# Technical Appendix

## AESO Project: Cluster Analysis of Alberta Electricity Generators

Keziah Chong, Clodagh Berg, Zhihao Wang, Sean Tan, Emily Deuchar

April 22, 2024

## 1 Research Question and Motivation

What insights can be gained from conducting cluster analysis on various characteristics of electricity generators to enhance our comprehension of the Alberta electricity grid? This is our overarching research question, but we are looking at more specific questions. For example, do generators that have more influence over the electricity price exhibit strategic behaviour in their bidding? Do some generators bid with a larger standard deviation than other generators? This initial analysis focused on combined cycle plants and then extended to other electricity generators that contribute to the grid.

There are two parts to answering these research questions. First, we created a tool that uses cluster analysis to analyze AESO's merit order data and then we were able to analyze bidding behaviour and market power using that tool. This helps us better understand the Alberta electricity grid and gives us insights into market power. Our tool allows us to identify potential instances of market power being exercised.

There are policy implications of this as well. By being able to better understand the behaviours of the electricity generators, policies can be developed to keep electricity affordable for Albertans.

## 2 Background

Every hour, generators submit bids indicating the price at which they are willing to supply electricity. The Alberta Electricity System Operator (AESO) is an independent operator in charge of operating the Alberta electricity grid and pricing information, ensuring a balance between both supply and demand sides. AESO manages and operates the provincial power grid, manages and plans the grid 24 hours day, plans and operates the market, plans the future of the system and its infrastructure, and connects customers to the grid.

A "merit order" is established by sorting bids from the lowest to the highest price. AESO dispatches electricity starting from the lowest-priced offers and moves upward until all demands are met. The price of the last dispatched offer is termed the system marginal price (SMP).

For example, if supply offer #3 is fully met all demands, then the SMP is at $40. If demand was increased by 40 MW (megawatt), AESO would move up the list and dispatch 40 MW electricity of supply offer #4, which would result in an SMP of $50.

| | | |
|---|---|---|
| Supply offer 5 | $ 80 | 25 MW |
| Supply offer 4 | $ 50 | 55 MW |
| Supply offer 3 | $ 40 | 50 MW |
| Supply offer 2 | $ 20 | 20 MW |
| Supply offer 1 | $ 10 | 10 W |

Figure 1: Merit Order

Generators bid based on their marginal cost, and the market price aligns with the point where supply meets demand. This SMP is determined on a minute-to-minute basis. The pool price, calculated as the average of all SMPs within an hour, is also the intersection point of supply and demand curves.

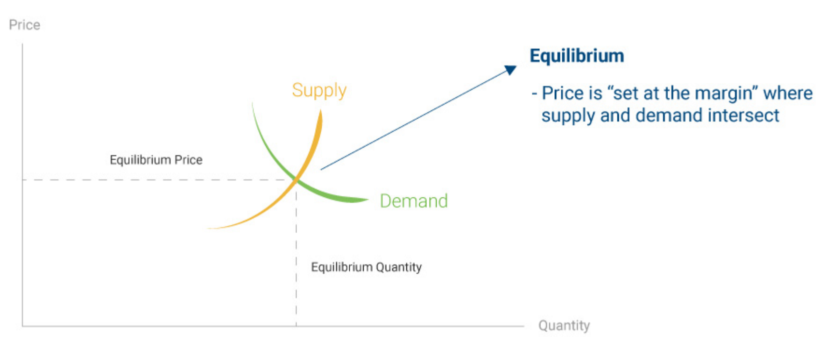


Figure 2: Equilibrium

Our project aims to analyze Alberta's electricity sources based on data from 2023 calendar year. The graph indicates that the primary sources in Alberta are natural gas and wind, with a minor contribution from coal, biomass, solar, and hydro. So far, we have exclusively looked at the six combined cycle plants but plan to extend that.

# 3 Data

For our analysis, we are utilizing the AESO's API to give us comprehensive data about all relevant aspects of Alberta's electricity market. This brings up hourly bidding data including block price, fuel type, asset name, etc. We have access to many years of data that can be used in the tool we created but we looked at 2023 and 2024 for our analysis.

AESO provided us with heat rate and variable O&M data as well, which we used in our marginal cost calculations. Carbon pricing data for the calculation came from the literature.

**Merit Order**: This data will be particularly important in understanding the marginal costs associated with different power plants, which is under the presumption that bidding at marginal cost is typically the most profitable choice, except for Peaker plants. The merit order will provide very useful insight into the variable costs of diverse electricity generators, which, in turn, can be applied to analyzing the effect of cost on the bidding patterns of plants.

**Marginal Price**: By analyzing the heat rate, we can assess the efficiency of various combined cycle generators. The marginal cost is calculated by considering the gas price, heat rate, and carbon tax. Gas price came from NGR, the carbon tax information came from Olmstead and Yatchew (2022), and both the heat rate and the variable O&M were given to us by AESO.

$$Marginal\ Cost\ of\ Nature\ Gas\ Plants = (Fuel\ Cost * Heat\ Rate) + Carbon\ Tax + Variable\ O\&M$$

# 4 Analysis/Empirical Strategy

We were given a lot of flexibility on this project, with the overarching idea aimed at understanding the market dynamics of electricity generators in Alberta. There are six combined-cycle generators that contribute to the grid. Some have seemingly erratic and/or unpredictable bidding behaviors while others appear to bid more consistently. We group the bids of individual generators by hour of day, peak/off-peak, and other groupings, and then look at all the bids from that one generator. We took those bids and by calculating the standard deviation, we could which generators bid more erratically and how they compare to the others.

The profit/market power analysis: The first analysis we did was a profit analysis taking in the bidding data from the AESO merit order data, as well as the marginal cost calculations from above. For this analysis, our team opted to focus on analyzing combined cycle energy plants due to their potential for diverse bidding behaviour, unlike wind and solar which bid right at marginal costs and thus limited strategic pricing actions. Additionally, considering the province's shift away from coal due to environmental concerns, insights into coal plant operations are becoming less relevant. Furthermore, we acknowledged that simple cycle plants tend to exert market power to cover their substantial fixed costs, a strategy not viable for renewable sources.

To do this analysis we compared the highest bid of each combined cycle plant with the calculated marginal cost. This comparison was done where each plant shares the same marginal cost. In our next analysis, we are going to follow the same analysis, but with an average of daily bids.

Cluster analysis: Clustering was done through k-means clustering. The value of k was determined using the elbow method.

Opting for a market power analysis seemed fitting, we believe this analysis could provide valuable insight, given that AESO rarely investigates the strategic behaviour of these entities. Among various types of generators, we chose to focus on combined cycle plants. Our tool now has been extended to include analysis of other generators such as simple cycle and renewables. We used the heat rate data to estimate the marginal cost for each of the six operational combined cycle plants, incorporating the carbon price and gas price into our calculations. The carbon tax and gas price variables will fluctuate, and the value used will be based on the values on the day of data analysis. After determining the estimated marginal costs, we analyzed the bidding behaviour for each combined cycle plant using the AESO merit order data. By plotting these bids on a graph and comparing the bid amount to the calculated marginal cost estimation, we will identify plants whose bids exceed their marginal cost. We also plan to look at the ownership of the generators to analyze whether companies owning multiple power plants are more incentivized to engage in strategic bidding.

# 5 Code

See .zip file also submitted for code and pdf of code. A brief **code description** is as follow:

1. Notebook intakes start and end date, as well as desired generation type (solar, gas, wind, hydro, etc.), then requests the corresponding merit order data from the AESO API.

2. Notebook defines multiple functions to preprocess the data, aggregate the data, construct additional features, conduct the cluster analysis and generate appropriate plots, and generate metrics and statistics about the data and analysis.

3. Plots are achieved through the plotly library, allowing for interactive and customizable 3D plots.

4. Each cluster analysis provides additional tables of information including weighted mean and variance of chosen features, size of defined clusters, dispatch ratio of each cluster, centroid values for each cluster, and all generators included in each cluster.

# References


Natural gas royalty. https://www.alberta.ca/alberta-natural-gas-reference-price.

Derek E.H. Olmstead and Adonis Yatchew. Carbon pricing and alberta's energy-only electricity market. *The Electricity Journal*, 35(4):1–7, 2022.

# A Very Brief Introduction to Cluster Analysis

Zhihao Wang

April 22, 2024

## 1 Idea

Cluster analysis is like sorting things into groups based on how similar they are to each other. Imagine you have a bunch of different fruits, and you want to put them into groups like apples, oranges, and bananas. Cluster analysis helps you do this by looking at similarities between the fruits, such as their shape, color, or taste, and then grouping them together based on those similarities.

## 2 Process

### 2.1 Visualization

In our scenario, we possess a collection of observations, denoted as $N_{i=1}^{n}$, representing various generators such as Sean's Power Plant, Emily's Power Plant, and others. Each observation is associated with a set of features $X_i = \{x_1, x_2, ...\}$, encompassing attributes like bid price, marginal cost, and others.

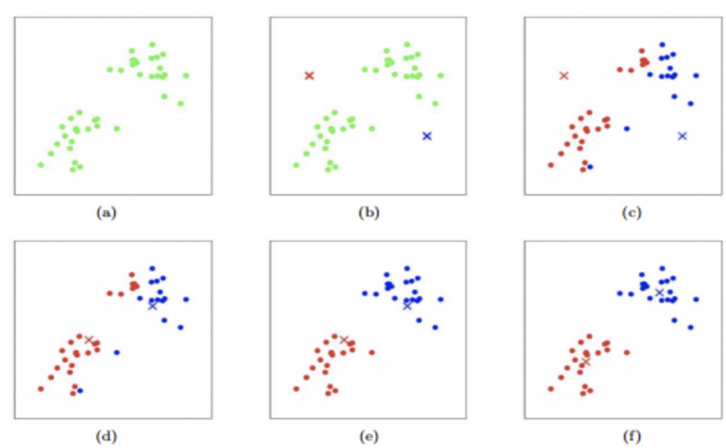


Figure 1: Process of Cluster Analysis (X=2, K=2)

(a) Setting-Up: We plot our observation in a scatter. The X-axis represents $x_1$ (bid price) and the

Y-axis represents $x_2$ (marginal cost). Each dot represents an observation (generator). Now, we try sort them into two clusters (groups).

(b) Initial Centroids: The machine randomly assigns a initial guess of centroid, which represents the center point of each possible cluster. As our goal is two groups, the two initial centroids are the red and blue crosses in the figure.

(c) Clustering: Based on the distance to each centroids, the dots go to the different groups. For example, if I am closed to the red cross rather than blue cross, I am going to the red group. After this process, the initial observations (all green dots) are sorted by two clusters (red and blue).

(d) Update Centroids: Recalculate the centroids of the clusters. This is done by taking the average of all the data points in each cluster along each dimension. You can see the red and blue crosses have new locations.

(e) Repeat Clustering: Reassign generators to clusters based on updated centroids.

(f) Repeat Centroid Updates: Recalculate centroids iteratively until they stabilize or a specified number of iterations is reached.

Hence, typically, we achieve a local optimization through the iterative process. Repeating the entire procedure becomes essential, as the initial guess in step (b) can influence the clustering outcome. This iterative approach ensures the exploration of different configurations, ultimately leading to a global optimization.

## 2.2 Example Results

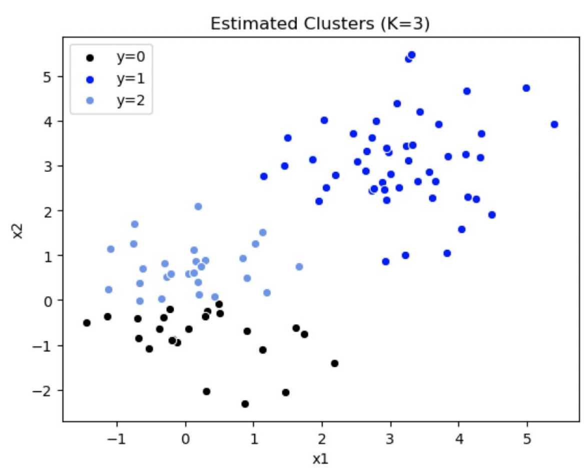


Figure 2: Two features and three clusters (X=2, K=3)

With higher dimensions, plotting becomes impractical, but we can extract meaningful information through a clustering table.

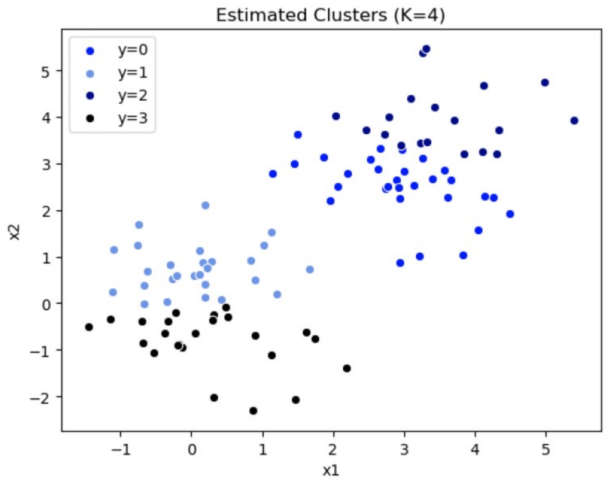


Figure 3: Two features and four clusters (X=2, K=4)

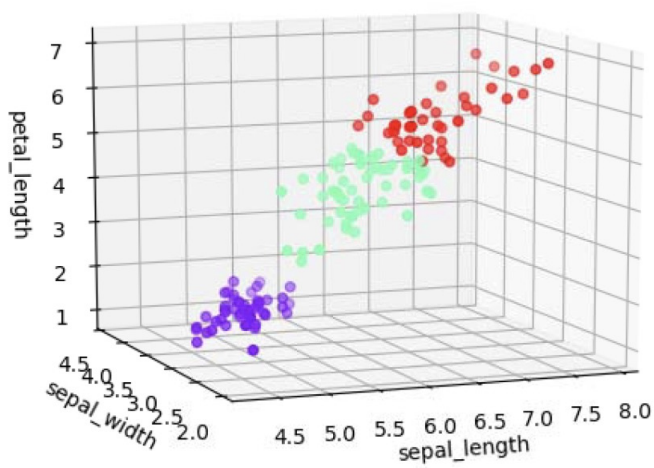


Figure 4: Three features and Three clusters (X=3, K=3)

Table 3. Cluster Analysis

| | Cluster 1 n = 132 | | Cluster 2 n = 129 | | Cluster 3 n = 148 | | Overall n = 409 | |
|---|---|---|---|---|---|---|---|---|
| | Average | S.D | Average | S.D | Average | S.D | Average | S.D |
| F1- Team capacity | 4.9762 | 0.4835 | 5.5105 | 0.3895 | 4.2027 | 0.9432 | 4.8648 | 0.8591 |
| F2 - Success in software development | 2.7178 | 0.8734 | 4.7578 | 0.6987 | 3.1824 | 1.2388 | 3.5293 | 1.298 |
| F3 - Relationship with external partners | 4.3617 | 0.8439 | 4.593 | 0.9897 | 2.8497 | 1.0276 | 3.8875 | 1.2399 |
| F4 – Culture | 4.7519 | 0.6841 | 5.345 | 0.5137 | 3.6436 | 1.0244 | 4.5379 | 1.0593 |
| F5 - Communication with customers | 4.6515 | 0.6314 | 5.1027 | 0.6371 | 3.397 | 1.0059 | 4.3399 | 1.076 |
| F6 - Environmental configuration | 4.9356 | 0.7677 | 5.2946 | 0.8674 | 4,0169 | 1.4456 | 4.7164 | 1.2153 |

Figure 5: The cluster table for higher dimension (X=6, K=3)

## 2.3 How to determine the K

The elbow method is a common technique used to determine the optimal number of clusters (k) in k-means clustering. Here's a step-by-step guide on how to use the elbow method:

(a) Run K-Means for different values of K: Perform k-means clustering for a range of values of k (e.g., from 1 to a certain maximum value). For each k, calculate the sum of squared distances (often called the "within-cluster sum of squares" or "inertia") between data points and their assigned cluster centroids.

(b) Plot the Elbow Curve: Create a plot with the number of clusters (k) on the x-axis and the corresponding sum of squared distances on the y-axis.

(c) Identify the Elbow Point: As we increase the number of clusters, the sum of squared distances will decrease. At some point, adding more clusters doesn't significantly reduce the sum of squared distances.The elbow point is where the reduction in sum of squared distances starts to slow down, creating an "elbow" shape on the plot.

(d) Choose the optimal K: The number of clusters at the elbow point is considered the optimal value for k. This is the point where increasing the number of clusters doesn't lead to a substantial improvement in clustering accuracy.

Examining this illustrative figure, the Elbow Point is identified at $K = 2$.

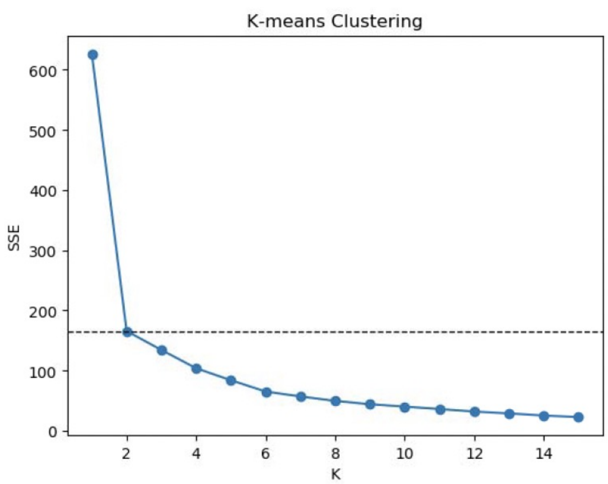


Figure 6: Elbow method

# Cluster Tool Guide

Sean Tan

April 23, 2024

1. Upon running, the notebook will prompt the user for a start and end date, in the format **DD-MM-YYYY**.

2. After this it will prompt the user to select the generator type(s), the possible choices are: **GAS**, **HYDRO**, **SOLAR**, **WIND**, and **Other**.

3. Next, it asks the user if they wish to run an aggregated cluster (aggregate the data), response is **yes or no (Y/N)**.

4. Once the elbow method is computed and plotted, it will prompt the user to select the optimal number of clusters, indicated by the kink in the elbow graph.

5. Finally, the notebook prompts the user to enter the features for the cluster algorithm. Possible selections are: **block_number**, **block_price**, **block_size**, **from_MW**, **to_MW**, **available_MW**, **dispatched**, **dispatched_MW**, and **flexible**.

6. For aggregated clusters, all the same features are available, with the addition of:
- **zero_offer_percent**: total percentage of generation offered in at a price of 0
- **sum_dispatched_block_price**:aggregated block price, given that the block was dispatched
- **sum_dispatched_block_size**:aggregated block size, given that the block was dispatched

7. The cluster plot is then generated, along with the metric tables. If 2 features are selected, then the plot is 2D, if 3 features are selected, then the plot is 3D, and if the number of features is $> 3$, then the metric tables are displayed instead.

# Econ 588 Literature review

Clodagh Berg and Emily Deuchar

March 2024

## 1 Project Idea

**Every hour, generators submit bids indicating the price at which they are willing to supply electricity. The Alberta Electricity System Operator (AESO) is an independent operator in charge of operating the Alberta electricity grid and pricing information, ensuring a balance between both supply and demand sides.**

**A "merit order" is established by sorting bids from the lowest to the highest price. AESO dispatches electricity starting from the lowest-priced offers and moves upward until all demands are met. The price of the last dispatched offer is termed the system marginal price (SMP).**

**For example, if supply offer #3 is fully met all demands, then the SMP is at $40. If demand was increased by 40 MW (megawatt), AESO would move up the list and dispatch 40 MW electricity of supply offer #4, which would result in an SMP of $50.**

**Generators bid based on their marginal cost, and the market price aligns with the point where supply meets demand. This SMP is determined on a minute-to-minute basis. The pool price, calculated as the average of all SMPs within an hour, is also the intersection point of supply and demand curves.**

**Our project aims to analyze Alberta's electricity sources based on data from 2023 calendar year. The graph indicates that the primary sources in Alberta are natural gas and wind, with a minor contribution from coal, biomass, solar, and hydro.**

## 2 Literature Review

**We intend to employ Brown and Olmstead (2017) to comprehend the market power dynamics in the Alberta electricity market. The study's findings, revealing a surge in market inefficiencies by 6.7%-19% during periods of high demand, coupled with the observation of a low degree of market power during hours of low demand. This will enhance our insight into the instances when firms are more prone to exercise market power. Building upon the content of this paper, we will incorporate the calculation of the marginal cost for various plants into our analysis of combined cycle plants. Specifically, the following equation will be used for the determination of the marginal cost of natural gas plants:**

$$Marginal cost of natural gas plants = (Fuel cost * Heat rate) + Carbon tax + Variable O\&M$$

**Brown et al. (2018) builds even more off the concepts outlined in Brown and Olmstead (2017) about the surge of market inefficiencies during peak hours. This paper analyzes whether firms utilize information revealed through the Historical Trading Report by looking at firms' bidding behaviour. They show that firms may be using bidding patterns to reveal their identities to their rivals to increase market prices. This paper gives us information on firms behaviour during peak hours as well as more information about potential behaviours that we will incorporate into our analysis.**

**Building upon the previous paper, we have obtained data from the AESO regarding Heat rate and Variable O&M. To acquire the necessary Carbon**

**pricing data and fuel cost information, we will be referring to the source Olmstead and Yatchew (2022). In this study, the projected carbon tax of \$65/MWh is applied, and the carbon intensity of combined cycle plants is 0.37 tCO2e/MWH. Leading to an estimated payment of approximately \$25/MWh for natural gas plants carbon payment. For the sake of this analysis, we will adopt this figure to compute the marginal cost. We will also use the value for fuel tax**

**We will be reviewing Ito and Reguant (2016) to better understand the role of market power in electricity pricing. This study looks at sequential markets (made up of forward and real-time markets). In theory, prices in sequential markets should converge over time. Though empirically, this doesn't happen and this paper looks at the role of market power in explaining this price differences. They find that dominant firms do withhold or undersell electricity production in the forward markets. Smaller or less dominant firms commit to selling more electricity than they can. Although the Alberta electricity market does not have forward markets, we can analyze these results to help inform ourselves of which firms or behaviours are likely to use market power to their advantage and where to direct our efforts in our analysis. Arbitrage is analyzed in this research and we will consider if arbitrage could be a factor in Alberta's electricity market. An important conclusion that is also relevant to policy is that price convergence should not be taken as evidence that a market is efficient.**

**In our review, we will include an analysis of the findings from Wolfram (1997). This study offers valuable insights into the strategic behavior associated with ownership. We found it essential to assess whether larger entities are prone to exercising market power in the power market's bidding process. The research method involved calculating the marginal cost for each power plant and comparing these costs with the bids submitted by the firms. This approach closely mirrors our intended methodology, albeit applied to the context of the UK instead of Alberta. The study's conclusions suggest that the largest firm, National Power, engaged in more strategic bidding practices compared to its smaller rivals.**

**We acquired peak hour data from the AESO website Alberta Electric System Operator (2022), which indicated that peak hours occurred from 7 am to 11 pm. This information guided our group in establishing the time frames for conducting our analysis on peak versus off-peak bidding.**

## References

Alberta Electric System Operator (2022). Annual market statistics 2022. Accessed on March 11, 2024.

Brown, D. P., Eckert, A., and Lin, J. (2018). Information and transparency in wholesale electricity markets: evidence from alberta. *Journal of Regulatory Economics*, 54:292–330.

Brown, D. P. and Olmstead, D. E. (2017). Measuring market power and the efficiency of alberta's restructured electricity market: An energy-only market design. *Canadian Journal of Economics/Revue canadienne d'économique*, 50(3):838–870.

Ito, K. and Reguant, M. (2016). Sequential markets, market power, and arbitrage. *American Economic Review*, 106(7):1921–1957.

Olmstead, D. E. and Yatchew, A. (2022). Carbon pricing and alberta's energy-only electricity market. *The Electricity Journal*, 35(4):107112.

Wolfram, C. (1997). Strategic bidding in a multi-unit auction: An empirical analysis of bids to supply electricity.