Exploratory Data Analysis: Emissions from Electric Power Generation in the US 2010

08 October 2021
Frederick Tan
CEE 322 - Lab 1

Introduction

Electricity generation is crucial to our daily lives. However, the process of generating electricity is often associated with many emissions; when primary fuel sources are turned into electricity at the power plants, emissions are an output of the process. One of the many activities monitored by the Environmental Protection Agency (EPA) involves air quality. In an attempt to plan allocation of resources, regulation of technologies, and identify opportunities for emissions reductions, this paper will analyze a data set containing emissions characteristics of electric power plants in the US for 2010.

The paper will work to identify the largest contributors to emissions within the electricity generation sector. It will then identify trends in the emissions data to determine areas to focus improvements and propose regulations that may impact emissions.

Dataset

The dataset used contains characteristics for electric power plants in the US for 2010. Basic power plant information includes a unique ID for each plant, the State, plant name, county, and latitude and longitude. Plant energy resource data includes combustion status (i.e. whether the plant is a combustion plant, partial combustion, or non-combustion), the primary fuel source, and the fuel generation category. Plant energy generation data includes plant capacity [MW], heat input [MMBtu], and net generation [MWh]. Emissions data include CO2, SO2, and NOX emissions in short tons as well as input and output rates for each emission type. Finally generation values based on the energy category (non-renewable, renewable, non-combustion, and combustion) are included.

Initial preparation of the data includes calculating various ratio parameters. The capacity factor is the ratio of actual output to maximum potential output per year. The heat rate ratio quantifies how efficiently primary fuel is converted into electric energy. And annual output emissions quantify how much gas pollutants are generated per unit of electricity. These ratios will be used later in the analysis to characterise the plants.

While the dataset includes information of renewable energy and mixed source power plants, this paper will focus on combustion plants for the following reasons:

1. 59.94% of the electricity output comes from combustion plants. Since the majority of electricity generated comes from combustion plants, the study will focus the study on the combustion plants and remove the non-combustion plants from the dataset.
2. Furthermore, only 2.14% of the plants are partial combustion plants and these partial plants only generate 2.87% of the US electricity . Thus the study will focus on the full combustion plants as they are the main type of combustion plant.

Additionally, plants typically utilize a mixture of fuel sources for generation. However this study will characterize each plant based on its primary fuel source. The five main types of primary fuel

source are oil, gas, biomass, coal, and “other fossil fuels”. For just the plants with oil as the primary fuel source, 97.28% of the generation comes from oil. For just the plants with gas as the primary fuel source, 96.72% of the generation comes from gas. For biomass, coal, and other fossil fuels, the percentage of generation within their categories are 89.35%, 96.22%, and 87.06% respectively. Since the plant’s primary fuel source always accounts for around 90%+ of the generation, the study will only focus on the emissions from the primary fuel sources of each plant.

Analysis

Characterization of Emissions based on Fuel Type

To begin the analysis, we first want to characterize which fuel sources produce the most emissions. Figure 1a and Figure 1b show that Coal produces a significant amount more of CO2, SO2, and NOX emissions compared to the other primary fuel sources. Additionally, the y-axis of the two figures show that CO2 emissions are three orders of magnitude larger than SO2 and NOX emissions.

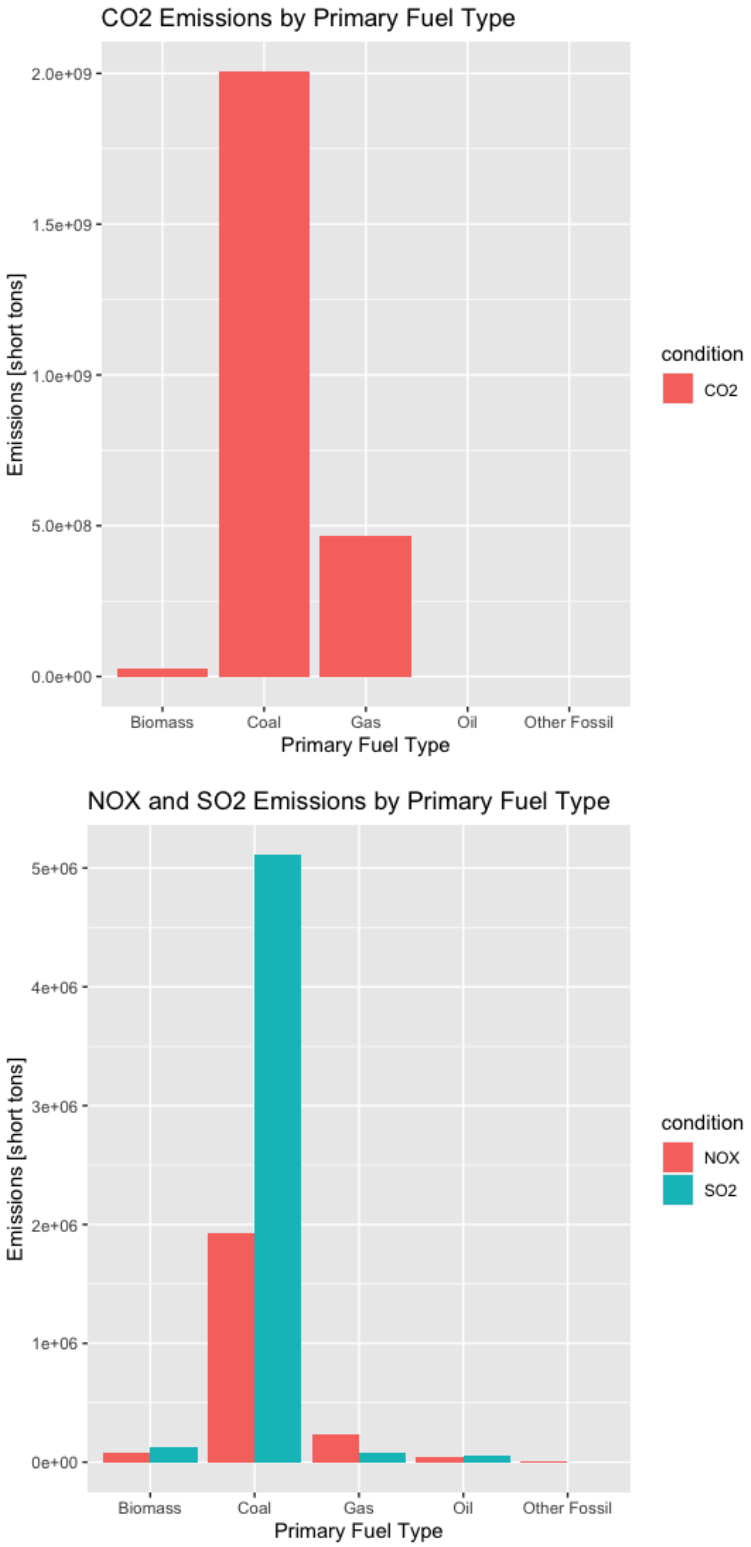

Figure 1. (a) CO2 emissions from primary fuel sources. (b) NOX and SO2 emissions from primary fuel sources.

Plant Characteristics

Emissions based on Net Generation

After determining CO2 is the main pollutant being produced, we then wanted to explore how plant characteristics other than the primary fuel type affects the emissions. First we looked at the net generation of the plant compared against the CO2 emissions for all combustion plants. Figure 2 shows there is a linear relationship between the net generation and the CO2 emissions; the larger the annual generation of a plant, the more CO2 emissions they emit. It should also be noted that there appear to be two relationships, denoted by the clustering of the data points into two lines with different slopes.

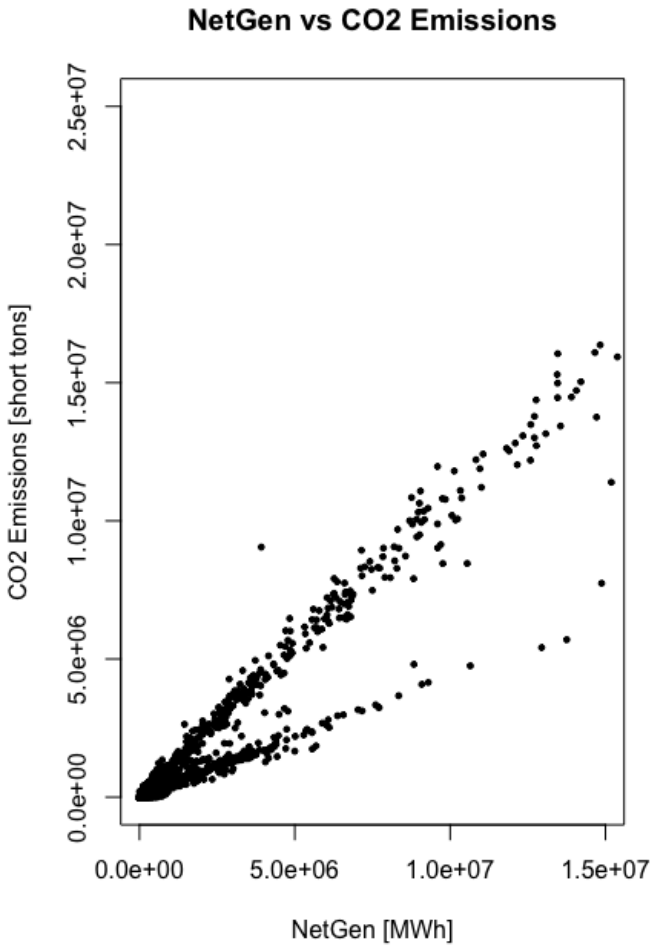

Figure 2. CO2 Emissions based on net generation of the platns

Net Generation and Location

Given the relationship between net generation and CO2 emissions, we then wanted to explore more characteristics of the net generation. We looked at the net generation relative to plant location, based on latitude. In Figure 3, we can see two groupings: group one between 20 - 45 degrees latitude and group two between 55-70 degrees latitude. These groups most likely correspond to mainland USA (group one) and then Alaska and Hawaii (group two). In group one we see a much larger range of net generation while group two has relatively low generation. This comparison, and the relationship between CO2 emissions and net generation, shows that efforts should focus on mainland America plants since they have the widest range of net generation and thus more room for improvements.

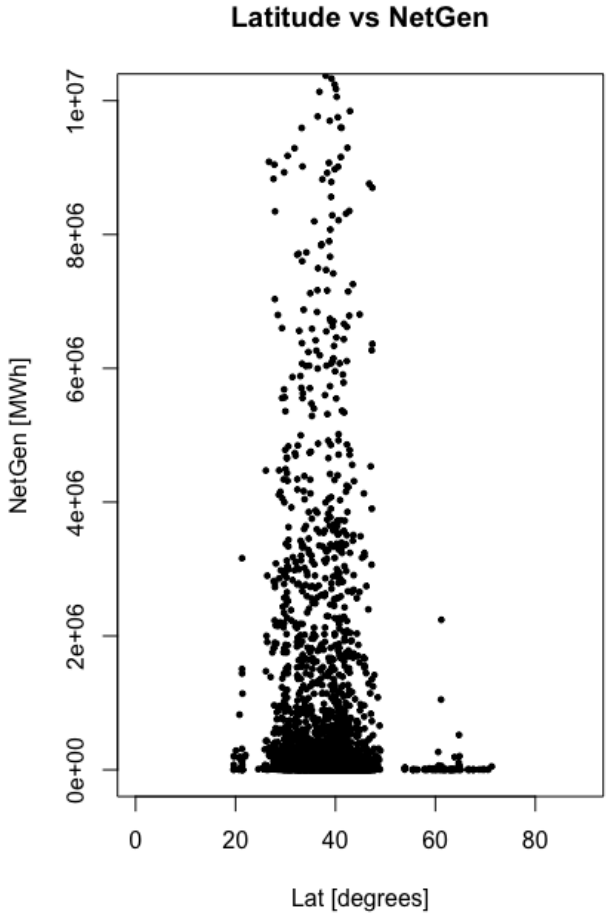

Figure 3. Net generation based on latitude of the plant

Emissions based on Heat Input and Heat Rate

Finally, we analyzed the relationship between heat rate and heat input versus CO2 emissions. The heat rate characterizes how efficiently the primary fuel is converted into electric energy.

Figure 4 shows the data skewed right, suggesting that higher heat rates (and thus lower efficiencies) do not necessarily correspond to higher emissions. Therefore the heat rate of a plant should not be a consideration for controlling emissions. On the other hand, Figure 5 shows that higher heat inputs corresponds to higher CO2 emissions. And similar to Figure 2, we see the plants separate into two clusters, denoted by the lines with differing slopes.

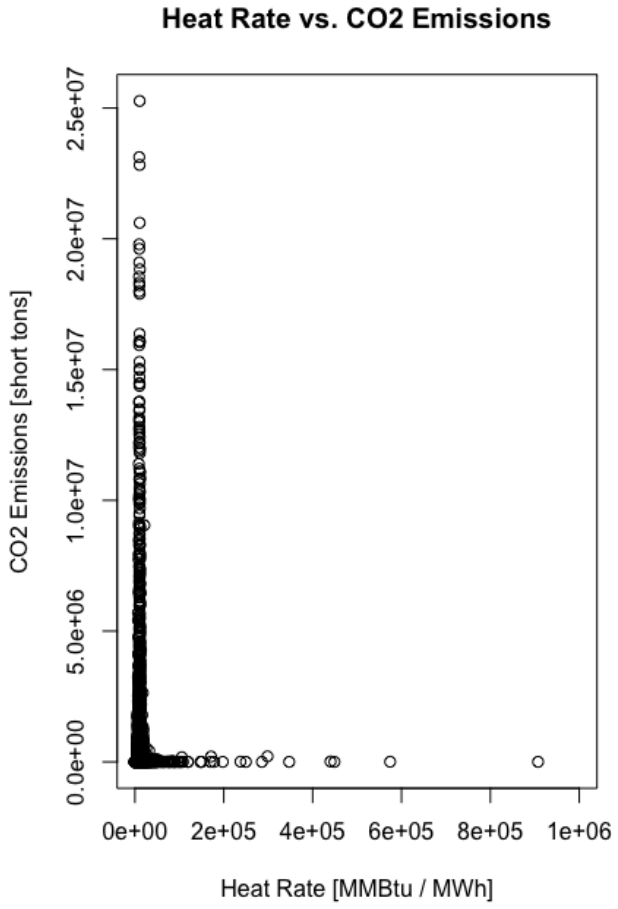

Figure 4. CO2 emissions based on heat rate of plants

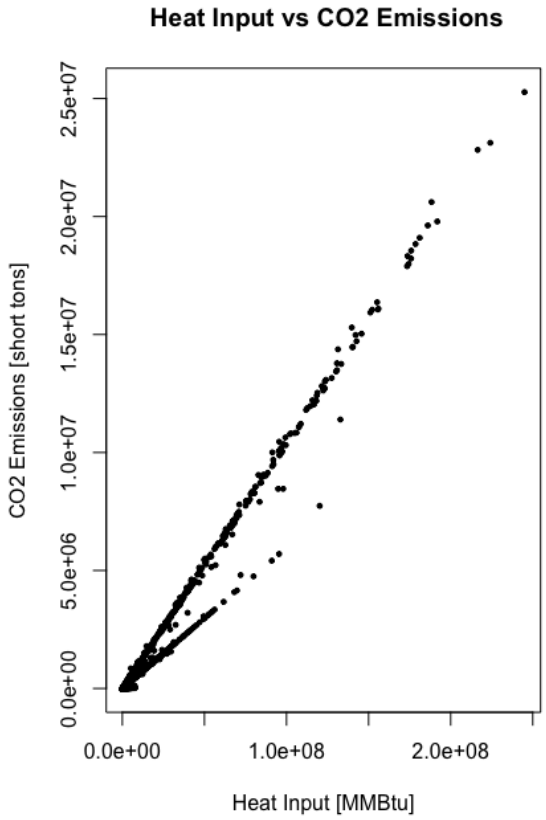

Figure 5. CO2 emissions based on heat input.

Conclusion

In this study we analyzed power plants, power generation, and emissions for power plants in the USA during the year 2010. From the exploratory data analysis, we found that plants that utilize coal as their primary fuel source produce the most emissions. Further, we found that CO2 is the main pollutant being emitted. Next, we showed a positive linear relationship between net generation of a plant and its CO2 emissions and showed that the plants with the highest net generation were in mainland America. Finally we found that heat rate does not necessarily affect CO2 emissions, but heat input does appear to have a positive linear relationship.

As mentioned in the analysis, the net generation versus CO2 emissions and heat input versus CO2 emissions showed two clusterings of power plants. Additional analysis should be done to determine the source of the clustering.

After an initial analysis, with suggestions for future analysis, we propose the following general conclusions:

1. The EPA should focus on coal burning power plants as they produce the most emissions

2. Metrics for emissions can focus on CO2 as they represent the highest short tons of emissions from plants
3. Efforts and policies should focus on high net generation and high heat input plants as they produce the most emissions and have the most to benefit from improvements.
4. Heat Rate, or plant efficiency, is not a useful metric to focus on improving in order to reduce emissions, but total heat input is useful.