The Forced Climate Experiment:

A Study of Power Plant Energy Generation and Pollution Rates (and the Effects of COVID-19)

Paul Natland & Chris Sternberger

Introduction and Motivation

In the realm of the modern electricity grid, there is no shortage of data collection and, subsequently, an endless amount of different data visualization possibilities. That said, the complexity of the grid and its corresponding labyrinth of data makes it difficult to parse out the order from the chaos. The current, seemingly global, grid system is a product of local electricity systems that were slowly strung together over time, which allows for greater reliability and efficiency of the whole system. This results in often irregularly shaped entities referred to as Balancing Authorities (BAs) that are responsible for maintaining the electricity balance within its region. This region only sometimes adheres to state boundaries and there is also transfer between BAs. This can prove difficult therefore to do something commonly done such as statewide analyses.

The purpose of these Balancing Authorities is to navigate keeping electricity flowing safely in accordance with ever changing demand within and between authorities while also making an effort to be accountable for the emissions produced by various cleaner and dirtier power sources (i.e. Solar, Nuclear, Coal etc.). This is where our original interest came from: What is the nature of the energy landscape? Is reducing your energy (carbon) footprint as simple as it seems? How does a unique global event like COVID-19affect these emissions and potentially the energy generated and needed?

As it turns out, energy is messy. This is probably due to the modern grid system being built out of the ruins of the first, isolated local grid systems...akin to how road networks in old cities appear stochastic while relatively newer city architectures may seem more logical (i.e. Boston vs. Las Vegas). Thus, analysis has proven to be challenging. This is both due to an overwhelming amount of data resources as well as the evidently obligatory dose of disjointed correspondence between them. While this made it impossible to do some comparisons we had originally surmised, it made for great exercises in some serious data manipulation sessions.

We set out to explore energy and emissions mostly as novices to the domain. The nature of our analysis was fundamentally observational, but we knew that there had to be hidden opportunities to merge rich data sets of COVID-19, energy use and energy emissions, and to explore possible relationships. In our pursuit, we learned of a relevant and relatively new metric referred to as MOER (Marginal Operating Emissions Rates) which was a main component of our exploration. MOER is concerned with optimizing the *timing* in which energy is used and how that demand load can be shifted with respect to time to reduce emissions from different energy sources. Additionally, we observed reduced energy generation across varying sectors over the first ⅔ of 2020 as a result of the COVID-19 response. While this is likely unsurprising to most, we were curious to see more precisely how much of a change occurred and what this could possibly teach us about future intentional energy demand reduction.

Data Sources

WattTime

- **Description**: WattTime is a nonprofit that offers data that can help make decisions regarding when devices should operate to reduce overall emissions (without additional cost). They provide MOER data associated with Balancing Authorities in and outside of the US. They have MOER data in 5-min intervals for every balancing authority dating back two years from today (and even have forecast values for up to 24 hours ahead of the current time). Interestingly, we had to sign a Non-Disclosure Agreement to get access to the MOER data as it is proprietary - and got approval to share any work we do with our instructors! Someone at WattTime also shared with us the spatial locations and attributes of all of the Balancing Authorities recognized by Wattime in GeoJSON format.
- **Size**: ~17,000,000 data points worth of MOER values for each of the over 60 US balancing authorities (which, in total, is ~860 Mb).
- **Location**: https://www.watttime.org/get-the-data/api-documentation/
- **Format**: JSON (HTML) for some data, CSV for other data (most of the data we used got downloaded as a *.zip file after a query to the API). The BA GeoJSONs were shared by a member of WattTime with us as a *.zip file.
- **Access Method**: API Query

Environmental Protection Agency (EPA) - EGrid Data

- **Description**: In coordination with the Energy Information Agency (EIA), the EPA releases an extensive report (seemingly every four years) of power plant energy generation and emissions data (broken down by power plant, balancing authority, state, etc.). We used the 2018 report to learn about the energy and emissions landscape and provide context to times before the COVID-19 lockdowns went into effect.
- **Size**: ~10.2 Mb
- **Location**:

  United States Environmental Protection Agency (EPA). 2020. “Emissions & Generation Resource Integrated Database (eGRID), 2018.” Washington, DC: Office of Atmospheric Programs, Clean Air Markets Division. Available from EPA’s eGRID web site: https://www.epa.gov/egrid/emissions-generation-resource-integrated-database-egrid
- **Format**: XLSX Excel Workbook
- **Access Method**: Data downloaded from the EPA website as an Excel Workbook.

US Energy Information Agency (EIA)

- **Description**: Net Energy Generation for all sectors by month by state
- **Size**: 126 KB
- **Location**: https://www.eia.gov/electricity/data/browser/
- **Format**: CSV’s
- **Access Method**: Direct click and download

COVID-19 Case & Lockdown Data

- **Description**: Institute for Health Metrics and Evaluation (IHME) “COVID-19 Mortality, Infection, Testing, Hospital Resource Use, and Social Distancing Projections” was used for our COVID-19 data. IHME provides projections as well as observed data for all of the metrics stated above. Their projections consist of a “Current” projection as well as a best and worst case projection. Since we only used observed data, picking between these individual files was of less concern. The file actually used is “Worse_hospitalization_all_locs.csv”. Again, the *observed* data was used and not the “worse case *projection*” data. The two data fields of most significance were the COVID-19 cases per 100k population and the State field as that allowed joining to other datasets. We had hoped to incorporate the lockdown dates provided in “Summary_stats_all_locs.csv”, but the data was ultimately grouped to monthly values which decreased daily differences such as lockdown start and end dates.

- **Size**: 43 MB
- **Location**: http://www.healthdata.org/covid/data-downloads
- **Format**: CSV
- **Access Method**: Direct click and download

State Shapefile

- **Description**: State boundary shapefile: cb_2018_us_state_500k
- **Size**: 2.2 MB
- **Location**: https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html
- **Format**:State boundary shapefile
- **Access Method**: Direct click and download

Data Manipulation, Analysis and Visualization

EPA’s eGRID Data (2018)

To see samples of the Data Manipulation performed to Analyze and Visualize the data, check out a sample Google Collaboratory Notebook HERE.

We used this data source to explore the energy and emissions landscape of the United States. This data set is seemingly published every four years and includes data that has been more scrutinized and reviewed, and is thus more reliable. It includes an abundance of data related to energy generation and emissions broken down by power plant, state, Balancing Authority, and more. These first four visualizations serve as a teaser to check out the interactive Google Colab Notebook linked to above!

FIGURE 1 shows the location of each power plant in the United States color-coded by the main fuel type at the power plant. Solar out outnumbers the other power plant types significantly with over 3500 different plants, but they are also much smaller in size.

FIGURE 2 plots the top 50 power plants by Net Energy Generation (MWh) in 2018 (encoded in the size of each circle). Notably, more than half of the top energy producers in the United States are nuclear power plants.

FIGURE 3 plots the top 50 power plants by CO_2 emission (lbs) in 2018 (encoded in the size of each circle). All 50 are coal-fired power plants.

FIGURE 4 plots the top 50 power plants by pounds of CO_2 emitted per MegaWatt-hour (lbs/MWh) in 2018 (encoded in the size of each circle). We will discuss similar units later in this report, but this is essentially a measure of efficiency of each power plant when it comes to harmful emissions. Interestingly, the worst CO_2 emitter is clearly a natural gas power plant in California. Though coal-fired power plants have the highest emissions, they also produce a lot of energy (which is part of why in the past it has been hard for the US to move away from them).

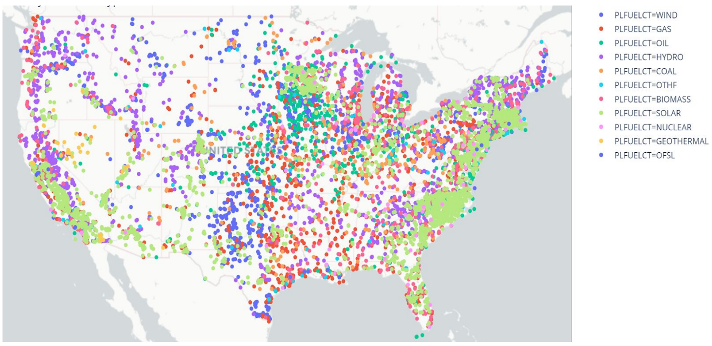


FIGURE 1

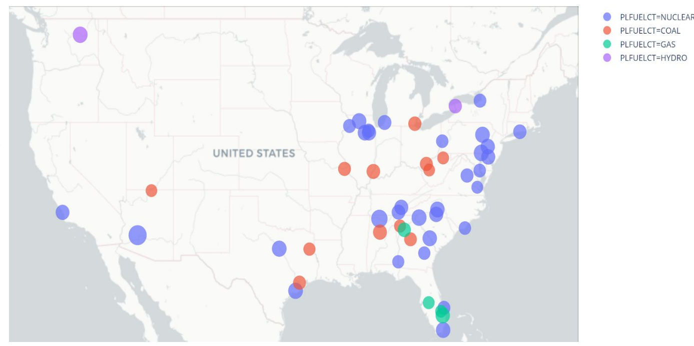


FIGURE 2

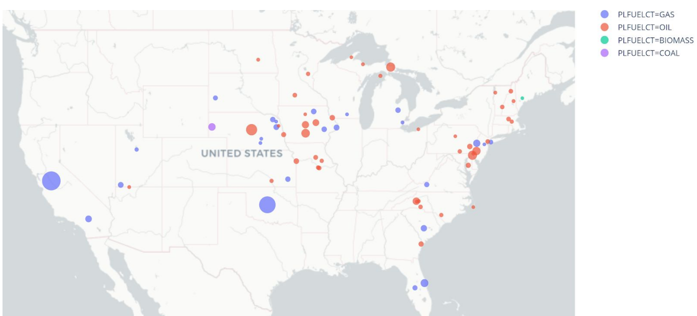


FIGURE 3

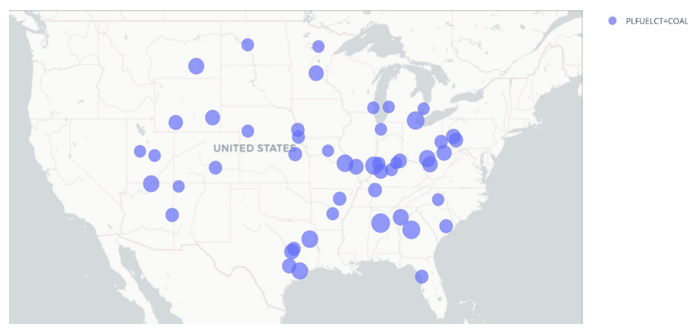


FIGURE 4

In **FIGURE 5** below, you can see that for combustion power plants like coal and natural gas (and to a lesser extent oil) there is a strong correlation between the net energy generated and CO_2 emitted. Nuclear power stands out in this graph as having a high energy output, and essentially no emissions for most plants.

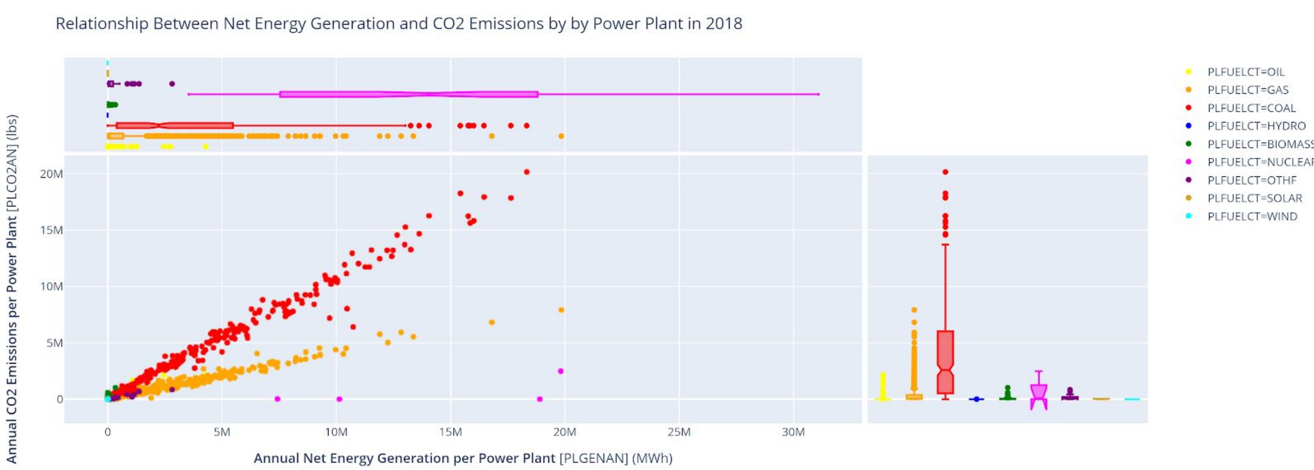


FIGURE 5

By aggregating the plant data by state (which was done by the EPA), in **FIGURE 6** you can see there is a fairly strong correlation between net energy generation and CO_2 emissions which is largely related to the presence of combustion fuels in each state (for example, the data in this graph colored by each state's energy generation by coal-fired power plants).

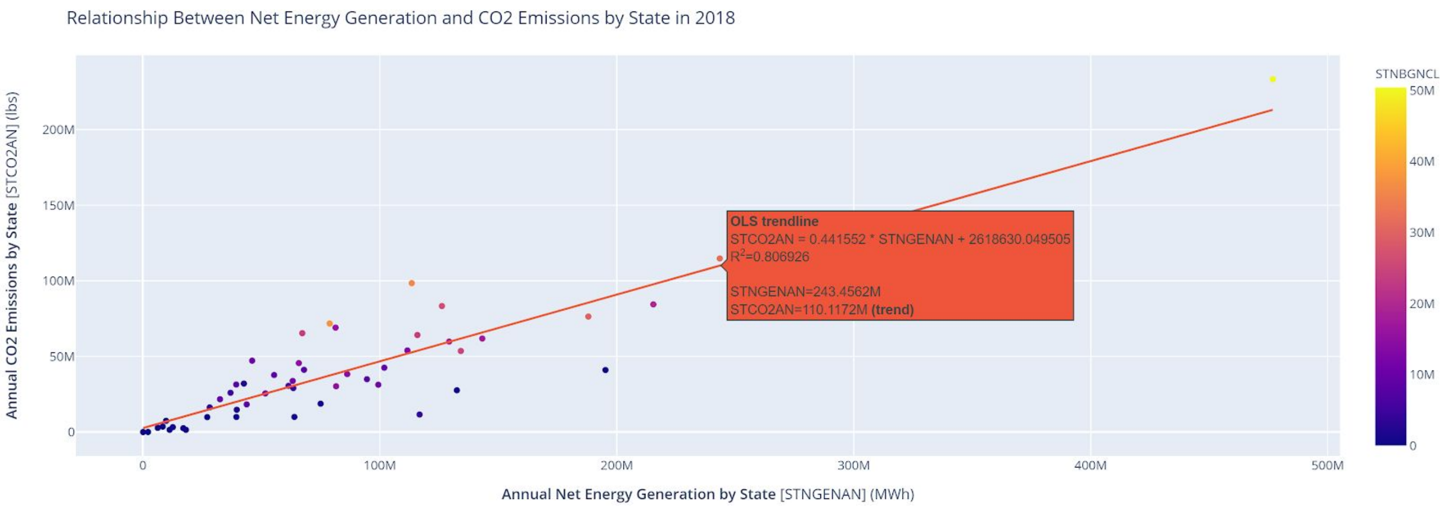


FIGURE 6

WattTime & Marginal Operating Emissions Rate (MOER) Data

To see samples of the Data Manipulation performed to Analyze and Visualize the data, check out a sample Jupyter Noteboopk in our roject google folder.

When WattTime was founded as an organization, a driving question of theirs was: "how clean is the electricity I am using right now?" They set out to find ways to empower people and organizations to use electricity during times of the month - and even times of the day - that reduce their carbon footprint. **Not all energy is created equally** - particularly from the standpoint of CO_2 emissions - and the WattTime founders are well aware of this and how the complexity of handling energy demands and corresponding emissions varies even at scales as minimal as 5 minute intervals.

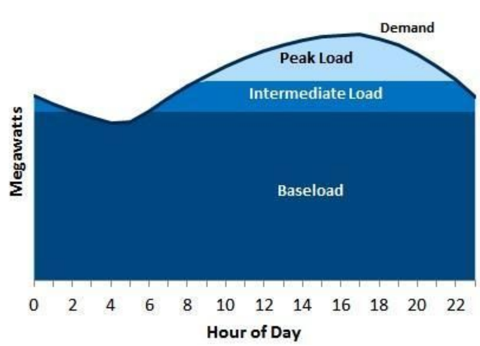


FIGURE 7 (LINK)

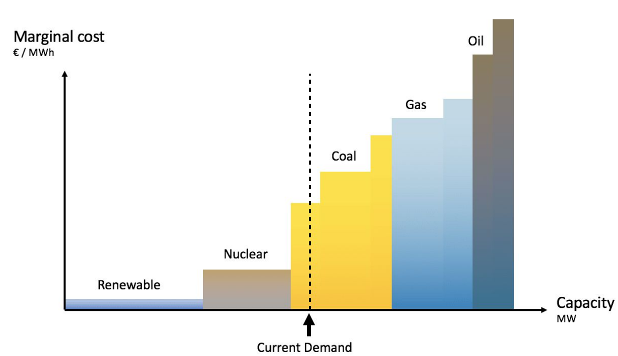


FIGURE 8 (LINK)

The above figures show graphical representations of two important concepts. During a given day, demands on the energy grid vary (illustrated in **FIGURE 7**). As the demands on the grid increase, the BA in charge of the region needs to respond by generating more energy, but power plants in their control have certain capacities, which means that as the demand increases, the BA effectively needs to bring into operation more power plants. Because of this, a BA will tend to use certain power plants for their “baseline” daily loads. These power plants tend to be the ones that are less expensive to operate, which - due to federal regulations and due to improvements in technology - are also usually the power plants that have lower overall emissions[1] (e.g. solar, nuclear, hydroelectric, etc.). As illustrated by **FIGURE 8,** as demand increases, BA’s will resort to operating power plants that are more costly and also have more harmful emissions. To capture this complex dance of energy demand and emissions - WattTime measures the **Marginal Operating Emissions Rate (MOER)** which is a reflection of how many pounds of CO_2 is emitted by the *marginal* power plant (effectively, the powerplant that either most recently came into operation or is about to) per MegaWatt-hour (MWh) of energy it generates. In **FIGURE 8**, this theoretical BA already has renewable power plant in operation (e.g. wind, solar, etc.) and a nuclear power plant, as well as a coal-fired power plant. The three levels in yellow effectively represent three coal-fired power plants of increasing cost of operation, and likely emissions. In this visual, the *marginal* plant is the first yellow bar (where the demand line is) because it is in use and, since that powerplant is not yet at its capacity, as demand increases, that power plant remains the one most recently “turned on”. Therefore, the MOER value would be a reflection of the emissions of this particular power generator (and is thus different from the simple average of all power plants currently in operation).

Our goal was to use the MOER values available through WattTime’s API to learn about real-time emission rates as they have MOER values (measured in lbs CO_2/MWh) generated every 5-min for each BA in the US (and even many around the world). The reason for this is the energy demand, and thus energy generation and emissions, varies noticeably throughout the day.

To illustrate this, we created a few graphs of MOER values for New Hampshire, which is part of the Balancing Authority called Independent System Operator New England (ISO-NE), on August 30[th], 2020. **FIGURE 9** shows how the MOER value varied significantly over just one hour. It would indicate that if you were to, say, run your water heater or charge your electric car, doing it after 4:35 PM would mean you have a lesser carbon footprint.

FIGURE 9

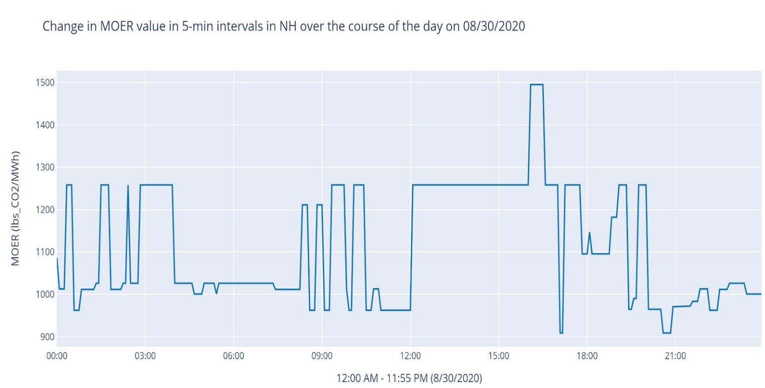


FIGURE 10

[1] The exception to this would be if there is a power plant that might have a slower or more-expensive start-up time, it might also make up what is considered a “baseload” power plant. (LINK)

FIGURE 10 shows how dramatically the MOER value can change during the day. Part of the reason the graph is not "smooth" is because, again, the MOER value is a reflection of the CO_2 emissions of the *marginal* power plant, not the current total emissions rate. Averaging these values daily over the month of August 2020 yields **FIGURE 11** below (the boundaries of the gray-scale indicates the daily max and min values over the course of the month.

FIGURE 11

The fact that the average value is often closer to the min value is a sign that the 'dirtier', marginal plants are likely not as needed due to energy demands (supported by the lone 'peak' of the MOER graph in **FIGURE 10**). WattTime has historical data dating back approximately 2.5 years, and plotting the month average min, max, and mean yields the graph shown in **FIGURE 12** below.

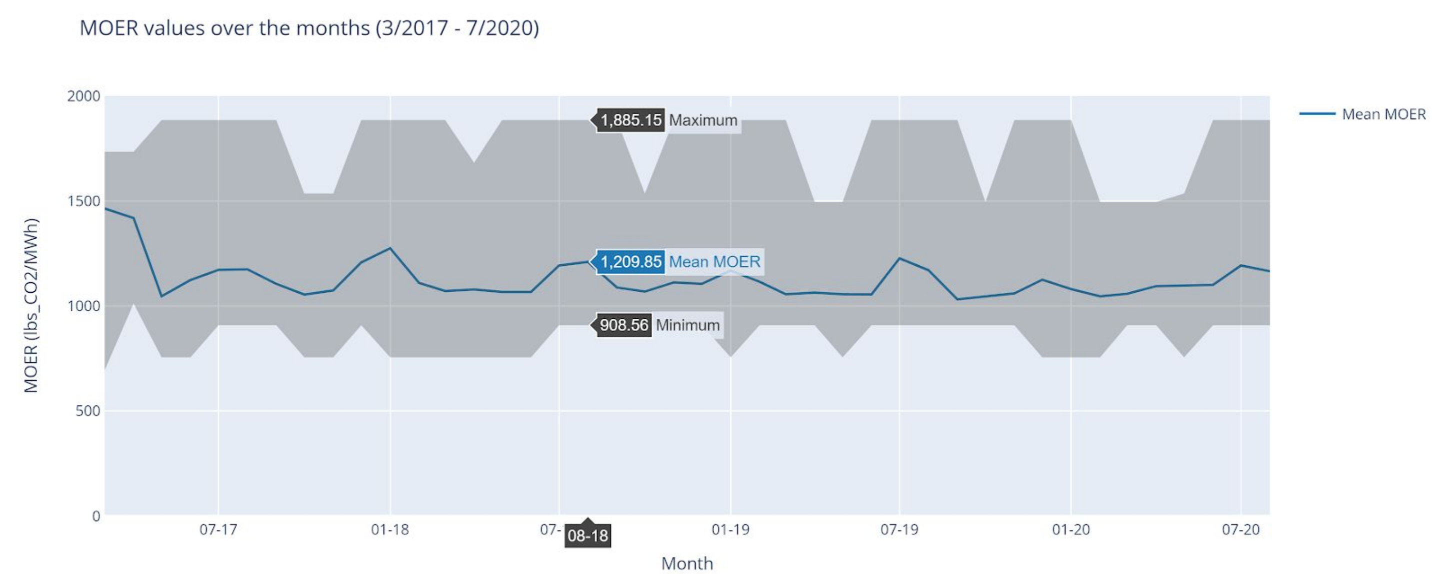


FIGURE 12

FIGURE 12 becomes especially interesting when compared to a similar timeframe for other BA's. For example, **FIGURE 13** shows a similar MOER graph for the BA 'Florida Power & Light" (FPL). Notably, the range for min and max is appreciably less.

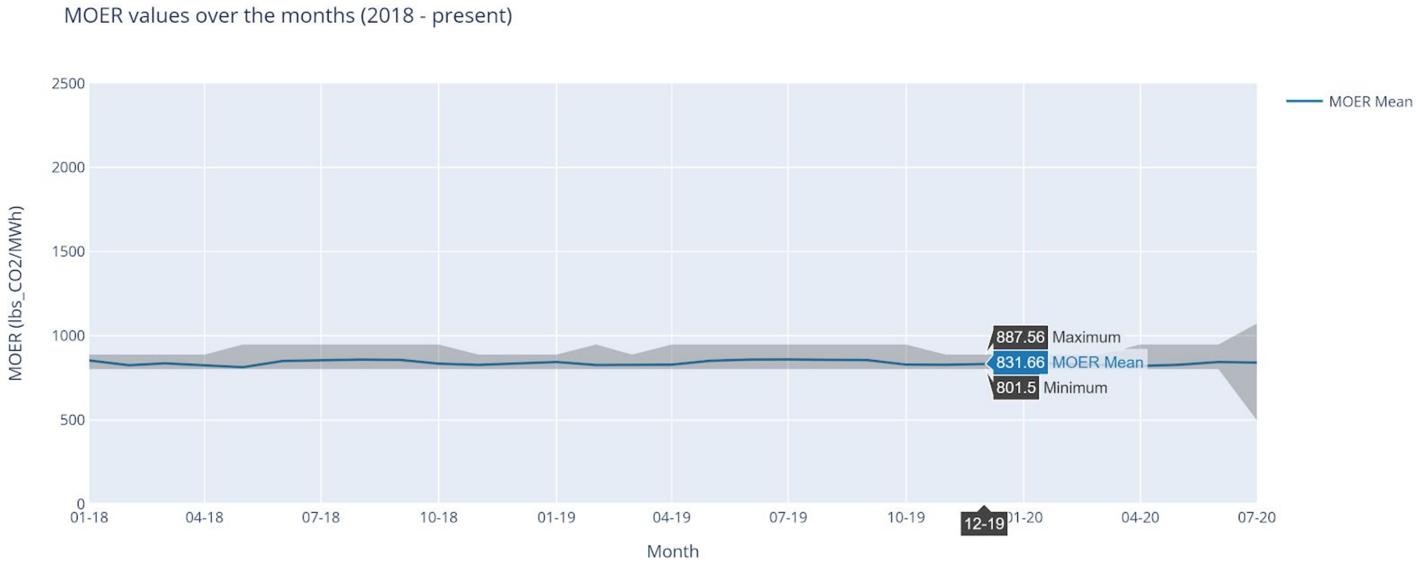


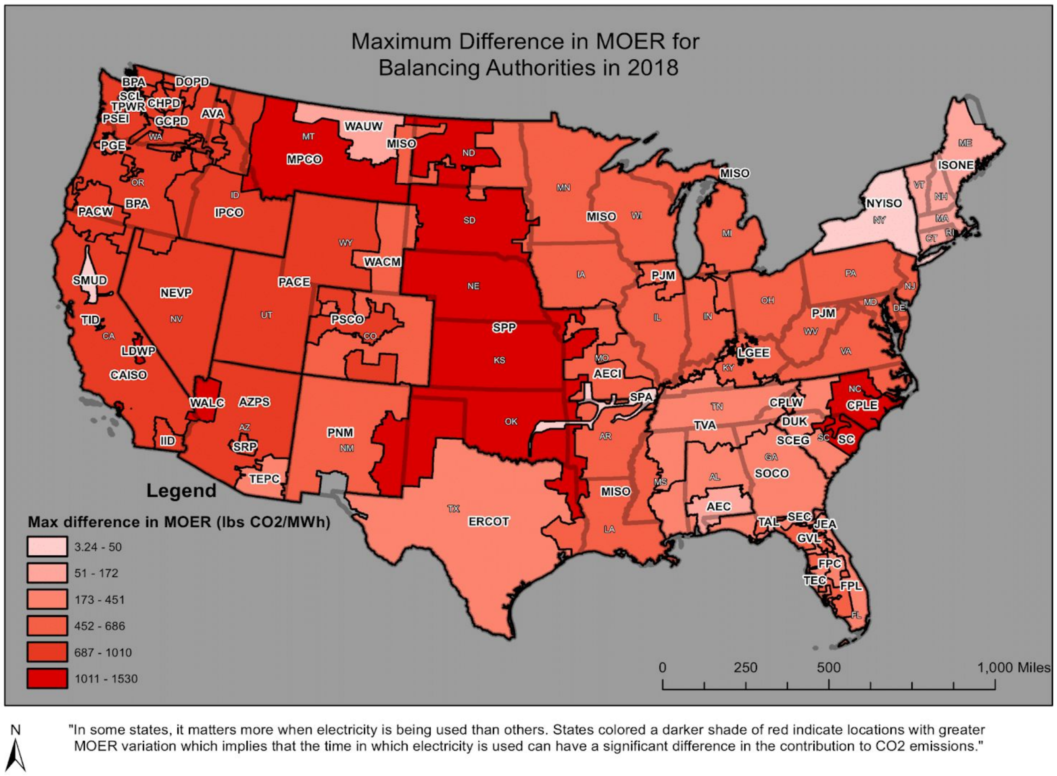
FIGURE 13

This is consistent with the fact that one of the main power plants that handles a good deal of energy load is "Turkey Point" containing a Nuclear Power generator which has effectively no emissions and it appears that some of their higher-emission, marginal plants are less needed. The 'dip' in min MOER value in July is consistent with a COVID-19 lockdown in Florida, but you can see the average is similar in the same month in another year. MOER value, because it is based on the emission of the marginal plant only, being not the same as the total emissions rate, did not decrease reliably during COVID-19 lockdowns (though it was more noticeable in some states with stricter lockdowns).

The stark difference between **FIGURE 12** and **13** got us wondering about which BA's having the highest and lowest swings in min & max MOER values. **BA's that have a higher difference** would ideally pay *more* attention to MOER values than BA's (and their corresponding states) that have minimal variation.

As a result, we wrote a function that would iterate overall all US BA's and created a DataFrame that contained the min and max values that occurred in 2018. We decided to plot the difference between the min and max MOER values by BA, but needed to map the corresponding balancing authority limits to a US map. We retrieved geojson files from Wattime's API that represented all of the balancing authorities as separate geojson files, which meant that they would need to be merged into one for visualization purposes. The individual geojson files were read into a list, tested for corruption (null geometry attribute) and added to another list. This list was then added as a value that pertained to a predefined dictionary key. Once this was done, all of the individual geojsons existed as their own nested dictionary within a list as a value to the original dictionary key. There were a lot of nested dictionaries and lists so it is a bit cumbersome to communicate in written language. **This can be observed in the Wattime_geojson.ipynb within the Data Manipulation folder.**

We encountered problems using the Plotly library to visualize the balancing authorities, so we instead took to ArcGIS to map the balancing authorities to a US map and their corresponding MOER Data. The result is shown in FIG. AAAA below. Notably, the BA's containing New York (NYISO), part of California (SMUD) and one juxtapositioned along the AR-MO border (SPA) show the lowest variance in MOER value while the BA's containing parts of SC, NC, NE, KS, etc. - those reddest on the graph - indicate those with swings in excess of 1000 lbs CO_2/MWh which can sometimes even occur within a given day. The residents and businesses in states that are the reddest in **FIGURE 14** would benefit *more* from considering MOER data to inform energy use, where flexible. (NOTE: We do not believe tornados are a factor in putting SPP in the red).



"In some states, it matters more when electricity is being used than others. States colored a darker shade of red indicate locations with greater MOER variation which implies that the time in which electricity is used can have a significant difference in the contribution to CO2 emissions."

FIGURE 14

Like net energy generation to CO_2 emissions in the eGRID data set, we wanted to see how the average MOER value correlated to the net energy generation. To explore this, we queried the Energy Information Administration's (EIA) API to get the hourly net energy generated for ISO-NE (the balancing authority for New England), which involved doing a similar aggregation of values by month to the MOER data. After merging the two data sources, **FIGURE 15** shows the interesting result, which is consistent with what we know about MOER values. It looks like net energy generation under 10,000 MWh shows no correlation, which is likely the range when power plants are using lower-emissions energy sources (also frequently in the winter months as indicated by the cyclic color scale indicating the month). After 10,000 MWh, there is a positive correlation of coefficient 0.64. This is not as strong as the correlations shown in **FIGURES 5** & **6**, but the change in MOER value should be somewhat 'step-wise' in that it should jump up or down based on the marginal power plant.

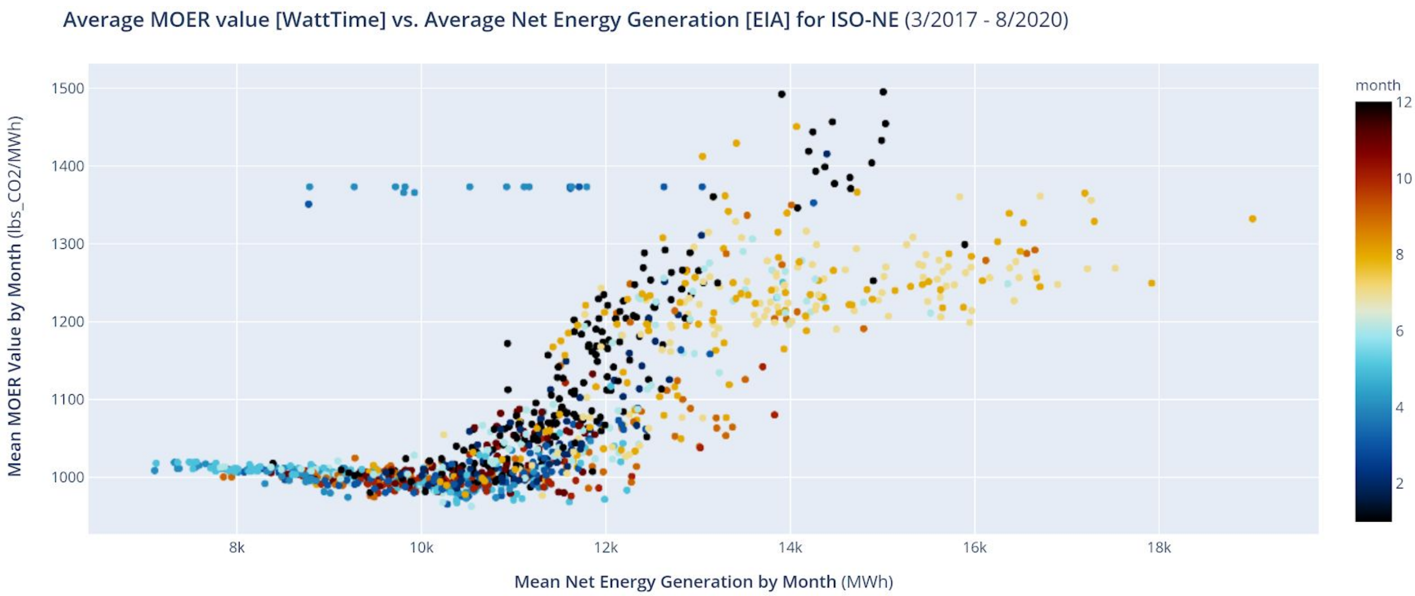


FIGURE 15

EIA Net Generation Data and COVID-19 Data

We wanted to also see if we could encode both COVID-19 data and energy data in the same visualization. We had intuitions from the beginning that these variables are likely related in some capacity. We ended up using COVID-19 infections per 100k population as our COVID-19 metric and the net generation for all sources across all sectors (residential, commercial, industrial, etc.) for our energy metric.

In terms of encoding, we thought COVID-19 infections were a great excuse to use something called a Cartogram. A cartogram scales some shape (often a true to proportion state) to a quantitative variable. This means that each state would be morphed in some fashion - either larger or smaller (sometimes about the same size) - in accordance with the COVID-19 cases per 100k population it had for the given month. An important property about Cartograms is that they do not compromise spatial relationships. This means that every state's neighboring states remain true. The color of each state is encoded to that state's net energy generation for the given month as a percentage of the average net generation for the given month compared to the most recent 5-year average.

Regarding data manipulation and cleaning, the COVID-19 data only existed from Feb-Sept of 2020, but it was daily data. Conversely, the net energy generation data consisted of data from January 2001-July 2020, but only had monthly data. Further, we assumed that it would be better to compare the net energy generation for 2020 to a more recent timeframe, so we created our "average energy usage" with respect to 2015-present.

The biggest hoop to jump through for this manipulation and subsequent visualization was to get both datasets in a form where each column pertained to an observed date and each row pertained to a particular state as well as a particular metric. The metric to be used for the cartogram was COVID-19 cases per 100k population, and the metric for net energy generation was a percentage of the generation observed during COVID-19 time (March-July 2020) computed for each month for each state as percentage of the average for that month in that state.

The COVID-19 data did not come with its dates pivoted out as columns, so that action had to be done via a Pandas pivot function. Further, the data needed to be aggregated from daily data to monthly. Conversely, the net generation data *did* come with its data pivoted out with columns as observed dates. In order to calculate my percent of average for each state, it was necessary to melt (un-pivot), groupby, and then pivot back out again.

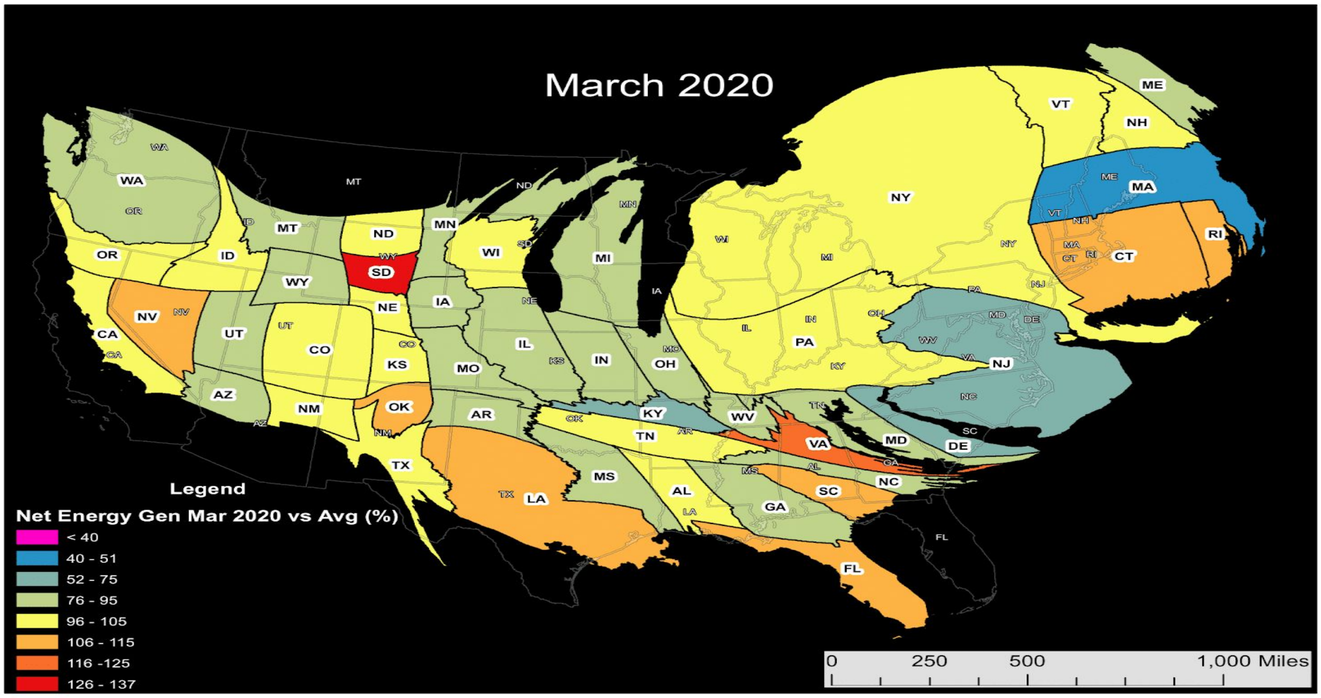
The final join between the two data sets occurred outside of a Python environment and simply as joined tables in ArcGIS. State name was used to join the COVID-19 data to the state shapefiles and the state abbreviation was used to join in the net generation data. This resulted in a table with states as rows as well as COVID-19 per 100k population and net generation percentages also as rows. This required “duplicate” columns.

One unforeseen challenge was having to go a step beyond exporting a pandas dataframe to a csv. We had to subsequently convert the csv to a dbf (dBASE table) for successful join into arcmap. This took a while to figure out, but we were able to use Geopandas to read in a csv and convert to a dbf.

NOTE: There were 5 different cartograms generated between March and July 2020. There are static images as well as an animated webm file located in the Visualizations folder. Please look at these to give yourself the proper perspective.

Cartograms can be tricky, so having the original state sizes subtly beneath is helpful for interpretation. It is important to recall that energy is sometimes generated in one state but is actually transferred to another state (or Balancing Authority). As such, looking at net energy generation with respect to state without controlling for generation imported and exported should be considered not entirely precise. Despite this, comparing these values over time in correspondence with a notion of COVID-19 severity presents a very intriguing comparison. As the cartogram progresses through time month by month, you can observe how the COVID-19 cases spiked and retreated in different locations. You can simultaneously see which states were generating less than average (blues), about average (yellows) and more than average (oranges and reds) energy.

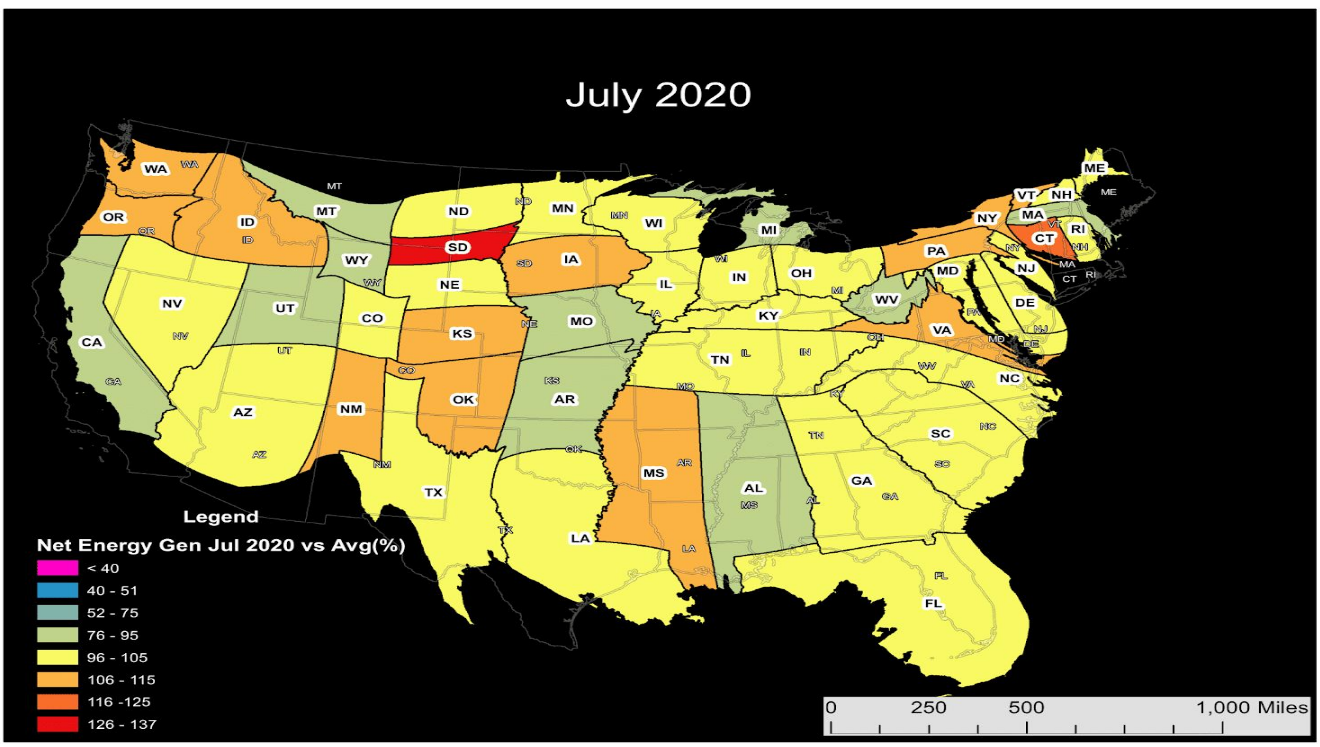
Although we did not find a correlation between the values of COVID-19 cases per 100k population and relative net generation percentages, it is clear that earlier in the year such as March (perhaps when lockdowns were most stringent), that there were more blues and greens and less oranges and yellows which indicates below average or average energy generation. The cartograms in June and July contain almost no blues and have a greater number of “average” and “higher than average” generators. This comparison is possible as seasonality has been taken into account. Since the values are a percentage of the 5 year average for each month as opposed to the absolute net generation, lower generation due to cooler spring temperatures and higher generation due to hotter summer temperatures is controlled for.



N

U.S. scaled to Covid19 Cases Per 100k Population
2020 VS Average Net Energy Generation
March 2020

Figure 16



N

U.S. scaled to Covid19 Cases Per 100k Population
2020 VS Average Net Energy Generation
July 2020

Figure 17

Summary

It does not take long when trying to digest and understand the energy landscape in the US to realize that it is largely a moving target. The complexity of such an intertwined system becomes apparent when trying to quantify things like energy generation, energy demand, and energy consumption etc. (which are all measured by the EIA). An additional dimension of complexity is added when interchange between balancing authorities is incorporated and then again when considering that not every Megawatt and Megawatt hour is equal in terms of its corresponding emissions. This certainly does not render analyses useless! Understanding energy (something that doesn’t come from simply flipping a light switch and receiving a bill) just takes an extra effort to communicate its nuances, often significant nuances on a large scale like environmental impacts pertinent to climate change. If you have an electric car that you charge every night, is a coal-fired power plant generating that electricity?

Our findings regarding a quick peak at COVID-19 merit further review. We were not able to obtain *daily* data with respect to state for the EIA net generation data. There was daily data available, but this was broken down by balancing authority, which meant it was not possible to merge and compare with the state-level data that defined the COVID-19 data. It is possible that we may see a more notable relationship between COVID-19 data (effectively dates when lockdown was more strict) and *daily* energy generation or consumption use. We will be continuing to dig for such data beyond the scope of this version of the report.

Our findings regarding the balancing authority’s MOER data differences, we feel, are worthwhile. We believe that studying a metric as simple as degree of MOER variation could serve as a method to select areas that would achieve a greater benefit by inheriting MOER data into their BA’s energy management approach, or even petitioning local government to change primary energy sources (because if people are making environmentally-conscious choices, but the supporting power plants make those decisions less impactful). We do not currently see such an analysis provided by WattTime who provided the data originally, and we plan on engaging them in a discussion based upon our analysis.

Statement of Work

Paul and Chris both have natural interests in energy and emissions given their scientific backgrounds in Physics and Geographic Information Systems, respectively. Paul was able to reach out to a former student at Wattime who provided access to their API which was used to access balancing authority data as well as MOER data used in several of our analyses. Paul worked extensively with the eGRID data, creating **FIGURES 1 - 6.** Paul also led the charge in terms of understanding the MOER data. He learned how it pertained to our project narrative, and performed the analyses via substantial data cleaning and manipulation for use in visualizations such as **FIGURES 9 - 13, 15**. Chris was able to leverage his knowledge and access to ArcGIS to create multilayered spatial analysis seen in **FIGURES 14, 16 and 17**. This consisted of some heavy data lifting and manipulation to get the various data sets into a format that could be joined to the US State shapefile in GIS. Additionally, the data was also morphed into a format that allowed encoding of multiple variables in a single visualization which provided a more unique perspective on the relevant data.