lower the use of fossil fuels

With the proper data, policymakers can create subsidies, tax credits, and penalties that could influence the country to wean off of fossil fuels. Furthermore, a forecast could create a tangible report that could be used to educate stakeholders as well as highlight the urgency of the climate change situation. Along with educating stakeholders, a report could garner public support for the response to climate change, serving as a catalyst for the policymakers to get elected and pass legislation.

The EPA constantly tracks the usage of fossil fuels and other energy sources, and a time series of the amount of energy being created by fossil fuels is included in their data sets. In order to provide policy makers with the information they need to make clean energy policy, the following report aims to evaluate the amount of electricity being created by fossil fuels in recent history as well as forecast the amount of electricity that will be generated by fossil fuels in the next 10 years.

Data Description

The data used for this project was found online from Our World in Data. The data set includes information about all types of energy use all over the world. The data selected for the present project is the United States' fossil fuels use. Specifically, the variable of interest is the amount of electricity being generated by fossil fuels in the United States. Rather than the amount of fossil fuels being consumed overall, the amount of electricity being generated is being used because it makes it comparable to other means of energy. This allows the forecast to be used practically as it represents the reliance of the country on fossil fuels as a means of powering our society, and furthermore, it can be used to create a benchmark that can be referenced as the country tries to lower reliance on unsustainable energy. To make a more accurate forecast, the data set was cut to only include the data points since 2013 as the recent years are likely more representative of current energy use.

Statistical Methods

Original Investigation

When the data was originally graphed (figure 1), it displayed a downward trend and some heteroskedasticity in the most recent years. There was a noticeable downward spike in 2020, likely due to the COVID pandemic, but in the years following it went back up. Overall, the use has lowered by around 400 terawatts in the last 20 years. In the original observation, the last 20 years was included, but given the patterns looked different in the second half of the data, it was shortened to include 2013 and on.

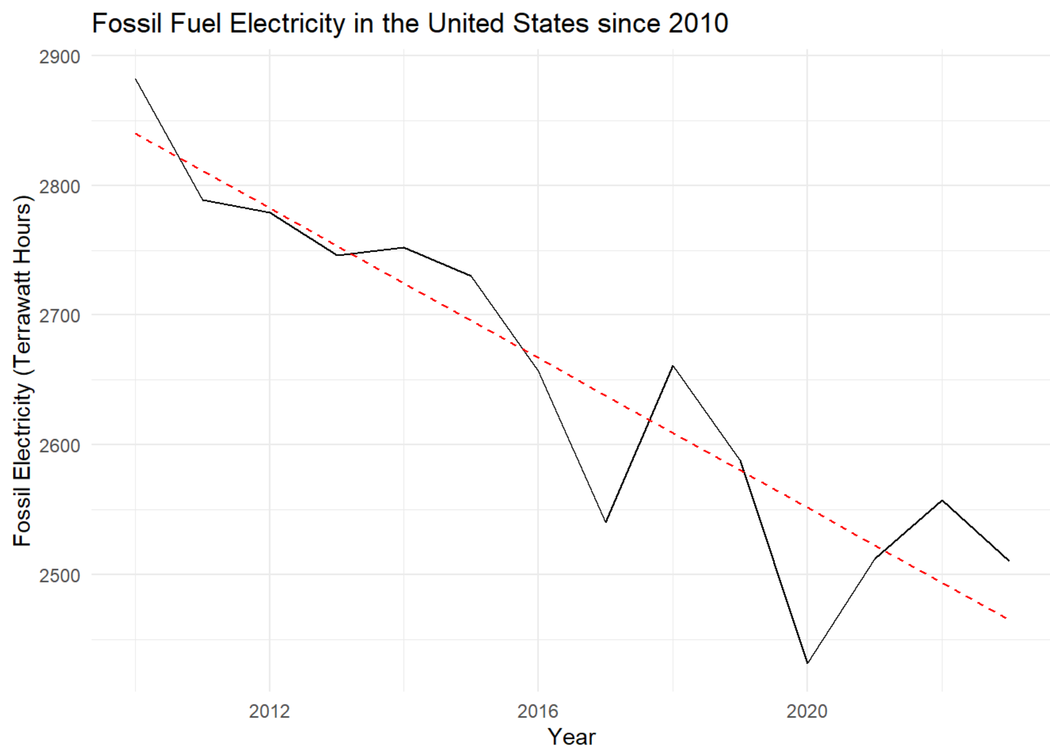

Figure 1

Auto-Correlation Function

Based on the correlogram (figure 2), fossil fuel electricity is displaying significant autocorrelation. When a variable displays autocorrelation, it means that it is correlated with lags in the time series. Due to this autocorrelation, it can be deduced that the current amount of electricity being produced is being influenced by past values. Government officials or energy companies may be interested in this trend because it could help them to predict the demand of fossil fuels or how much fossil fuels are going to be used in the near future.

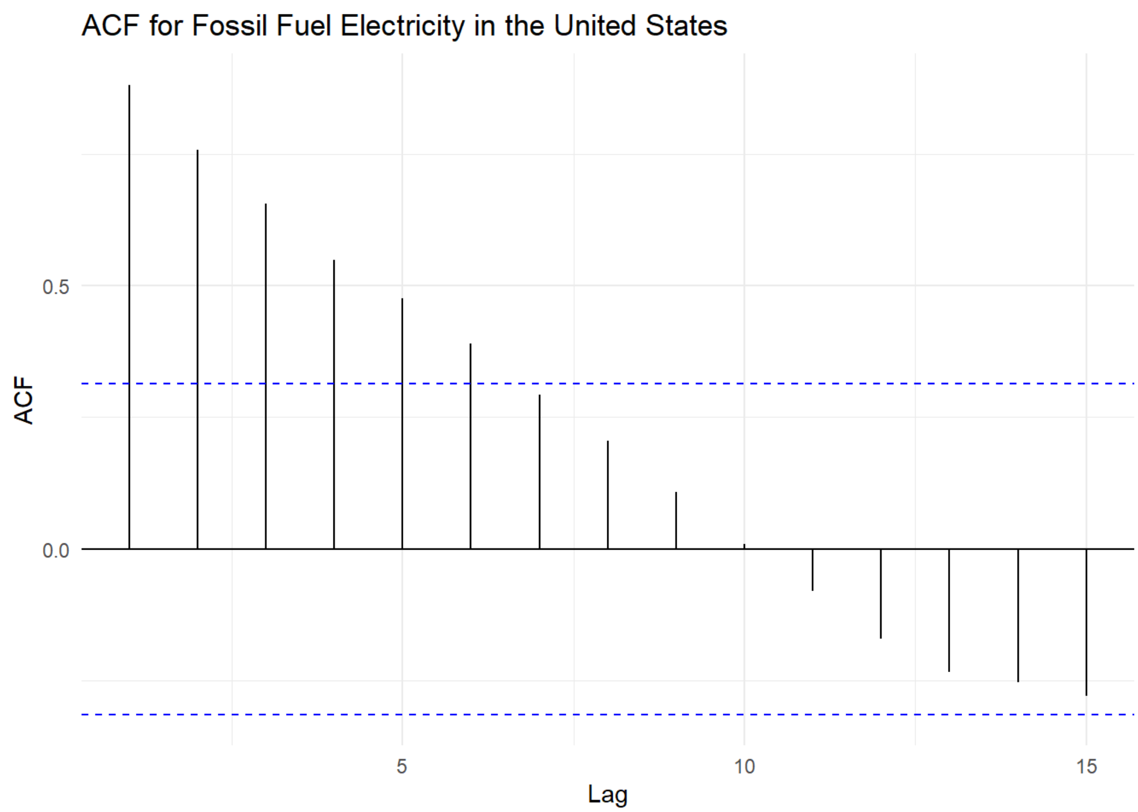

Figure 2

Transformation

In order to account for heteroskedasticity in the later years of the series, a box cox transformation was made (figure 3). Transforming the data using a box cox function allows for a better model due to the fact that it accounts for the variance and normalizes the data. By making the data more linear and stabilizing the greater variance in the later years, more accurate conclusions can be drawn about the use of fossil fuels.

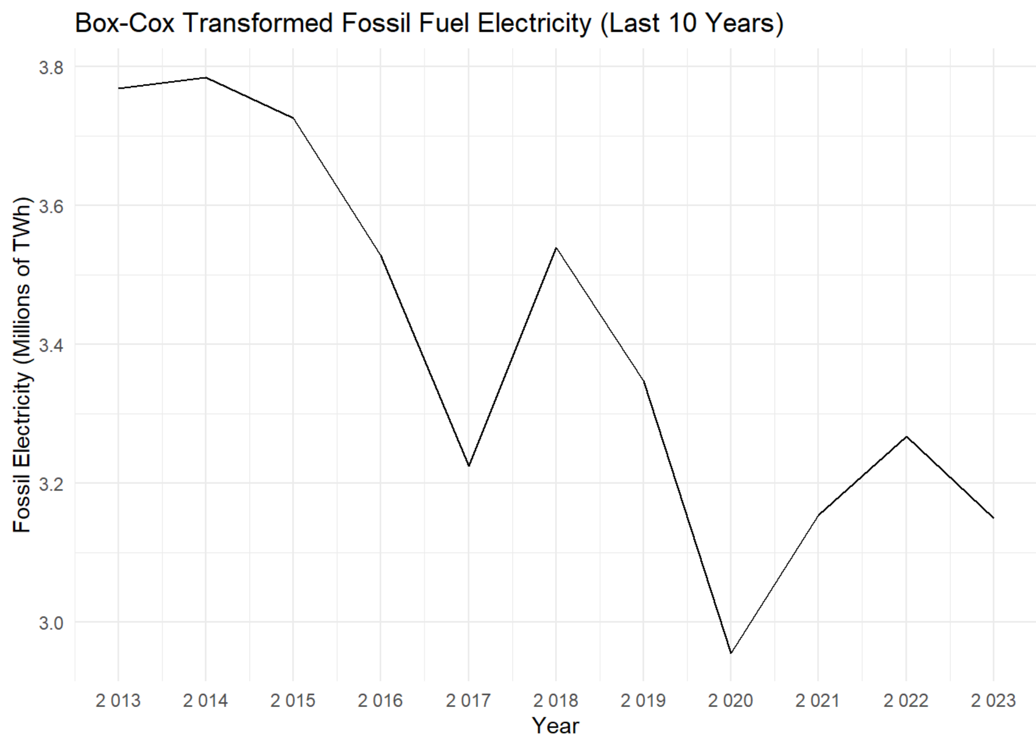

Figure 3

Decomposition

First, the decomposition of the fossil fuels (figure 4) data shows that there is a large downward trend in the data. The downward trend likely reflects recent efforts to become independent from fossil fuels, limit emissions in general, and the movement other energy sources. Second, the remainder displays the random fluctuations in the data. The explanation for this piece of the data is harder to understand. However, it does show that the use of fossil fuels may have been affected by random, or unexpected, events, which caused noise in the data. The does not display any sort of seasonality, which reflects the idea that the demand for fossil fuels has been steady over the years. The fossil fuels demanded does not depend on the year or a cycle within the years.

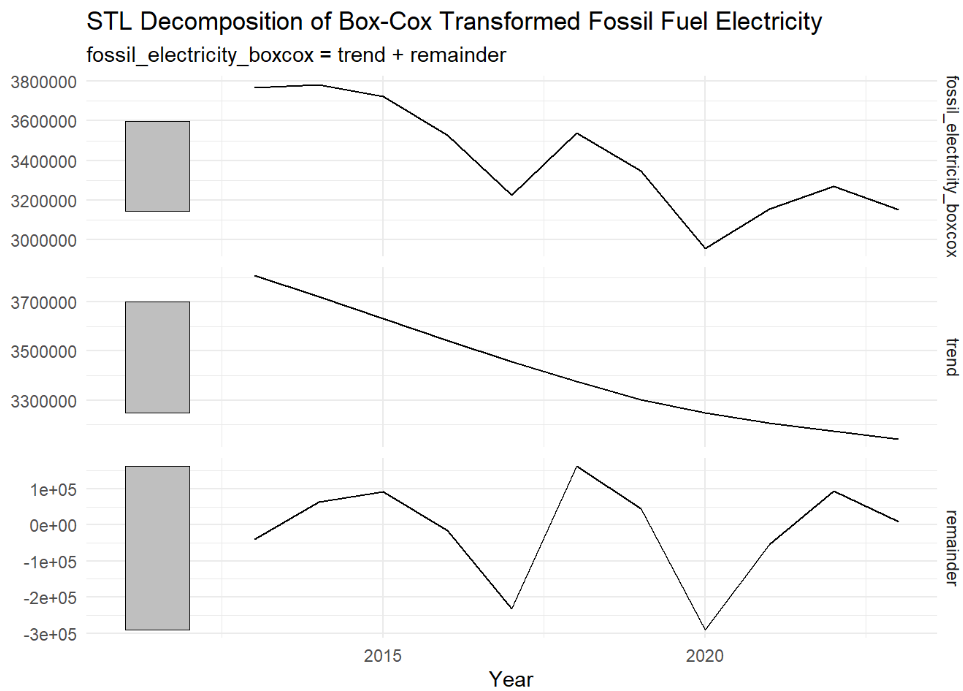

Figure 4

Forecasting:

Benchmark Models

The benchmark forecasting methods, the naive, mean, and drift models, were used to create the original forecasts. Using these methods can create a good baseline for more complex models that are used later. The accuracy results of the benchmark models are shown in table 1.

A tibble: 3 × 11

	.model	country	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Drift	United Stat…	Test	4.67e5	4.80e5	4.67e5	14.6	14.6	2.22	1.95	-0.172
2	Mean	United Stat…	Test	-2.94e5	2.99e5	2.94e5	-9.24	9.24	1.39	1.21	-0.666
3	Naïve	United Stat…	Test	2.35e5	2.41e5	2.35e5	7.34	7.34	1.12	0.980	-0.666

Table 1

The naive method is the best of the benchmark models because it has the lowest mean absolute percentage error (MAPE) of 7.33. The MAPE measure is used to assess the accuracy of a forecasting model by calculating the percent error on average, or how off the predictions are compared to the actual values. The naive model may work best because the use of fossil fuels to generate electricity has a relatively stable downward trend. The naive model can forecast stable models better, especially in the short term, such as the 11 years contained in the energy data set or the 10 future years that are being forecasted.

Testing the Naïve Method

In order to test the Naïve model further, the residuals were tested to ensure that it accounted for the trend in the data. There does not appear to be any significant autocorrelation in the data, which depicts that the model is a good fit for the data. Since the residuals are random, or white noise, the model has accounted for all the trend in the data. This also implies that the forecast is accurately fitting the data while capturing the main components. Furthermore, the p-value of 0.41, which is not significant, indicates that the residuals are white noise. The results of this test are displayed in table 2 and figure 5.

```
# A tibble: 1 × 3
  .model lb_stat lb_pvalue
  <chr>    <dbl>     <dbl>
1 Naïve     6.07     0.416
```

Table 2

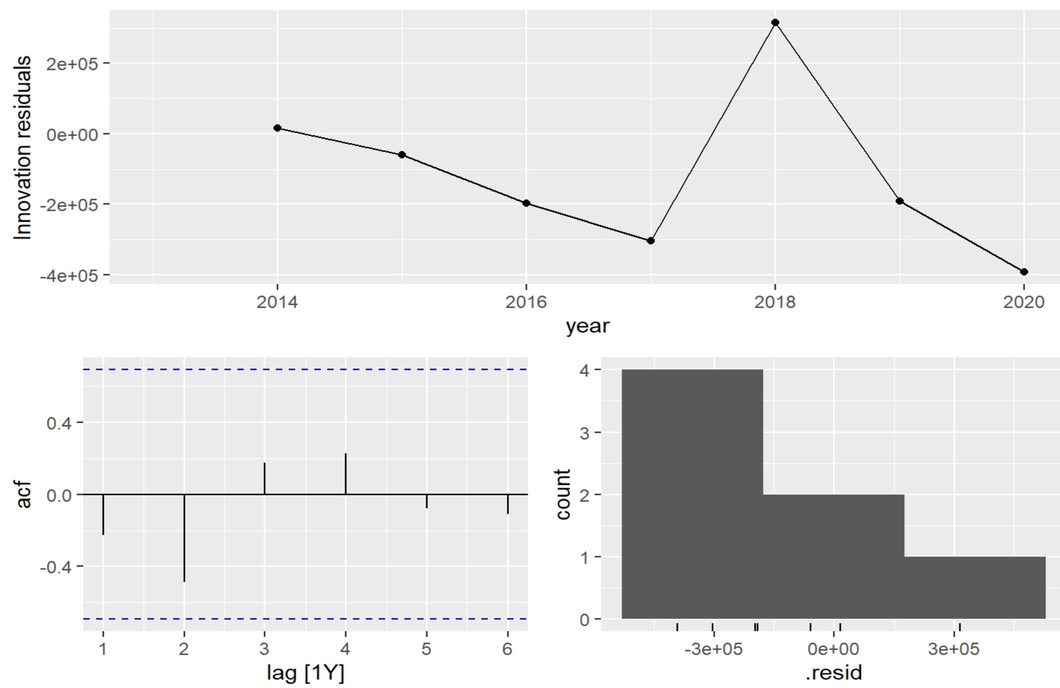

Figure 5

The forecast made using the naïve model is shown in figure 6.

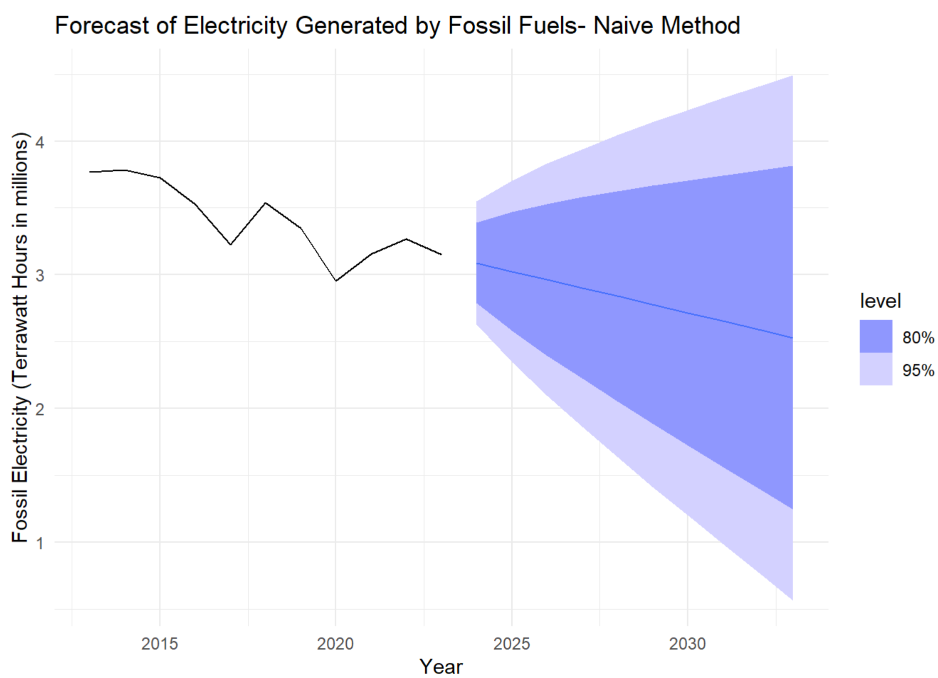

Figure 6

The resulting naïve forecast displays a consistent downward trend in the next ten years. While the naive forecast makes sense, more complicated models could be explored to seek a more accurate forecast. Some more complicated models include ARIMA models and creating an ensemble that includes combining the best forecasting methods.

Exponential Smoothing Models (ETS)

Exponential Smoothing Models (ETS) were used to create forecasts as well. ETS models are a good choice for this data because they would likely work well with this data due to them being well suited with trend yet no seasonality. The best of the ETS models was the Holt Damped model, which had a MAPE of 5.3. The results of the ETS Models is shown in figure 7 and a table with the accuracy results are in table 3.

A tibble: 5 × 11

	.model	country	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Drift	United …	Test	4.67e5	4.80e5	4.67e5	14.6	14.6	2.22	1.95	-0.172
2	ETS	United …	Test	1.83e5	1.91e5	1.83e5	5.70	5.70	0.867	0.775	-0.666
3	Holt	United …	Test	2.78e5	2.95e5	2.78e5	8.69	8.69	1.32	1.20	-0.203
4	Holt_Damped	United …	Test	1.70e5	1.91e5	1.70e5	5.31	5.31	0.808	0.775	-0.271
5	ses	United …	Test	1.83e5	1.91e5	1.83e5	5.70	5.70	0.867	0.775	-0.666

Table 3

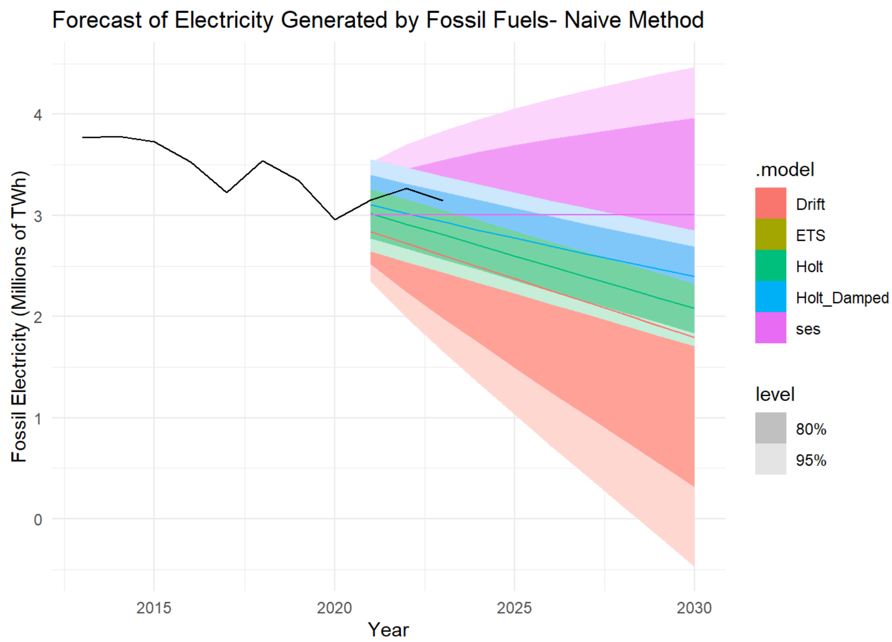

Figure 7

ARIMA Models

Next, Auto Regressive Integrated Moving Average (ARIMA) models were used to create forecasts of the data. The auto regression part of the ARIMA uses past values of the data to calculate future values, the integrated part accounts for trend and seasonality, and the moving average piece accounts for past forecasting errors. ARIMA models are useful because they are very adaptable to different data sets and they can capture trend/seasonality well. By using different values of p (auto regression), d (integration), and q (moving average), the different effects can be accounted for. Due to the fossil fuel data having a downward trend, using the ARIMA model could present better accuracy results. Initial models of the ARIMA function were calculated using the auto ARIMA function in R as well as initial attempts at manually plugging in values of p, d, and q based on the ACF and Partial autocorrelation function (PACF) located in figure 8. The accuracy results of the initial ARIMA models are located in table 4, and the graph of these forecasts is included in figure 9.

```
# A tibble: 3 × 11
  .model     country .type        ME   RMSE    MAE   MPE  MAPE  MASE RMSSE    ACF1
  <chr>      <chr>   <chr>     <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>
1 arima1     United… Test    1.28e5 1.38e5 1.28e5  4.00  4.00 0.609 0.559 -0.360
2 arima2     United… Test    5.25e4 6.08e4 5.25e4  1.64  1.64 0.249 0.247 -0.0998
3 auto_arima United… Test   -2.94e5 2.99e5 2.94e5 -9.24  9.24 1.39  1.21  -0.666
```

Table 4

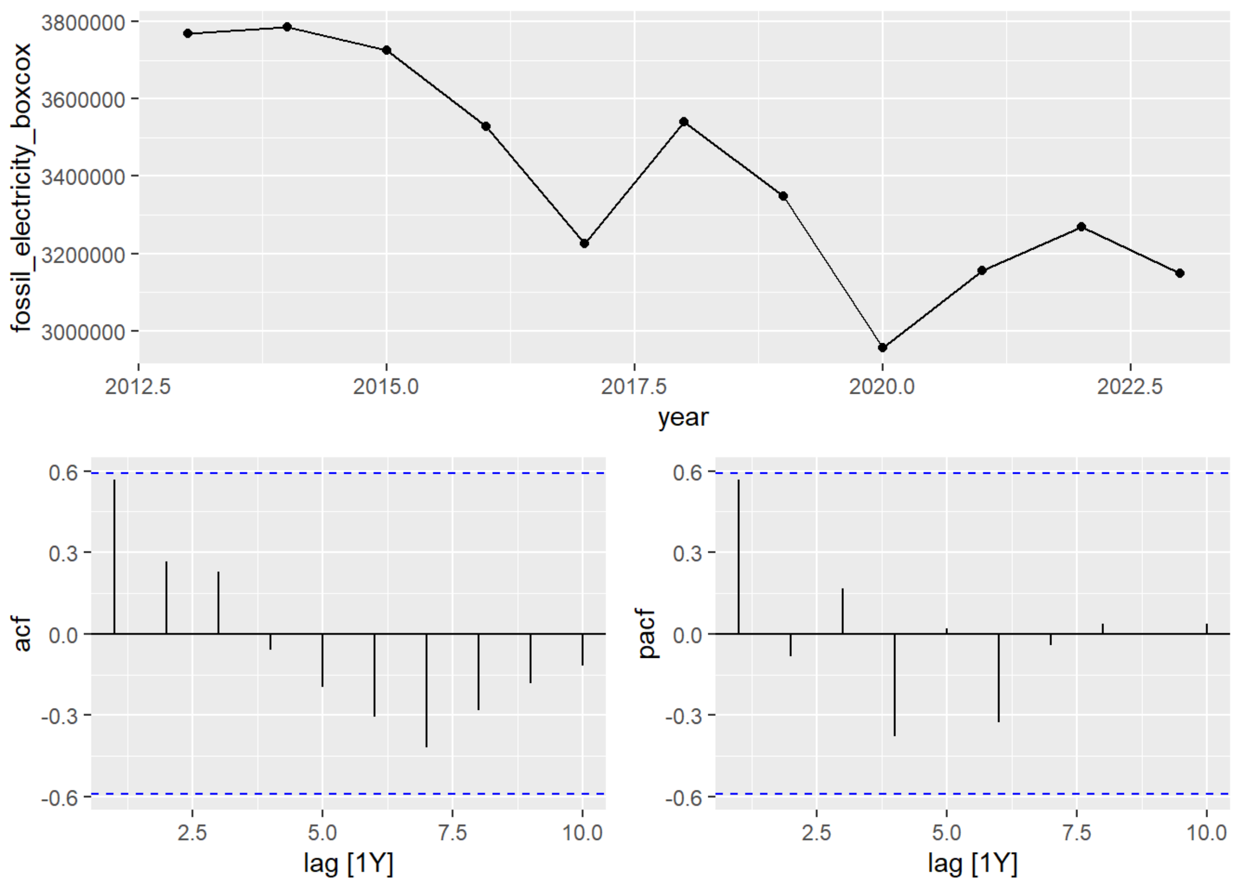

Figure 8

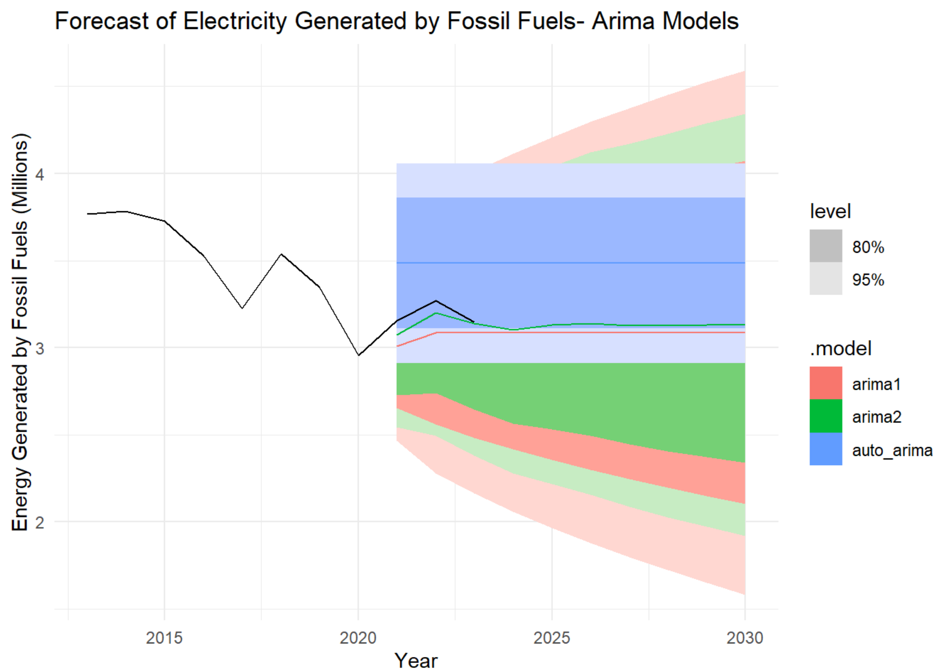

Figure 9

As can be seen in the table, the ARIMA models seem to be more accurate than the benchmark methods, displaying better measures of accuracy. Of the ARIMA models, the model named "arima2", with pdq(2,1,0) and PDQ(0,1,0) was the most accurate (MAPE = 1.63).

For the last part of the analysis, the best 4 models, which included the arima2, arima1, Holt Damped and naïve models, were used to create an ensemble. The ensemble was created by using different combinations of the forecasts to make new forecasts, and it was created to try and find a more accurate forecast. After completing the ensemble, the most accurate model was the arima2 model (MAPE = 1.63), with the combination of the arima1 and arima2 models being the second most accurate (MAPE = 2.82). The results of the can be found in table 5 and a graph is displayed in figure 10.

A tibble: 11 × 11

	.model	country	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Holt_Damp…	United…	Test	1.70e5	1.91e5	1.70e5	5.31	5.31	0.808	0.775	-0.271
2	Naive	United…	Test	2.35e5	2.41e5	2.35e5	7.34	7.34	1.12	0.980	-0.666
3	arima1	United…	Test	1.28e5	1.38e5	1.28e5	4.00	4.00	0.609	0.559	-0.360
4	arima2	United…	Test	5.25e4	6.08e4	5.25e4	1.64	1.64	0.249	0.247	-0.0998
5	comb1	United…	Test	1.82e5	1.88e5	1.82e5	5.67	5.67	0.862	0.764	-0.578
6	comb2	United…	Test	1.39e5	1.44e5	1.39e5	4.33	4.33	0.658	0.586	-0.481
7	comb3	United…	Test	1.44e5	1.48e5	1.44e5	4.49	4.49	0.682	0.602	-0.552
8	comb4	United…	Test	9.04e4	9.85e4	9.04e4	2.82	2.82	0.429	0.400	-0.252
9	comb5	United…	Test	1.17e5	1.22e5	1.17e5	3.65	3.65	0.555	0.495	-0.665
10	comb6	United…	Test	1.11e5	1.18e5	1.11e5	3.47	3.47	0.528	0.479	-0.502
11	comb7	United…	Test	1.53e5	1.58e5	1.53e5	4.76	4.76	0.724	0.643	-0.612

Table 5

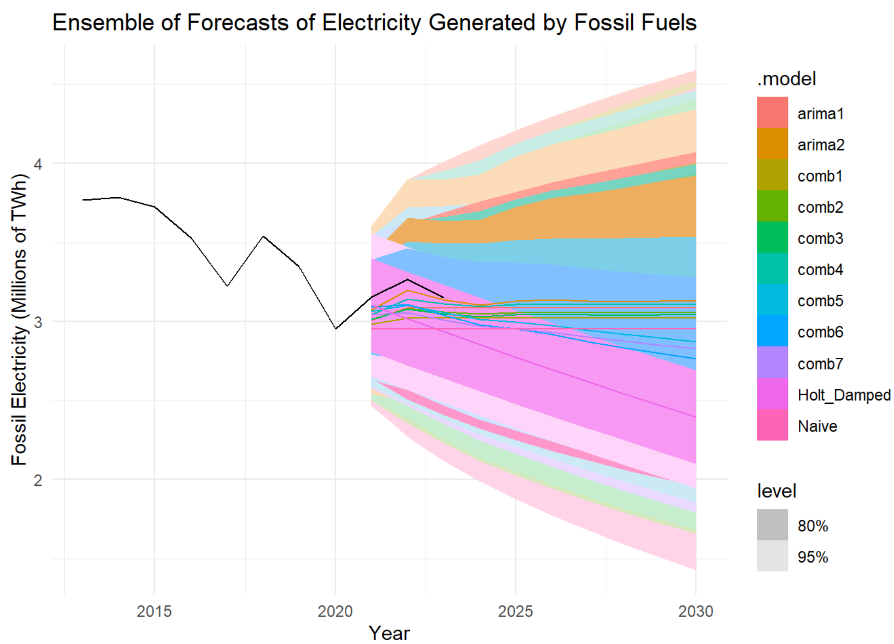

Figure 10

None of the forecasts created using the ensemble were better than the arima2 model, so that is the model used to make the final forecast, which is displayed in figure 11. Furthermore, after testing the residuals of the arima2 model, it displayed that it is a good fit in that its residuals are white noise, which are displayed in figure 12 and table 6.

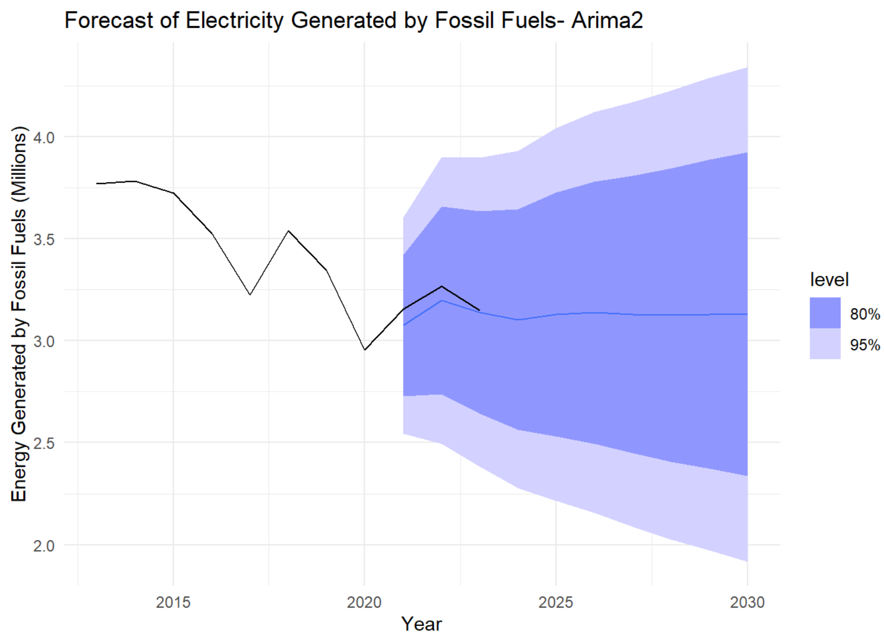

Figure 11

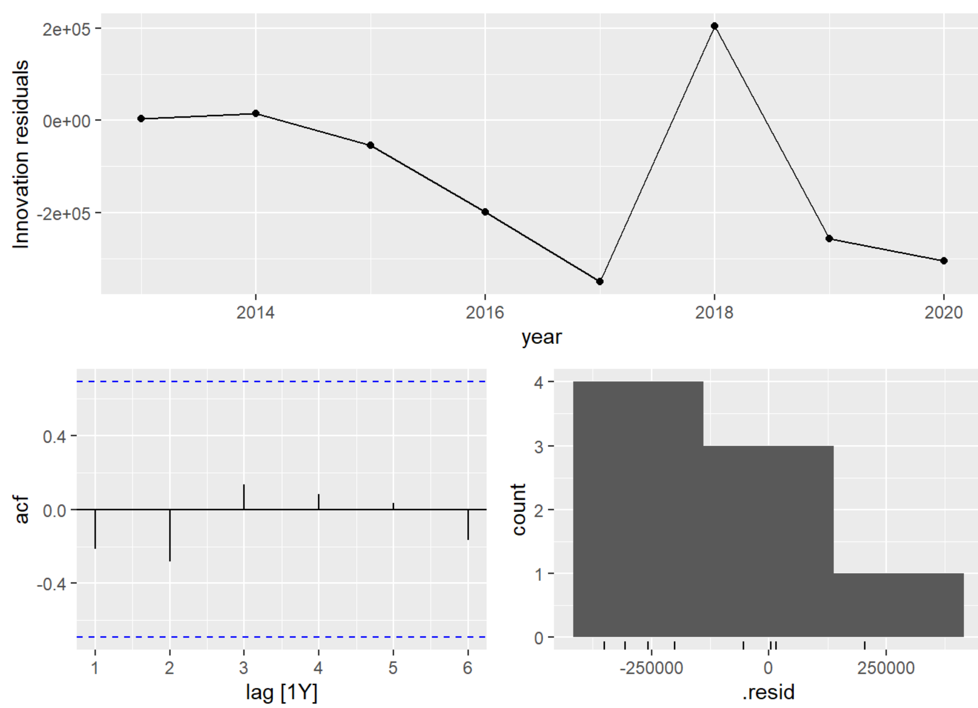

Figure 12

```
# A tibble: 1 × 4
  country       .model lb_stat lb_pvalue
  <chr>         <chr>    <dbl>     <dbl>
1 United States arima2    3.14     0.792
```

Table 6

Results

The ARIMA2 (2,1,0)(0,1,0) forecast suggests that, while the use of fossil fuels is not projected to rise significantly, it also does not show a decline, raising concerns about the stagnation in the adoption of cleaner energy sources. The ARIMA 2 model was most accurate because the *pdq* value of (2,1,0) and *PDQ* values of (0,1,0) capture the model best. The *p* value of 2 accounts for the autoregressive tendencies of the data, especially because the data was significantly autocorrelated with itself. Furthermore, the *d* value of 1 removed the linear trend in the data, making it stationary and easier to analyze. The *D* value of 1 captures the periodic trends in the data as well as the short-term noise and variability. The concerning idea here is that the forecast suggests that electricity generated by fossil fuels in 2033 will be 3,130,710 TWh whereas the number generated in 2023 was 3,149,126 TWh. The numbers here highlight the need to create sustainable energy policies.

Discussion

The forecast predicts a stable amount of electricity being generated by fossil fuels in the next ten years. The leveled off forecast suggests that there may be a slow adoption of alternative sources of energy in general. Policies may not be working or passing in the government, and companies may not be pursuing other sources of energy due to that. The economy also still heavily relies on fossil fuels, so the projection that its use stays steady makes sense. forecast suggests that there is an urgent need for policies to be passed. However, it does provide a benchmark that can be used to pursue those policies.

Assuming that the demand is consistent with that trend, policy makers can attempt to create legislation that can encourage a decrease in the amount energy that is generated using oil, gas, etc. If the government and energy companies are aware of a steady trend of electricity demand and generation, then they could start replacing some of the demand with energy generated through other means through policy.

One possible means of legislation is to consider policy that creates a cap within a margin of error around the current demand that could prevent the use of fossil fuels from spiking again, and could encourage energy companies to look for new ways to supply energy to meet demand if it does increase. Moreover, subsidies and tax credits could influence companies to begin developing new green energy solutions. If some of the reliance can be shifted towards other means by companies that America relies on, then the use of fossil fuels could start to trend downwards.

According to the forecast created in the present paper, fossil fuel use is projected to remain steady over next decade. This is significant in that it suggests that the United States may not be headed where it needs to or moving fast enough to reach sustainability goals. In the short term, the policy makers need to use this forecast to start creating action to oppose climate change. In the long term, the bare minimum is to maintain the benchmark provided by this forecast, but the goal should be to create a downward trend. However, while the trend appears static or not declining fast enough, it could be disrupted with more efforts from policy makers and society.