Overpowered

Connecting Energy to the Grid, Faster
The Team
UC Berkeley Masters of Information and Data Science

Paul Cooper
Hailee Schuele
Zhifei Dong
Adam Kreitzman
Our Mission

Provide a faster, more efficient way of connecting new energy sources to the power grid.
The Space
The Power Grid

Generation → Transmission → Demand

Storage
Our Focus: New Power Generation

Generation → Transmission

Demand

Storage
Geographic overview of RTOs and ISO. Source: FERC.gov
When a new power generation facility wants to come online, they submit an application to the interconnection queue.
Each interconnection request is subject to long, technical studies:
- transmission capacity
- market conditions
- incident analysis
- feasibility and impact

One request may be dependent on other requests.

Additional infrastructure investment is often needed to accommodate a new power generation source.
The Problem
Time, Money, and Dropout

1. 4+ year wait times
2. Unfeasible infrastructure costs
3. Wasted time and resources
The Power Grid's Big Bottleneck

The queue process is outdated, inefficient, and expensive.
The Solution
Cluster Processing

- New federal ruling allows for processing multiple applications together as a cluster
- Clusters allow:
  - Shared development investment
  - Reduced queue wait times
  - Fewer withdrawn applications
Improving Clusters

- If we can:
  - Identify efficient clusters
  - Understand how likely a given cluster is to be successful
- Then we can:
  - Speed up the queue
  - Reduce dropout
  - Save time and resources
Our Product

- Recommend **clusters of interconnection requests** to be studied together
- Provide **scoring mechanisms** to understand how strong the candidate batch is
- Provide tools for expert users to customize their results
- Provide an **integrated platform** for data querying, visualization and decision support

---

**How to Use the Querying Map**

Currently, the querying map only supports CAISO database.

1. To start with, choose a California county to zoom into. The county boundary is filled with "yellow" color.
2. Select a zoom-in scale to display the transmission lines in "blue" color.
3. Choose an extra data layer to display as "black dot". The available data layers are substations, power plants and retired generators.
4. This is optional. User can enter a location coordinate by latitude and longitude. The location will be added to plots as "red diamond"
The Data
Piecing the Data Together

- CAISO interconnection queue data
- California substation GIS data
- California transmission line GIS data
- California power plant GIS data (including retired ones)
- CAISO future transmission projects
- PowerWorld - Power flow simulation software used by ISOs
- Simulated line capacity and load
Key Data Takeaways - Completed Queues

Histogram of Waiting Time

- Frequency
- Years

0 2 4 6 8 10 12 14

0 5 10 15 20 25
Key Data Takeaways - Completed vs. Withdrawn

- Total Completed MWs: 6.9%
- Total Withdrawn MWs: 93.1%
Key Data Takeaways - Current Queue

Cumulative MWs by Year

Year:
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013
- 2014
- 2015
- 2016
- 2017
- 2018
- 2019
- 2020
- 2021
- 2022
- 2023

MWs:
- 0
- 2000
- 4000
- 6000
- 8000
- 10000
- 12000

Graph shows the cumulative MWs increasing significantly from 2020 onwards.
The Technical Approach
## Concept Abstraction & Operationalization

<table>
<thead>
<tr>
<th>Concept</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>● Geolocation of Interconnection Point</td>
</tr>
<tr>
<td></td>
<td>● Proximity to Existing Infrastructure (retired plants)</td>
</tr>
<tr>
<td></td>
<td>● Proximity to Planned Infrastructure</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>● Included Storage</td>
</tr>
<tr>
<td></td>
<td>● Transmission Availability</td>
</tr>
<tr>
<td></td>
<td>● Generator Type(s)</td>
</tr>
<tr>
<td></td>
<td>● Amount of Energy Output</td>
</tr>
<tr>
<td>Process</td>
<td>● Permit Status</td>
</tr>
<tr>
<td></td>
<td>● Expected On-line Date</td>
</tr>
<tr>
<td></td>
<td>● Queue Position</td>
</tr>
<tr>
<td></td>
<td>● Utility company</td>
</tr>
</tbody>
</table>
## Output Variables

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Scores</td>
<td>● Likelihood of success based on features learned from historical data</td>
</tr>
<tr>
<td>Cluster Strength</td>
<td>● The strength of the recommended cluster as a whole using aggregated similarity calculations</td>
</tr>
<tr>
<td>Total MegaWatts</td>
<td>● The total MWs provided to the grid by the cluster minus the current available transmission capacity</td>
</tr>
</tbody>
</table>
Data Pipeline

- Databricks & Blob Storage
- Feature Selection/Engineering
  - SME feedback
  - Inference from past interconnection requests
- Preprocessing for Machine Learning:
  - VectorAssembler + Scalers
  - Generating a numeric representation
Likelihood of Approval

- A method of judging the strength of an application based on supervised learning past data and future grid infrastructure development.
- Combines coefficients from supervised learning with proximity to retired power plants and future transmission projects to gauge relative strength of a grid applicant.
- Allows a decision maker to consider the probability of application success in cluster-building process.
Custom Clustering Algorithm

- Crafting a similarity algorithm that involves group subsets + process subtleties – not a simple or intuitive process to create.
- Variable-centric similarity inspired by Cosine Similarity.
- Custom similarity algorithms built to handle all variables.
- Only considering projects downstream for each project (i.e in a FIFO approach, what projects could be added to this one to form a cluster, maintains fair process)
### Sample Results

#### MONTEZUMA (HIGH WINDS III) Suggested Cluster

<table>
<thead>
<tr>
<th>Project</th>
<th>Net MWs to Grid</th>
<th>Likelihood of Approval</th>
<th>Location</th>
<th>Process</th>
<th>Infrastructure</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOLANO 4 WIND</td>
<td>90.8</td>
<td>0.7049</td>
<td>0.7832</td>
<td>0.8736</td>
<td>0.73</td>
<td>0.7770</td>
</tr>
<tr>
<td>RECLAIMED WIND</td>
<td>90.7</td>
<td>0.8176</td>
<td>0.5267</td>
<td>0.8737</td>
<td>0.75</td>
<td>0.7424</td>
</tr>
<tr>
<td>MULQUEENEY FAN...</td>
<td>20</td>
<td>0.7431</td>
<td>0.5349</td>
<td>0.8587</td>
<td>0.79</td>
<td>0.7317</td>
</tr>
<tr>
<td>SANDSTORM WIND...</td>
<td>150</td>
<td>0.4227</td>
<td>0.1235</td>
<td>0.5988</td>
<td>1</td>
<td>0.6391</td>
</tr>
<tr>
<td>SALOON ENERGY S...</td>
<td>150</td>
<td>0.7041</td>
<td>0.4655</td>
<td>0.8788</td>
<td>0.5</td>
<td>0.6371</td>
</tr>
<tr>
<td>GONZAGA WIND FA...</td>
<td>76.35</td>
<td>0.254</td>
<td>0.3939</td>
<td>0.881</td>
<td>1</td>
<td>0.6322</td>
</tr>
<tr>
<td>WINDWALKER OFF...</td>
<td>1000</td>
<td>0.6066</td>
<td>0.2721</td>
<td>0.8399</td>
<td>0.75</td>
<td>0.6217</td>
</tr>
<tr>
<td>PROXIMA SOLAR</td>
<td>390</td>
<td>0.6590</td>
<td>0.4394</td>
<td>0.9015</td>
<td>0.5</td>
<td>0.625</td>
</tr>
<tr>
<td>MONTEZUMA II</td>
<td>76</td>
<td>0.6128</td>
<td>0.802</td>
<td>0.7098</td>
<td>1</td>
<td>0.6334</td>
</tr>
<tr>
<td>SEAGLASS OFFSHO...</td>
<td>600.1024</td>
<td>0.6339</td>
<td>0.2709</td>
<td>0.8012</td>
<td>0.75</td>
<td>0.022</td>
</tr>
<tr>
<td>KEYHOLE WIND</td>
<td>190</td>
<td>0.6906</td>
<td>0.2025</td>
<td>0.8361</td>
<td>0.75</td>
<td>0.6190</td>
</tr>
<tr>
<td>CUERING GRANDE...</td>
<td>150</td>
<td>0.7289</td>
<td>0.1569</td>
<td>0.782</td>
<td>0.75</td>
<td>0.6044</td>
</tr>
<tr>
<td>WINDSTAR I ALTER...</td>
<td>120</td>
<td>0.2282</td>
<td>0.1829</td>
<td>0.5989</td>
<td>0.75</td>
<td>0.5026</td>
</tr>
</tbody>
</table>

#### Cluster Metrics
- **Cluster Strength:** 0.653
- **Net Transmission Capacity:** -18942.0
- **Likelihood of Approval:** 0.6073

#### Aggregate Score
- Overall: 0.6217
The Demo
Demo - Clustering Tool

Overpowered - Connecting Renewable Energy to the Grid Faster

Let’s get to clustering!

Studying a single application at a time makes for a slow going. Overpowered’s clustering tool helps you determine which projects make sense to study together. This tool focuses on the California grid operator (CAISO) Interconnection Queue.

Set Custom Weights

Overpowered provides a structured scoring mechanism to determine the best groups of applicants to study together. We also recognize that expert energy users have insights into how Queue applications are successful in their unique operating regions. Feel free to start with the default equal split, or configure the weighting parameters as you see fit!

Pick a Project

Now that you’ve set your weights, click a base project in the CASQ Queue below to generate cluster recommendations.
Overpowered - Connecting Renewable Energy to the Grid Faster

Interactive Map

Each application in the queue data indicates the county where the project is to be built. The application also briefly describes the station or transmission line it plans to connect to. Therefore, the interactive map allows the user to explore the available datasets (transmission lines, substations, retired power plants, and future infrastructure projects) by California counties.

Here are some examples of using the interactive map:

- Scenario 1: The application indicates the transmission line it plans to connect to. The user can load the current queue and quickly find the shortest distance to the nearby transmission line. Additionally, the user can compare the proposed power with the remaining line capacity to determine if the transmission line has enough capacity for the application.
- Scenario 2: The application includes a power storage unit. The user can load the current queue and the retired power plants to check the availability of the nearby plants as ideal storage units.
- Scenario 3: The project location of the application is far away from the existing infrastructure. The user can load the current queue and the future infrastructure to determine if an infrastructure project is to be built near the site.

Currently, the querying map only supports California database.

Available Datasets

- US state boundaries: base map filled in gray color
- California county boundaries: base map filled in yellow color
- California transmission lines: base map lines in blue color. The transmission lines are labeled by their names and simulated remaining capacity. Hover over a transmission line to view.
- Additional datasets
  - California substations: add-on points in "red triangle". The substations are labeled by their names. Hover over a substation to view.
  - Retired power plants: add-on points in "black cross". The retired power plants are labeled by their names and dates of retirement. Hover over a plant to view.
  - Current queue: add-on points in "green diamond". The queue applications are labeled by their names and station/transmission line to connect to. Hover over an application to view.
Key Takeaways

TULE WIND Suggested Cluster

Set Parameters

Assign relative weights

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Location</th>
<th>Infrastructure</th>
<th>Process</th>
<th>Likelihood of Approval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Cluster Strength: 0.6676
Net Transmission Capacity: -18709.1
Likelihood of Approval: 0.7159

<table>
<thead>
<tr>
<th>Project</th>
<th>Net MWs to Grid</th>
<th>Likelihood of Approval</th>
<th>Location</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOUNT LAGUNA WIND 2</td>
<td>400</td>
<td>0.7134</td>
<td>0.8203</td>
<td>0.1</td>
</tr>
<tr>
<td>RUGGED SOLAR FARM</td>
<td>71.88</td>
<td>0.718</td>
<td>0.8368</td>
<td>0.1</td>
</tr>
<tr>
<td>STARLIGHT SOLAR</td>
<td>20</td>
<td>0.6611</td>
<td>0.837</td>
<td>0.1</td>
</tr>
</tbody>
</table>
The Evaluation
Measuring Impact

- Our ultimate goal is to **improve queue efficiency**
- We want to measure:
  - Cluster strength
    - How compatible are the projects within a cluster?
    - How likely is the cluster to be approved?
  - Improved wait times
    - Are we speeding up the queue?
The Tool's “Goodness”
SME Feedback

“The old way has its limitations and everyone complains about it”

“When we’re trying to advise clients, the want to know where the best place to connect would be... and we really don’t know”

“Having historical data boiled down is the value of your tool. You don’t have to rerun the study every time (you assess a new applicant)”
The Overpowered Value
Breaking Down a Complex Problem

- **Technical Challenges**
  - Piecing together various data sources of different types
  - Calculating unknown variables such as transmission line capacity and usage load

- **Human-Centric Challenges**
  - Abstracting technical and human-centric process considerations
  - How to allow tool flexibility in an inexact problem-space
  - Machine learning expertise in an antiquated system
Roadmap

- Extend the application to developers
- Incorporate additional data points that grid operators use
- More SME feedback
- Additional data outside of California
- Incorporate outcomes from initial cluster studies
- More supervised ML-focused approaches to grid expansion built on top
The Mission
Impact

Provide a faster, more efficient way of connecting new energy sources to the power grid.
Acknowledgements

- Damian Berger (SME at Atwell Group)
- Annie Polakowski (SME at Atwell Group)
- Dr. Roy McCann (Principal Investigator at GRAPES)
- Thomas Dempsey (SME at MYNO Carbon)
- CAISO Customer Support
- UC Berkeley MIDS Capstone class and professors

"Striking a balance between a slow and certain approach and considering all-comers fairly is really the heart of the solution."

- Damian Berger (SME at Atwell Group)