Electricity Price Volatility Prediction

Adam Goldstein, David Harper, Miles Franklin, Himanshu Ghritalhre

Abstract

Our project aims to understand the volatility of electricity prices across the U.S. aggregated by state and years 2001-2020 across features encompassing fuel sources (MWh and powerplant count), fuel consumption by residential, commercial, and industrial sectors (normalized by state population and number of electric accounts), weather (temperature and drought index), and historical futures contracts (open, high, low, close, and volume for Crude Oil, Natural Gas, and Heating Oil futures contracts) to explore this question. Our target variable is the coefficient of variation in electricity price. The model that best explains our target variable is an XGBoost Regressor model. This model achieved an adjusted R^2 of 0.77 and heavily relies on the coefficient of variation of the previous year, state, the Brents Crude Oil futures contract data, and features that reflect the supply and demand of electricity (ex. Total plant count). This model performs consistently, with exceptions to LA, OK, RI, and HI. A particular anomaly is NV which trains well, but struggles with the test data.

Introduction

One of the most common approaches to modeling price action in any sector is by using a time-series algorithm like autoregression. The advantage of this approach is that it allows the model to learn the importance of sequential data when making the prediction for the next value in the series. In our project we decided to approach this problem a bit different, predicting the volatility of the price action, rather than the price itself. This approach introduces some trade-offs with the hope of attaining a more robust understanding of modeling energy prices.

For one, this approach sacrifices the exposure of sequential values for the introduction of other variables of interest. At the core these features include measures of state electricity sales by sector, net energy generation by plant, as well as complementary metrics of weather and market conditions at respective times.

Another benefit that we hope to gain from this approach is developing a model that can learn from the history of other states. This will vastly expand the training data our model will have access to, and ideally yield different insight from the traditional autoregressive models.

Data Description

Target Variable

Due to the constraints of the data we had access to, our electricity price data is separated by state, and aggregated on a yearly basis. Thus we measured our target variable, price volatility, as the Coefficient of Variation (COV) for the following year for each state. This metric allows us to compare volatility of each state while adjusting for the magnitude of the overall price action.

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Mean}}$$

For example, a \$0.05 swing in a state where electricity prices are around \$0.15 per kWh will be felt much more substantially than in states where electricity prices are closer to \$0.30 per kWh, and COV captures this nuance well.

Weather Data

Two weather variables are used in this project. The first is the monthly average temperature (at 2 meters above the surface of land). The second is the Palmer Drought Severity Index (PDSI) which is derived using temperature, precipitation, and measures relative dryness . Both of these variables were extracted from Earth Engine, gridded at a 60km X 60km granularity. The data was pulled from 1981 through 2020. The average and standard deviation for 1981 through 2000 was calculated at each of the 60km X 60km locations and used to standardize the data at these locations for years 2001 through 2020. The following features were created using these two variables:

1. Hot Summer Months - The number of summer months (June, July, August, September) with an average standardized temperature above 1.
2. Cold Winter Months - The number of winter months (December, January, February, March) with an average standardized temperature below 1.

3. Dry Summer Months - The number of summer months (June, July, August, September) with an average standardized PDSI below 1.

Futures Data

Historical open, high, low, close, and volume for futures contracts, at both a daily and weekly level, was gathered through yahoo finance for NG (natural gas), CL (crude oil), and HO (heating oil). Price movement was derived using the range in trading price for each day/week divided by the open price. The average, standard deviation, and coefficient of variation of the price movement and volume was calculated for each month of each year. These figures were averaged to get one value per year. In addition, the following features were created using these variables for each futures contract:

1. Months With Above Average Price Movement - number of months in the year with price movement above the 6 month moving average price movement.
2. Months With Above Average Volume - number of months in the year with volume above the 6 month moving average volume.

Energy Production Data

We hypothesized that understanding the diversity of states' power plant fuel types would impact COV for each state. The Energy Information Agency (EIA) contains plant net generation MWh across 37 fuel types. This data was provided as a monthly aggregate.

Energy Consumption Data

We hypothesize that different sectors' electricity usage or the totality of usage by the state would impact the volatility of the electricity price. To capture this detail, electricity sales by state and sector from EIA was used. The sectors used in this analysis are Residential (RES), Commercial (COM), and Industrial (IND) as the other sectors, Transportation and Other, were sparse. Since usage will be skewed by state population, two per capita measurements were created. The two calculations are below.

$$\text{kWh\_percapita} = \frac{\text{kWh}}{\text{electric\_accounts}}$$

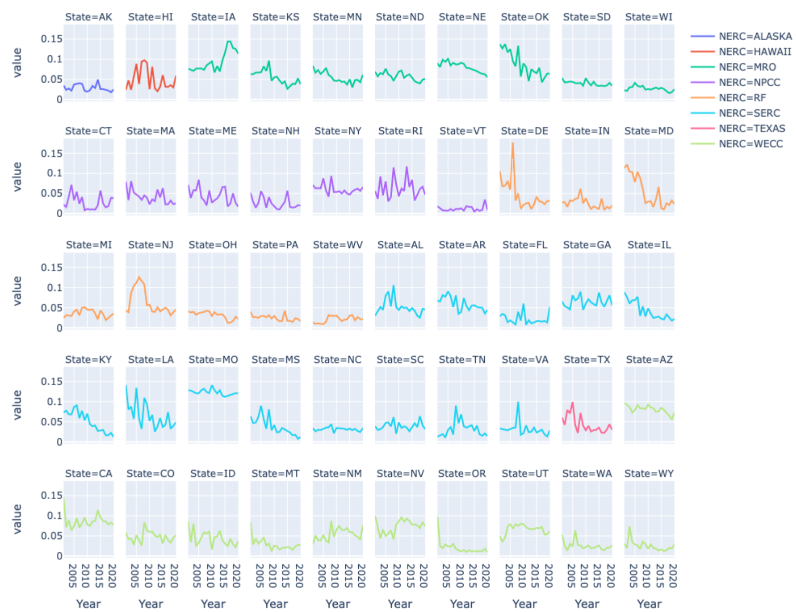
$$\text{kWh\_percapita} = \frac{\text{kWh}}{\text{US\_statepop}}$$

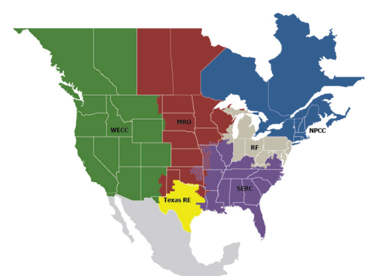
Electric accounts, supplied by EIA, are supplied for 2008 through 2021 and the US census state population estimates are supplied by the US census for 2001 through 2020.

EDA

COV

As seen below, states have a variety of COV rates over 2001 to 2020. For more clarity on what region each state was colored with their respective NERC region.



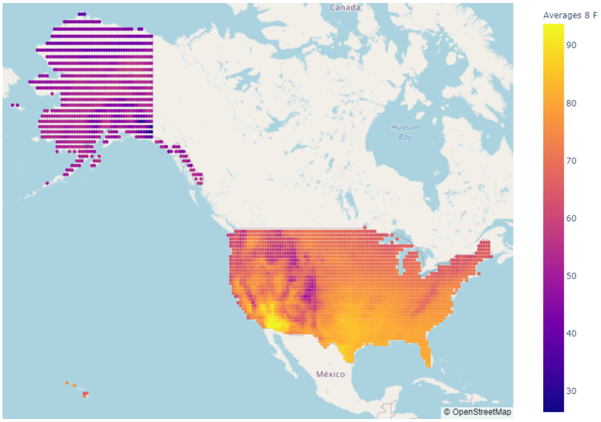


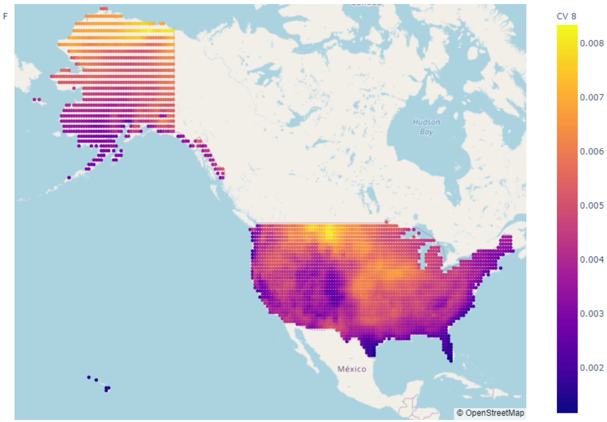
Weather Data

Temperature Data

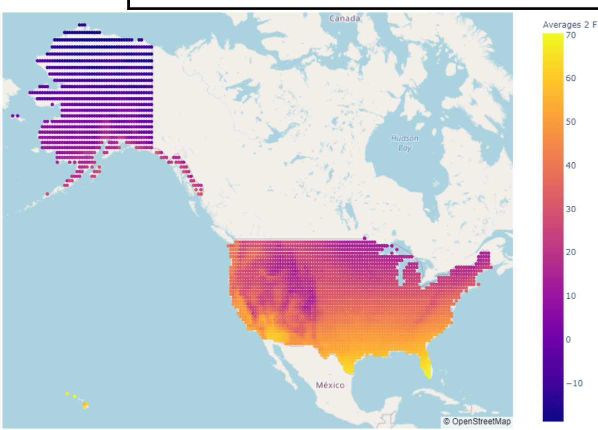
First, we were interested in looking at what the temperature averages and variation look like for a summer month and a winter month across the U.S. The first set of plots below shows the average temperature and COV across the U.S. for the month of August (using data from 1981-2000). The second one shows the same information but for February.

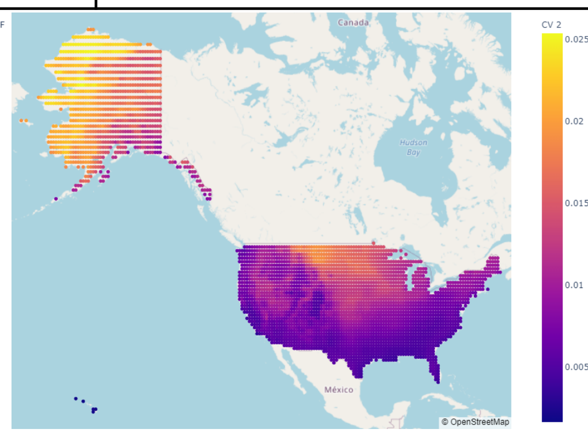
Mean Temperature for August	Coefficient of Variation in August Temperat





Mean Temperature for February	**Coefficient of Variation in February Temperature**

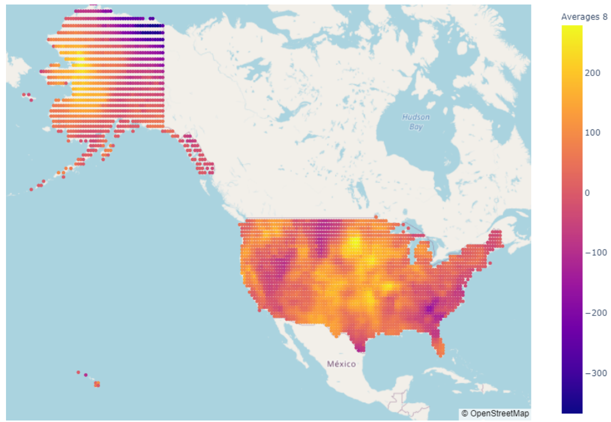




These plots show that the north-central region of the United States and Alaska have the highest variability in temperature for August and February.

Drought Data (PDSI)

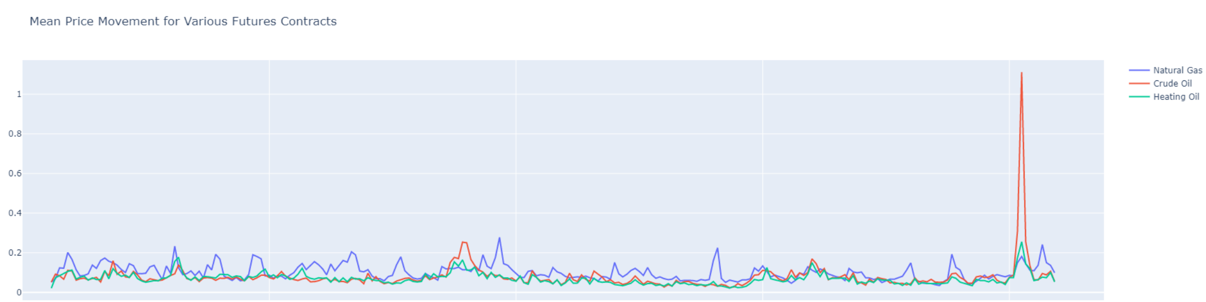
Our other weather hypothesis was that droughts have an impact on COV since a source of power is hydroelectric. Therefore, we were interested in seeing which regions of the U.S. tend to have the worst droughts. The figure below shows this.

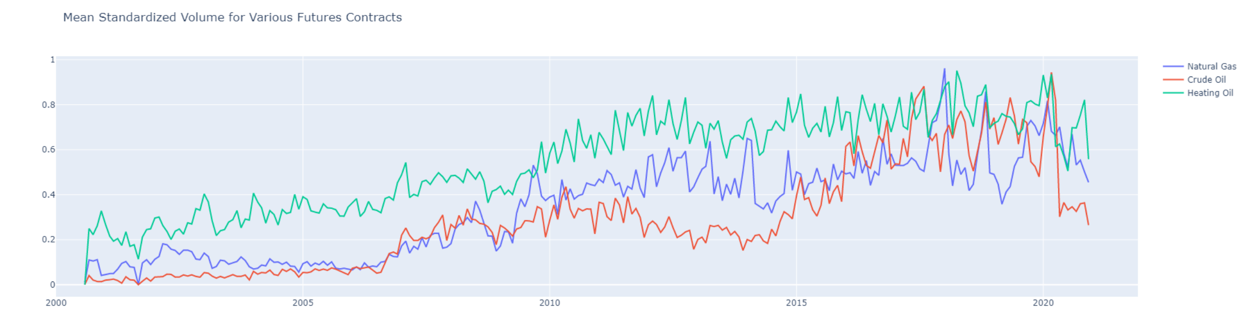


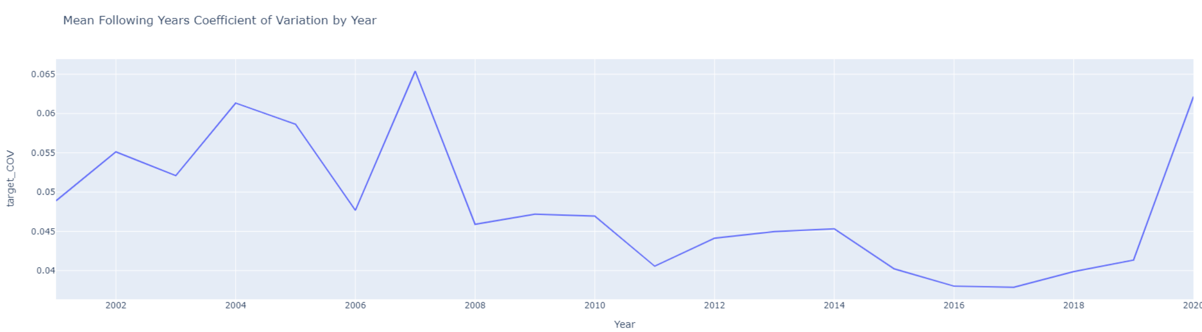
Since low values suggest worse drought conditions, southern Texas, the southeast, northern Montana, northern Alaska, and Nevada have the worst droughts, on average, for the month of August. However, it is important to note that the worst time of year for droughts is regional. For example, August is monsoon season in the four corners region, which agrees with that region having high PDSI values.

Futures Data

Our hypothesis for futures data was that the price action would have some relationship with the volatility of the electricity prices. The first figure below shows the mean price movement in the various futures contracts as a function of time, and the second figure shows the mean volume. Lastly, the bottom figure shows the average next years' COV for comparison with the futures data.



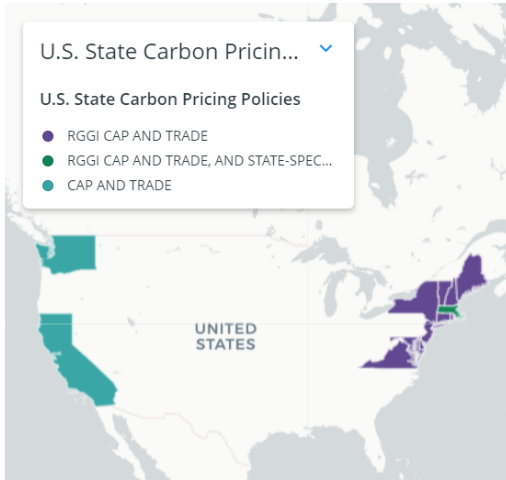




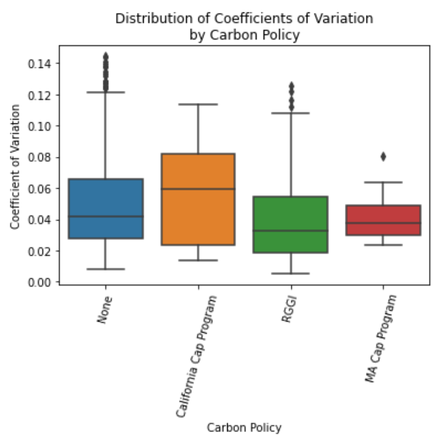
The mean price movement is very cyclical, which makes sense. However, these cycles are occuring at a smaller time frame than a year, so they aren't very useful to our time frame. However, the volume of these contracts have increased over time, showing long term trends that appear negatively correlated with the target coefficient of variation in the bottom plot.

Carbon Credit States

Using the following map, three carbon credit program binary variables were created. These include RGGI Cap and Trade program, state specific program, cap and trade program, and no program.



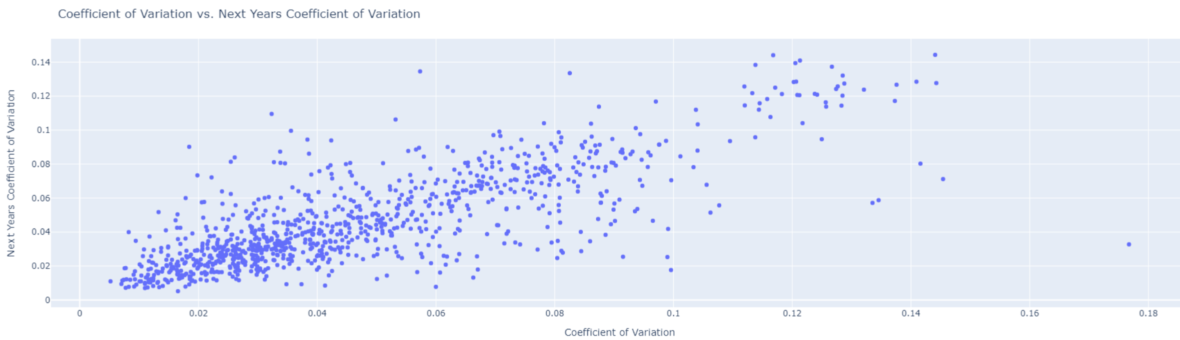
The two states that participate in the California Cap and Trade program (CA and WA) have the highest coefficient of variation on average, while the states that participate in the RGGI program (ME, NH, VT, MA, CT, RI, NJ, MD, DE, and VA) have slightly lower coefficients of variation on average compared to states with no carbon policy.



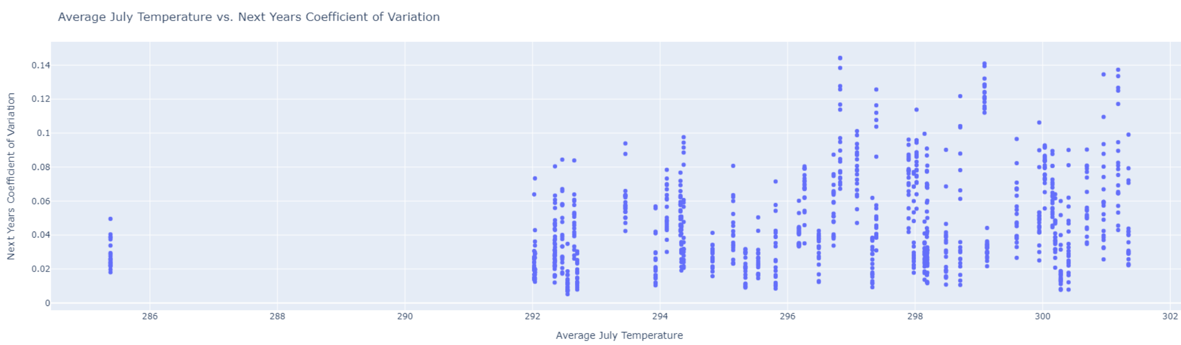
Variables Most Correlated with Target Coefficient of Variation

Prior Years Coefficient of Variation

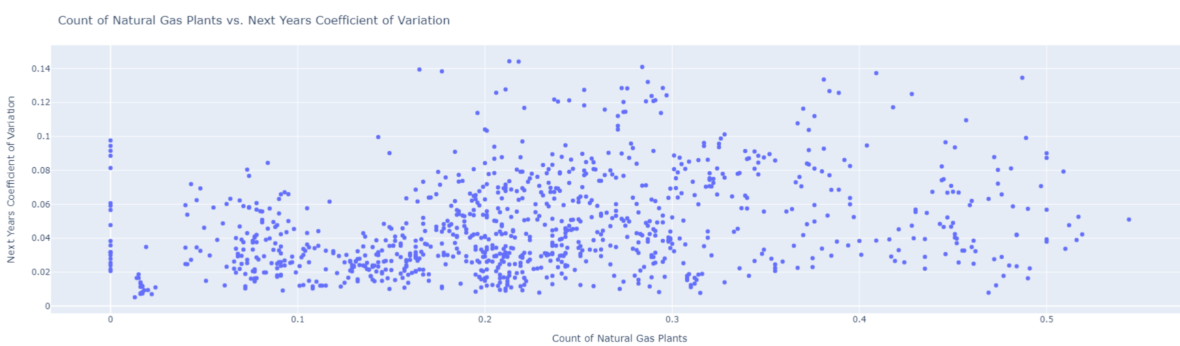
This was by far the most correlated variable, yielding a correlation of 78%. The scatterplot below shows this strong relationship.



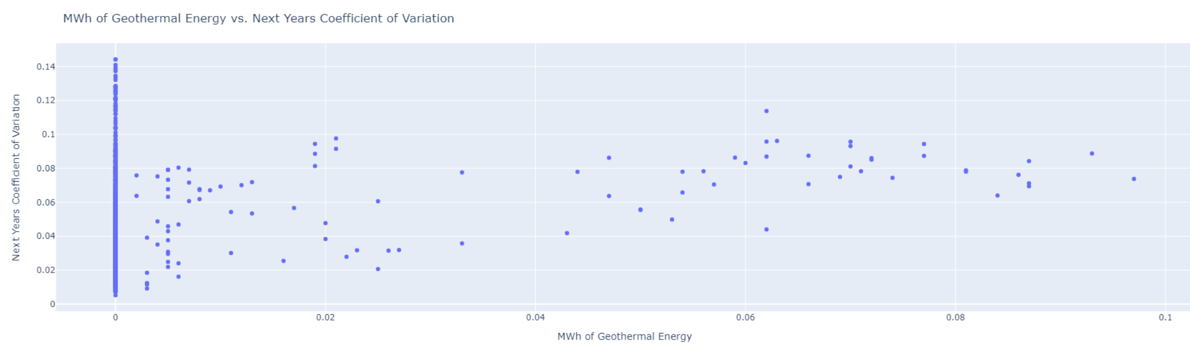
The average July temperature has a correlation of 32% and is shown in the scatterplot below.



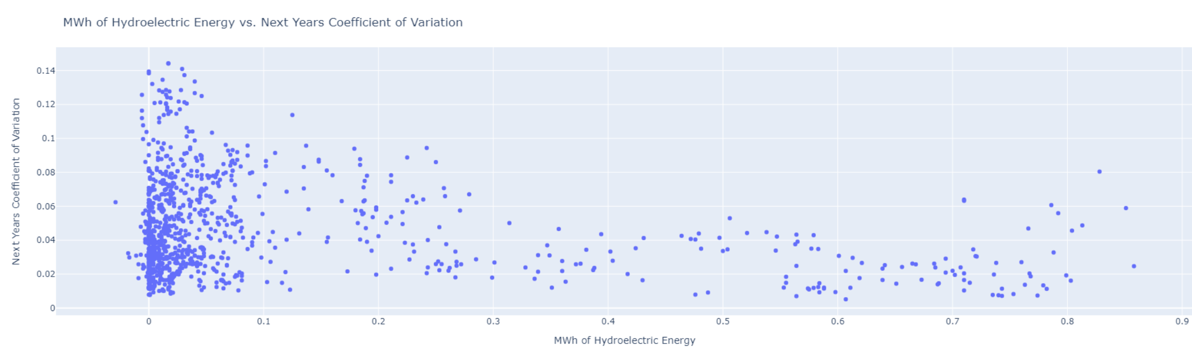
The total count of natural gas related plants has a correlation of 31% and is shown in the scatterplot below.



The MWh of Geothermal energy has a raw correlation of 23% and is shown in the scatterplot below. However, it is important to note that if the state and year combinations absent of Geothermal were removed, this correlation would be much higher.



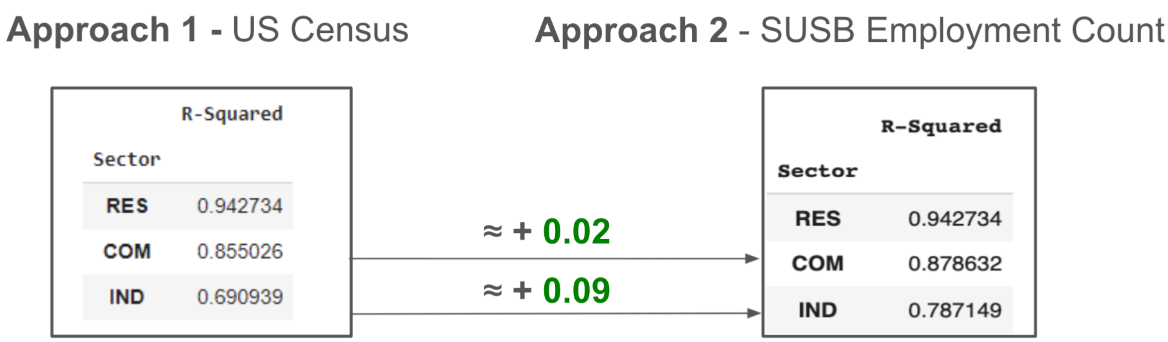
The last notable variable is the MWh of Hydroelectric energy which has a correlation of -23% with the target variable. The scatterplot is very noisy at the lower end of MWh, but past this, there is an obvious negative correlation.



Feature Engineering

As mentioned above, we had several features related to the number of energy accounts in a given state that were only documented as far back as 2008. This presented an issue because there was a noticeable decrease in model effectiveness when we simply tried to remove these features from the analysis in exchange for 8 additional years of training data. Our solution to this problem was to make separate linear regression models for RES, COM, and IND electricity accounts and backfill the data with the predicted values. The goal was to build a linear regression model that yielded a $R^2 >= 0.75$. The first approach was to leverage US census state populations to fill in electricity accounts. This worked well for

RES and COM, but IND was under the threshold so the second approach leveraged employment count from the Statistics of U.S. Businesses (SUSB), which yielded an acceptable R^2.



For the final data frame, we had mostly continuous variables, and chose to use min-max scalers to constrain the data to values within [0,1] because most models are sensitive to scale. We also chose to drop the year from our analysis and represented the state, as a categorical column with a numerical code for each state. We concluded that adding the previous year's measurement of COV provided enough of a time element into the model, without taking on too much of overfitting the model.

Modeling

We constructed a pipeline for our models to be logged in Neptune AI. This enabled a holistic view on the overall progress of the experiments and made for simpler comparisons between experiments. We carried out 217 unique experiments, 15 feature pipelines, 17 unique models (7 Ensemble, 10 Linear).

Since all of our models in the pipeline were regression algorithms, we evaluated the mean squared error (MSE), and then picked our final model based on Adjusted R^2. We chose Adjusted R^2 as a metric because R^2 is sensitive to wider data frames and would be unable to accurately describe how much variance our model accounts for. On the other hand, Adjusted R^2 does account for this, and as a result proved to be a more accurate measure for this project.

Results/Conclusion

Final Model

The top experiments are shown below.

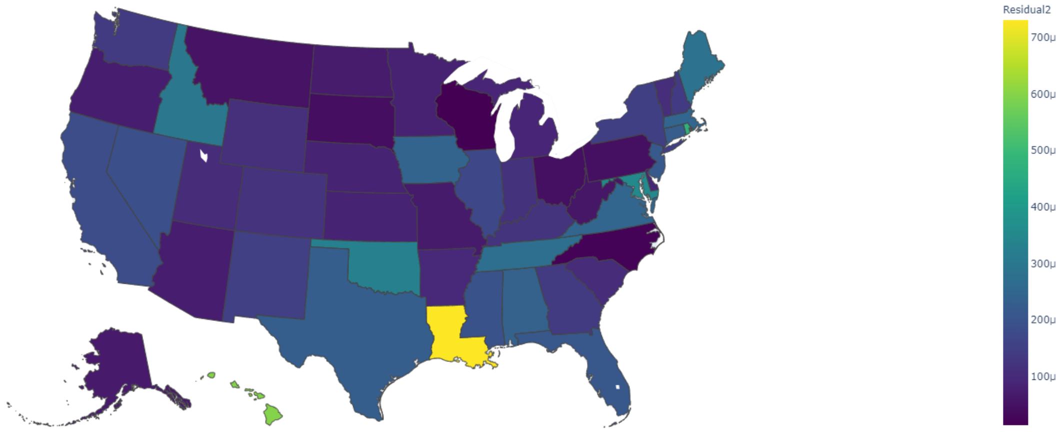
Cust_Model	base_estimator	Adj_R2_test
XGBRegressor	nan	0.768890
HistGradientBoostingRegressor	nan	0.766902
BaggingRegressor	HistGradientBoostingRegressor	0.763546
AdaBoostRegressor	LassoLarsIC	0.763318
ExtraTreesRegressor	nan	0.713314
GradientBoostingRegressor	nan	0.695125
RandomForestRegressor	nan	0.685285

Our final model was an XGBoost regressor model with 400 estimators, a learning rate of 0.1, and a max depth of 1. The rest of the hyperparameters were kept at their defaults. This final model achieved the following performance:

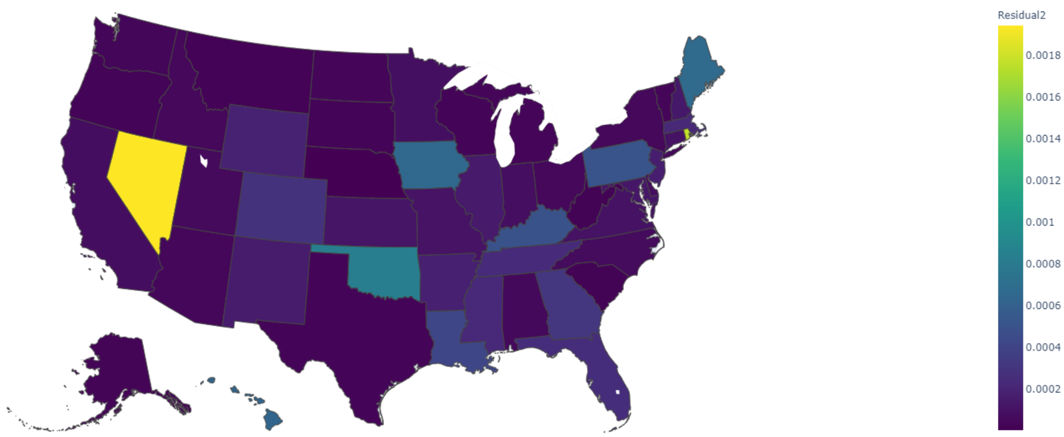
Adj. R^2 (training)	R^2 (training)	R^2 (test)	MSE (training)	MSE (test)
0.77	0.77	0.777	0.00017	0.00017

In addition to seeing that the overall metrics were consistent between the training and test sets, we wanted to investigate if the model is performing well/poorly in the same states when looking at training vs. test data. Below is a choropleth plot which shows the MSE of the model by state, both on the training data and the test data.

MSE by State on Training Data



MSE by State on Test Data

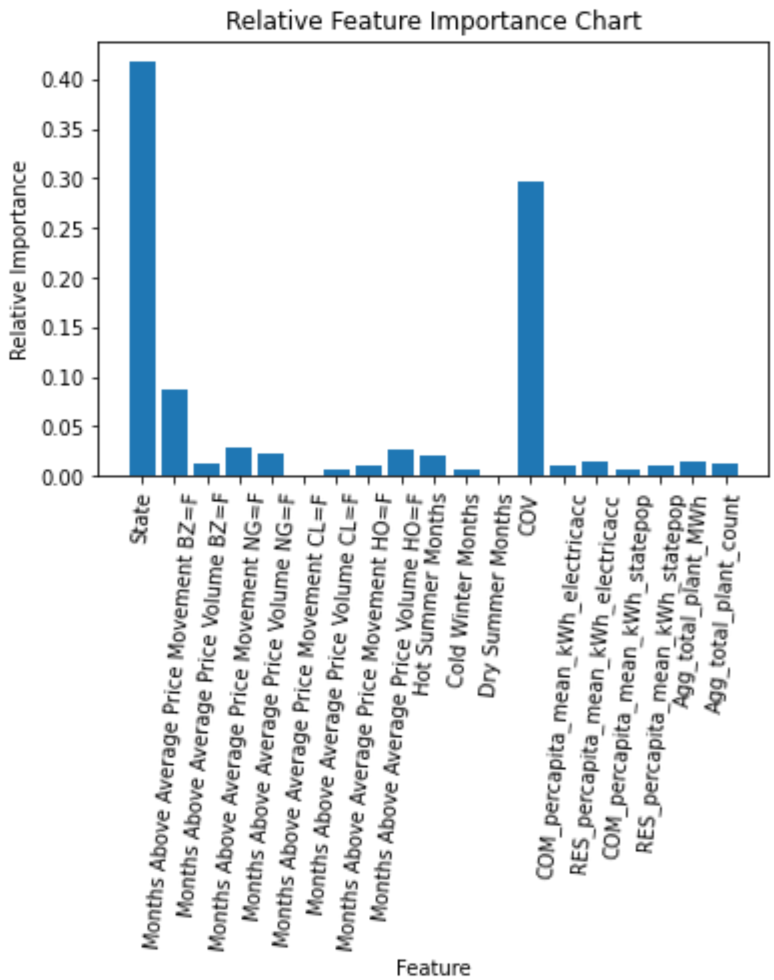


The model performs well in most states, but doesn't perform well in LA, OK, ID, HI and RI. When this model is used on test data, the states mentioned still do not perform as well with the exception of ID. In addition, the model performed poorly on NV with the test data.

Feature Importance

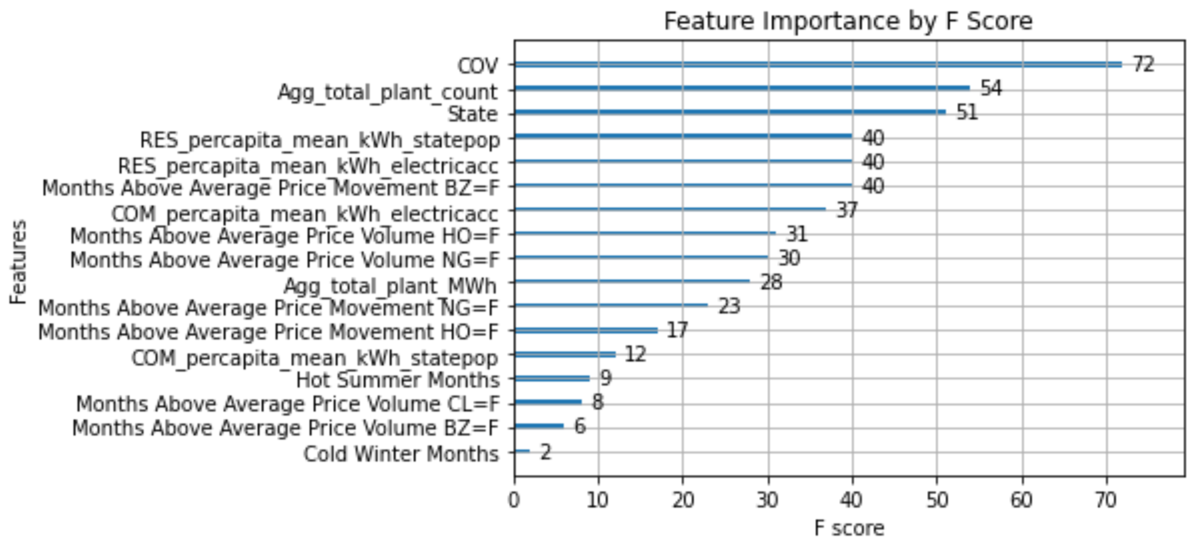
Evaluation of top feature importance for predicting COV were identified using two methods, Relative Importance and F Score.

Relative Importance



This relative feature importance takes into account the contribution of each feature to each model. It is obvious that previous year's COV is the most important feature in our model.

F Score Importance



The F Score shows the importance of each feature by showing how many times each feature was split on across all the underlying models. This is different from relative importance because it only considers whether the features were used, and not how much the features actually contributed to the models. This shows that the four most used features are State, COV, total plant count, and total residential KWh per person (using state populations).

Future Additions

Different Data Sources

The majority of our data was pulled from the US Energy Information Agency (EIA). This was the most comprehensive collection of data we could find related to our topic, however it is certainly a limiting factor in our analysis. For the majority of features explored, the data was only available from 2001 to present, with a few features that were only recorded as far back as 2008. This included the monthly cost of energy for each state, which had the additional limitation that electricity prices were already aggregated to a monthly level.

This project would likely benefit the most from having additional years of data for the model to learn on. For some of the features, we were able to overcome this limitation with regression techniques described above. In addition, verifying US electricity prices from additional sources would help limit the potential biases.

Feature Reduction

One of the other limitations of this project was the number of features we considered in our final models. To reduce the complexity of the model, future work can explore PCA or other feature reduction techniques. The complexity of the model can be best observed in our use of Adjusted R^2 as our evaluation metric of choice. Adjusted R^2 takes the number of features into account in its calculation, and therefore makes it a less sensitive metric than R^2 when working with wider datasets.

Transforming into a Classification Problem

We had spent some effort looking at the problem from a classification perspective, however we did not have the time to see this through to the end. Our plan included two potential methods for bucketing the target variable: the first being a basic increase or decrease in COV from the previous year, and the second being a measure of low, medium and high COV. Both approaches would simplify the task for the model and could result in accuracy improvements considering the shallowness of the data we have to work with. Moreover, we still believe the data is aggregated at an appropriate level because the task is more interested in the overall effects of the features rather than a precise variance measurement for a given year.

<u>*Additional note for the reader***</u>**

After further investigation of the chosen XGBoost model's performance on the train data across states compared to its performance on test data across states, we found that a large reason is that the train/test split was not stratified by state. Therefore, some states, such as NV, only had one observation in the test data while others had more than 4. We did additional modeling experimentation with a train/test split stratified by state and were not able to find a comparable model. The best model we could find in the amount of time we have been granted is found below.

```
----------------------------------------------------------------
VotingRegressor(estimators=[('BR', BayesianRidge()),
                            ('LIC', LassoLarsIC(max_iter=1000)),
                            ('RF',
                             RandomForestRegressor(max_depth=4,
                                                   max_leaf_nodes=30,
                                                   min_samples_split=10,
                                                   n_estimators=50))])
Train R2: 0.7016935054593995
Train Adjusted R2: 0.6872398933416852
Train MSE: 0.00023707870340786982
Test R2: 0.6418975234523895
Test Adjusted R2: 0.6245466632158969
Test MSE: 0.0002462676966548984
```

References/Citations

Energy Data

United States. (2010) U.S. Energy Information Administration EIA. United States. [Web Archive] Retrieved from the Library of Congress, https://www.loc.gov/item/lcwaN0015422/.

Temperature Data

Muñoz Sabater, J., (2019): ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (November 29, 2021). doi:10.24381/cds.68d2bb30 https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_MONTHLY?hl=en

PDSI Data

Abatzoglou, J.T., S.Z. Dobrowski, S.A. Parks, K.C. Hegewisch, 2018, Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015, Scientific Data 5:170191, doi:10.1038/sdata.2017.191

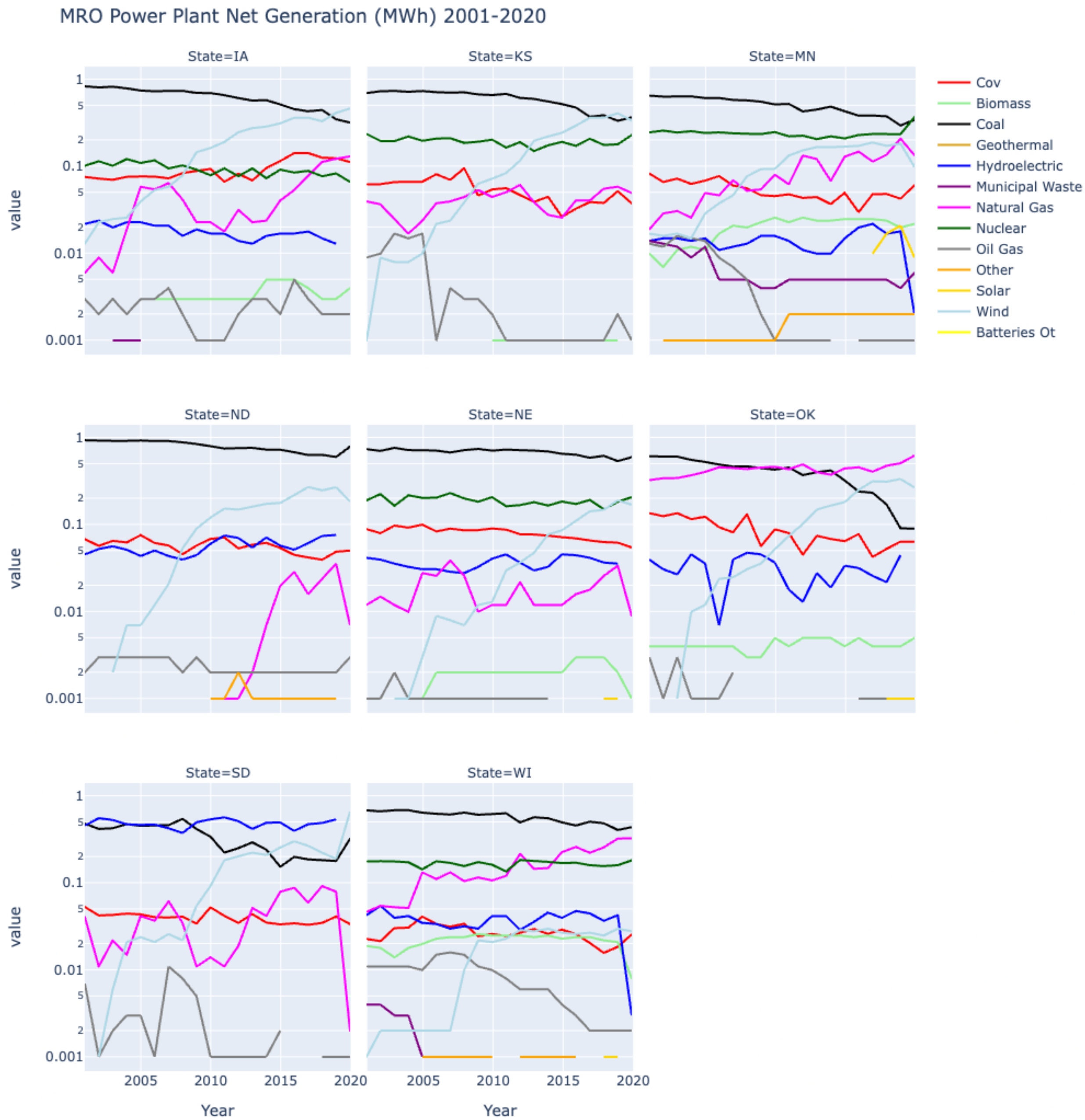
https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE?hl=en

Appendix

EDA

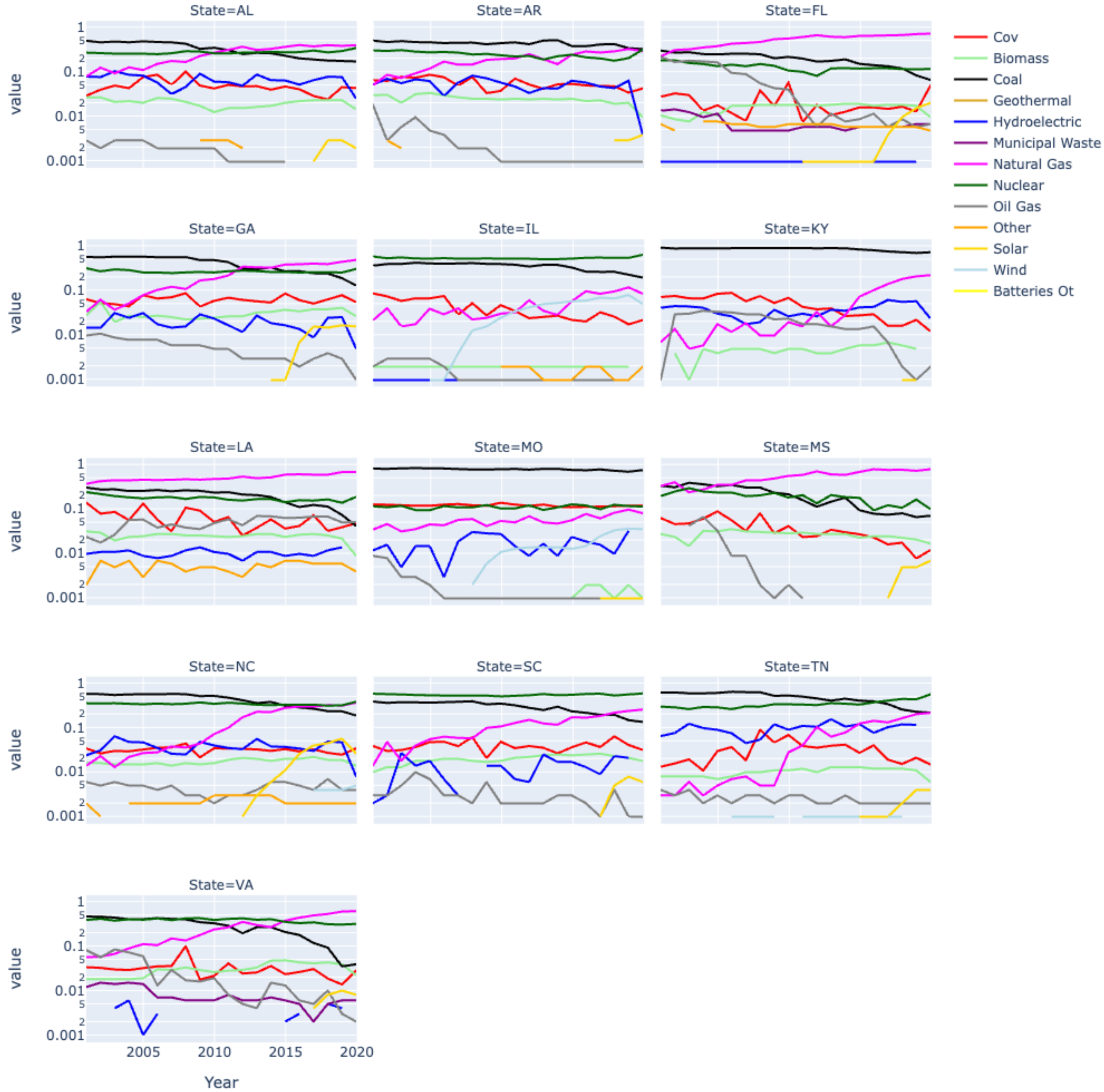
RF Power Plant Net Generation (MWh) 2001-2020
State=DE
State=IN
State=MD
State=MI
State=NJ
State=OH
State=PA
State=WV
value
1
0.1
0.01
0.001
Cov
Biomass
Coal
Geothermal
Hydroelectric
Municipal Waste
Natural Gas
Nuclear
Oil Gas
Other
Solar
Wind
Batteries Ot

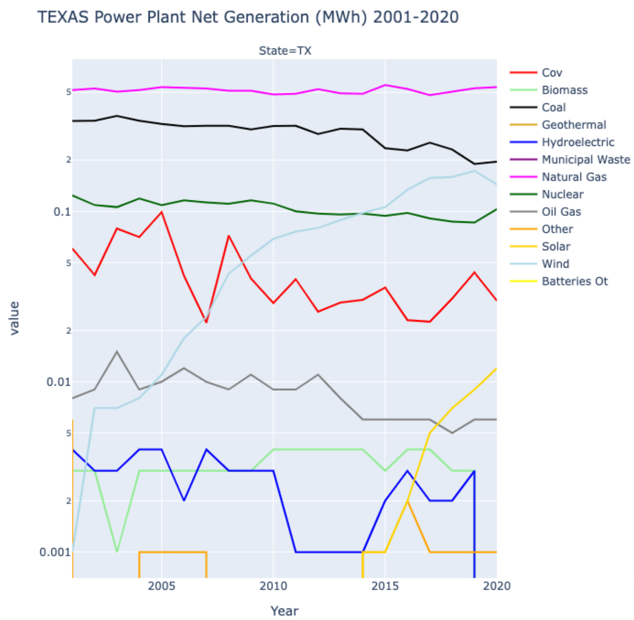
NPCC Power Plant Net Generation (MWh) 2001-2020
State=CT
State=MA
State=ME
State=NH
State=NY
State=RI
State=VT
value
Year
2005
2010
2015
2020
1
0.1
0.01
0.001
Cov
Biomass
Coal
Geothermal
Hydroelectric
Municipal Waste
Natural Gas
Nuclear
Oil Gas
Other
Solar
Wind
Batteries Ot

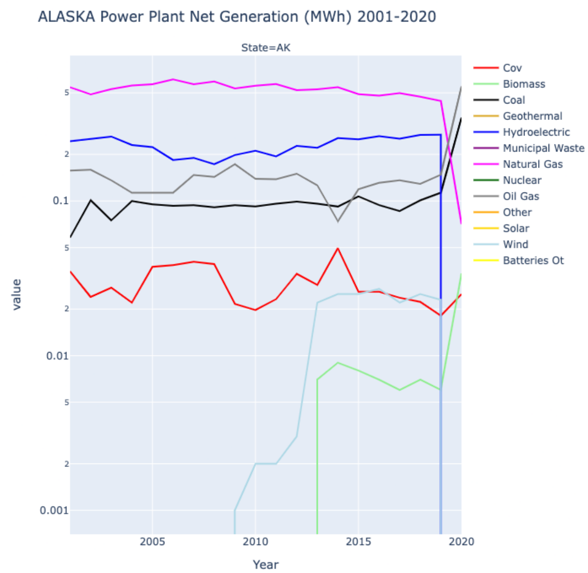


MRO Power Plant Net Generation (MWh) 2001-2020
State=IA
State=KS
State=MN
State=ND
State=NE
State=OK
State=SD
State=WI
value
Year
Cov
Biomass
Coal
Geothermal
Hydroelectric
Municipal Waste
Natural Gas
Nuclear
Oil Gas
Other
Solar
Wind
Batteries Ot

SERC Power Plant Net Generation (MWh) 2001-2020







Spearman Correlation