Environmental Impact and Industrial Revenue vs. Energy Generation: A Predictive Analytics Approach On Time Series

Rohan R. Kulkarni, Rojeen B. Farkhoor, Jasmine Cui, Spencer Austin

April 23, 2023

1 Introduction

As the world continues to experience the effects of climate change, there is a growing need to explore the relationship between energy generation needed to sustain communities and the industry's impact on environmental health. This relationship is crucial in any discussion regarding energy policy and planning. Our study contextualizes the environmental and economic outcomes of various sectors of the energy industry in the United States over time. These two factors involve a trade-off: more energy generation means more industrial revenue, but also can result in more detrimental environmental impact, especially for fossil fuels. We investigate per capita energy generation, key environmental health factors (water quality, air quality, drought severity), and per capita revenue from electricity sales monthly time series data from 2001-2022 for certain states in the United States, attempting to illuminate how industrial patterns have affected economic development and/or disturbed local environments. Through this, we hope to be able to answer the following questions, with notable significance in public policy: Did a given U.S. state value industrial output or environmental health more in the past 20 years? How did the state trade-off the two factors? How accurately can the state predict those two factors given their current trade-off strategy? How can they quantify the effect of alternative trade-off strategies in the future?

2 Data Collection and Cleaning

States. In our analysis, we examined the states that have been shown to have the most significant role in the United States' energy sector. According to the U.S. Energy Information Administration, the following six states accounted for over half of the primary energy produced in 2019: Texas, North Dakota, Wyoming, Pennsylvania, Oklahoma, and West Virginia [25]. According to the same source, in 2000, the top six energy-producing states were responsible for 39% of the total primary energy production in the United States, which suggests that there has been a trend towards greater concentration of primary energy production in these states [25]. On the other hand, according to Statista, the following six states led the nation in electricity consumption in 2021: Texas, California, New York, Florida, Ohio, and Pennsylvania [13]. Although these rankings have changed over time, these selected states have consistently ranked in the top 10 of their respective category, and we believe them to be an appropriate sample for past and future trends.

Per Capita Energy Generation by Sector and Revenue from Electricity Sales. We gathered state energy generation monthly data in thousand megawatt-hours from EIA (U.S. Energy Information Administration) database. The sectors we consider are coal, biomass, wood, natural gas, wind, hydroelectric, and nuclear energy generation [21-24]. This time series data will be the input to our predictive model. We also obtained the monthly time series of revenue in millions of dollars from gross sales of electricity by state from the EIA database [21-24]. This will be one factor in our response variable. Population data for each state was collected from the United Census Bureau to generate per capita data. We normalized each month's value of the industrial generation and revenue by the Census-provided yearly estimate of state population [15]. This was our best estimate of per-capita energy generation and revenue.

Water Quality. One aspect of environmental health quality is water quality. It is measured by various factors including pH, temperature, turbidity, salinity, and dissolved oxygen [9,17]. In order to determine the water quality index for various states, we retrieved daily data from the USGS' National Water Information System (NWIS) using the `hydrofunctions` library in Python for each state from 2000 to 2022 [10]. We found that these states all had pH, temperature, and turbidity data while the other factors were not available for every state. We aggregated these daily rows into months by taking the average of each measurement for each month. We created CSVs for all of these factors with all of the monthly values for each state.

Air Quality. Another important measure of environmental health is air quality. Not limited to fossil feuls, according to the U.S. Department of Energy's Office of Policy, "renewable energy systems can have some impacts on air quality during their construction and maintenance phases, such as the release of dust and other particulate matter" [19]. The prevalent measure of air quality is the AQI index which incorporates the main pollutants' level in the air–CO, Ozone, PM10 (particles 10 micrometers or less in diameter), PM2.5, NO2. We downloaded calculated daily AQI data from the EPA for selected states by county [1], then took the daily county average data into the state AQI figures, merged all states to aggregate by month and generated monthly air quality data CSV file by state.

Drought Severity. Droughts can significantly impact the health of the environment in many ways, including devastating crops, reducing food and water, causing seawater intrusion, and contributing to land subsidence [8,18]. To evaluate drought levels, we looked at the Drought Severity and Coverage Index (DSCI) calculated by the U.S. Drought Monitor–developed through partnership by the U.S. Department of Agriculture, the National Oceanic and Atmospheric Administration (NOAA), and the National Drought Mitigation Center at the University of Nebraska-Lincoln [3]. This index is calculated by converting drought levels for each geographical area to one value between 0 (none of the area is "abnormally dry or in drought") and 500 (all of the area is in "exceptional drought"). The raw data consisted of weekly, state-wide DSCI observations from 2000 to 2023. To clean the data, we averaged the weekly data to yield monthly values ranging from 2001 to 2022 for the states that we considered. Lastly, we scaled the DSCI values to range between 0 and 100.

3 Exploratory Data Analysis

Time Series Visualization. We plotted the input data to our model, the energy generation data for each sector, gross and per capita (see a coal generation plot in Figure 1). We can clearly see a difference when the energy generation when weighted per capita. High populated states like California and Texas will naturally have higher energy generation, electricity revenue, and climate impact simply because they service more people. In order to fairly measure the impact by state, we decided to proceed with per capita generation in our analysis. All gross and per capita energy generation plots are shown in Appendix B.

STL Decomposition. To understand the nature of the time series data, we plotted an STL (Seasonal and Trend using Loess) decomposition of that data (see the coal generation example extended in Figure 2) [5]. We hypothesized a 12-month seasonality (over 12 consecutive datapoints) as this seemed intuitive with calendar year patterns seen in the time series plot (Figure 1). The results showed a high seasonality term and clear trend in the energy generation over time (coal generation going down), which implies that the time series data is not stationary [5]. Such STL decompositions with high seasonality were observed for all of the time series data (see more examples in Appendix C). This means that the average moves as time progresses, and leads to the conclusion that averaging and smoothing time series forecasting techniques (i.e. moving averages, exponential smoothing, etc.) are preferable to autoregressive techniques (i.e. ARIMA) [2,5,11]. We decided to implement a multiple linear regression model and a rolling windows regression model.

4 Methodology

After collecting, cleaning, and conducting EDA on the data, we describe the model architecture and perform preprocessing. The predictive model makes time series forecasts for a given state. The input to the

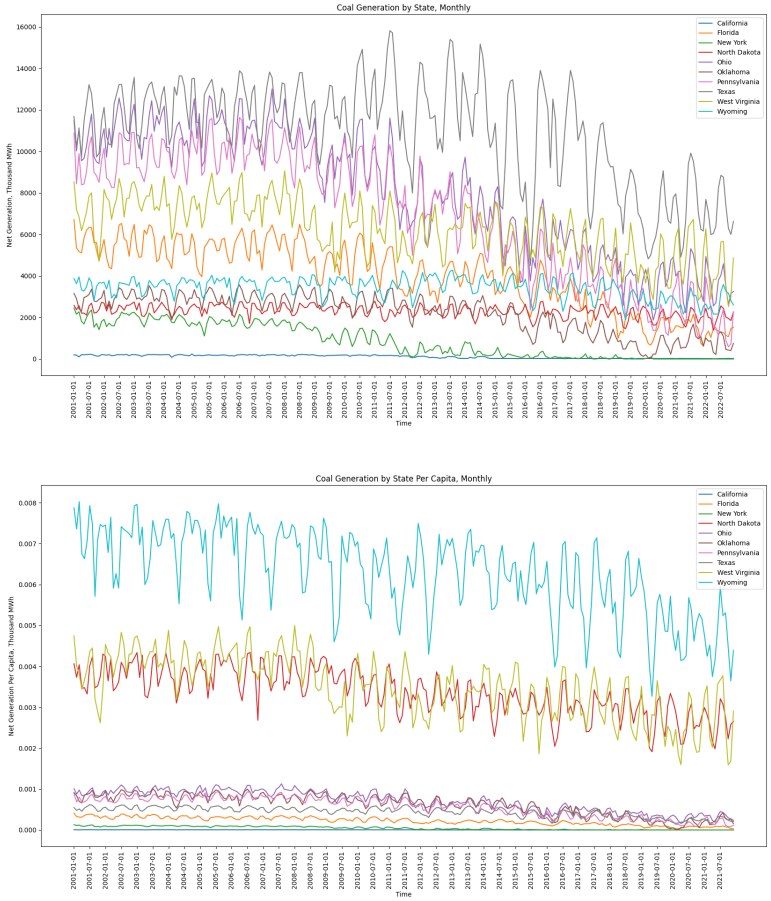


Figure 1: Coal Generation, Gross vs. Per Capita

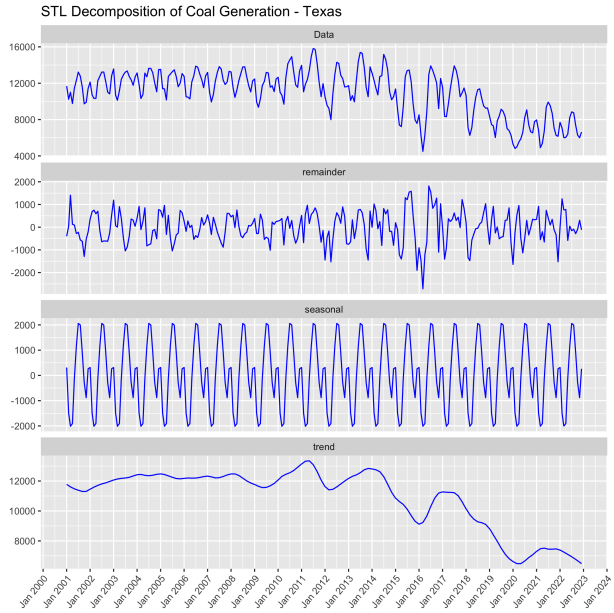


Figure 2: Texas Coal Generation STL Decomposition

model is the state's monthly energy generation in the considered sectors, and the output must be some combination of the economic and environmental impact. We define this response as a weighted combination of the per capita retail sales of electricity due to energy generation and a Environmental Health Index (EHI).

Environmental Health Index (EHI). There are many ways to quantify the health of a local environment, and many times the factors considered is application dependent. We consider three factors: the water quality, air quality, and drought severity index. To build an EHI from these factors, we use a min-max scaler to scale each of the factors, take a weighted combination of them, and then use the min-max scaler once more to obtain a score between 0 and 100.

$$EHI \sim w_1 \times (scaled.water.quality) + w_2 \times (scaled.air.quality) + w_3 \times (scaled.drought.severity)$$

The weights of the combination of the factors is a user-defined parameter; in this analysis we use a equal weighting of the factors $w_1 = w_2 = w_3 = \frac{1}{3}$. The EHI is thus constructed as a monthly time series for each state over the entire time range.

Response Variable. Now that we have quantified an EHI, we need to somehow express the trade-off between the economic output and environmental impact. This is a convex combination of the state's per capita revenue from retail sales of electricity and its EHI time series:

$$Response \sim \lambda \times (energy.revenue) + (1 - \lambda) \times (EHI)$$

where λ is what we define as the state's "care parameter". This parameter demonstrates how much the state cares about the economic output of its energy industry and the environmental impact of the same. A λ value closer to 1 reflects a state energy policy that priorities revenue from energy generation and a λ closer to 0 reflects a policy that focuses on environmental impact. This convex combination is a monthly time series that is the response or output of our regressive model. The response variable was shifted up by 1 month during preprocessing to avoid data leakage. In other words, this month's energy generation input would be used to predict next month's combination of revenue from electricity sales and impact on environmental health index.

Multiple Linear Regression Model. The first model we utilized is a standard linear regression of multiple input variables: the energy generation time series for each industrial sector (i.e. coal, biomass, wind, etc.). The model would be run on every state independently. However, the obvious question was, for a given state, what care parameter should be chosen to generate the appropriate response? The care parameter is a user-defined hyperparameter, but we choose the the most suitable care parameter for a given state using time series cross validation [4,12]. Time series cross validation works slightly differently than k-fold cross validation as there is the issue of data leakage; we cannot move the validation fold as we cannot train on future data points to predict the past. Instead, the training set expands by a fixed sliding window size in each iteration, and the validation set is pushed forward. A prediction is made using the linear model on the validation set, and mean squared error for that "fold" is saved. Then, just as in k-fold cross validation, the optimal setting of the care parameter is that which minimizes the average error over all the "folds" [4,12]. First, we split the data into a training set (50%), a validation set (25%), and a test set (25%). We used time series cross validation with a sliding window of 3 months (data points) to evaluate each parameter, and the best one was chosen. Note that this care parameter may differ for each state depending on their energy industry's patterns. Then, the linear model was retrained on the entire training and validation set (75%) against a response variable made using the optimal choice of the care parameter, and evaluated out-of-sample (OOS) on the test set (25%). The results from the OOS are shown in the Results section.

Rolling Window Regression Model. We know from the STL decompositions that the time series is not stationary and so a moving average prediction approach may yield good results. Hence, we also implement a rolling window model, which is also a linear regression of the same energy generation that outputs the response variable but the data is smoothened by taking a moving average of the time series data. We fix window size of 3 months (data points) and find the mean of the time series data in that window. Then the window is "rolled" to the next three observations. This rolling windows calculation is done for all of the data, input and response, as they are all nonstationary time series. Then, we conduct a linear regression on

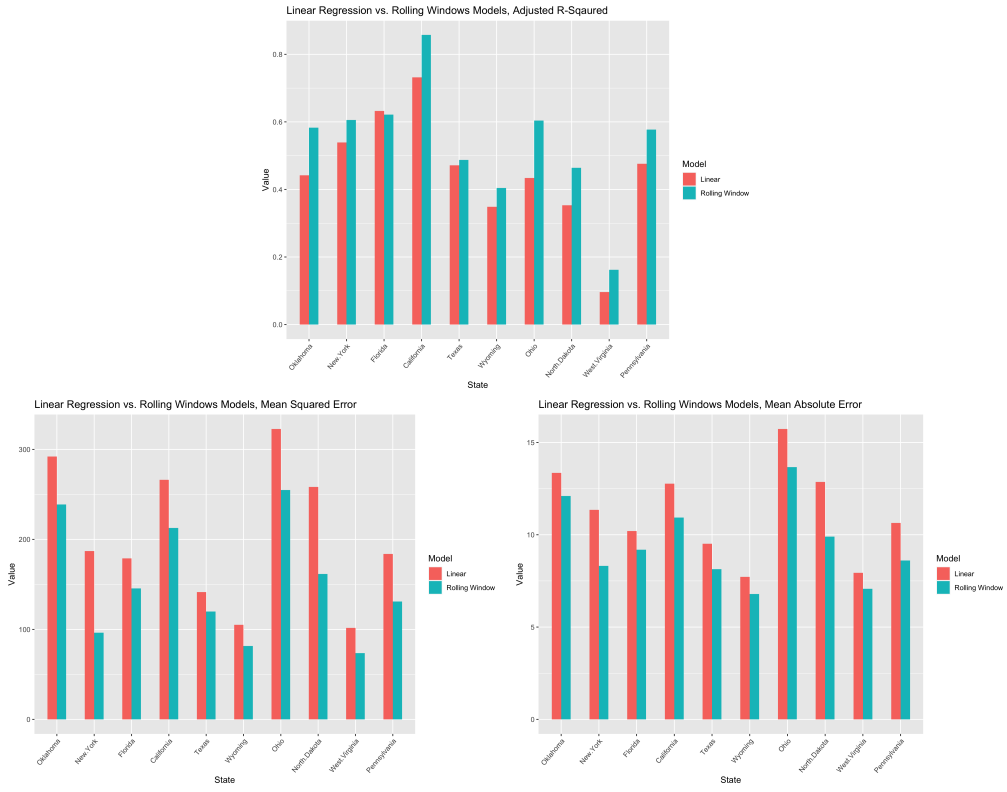


Figure 3: Training Adjusted R^2, Mean Absolute Error, Mean Squared Error

the data smoothened by moving average on rolling windows. The same time series cross validation approach is used to find the best choice for the care parameter for a given state. Once again, the model is tested out of sample on a separate test set.

5 Results

We compare the linear regression and rolling windows regression models for each of the states in Figure 3. It is clear that the rolling windows approach indeed does do better than the multiple linear regression, with lower errors and a higher adjusted R^2. The R^2 results indicate that for the most part the model is able to capture the behavior of how the energy generation relates to the response variable, and the mean absolute error shows that predictions for states' response variables were anywhere between 5 and 17 units off on average. We can actually visualize the prediction versus the actual response time series, as shown for Texas in Figure 4.

However, we can also conduct some basic prescriptive analytics if we want to shed light on what could have happened if state energy policy had dictated a different care parameter value. This is done by reconstructing the response variable with that user-defined care parameter. We chose 0.1, indicating a highly environmentally-conscious energy policy. Plotting this as well, as shown for Texas in Figure 4 we can see the prediction, the actual response, and the desired response if the care parameter is 1. We can pinpoint time periods where the actual response underachieves or overachieves the desired response. Since the response is mostly dependent on the EHI, we can assert those regions as being unaligned and aligned, respectively, with the environmentally-conscious policy. All predictive/prescriptive visualizations are shown in Appendix A.

The outputted results of our better model, the rolling windows regression, are shown in Figure 5. The additional information here are the top 3 lowest p-valued features for each state's regression, which can be indicative of which industrial sector's energy generation was most "important" towards making the prediction. The best care parameter chosen for each state during time series cross validation is also given.

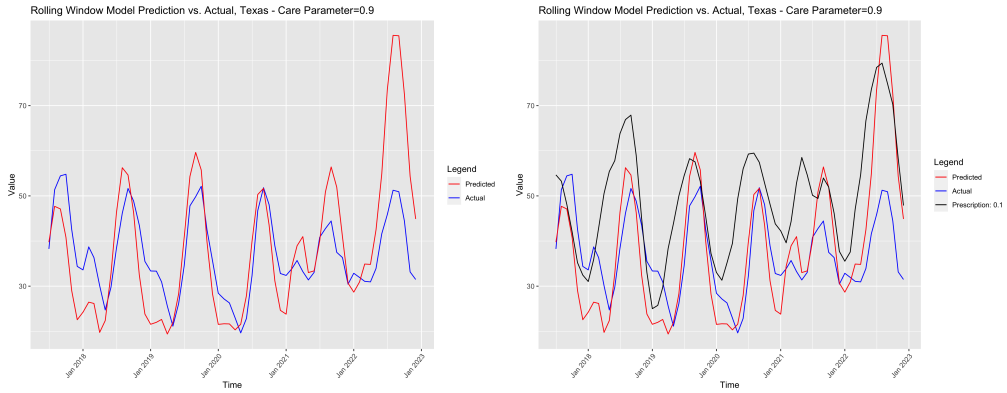


Figure 4: Predicted, Actual, and Desired Response, Texas

res_windows_model

		Oklahoma	New.York	Florida	California	Texas
	MAE	12.1012424656019	8.31442209685132	9.19131476626386	10.933581954838	8.1362119950079
	MSE	238.867065640465	96.2045571677932	145.460588310017	212.638273588707	119.914054194136
	best.careparam	0.8	0.6	0.7	1	0.9
	adj.r.sq	0.582696557916593	0.605683501284896	0.621382744704956	0.857469607108342	0.487601098481078
	t3.pvals	c("naturalgas", "coal", "hydroelectric")	c("naturalgas", "hydroelectric", "wood")	c("naturalgas", "coal", "nuclear")	c("wind", "naturalgas", "hydroelectric")	c("naturalgas", "coal", "wind")
		Wyoming	**Ohio**	**North.Dakota**	**West.Virginia**	**Pennsylvania**
	MAE	6.7903914791023	13.6673563836729	9.89874726312191	7.07554346651387	8.60591456741141
	MSE	81.5979268982199	254.776479721587	161.625473681239	73.7176899384744	131.00954319957
	best.careparam	0.5	0.9	0.4	0.5	0.6
	adj.r.sq	0.40416589800955	0.60414565822775	0.46399866021015	0.161827638256745	0.577068797671948
	t3.pvals	c("wind", "hydroelectric", "naturalgas")	c("biomass", "naturalgas", "coal")	c("coal", "hydroelectric", NA)	c("coal", "naturalgas", "hydroelectric")	c("coal", "naturalgas", "biomass")

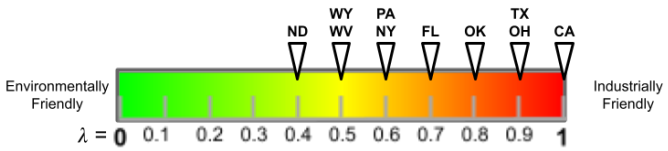


Figure 5: Model Results Output and Care Parameter Comparison

Visualizing these, we can compare across states to see trade-offs of revenue and environmental impact due to the energy industry.

6 Conclusion

Our study findings reveal that different states have placed varying levels of importance on industrial revenue and environmental impact. Some have prioritized one over the other while others have achieved a more even balance between the two factors. With our model, states will be able to use the trade-off they have determined to predict the sales of electricity and environmental health (water quality, air quality, and drought severity/coverage) and see if it aligns with what they hope their policies will achieve. Then, they will be able to decide how they would want to change their trade-off strategy to achieve their goals.

References

1. "Air Data - Multiyear Tile Plot." Environmental Protection Agency, 9 February 2023, https://www.epa.gov/outdoor-air-quality-data/air-data-multiyear-tile-plot.

2. Bajaj, Aayush. "Time Series Prediction: How Is It Different From Other Machine Learning? [ML Engineer Explains]." Neptune.ai, 19 April 2023, https://neptune.ai/blog/time-series-prediction-vs-machine-learning.

3. Drought Monitor, University of Nebraska-Lincoln, U.S. Department of Agriculture, and National Oceanic and Atmospheric Administration (NOAA). "Drought Severity and Coverage Index." U.S. Drought Monitor, 2023, https://droughtmonitor.unl.edu/DmData/DataDownload/DSCI.aspx.

4. "5.10 Time series cross-validation." Forecasting: Principles and Practice, OTexts, https://otexts.com/fpp3/tscv.html.

5. Hartmann, K., et al. "STL decomposition." E-Learning Project SOGA: Statistics and Geospatial Data Analysis, Department of Earth Sciences, Freie Universitaet Berlin, 2018, https://www.geo.fu-berlin.de/en/v/soga/Geodata-analysis/time-series-analysis/Seasonal-decompositon/.

6. "Index of /programs-surveys/popest/datasets." Census.gov, https://www2.census.gov/programs-surveys/popest/datasets/.

7. National Drought Mitigation Center at the University of Nebraska-Lincoln, et al. "Drought Severity and Coverage Index — U.S. Drought Monitor." U.S. Drought Monitor, 2023, https://droughtmonitor.unl.edu/About/AbouttheData/DSCI.aspx.

8. National Integrated Drought Information System. "Drought Impacts." Drought.gov, https://www.drought.gov/impacts.

9. National Marine Sanctuaries, and National Oceanic and Atmospheric Administration. "What is water quality?" Florida Keys National Marine Sanctuary, https://floridakeys.noaa.gov/ocean/waterquality.html.

10. Roberge, Martin. "Writing Valid Requests for NWIS." Hydrofunctions 0.2.3 documentation, https://hydrofunctions.readthedocs.io/en/master/notebooks/Writing_Valid_Requests_for_NWIS.html. Accessed 23 April 2023.

11. Ruiz, Pablo. "ML Approaches for Time Series. In this post I play around with some... — by Pablo Ruiz." Towards Data Science, 19 May 2019, https://towardsdatascience.com/ml-approaches-for-time-series-4d44722e48fe.

12. Shrivastava, Soumya. "Cross Validation in Time Series. Cross Validation: — by Soumya Shrivastava." Medium, 14 January 2020, https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4.

13. Statista. "U.S. electricity consumption by leading state." Statista, 25 January 2023, https://www.statista.com/statistics/560913/us-retail-electricity-consumption-by-major-state/.

14. Statista Research Department. "U.S. electricity consumption by leading state." Statista, 25 January 2023, https://www.statista.com/statistics/560913/us-retail-electricity-consumption-by-major-state/.

15. United States Census Bureau. Census.gov, https://www2.census.gov/programs-surveys/popest/datasets/.

16. United States Environmental Protection Agency. "What is Being Done?" Acid Rain Students Site, https://www3.epa.gov/acidrain/education/site_students/beingdone.html.

17. United States Geological Survey (USGS). "USGS Water-Quality Data for the Nation." U.S. Geological Survey, 23 April 2023, https://waterdata.usgs.gov/nwis/qw.

18. U.S. Climate Resilience Toolkit. "Drought." U.S. Climate Resilience Toolkit, 28 September 2021, https://toolkit.climate.gov/topics/water/drought.

19. U.S. Department of Energy, Office of Policy. "2021 U.S. Energy and Employment Report." Energy.gov, U.S. Department of Energy, 2021, https://www.energy.gov/policy/2021-us-energy-and-employment-report.

20. "Our World in Data." Energy Production and Consumption. Oxford Martin Programme on Integrating Renewable Energy, University of Oxford, 2021, https://ourworldindata.org/energy-production-consumption.

21. U.S. Energy Information Administration. "Coal Data Browser." EIA, https://www.eia.gov/coal/data/browser/.

22. U.S. Energy Information Administration. "Crude Oil Production." EIA, 3 31 2023, https://www.eia.gov/dnav/pet/pet_crd_crpdn_adc_mbbl_m.htm.

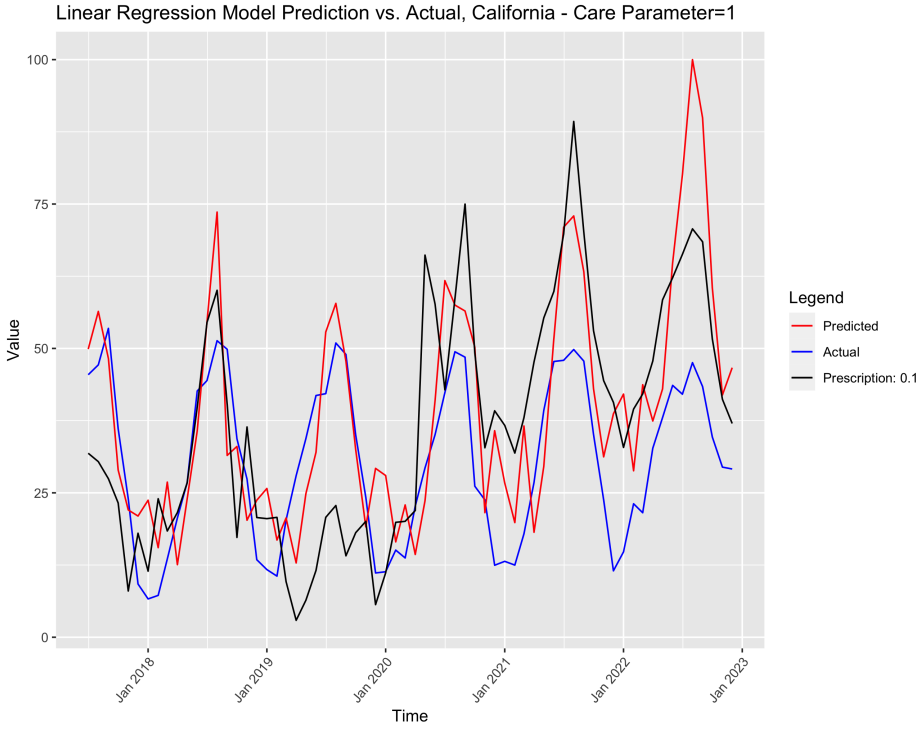
23. U.S. Energy Information Administration. "Electricity explained: Electricity generation, capacity, and sales in the United States." EIA, https://www.eia.gov/energyexplained/electricity/electricity-in-the-us-generation-capacity-and-sales.php.

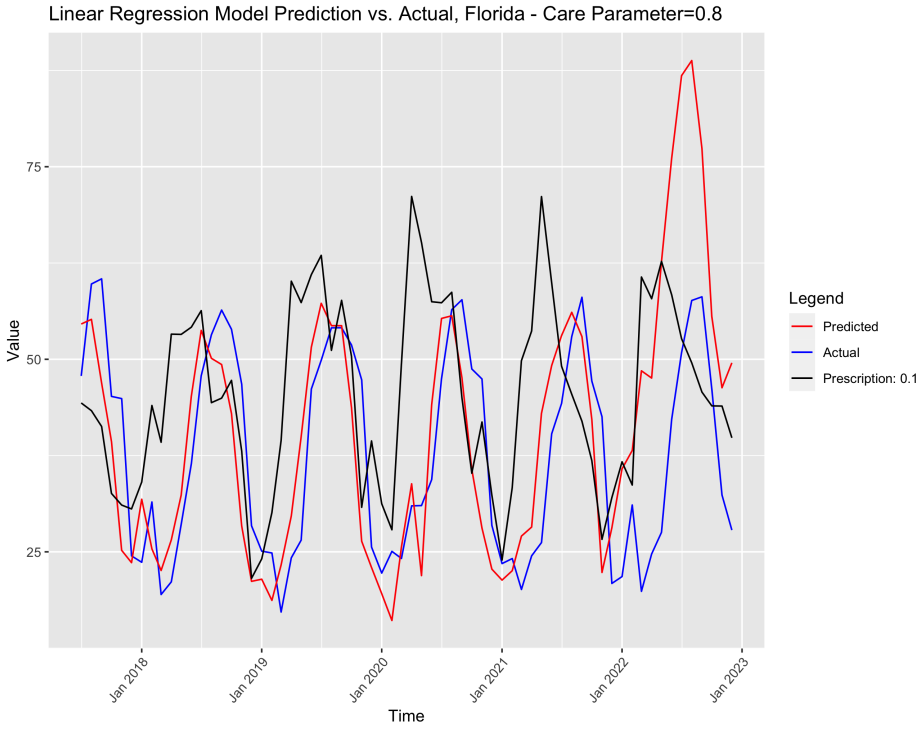
24. U.S. Energy Information Administration. "Gasoline and Diesel Fuel Update." EIA, https://www.eia.gov/petroleum/gasdiesel/.

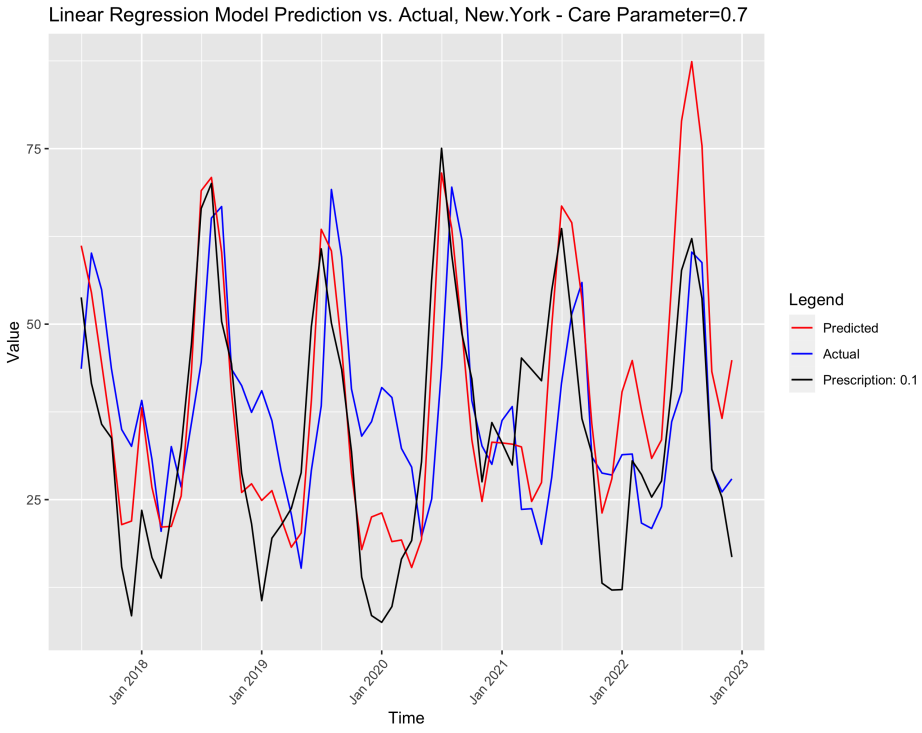
25. U.S. Energy Information Administration. "Six U.S. states accounted for over half of the primary energy produced in 2019." EIA, 31 August 2021, https://www.eia.gov/todayinenergy/detail.php?id=49356.

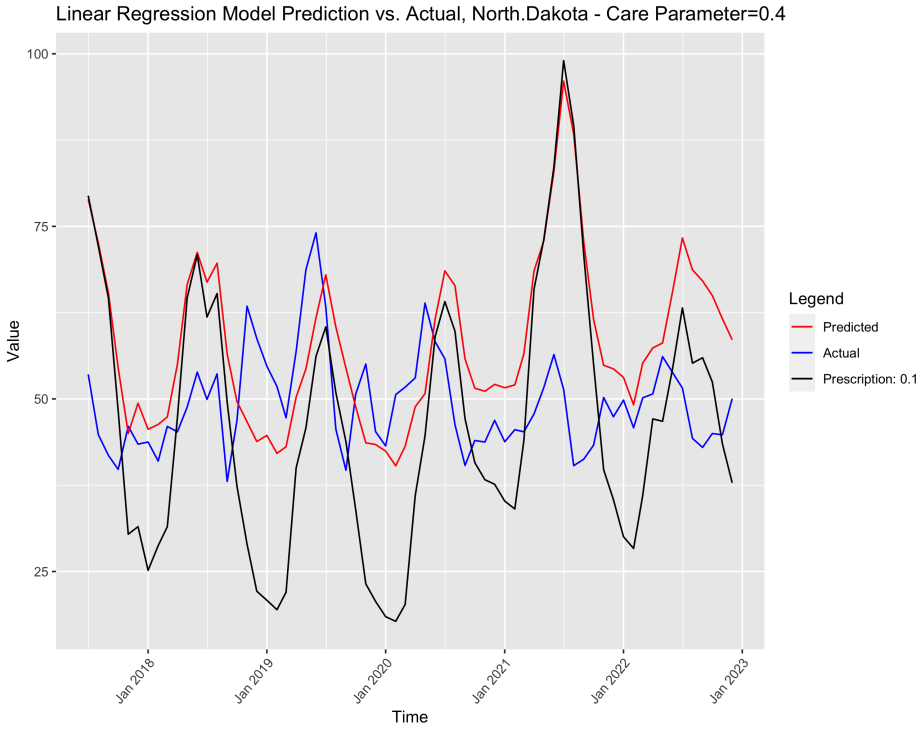
Appendix A - Model Results

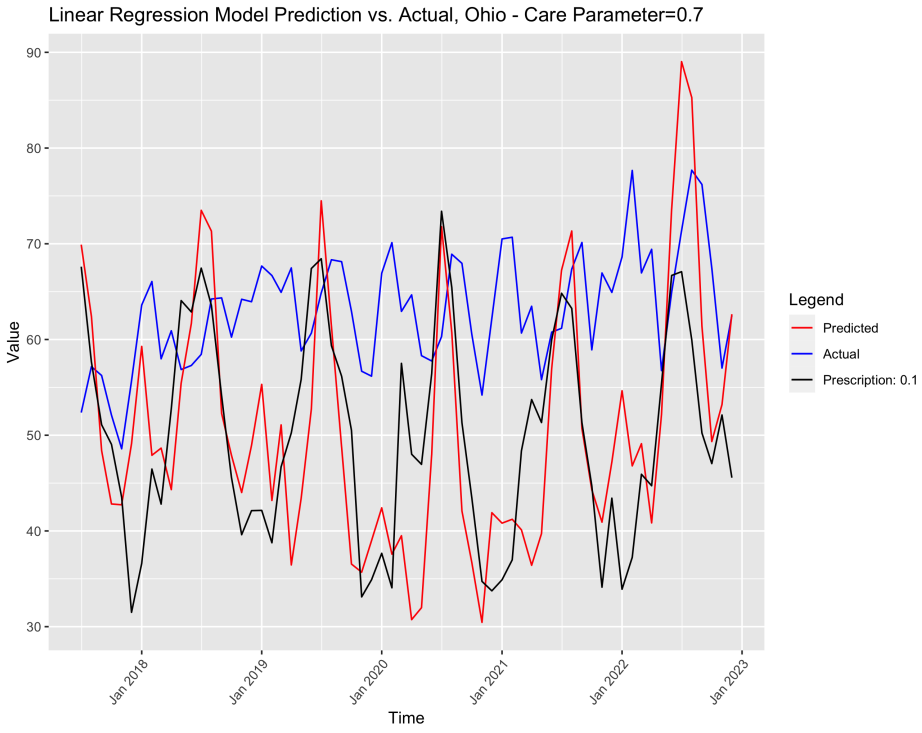
Linear Regression Model:
Predictive and Prescriptive Results with Care Parameter=0.1

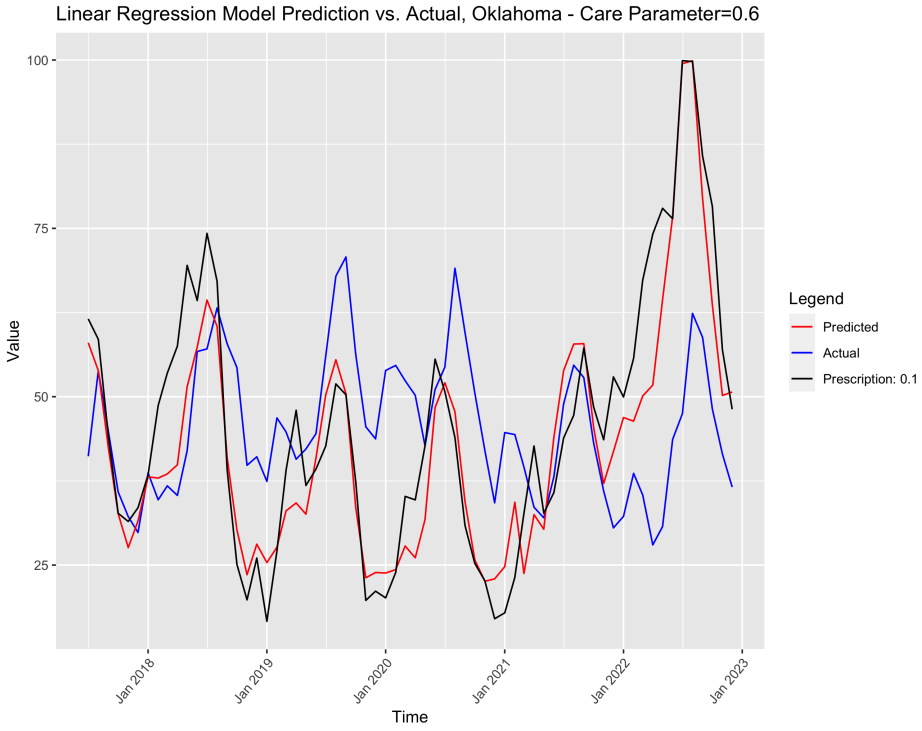


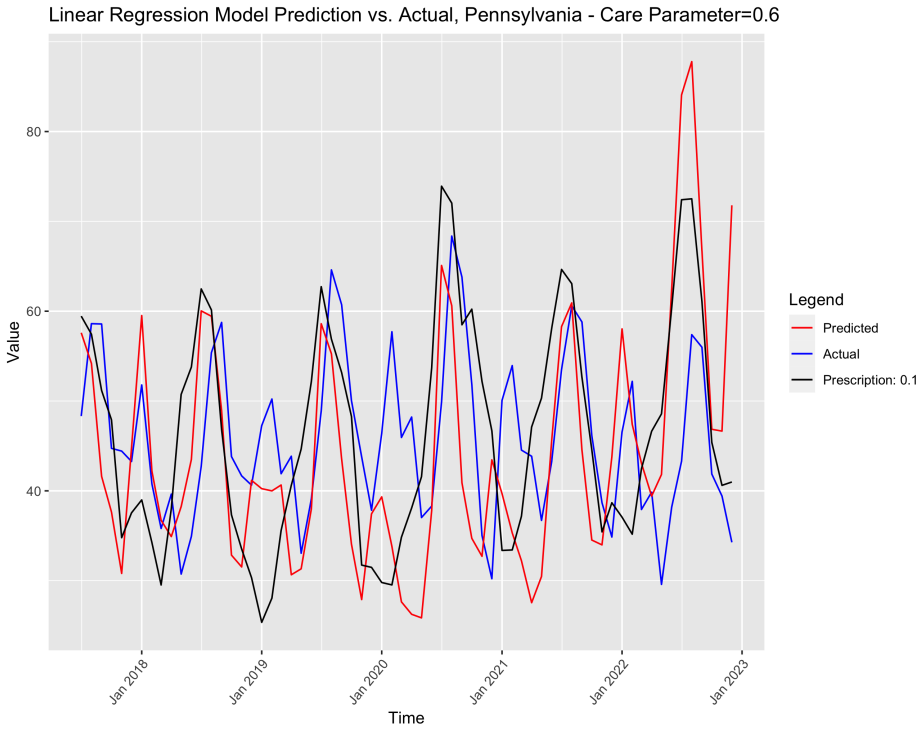


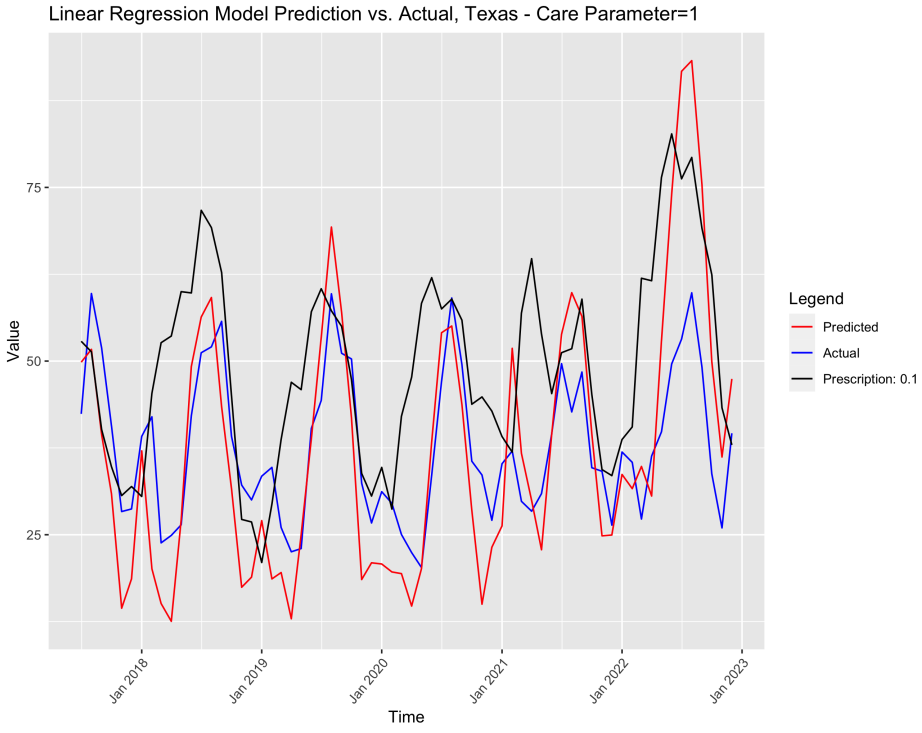

Linear Regression Model Prediction vs. Actual, New.York - Care Parameter=0.7
Value
75
50
25
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

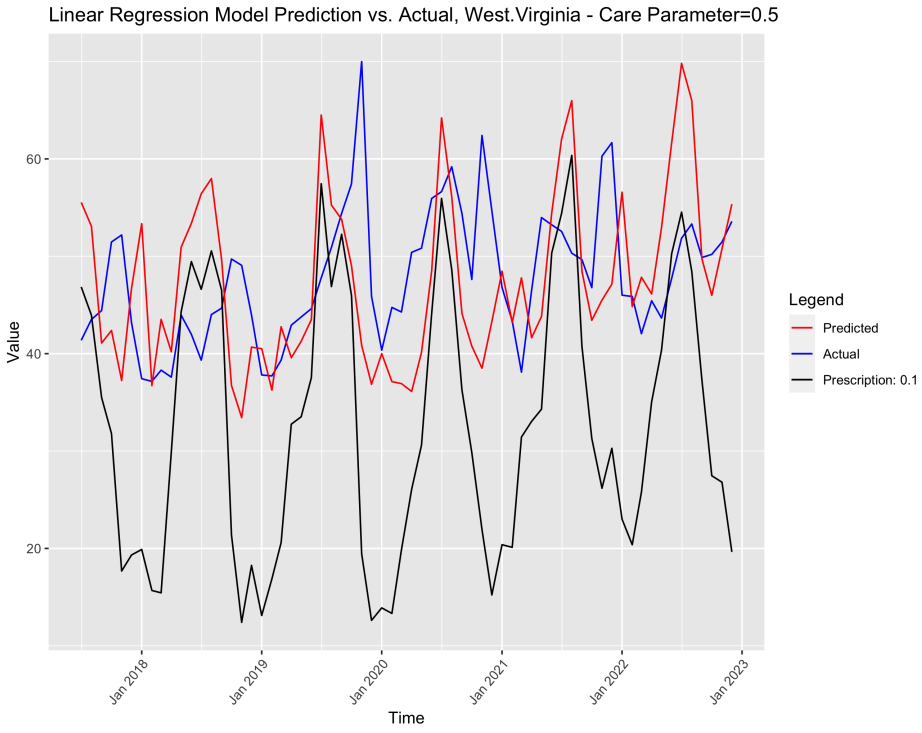

Linear Regression Model Prediction vs. Actual, North.Dakota - Care Parameter=0.4
Value
100
75
50
25
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

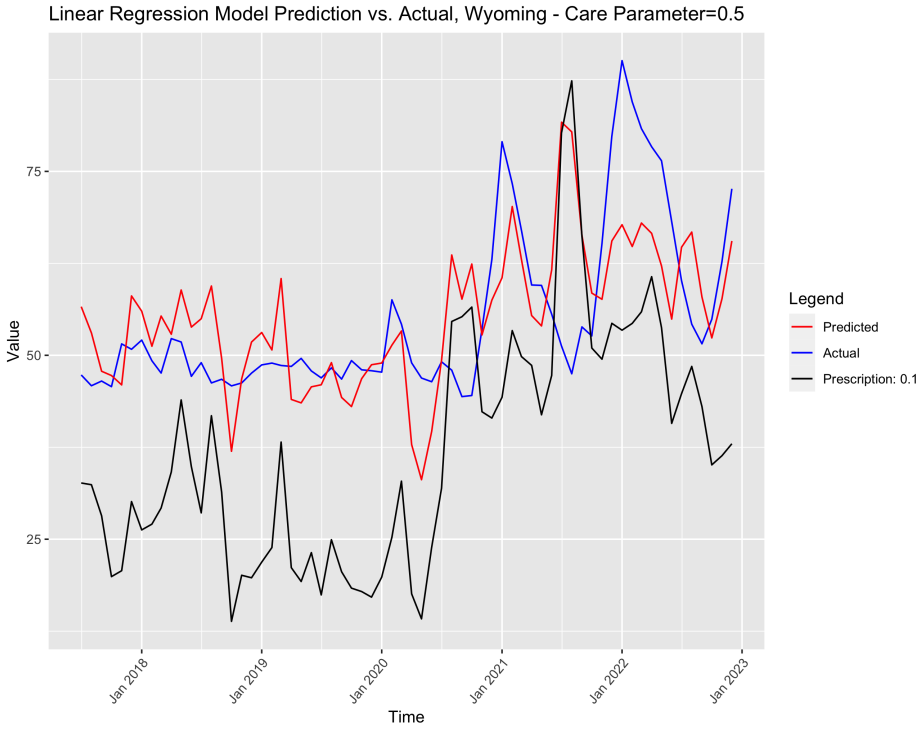

Linear Regression Model Prediction vs. Actual, Ohio - Care Parameter=0.7
Value
Time
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Legend
Predicted
Actual
Prescription: 0.1


Linear Regression Model Prediction vs. Actual, Oklahoma - Care Parameter=0.6
Value
Time
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Legend
Predicted
Actual
Prescription: 0.1

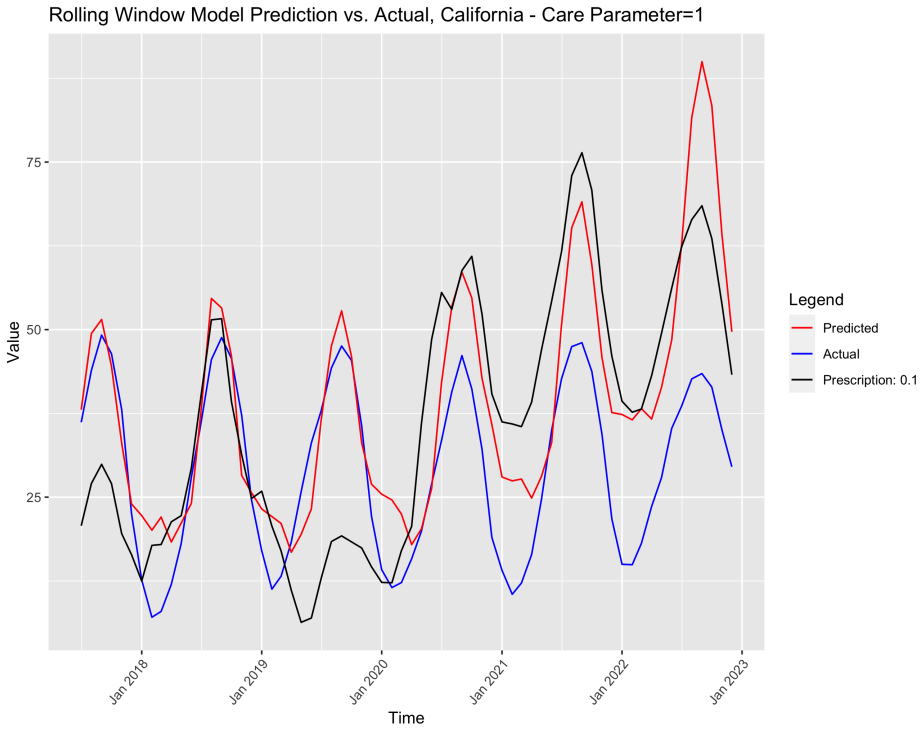

Linear Regression Model Prediction vs. Actual, Pennsylvania - Care Parameter=0.6
Value
80
60
40
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

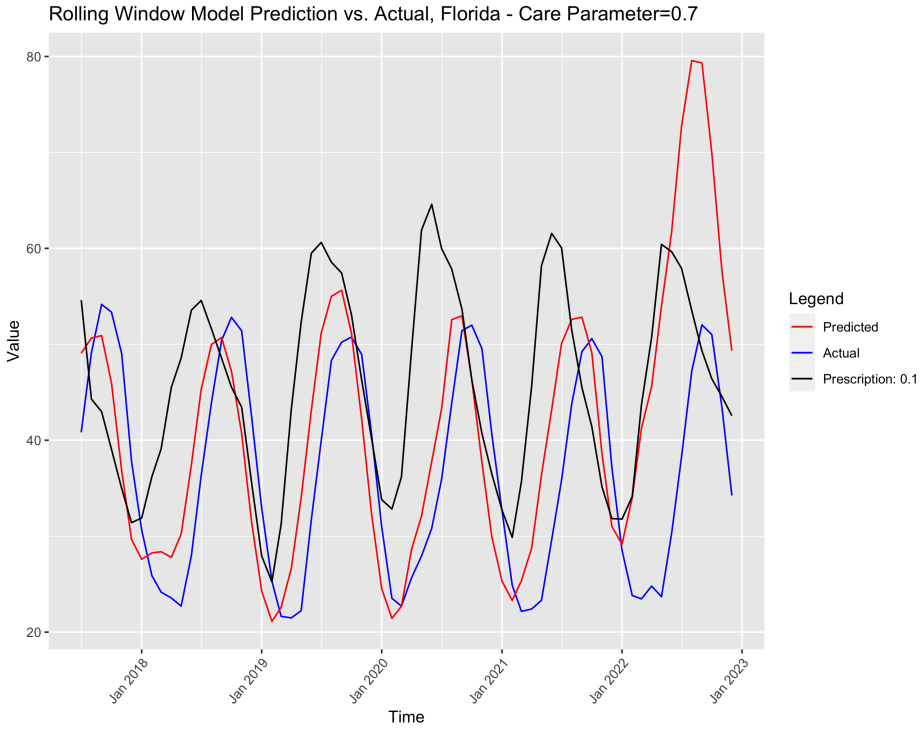

Linear Regression Model Prediction vs. Actual, Texas - Care Parameter=1
Value
75
50
25
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

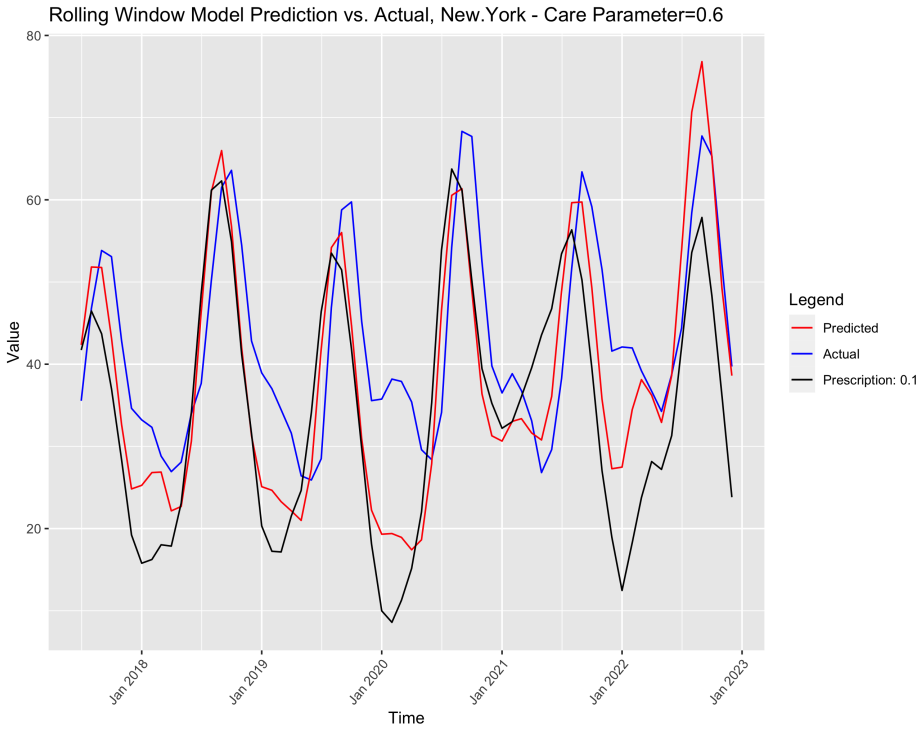


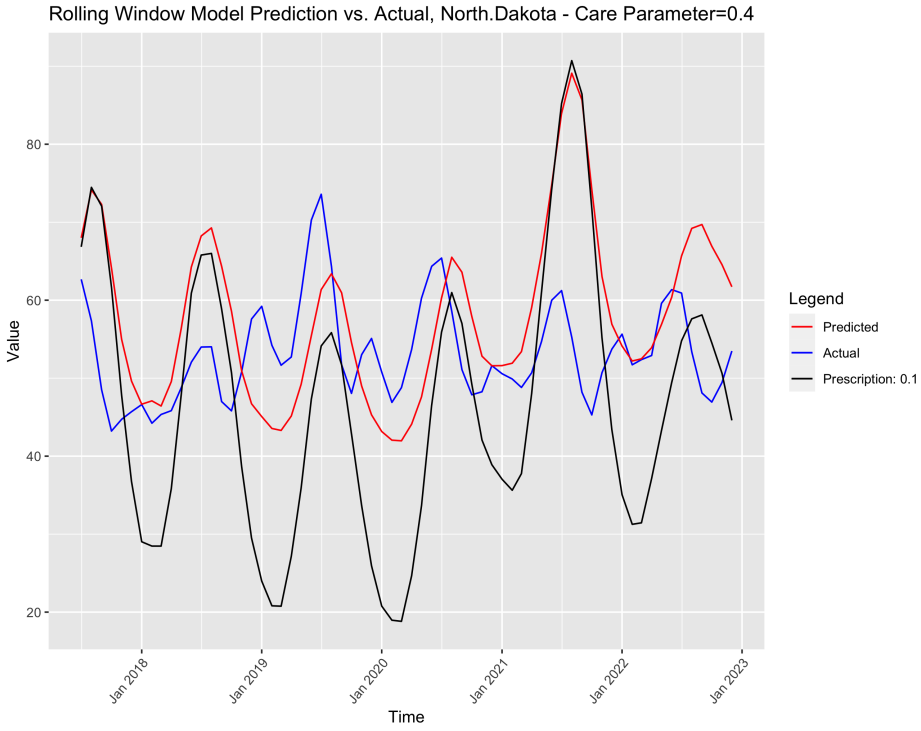


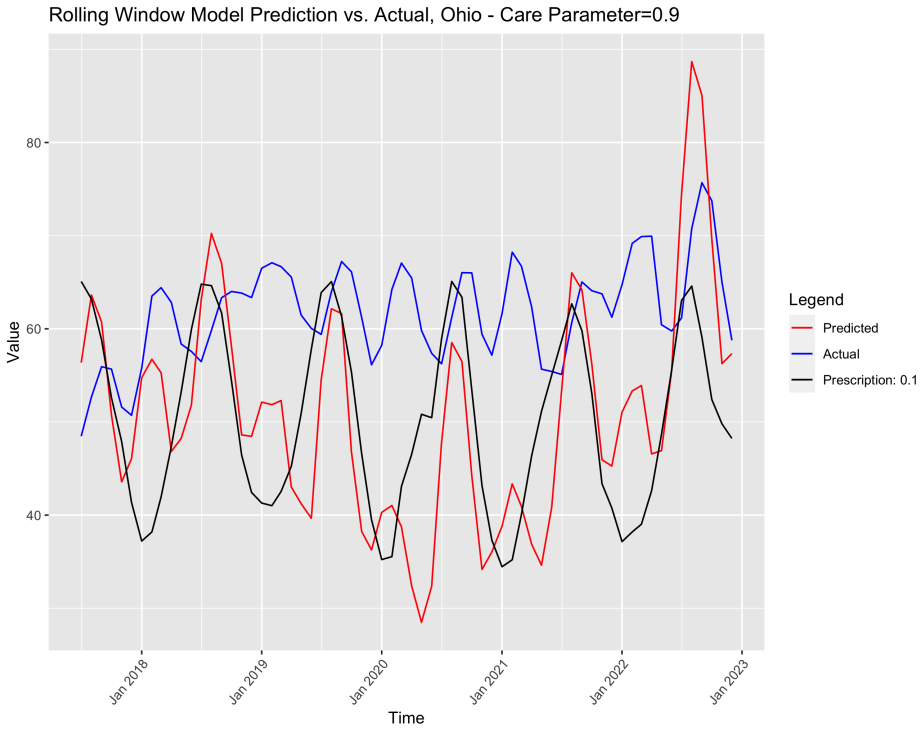
Rolling Window Model:
Predictive and Prescriptive Results with Care Parameter=0.1

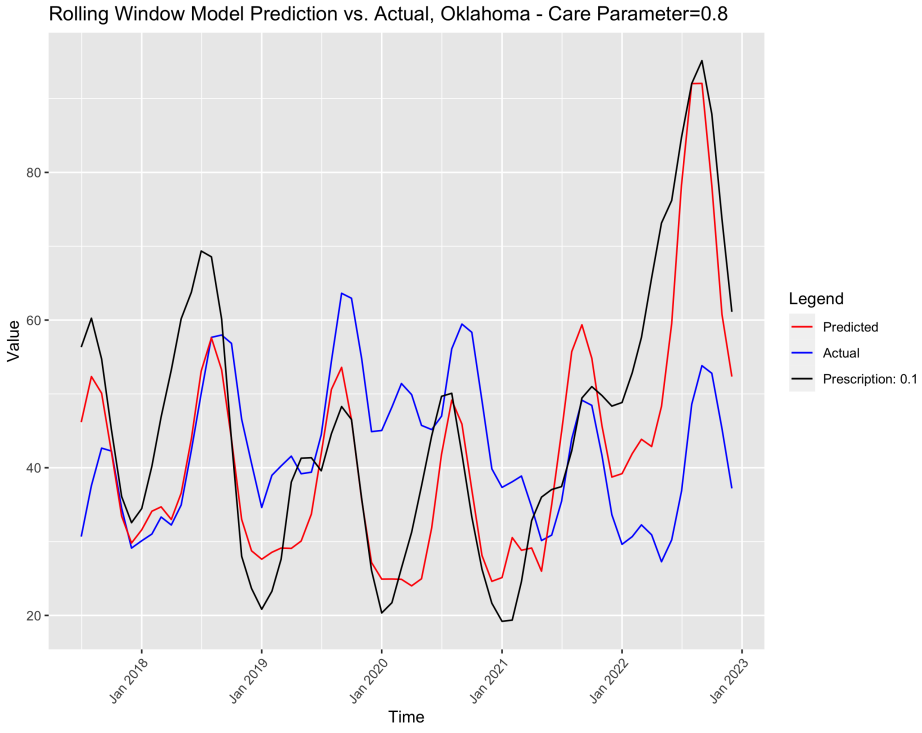

Rolling Window Model Prediction vs. Actual, California - Care Parameter=1
Value
75
50
25
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

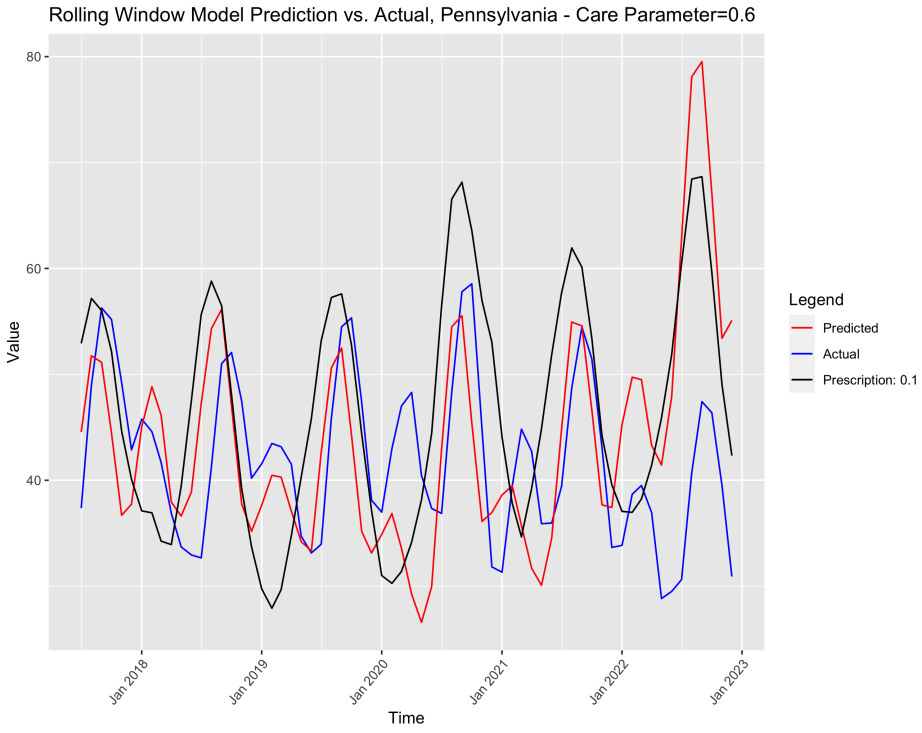

Rolling Window Model Prediction vs. Actual, Florida - Care Parameter=0.7
Value
80
60
40
20
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

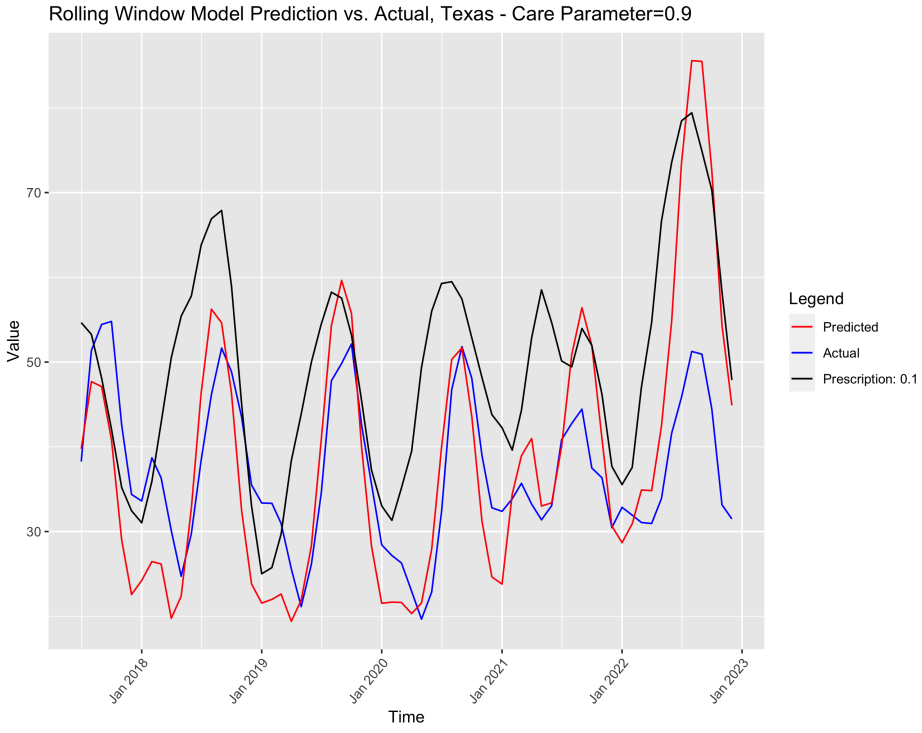

Rolling Window Model Prediction vs. Actual, New.York - Care Parameter=0.6
Value
Time
Legend
Predicted
Actual
Prescription: 0.1
80
60
40
20
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023

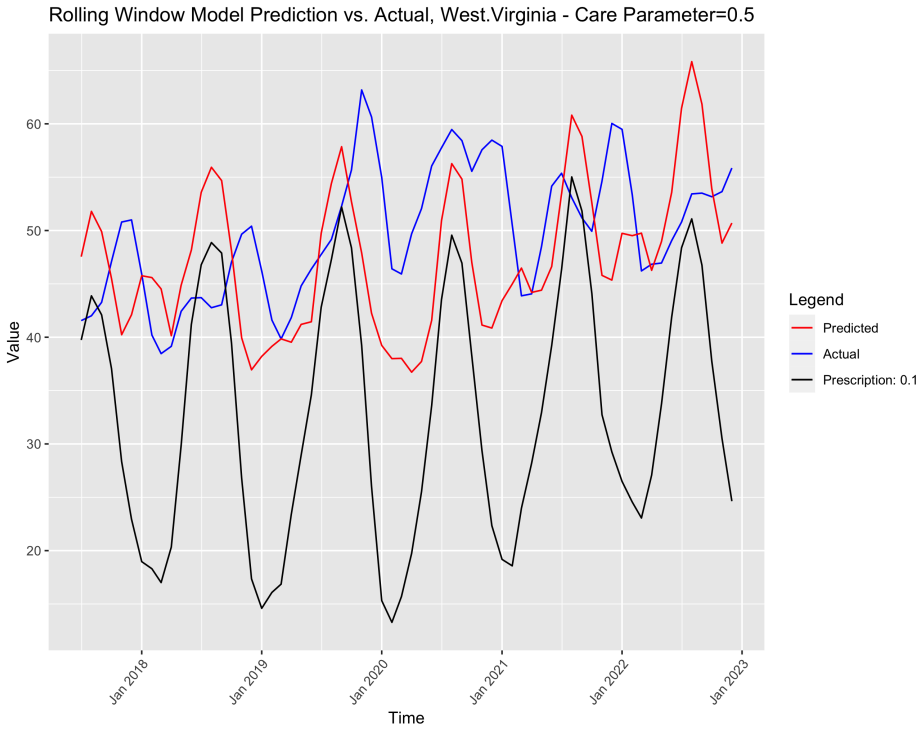

Rolling Window Model Prediction vs. Actual, North.Dakota - Care Parameter=0.4
Value
Time
Legend
Predicted
Actual
Prescription: 0.1
80
60
40
20
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023

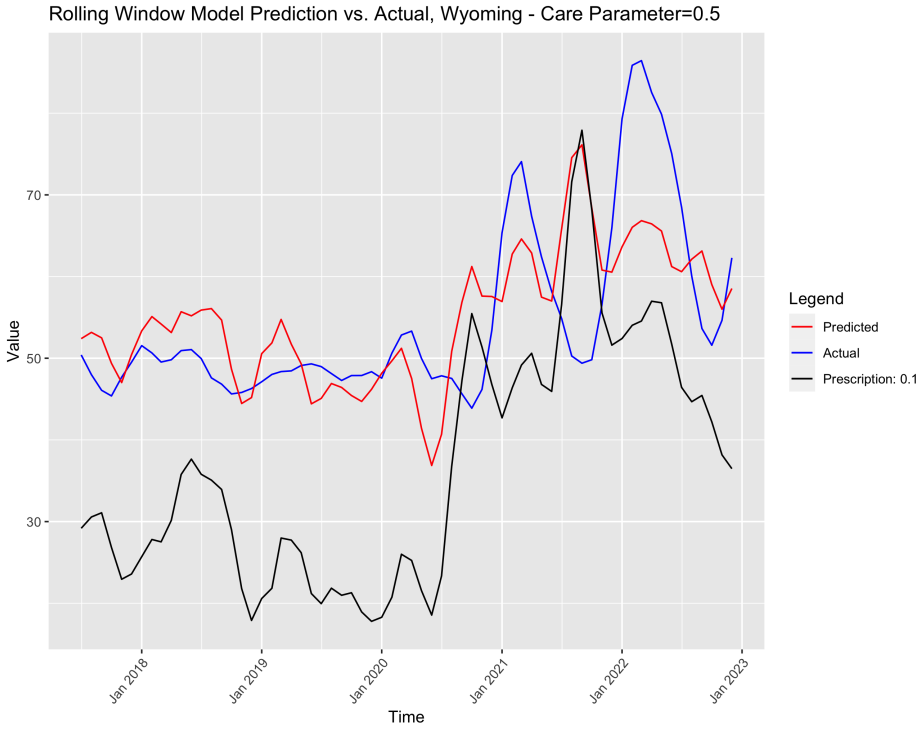

Rolling Window Model Prediction vs. Actual, Ohio - Care Parameter=0.9
Value
80
60
40
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1


Rolling Window Model Prediction vs. Actual, Oklahoma - Care Parameter=0.8
Value
80
60
40
20
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1


Rolling Window Model Prediction vs. Actual, Pennsylvania - Care Parameter=0.6
80
60
40
Value
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

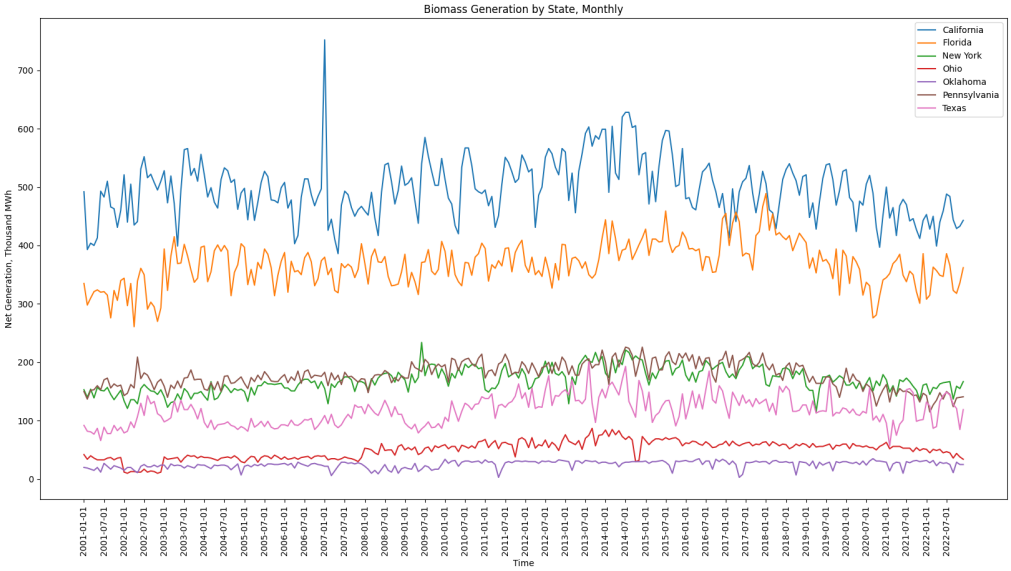

Rolling Window Model Prediction vs. Actual, Texas - Care Parameter=0.9
70
50
30
Value
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

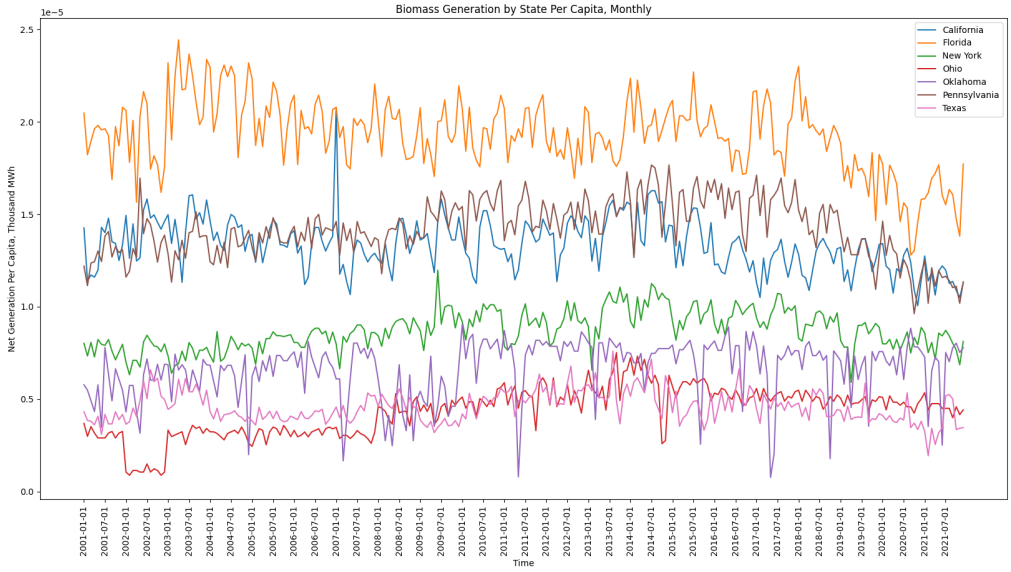

Rolling Window Model Prediction vs. Actual, West.Virginia - Care Parameter=0.5
Value
60
50
40
30
20
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

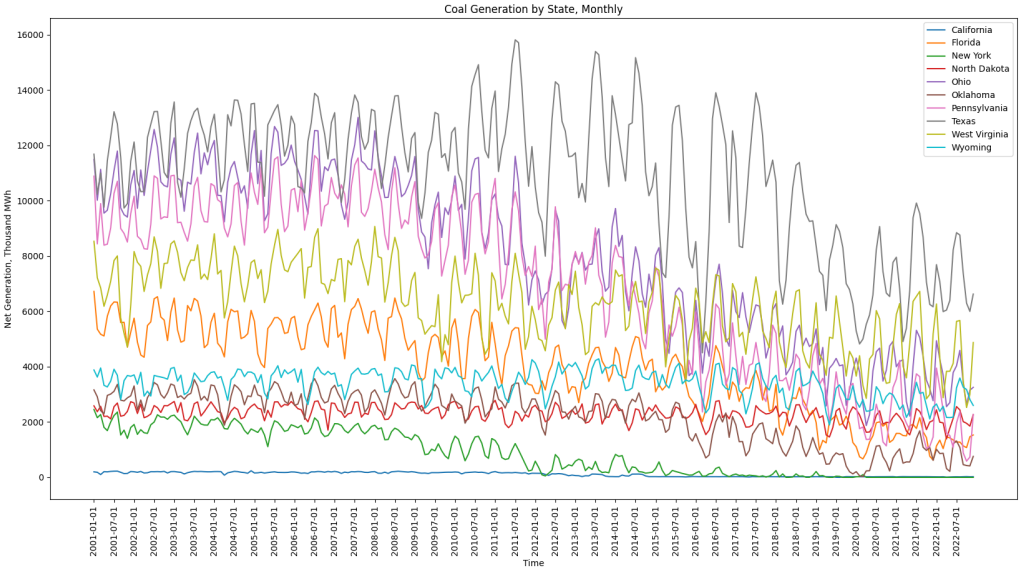

Rolling Window Model Prediction vs. Actual, Wyoming - Care Parameter=0.5
Value
70
50
30
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Time
Legend
Predicted
Actual
Prescription: 0.1

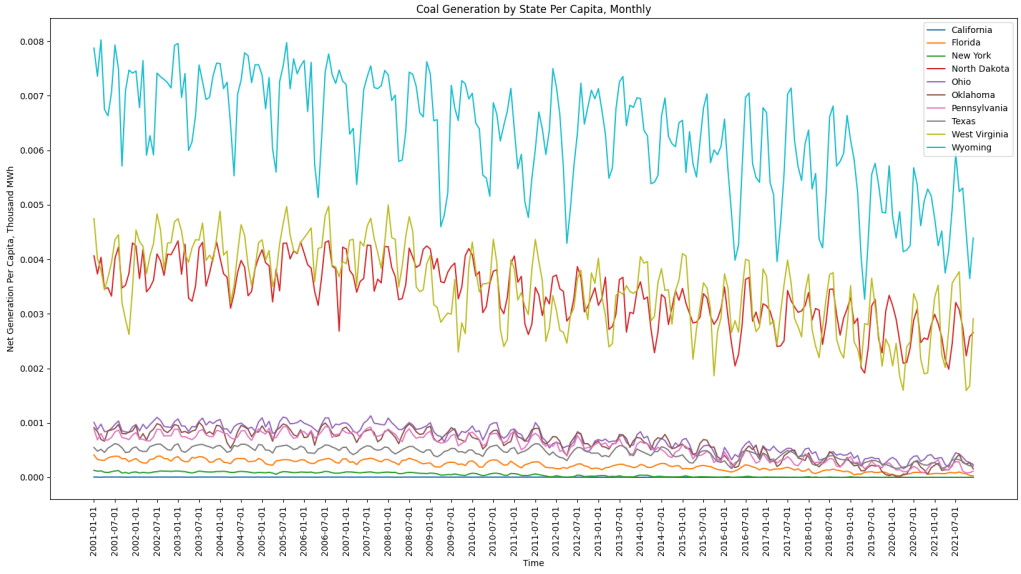
Appendix B - Energy Generation Plots

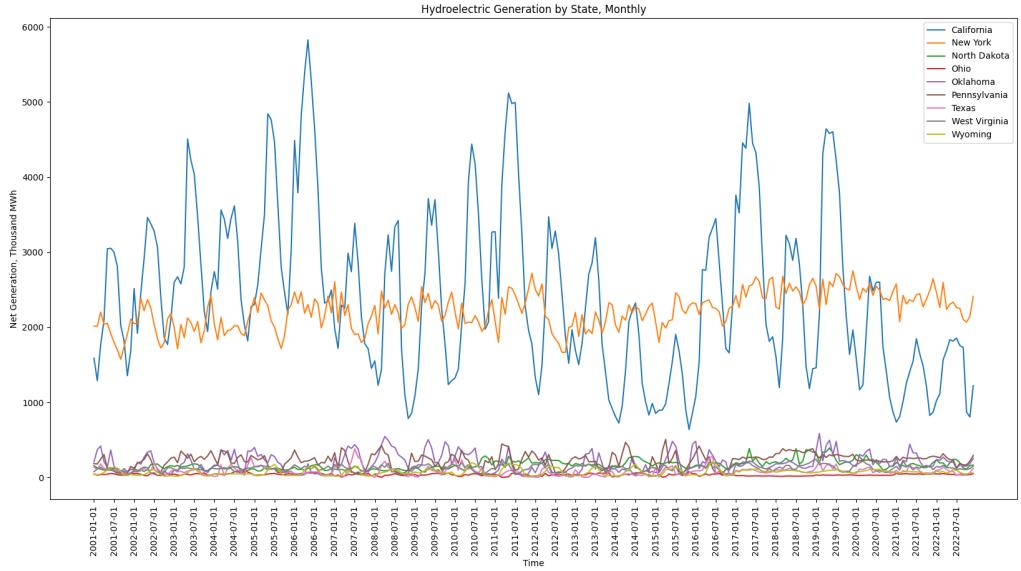
Gross vs. Per Capita Time Series by State

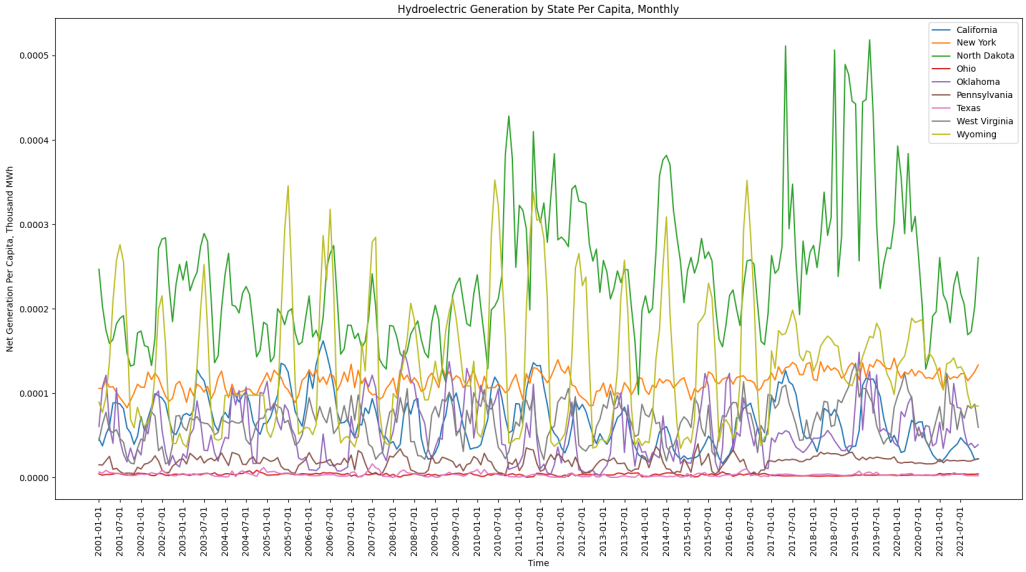

Biomass Generation by State, Monthly
California
Florida
New York
Ohio
Oklahoma
Pennsylvania
Texas
Net Generation, Thousand MWh
Time

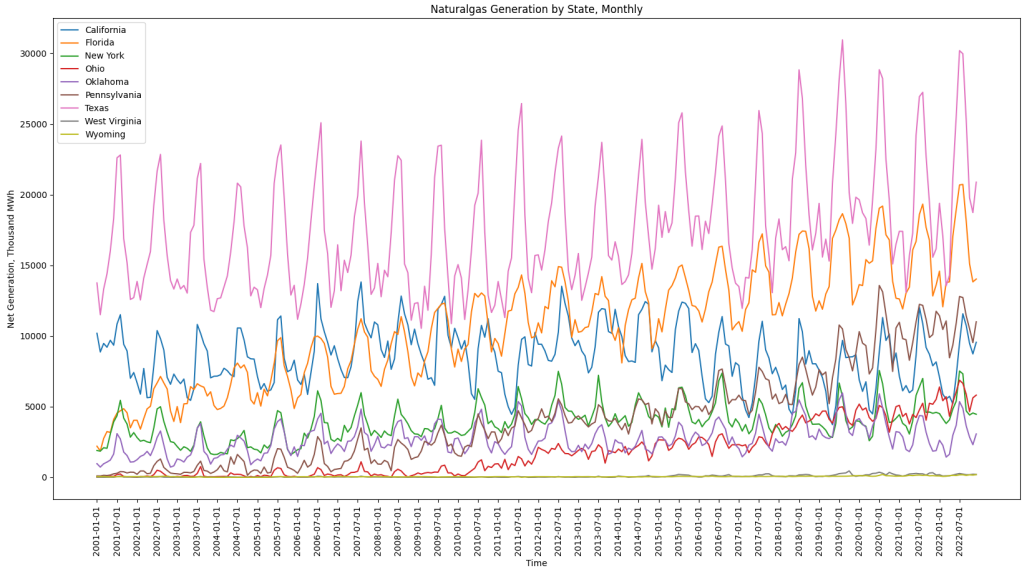

Biomass Generation by State Per Capita, Monthly
1e−5
California
Florida
New York
Ohio
Oklahoma
Pennsylvania
Texas
Net Generation Per Capita, Thousand MWh
Time

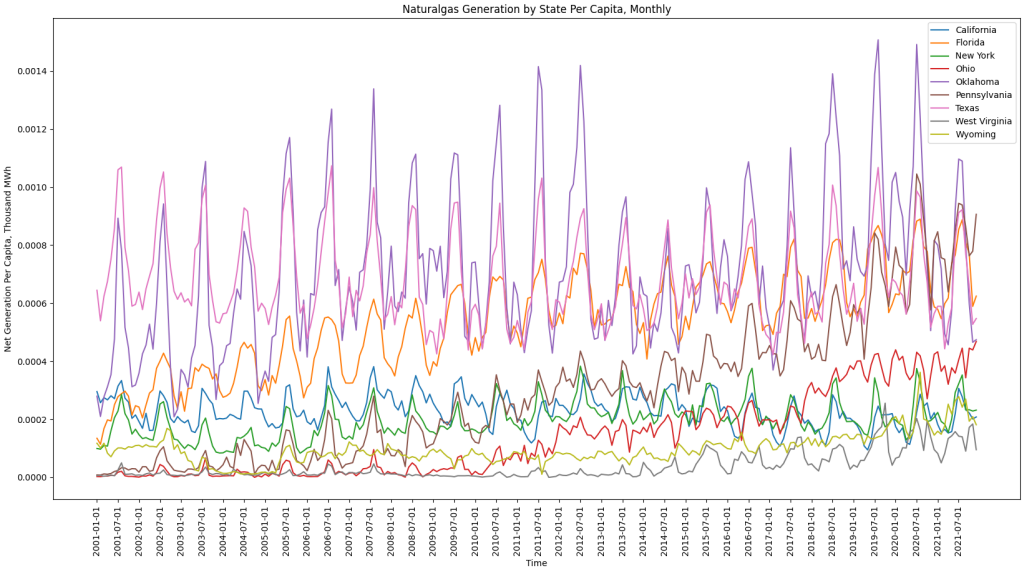

Coal Generation by State, Monthly
Net Generation, Thousand MWh
Time
California
Florida
New York
North Dakota
Ohio
Oklahoma
Pennsylvania
Texas
West Virginia
Wyoming


Coal Generation by State Per Capita, Monthly
Net Generation Per Capita, Thousand MWh
Time
California
Florida
New York
North Dakota
Ohio
Oklahoma
Pennsylvania
Texas
West Virginia
Wyoming

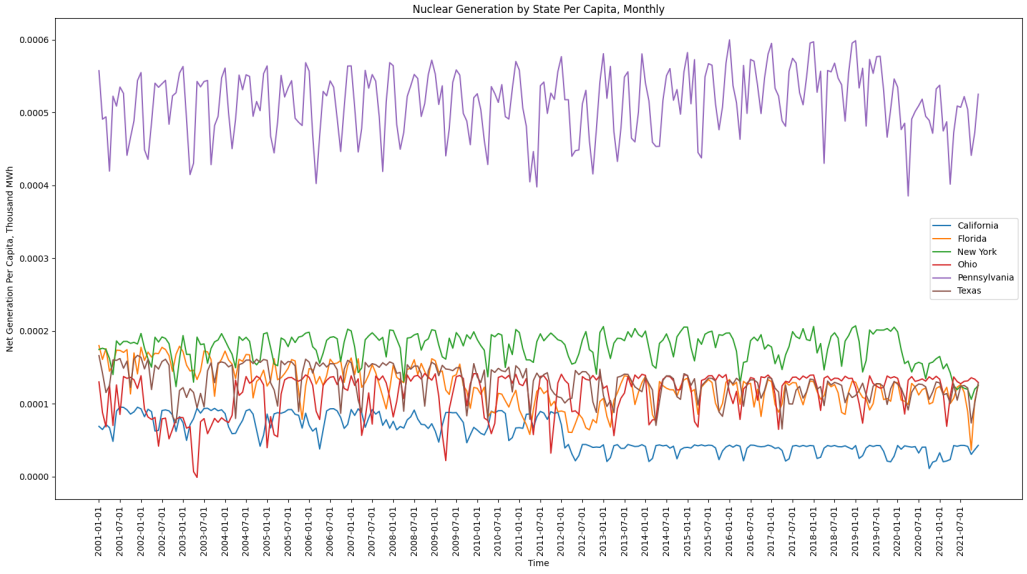

Hydroelectric Generation by State, Monthly
Net Generation, Thousand MWh
Time
California
New York
North Dakota
Ohio
Oklahoma
Pennsylvania
Texas
West Virginia
Wyoming

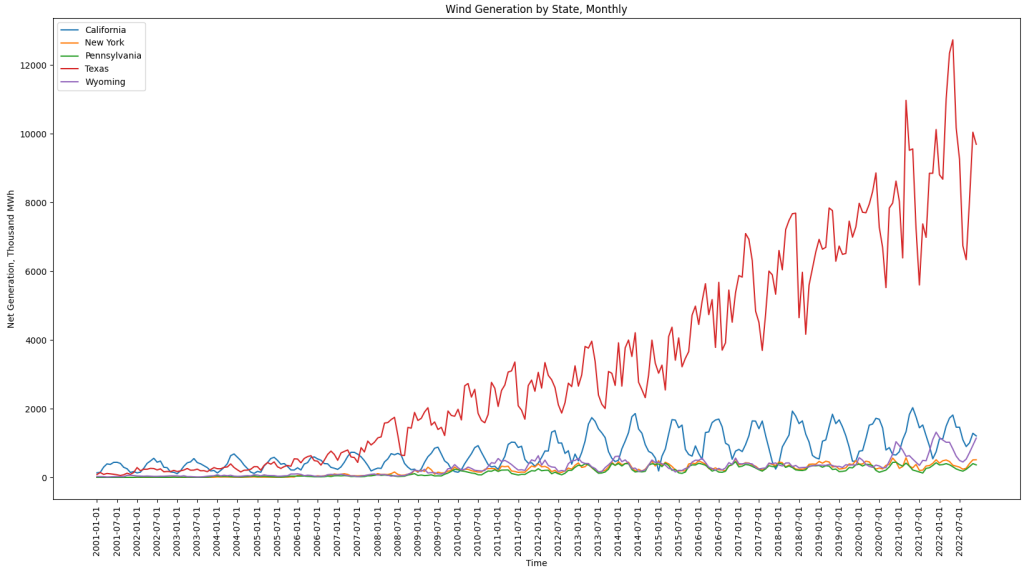

Hydroelectric Generation by State Per Capita, Monthly
Net Generation Per Capita, Thousand MWh
Time
California
New York
North Dakota
Ohio
Oklahoma
Pennsylvania
Texas
West Virginia
Wyoming

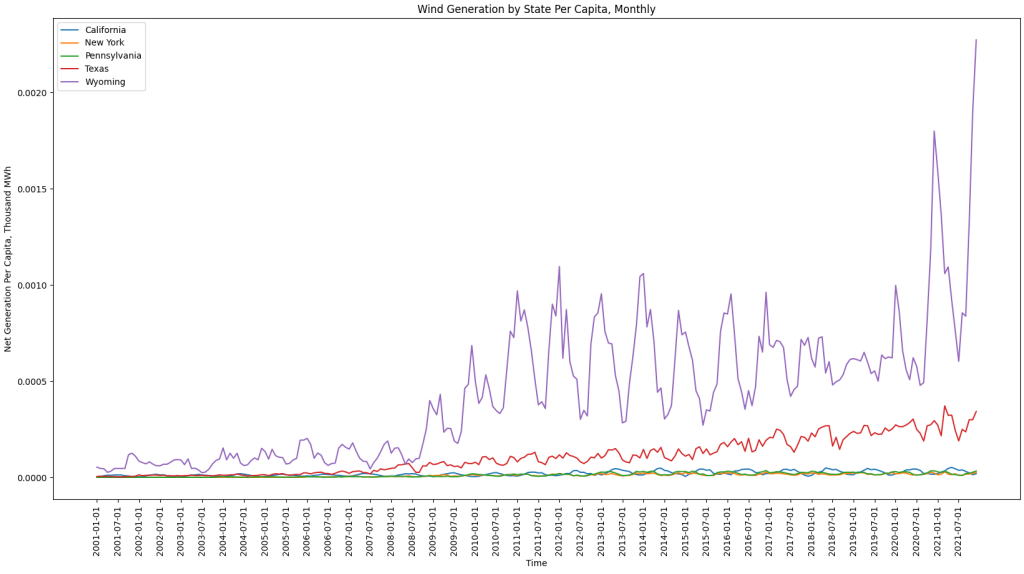

Naturalgas Generation by State, Monthly
California
Florida
New York
Ohio
Oklahoma
Pennsylvania
Texas
West Virginia
Wyoming
Net Generation, Thousand MWh
Time

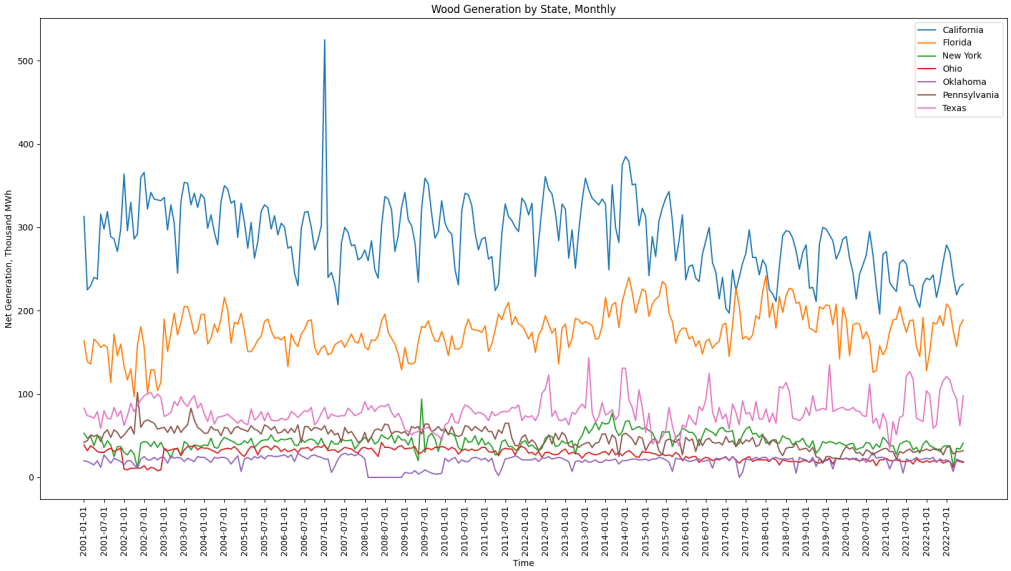

Naturalgas Generation by State Per Capita, Monthly
California
Florida
New York
Ohio
Oklahoma
Pennsylvania
Texas
West Virginia
Wyoming
Net Generation Per Capita, Thousand MWh
Time

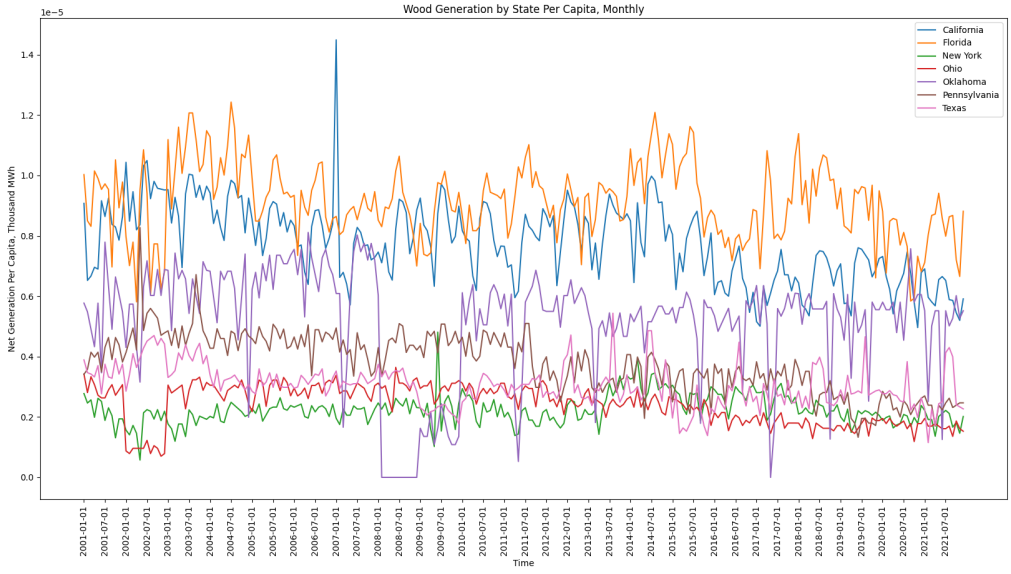

Nuclear Generation by State, Monthly
California
Florida
New York
Ohio
Pennsylvania
Texas
Net Generation, Thousand MWh
Time


Nuclear Generation by State Per Capita, Monthly
California
Florida
New York
Ohio
Pennsylvania
Texas
Net Generation Per Capita, Thousand MWh
Time


Wind Generation by State, Monthly
California
New York
Pennsylvania
Texas
Wyoming
Net Generation, Thousand MWh
Time

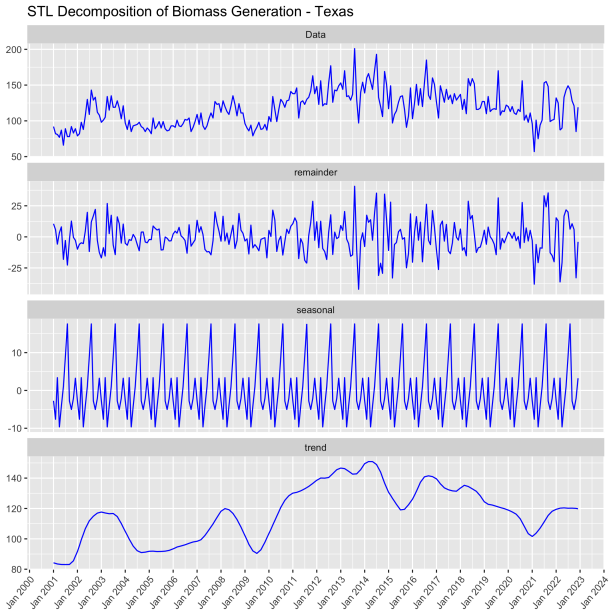

Wind Generation by State Per Capita, Monthly
California
New York
Pennsylvania
Texas
Wyoming
Net Generation Per Capita, Thousand MWh
Time

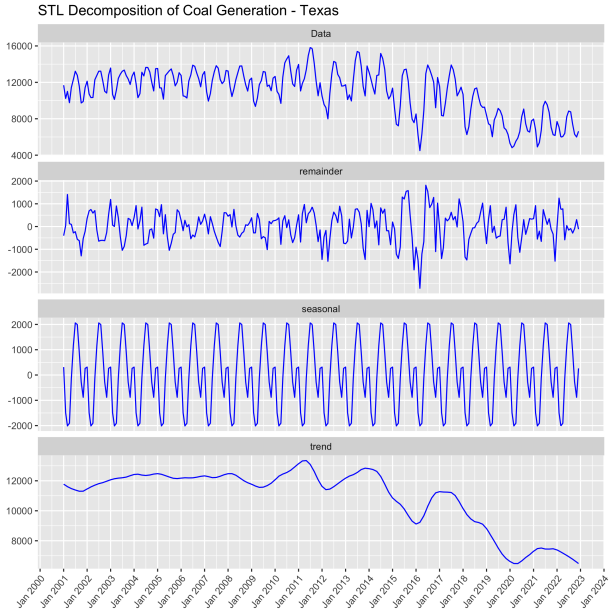

Wood Generation by State, Monthly
California
Florida
New York
Ohio
Oklahoma
Pennsylvania
Texas
Net Generation, Thousand MWh
Time

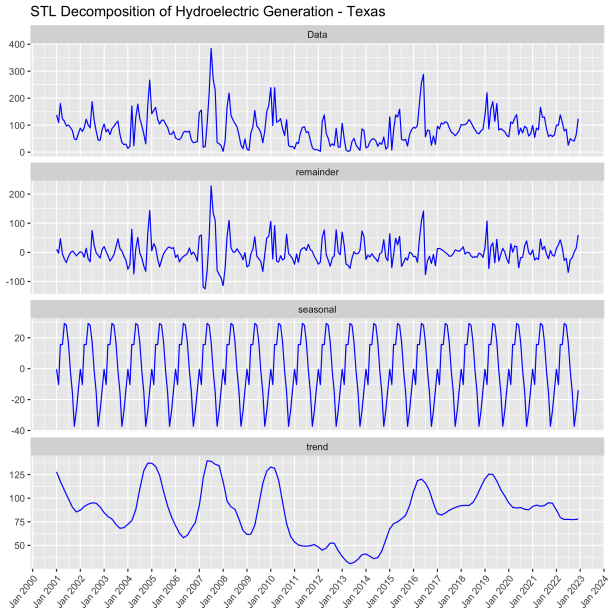

Wood Generation by State Per Capita, Monthly
1e−5
California
Florida
New York
Ohio
Oklahoma
Pennsylvania
Texas
Net Generation Per Capita, Thousand MWh
Time

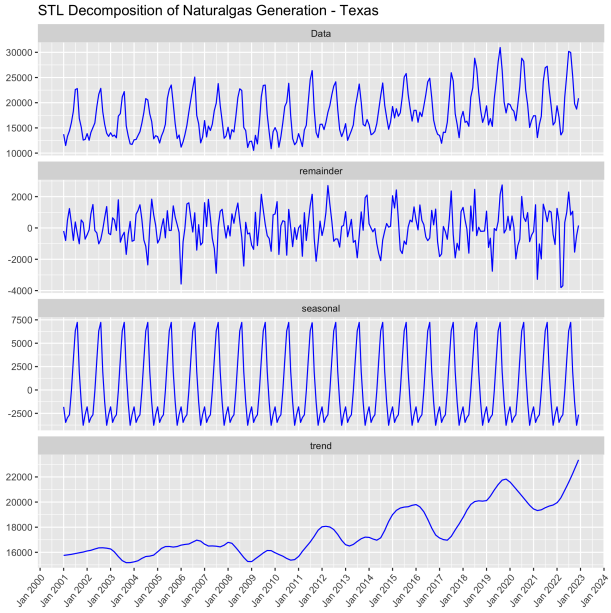
Appendix C - STL Decomposition

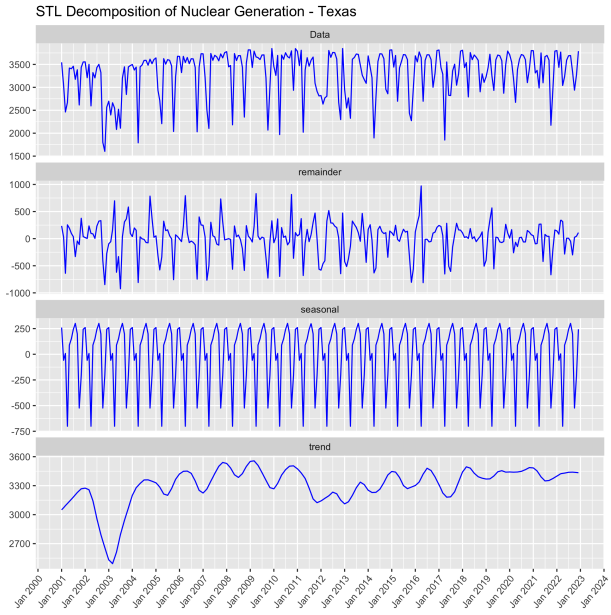
We conducted the STL decomposition of the time series over various sectors of energy generation for Texas. Similar patterns can be observed for other states as well. The seasonality chosen was 12 as there are 12 months in a year.

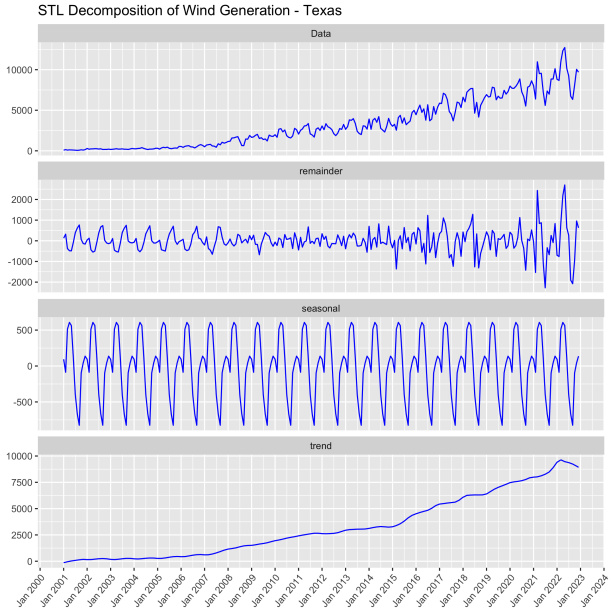


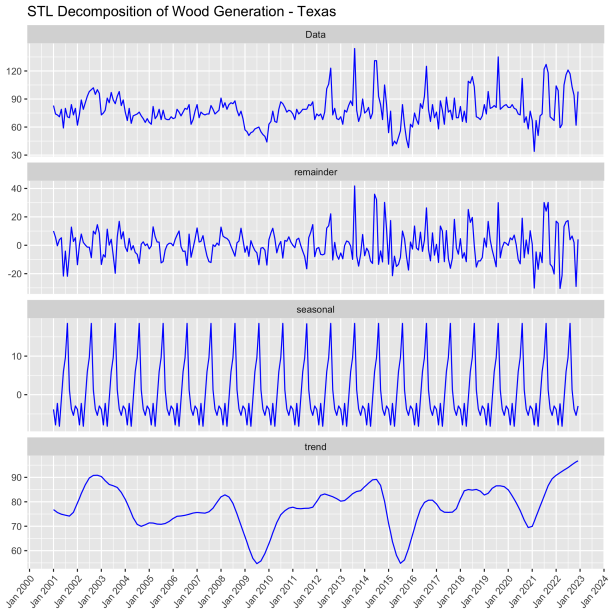



STL Decomposition of Hydroelectric Generation - Texas
Data
remainder
seasonal
trend
Jan 2000
Jan 2001
Jan 2002
Jan 2003
Jan 2004
Jan 2005
Jan 2006
Jan 2007
Jan 2008
Jan 2009
Jan 2010
Jan 2011
Jan 2012
Jan 2013
Jan 2014
Jan 2015
Jan 2016
Jan 2017
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Jan 2024


STL Decomposition of Naturalgas Generation - Texas
Data
remainder
seasonal
trend
Jan 2000
Jan 2001
Jan 2002
Jan 2003
Jan 2004
Jan 2005
Jan 2006
Jan 2007
Jan 2008
Jan 2009
Jan 2010
Jan 2011
Jan 2012
Jan 2013
Jan 2014
Jan 2015
Jan 2016
Jan 2017
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Jan 2024


STL Decomposition of Nuclear Generation - Texas
Data
remainder
seasonal
trend
Jan 2000
Jan 2001
Jan 2002
Jan 2003
Jan 2004
Jan 2005
Jan 2006
Jan 2007
Jan 2008
Jan 2009
Jan 2010
Jan 2011
Jan 2012
Jan 2013
Jan 2014
Jan 2015
Jan 2016
Jan 2017
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Jan 2024


STL Decomposition of Wind Generation - Texas
Data
remainder
seasonal
trend
Jan 2000
Jan 2001
Jan 2002
Jan 2003
Jan 2004
Jan 2005
Jan 2006
Jan 2007
Jan 2008
Jan 2009
Jan 2010
Jan 2011
Jan 2012
Jan 2013
Jan 2014
Jan 2015
Jan 2016
Jan 2017
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Jan 2024


STL Decomposition of Wood Generation - Texas
Data
120
90
60
30
remainder
40
20
0
-20
seasonal
10
0
trend
90
80
70
60
Jan 2000
Jan 2001
Jan 2002
Jan 2003
Jan 2004
Jan 2005
Jan 2006
Jan 2007
Jan 2008
Jan 2009
Jan 2010
Jan 2011
Jan 2012
Jan 2013
Jan 2014
Jan 2015
Jan 2016
Jan 2017
Jan 2018
Jan 2019
Jan 2020
Jan 2021
Jan 2022
Jan 2023
Jan 2024