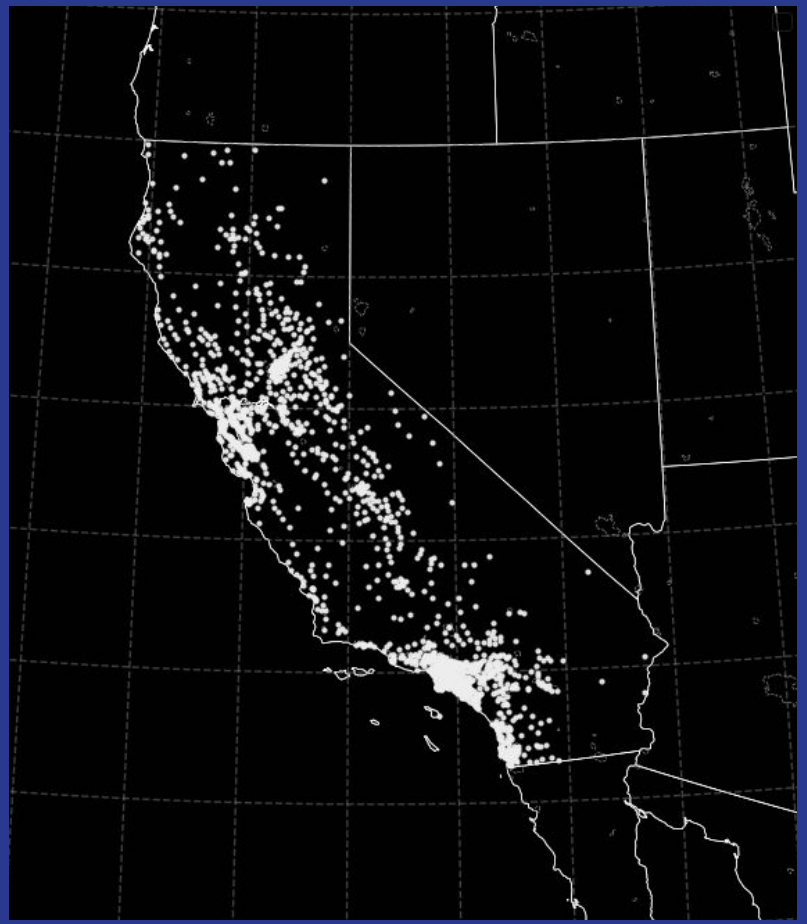
California Energy Storage

Predictive Modeling

Introducing the Data

Exposition for the Data Analysis to Follow



Data Structure

Key Columns:

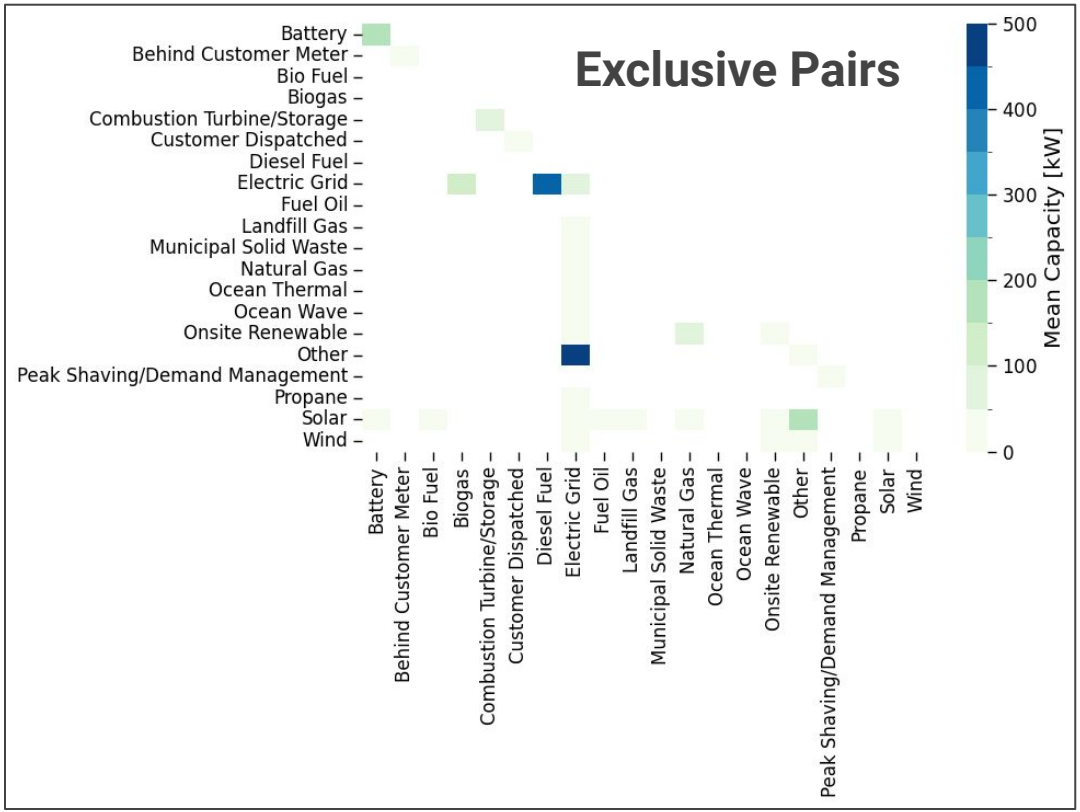
- **Approval Date**: Date of regulatory approval.
- **Nameplate Capacity**: The nominal capacity of the storage asset.
- **Fuel Types**: List of energy sources which fall under the approval and capacity.
- **CAISO Flag**: Indicates if the facility is within CAISO's regulatory control.
- **Facility Zip Code / City / County**: Various geographic indicators of each approved source.

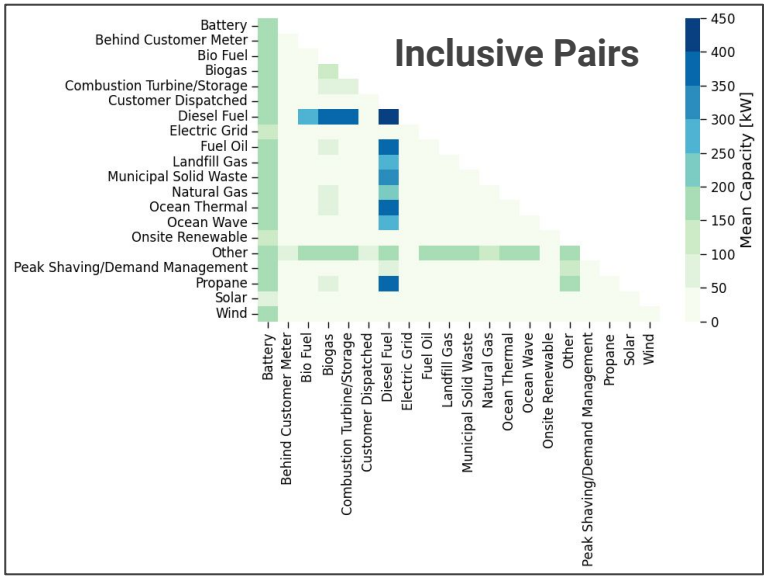
This dataset is fundamentally time-series data, where each entry corresponds to an addition of capacity from a set of approved energy sources (fuel types).

The time-series structure emerges when the data is grouped by attributes such as location, utility, or sector.

More information on the California Energy Storage System Survey.

Fuel Type Combinations



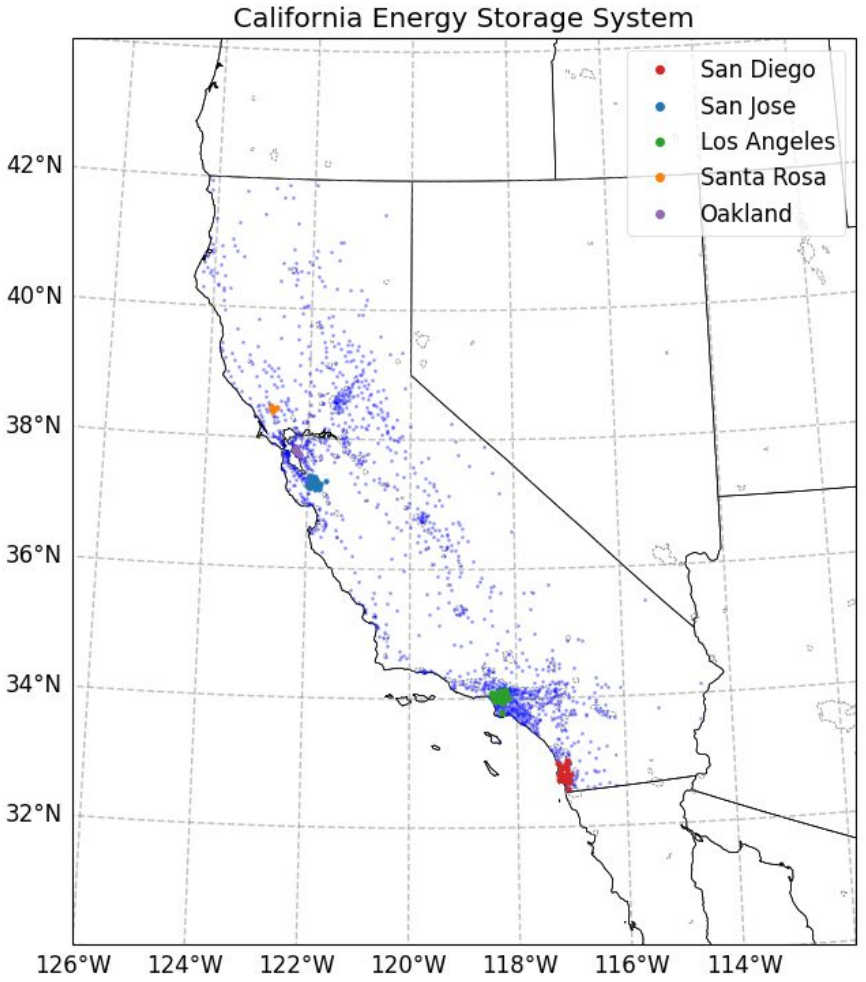


Exclusive pairs show two fuel types used alone, with diagonal elements showing fuel types unpaired. ***Inclusive pairs*** include additional types with each pair element.

Geographic Arrangement of Energy Storage

The zip code of each entry, along with libraries such as *pgeocode* allows us to map each entry to a geographic location.

In this case, the data is narrowed to the *Residential* data and grouped by city, from which the five highest cities by net capacity are shown here. The light blue, small data points indicate the full data set by unique zip code.



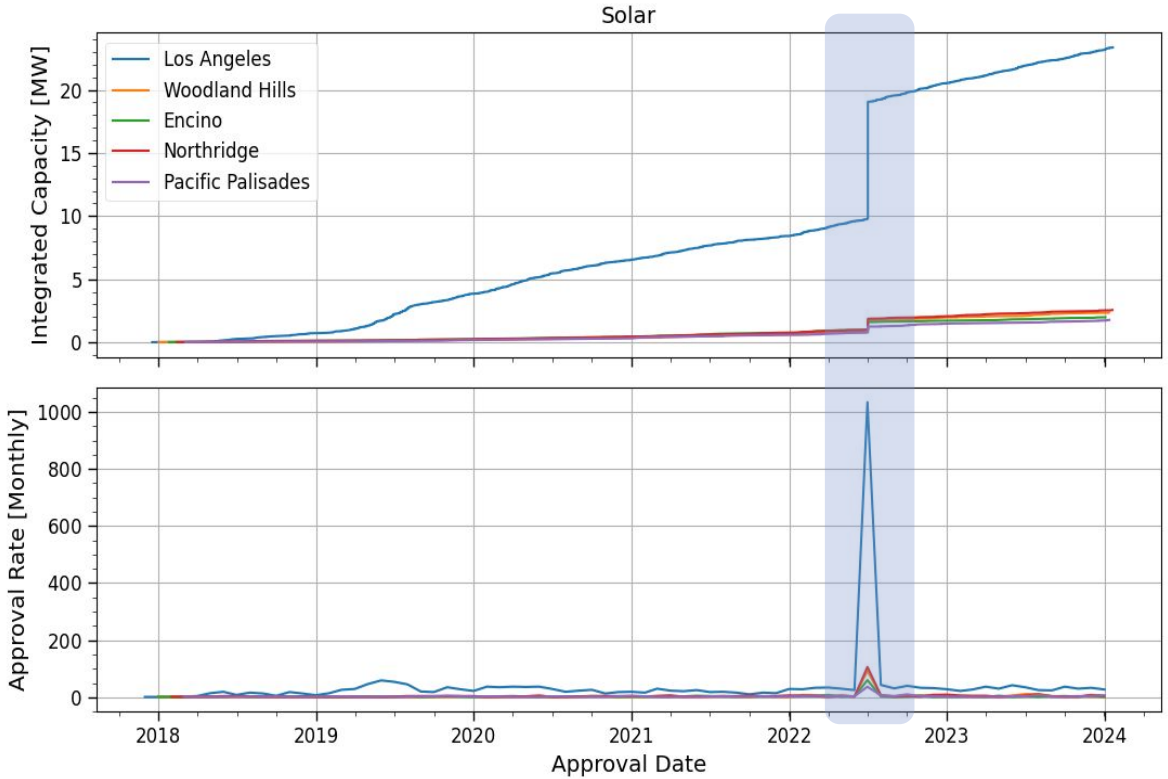
Cumulative Capacity Series

On June 30, 2022, Governor Gavin Newsom signed Assembly Bill (AB) 205 into law.

The nameplate capacity at a given timestamp (not shown here) represents individual contributions to the capacity of a system (city, count, zip code).

The **cumulative capacity** (top) is the increased total capacity up to that point in time.

The **approval rate** (bottom), take monthly here, indicates the number of approvals of that system.



Modeling Energy Capacity Growth

Capacity Growth

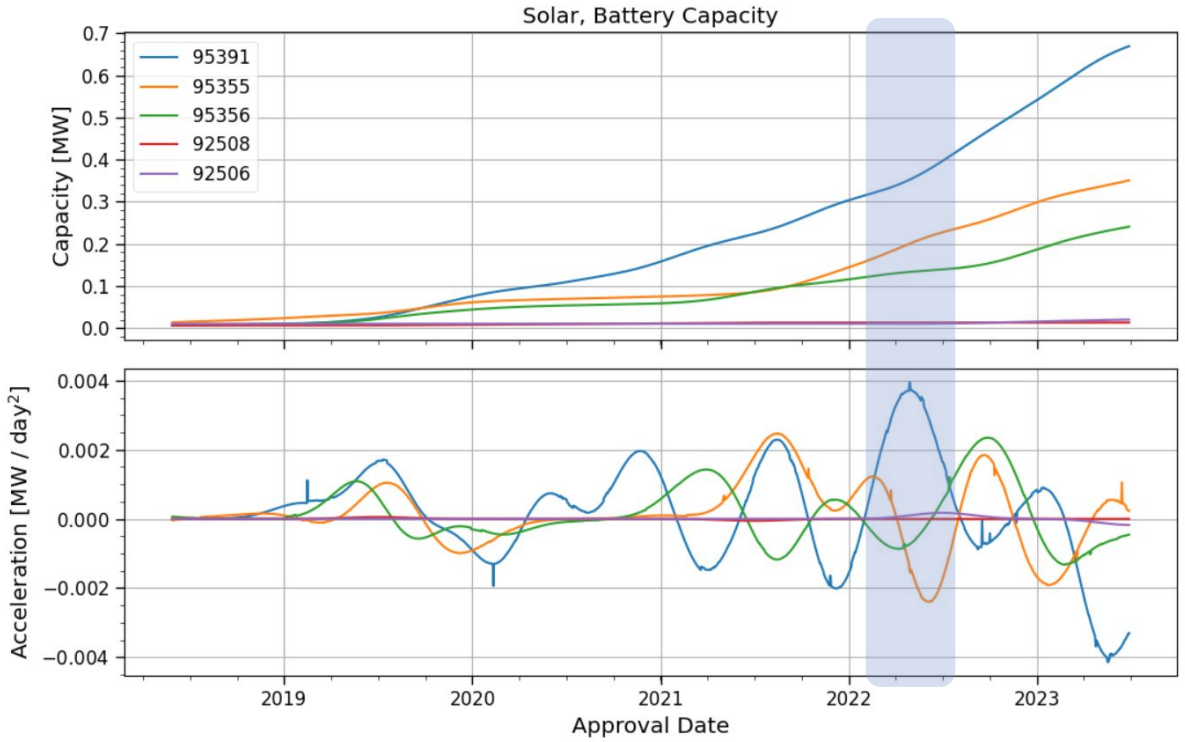
The capacity (cumulative capacity) for a given source *n* (fuel, location) is a function of time

$$c_n(t)$$

and the acceleration is the function of how that capacity growth changes over time. It is the change in the slope of the curve is

$$a_n(t) = \frac{d^2 c_n(t)}{dt^2}$$

The peaks in acceleration correspond to bends in the capacity curve.



Questions and Predictions

For a set of systems (e.g., by city, county, or zipcode) of a given fuel type or combination:

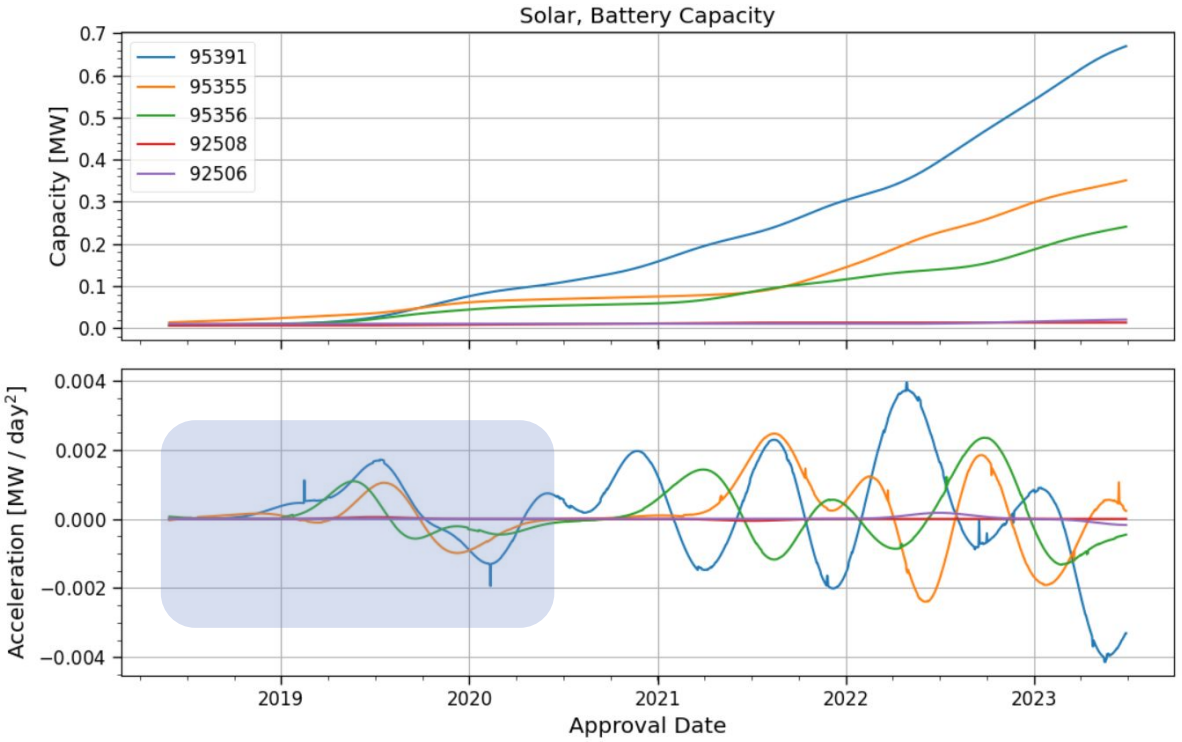
- ***Lead-Lag Relationships***: Are the growth rates of some systems consistently leading or lagging behind others in the group? ***Can an increase in growth rate in one system reliably predict a future increase in another?***
- ***Correlated Acceleration***: Are there identifiable subsets of systems with correlated acceleration patterns? ***Does an increase in capacity growth rate in one system indicate a similar trend (or opposing trend) in another at the same time?***

Lead-Lag Relationships

*To make this more concrete, notice the nameplate capacity acceleration trends of the zip codes **95356** and **95355**, of which at certain points, the acceleration of **95356** seems to lead **95355**.*

The way to look for this, is to use convolution of the portion of the reference signal in question, and find high overlaps.

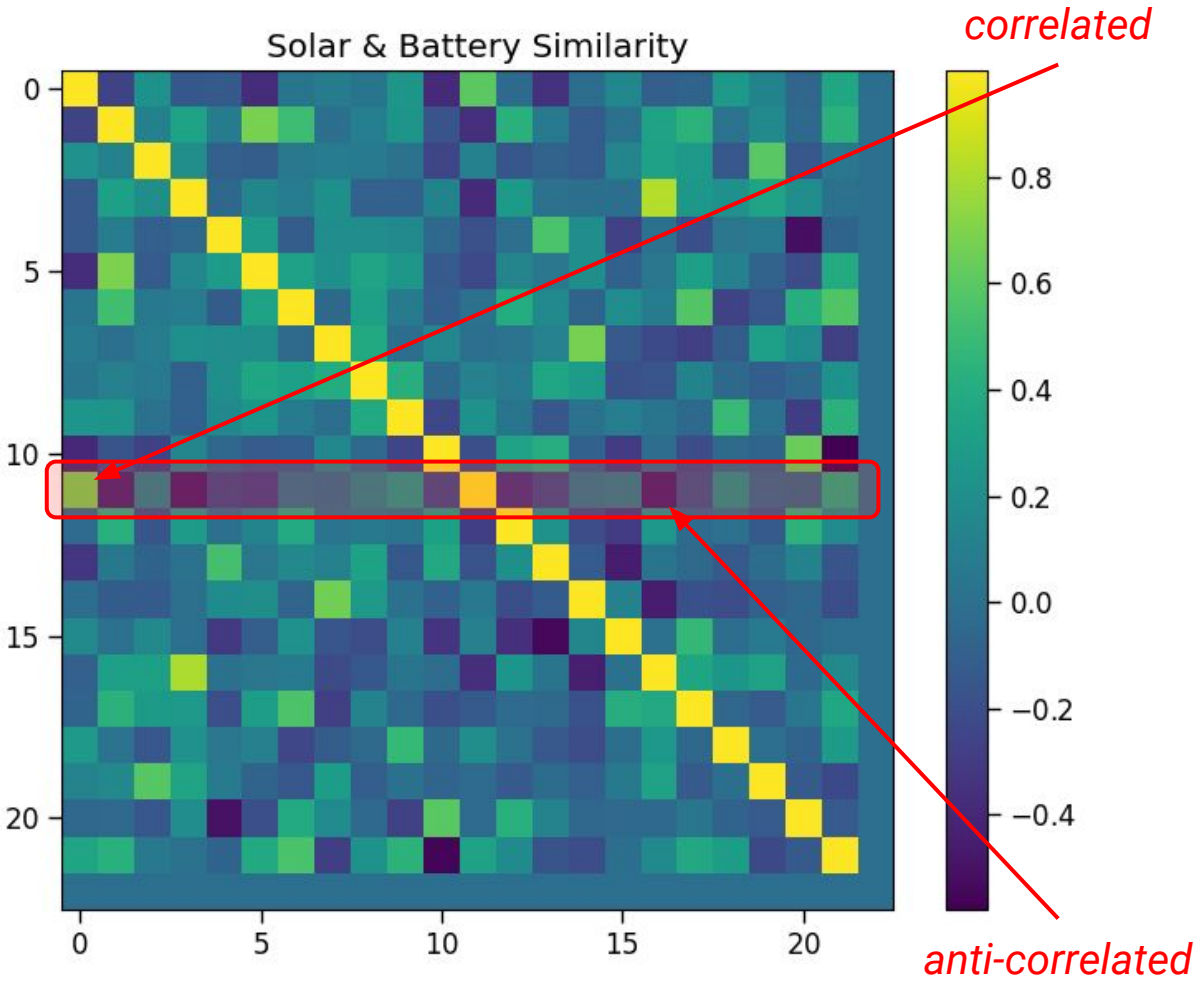
This was the problem I was most excited about working out, but fell short by several steps along the way.

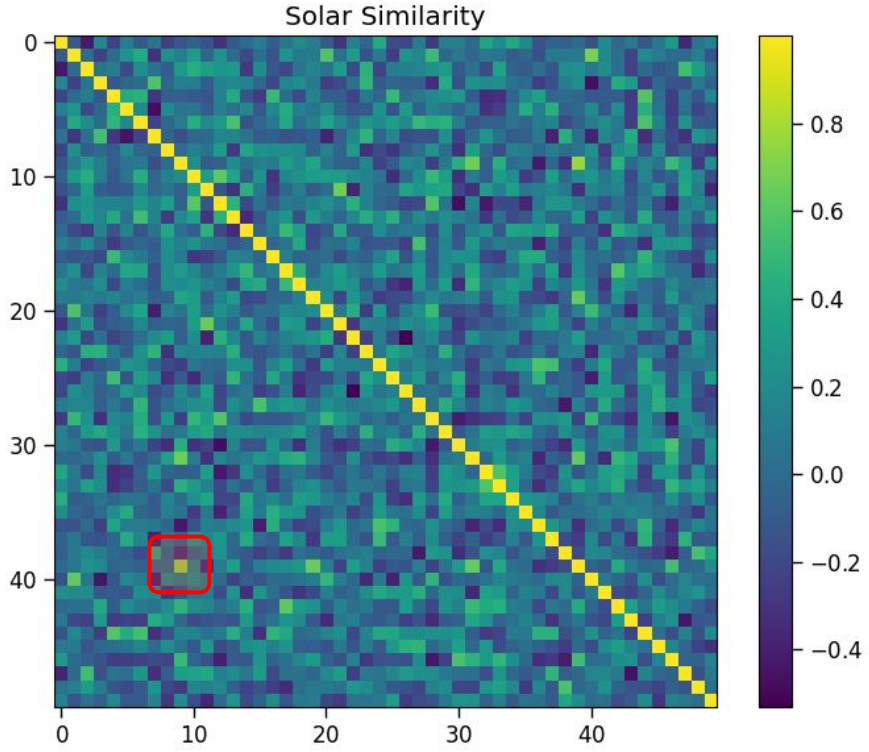


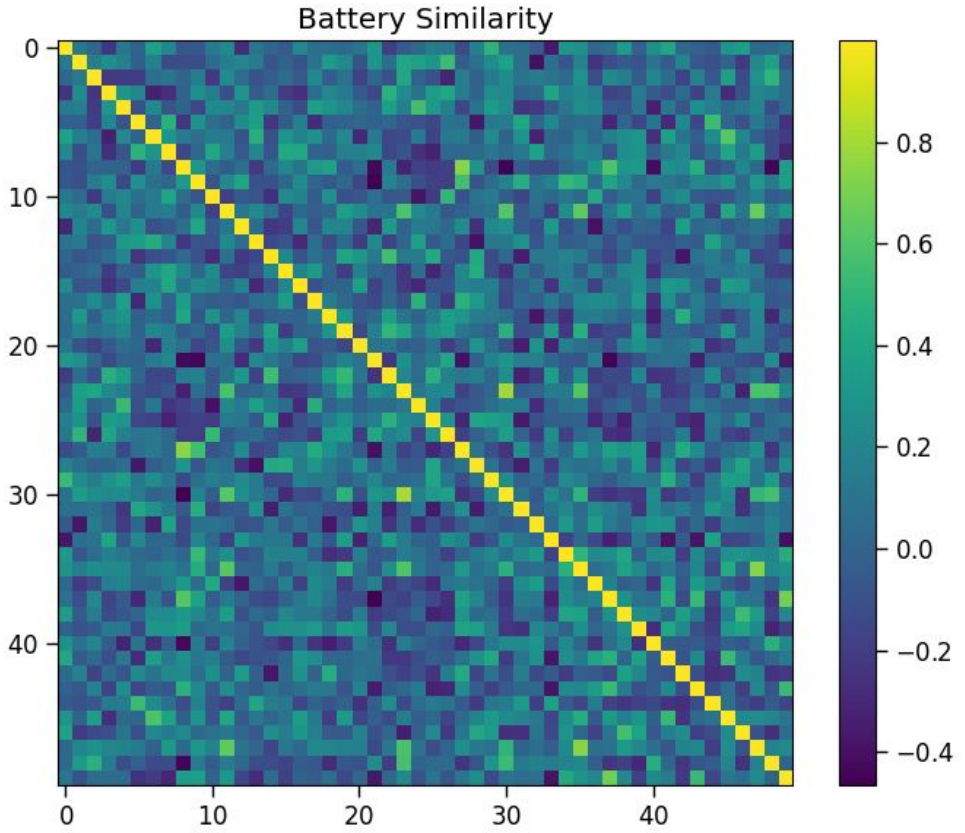
Similarity Matrices

The inner product of the capacity acceleration is computed between all pairs (m, n) from the defined set. From which an angle is computed.

Accelerations $a_m(t)$ and $a_n(t)$ which are trend with each other will tend toward similarity values of 1.0, whereas those that trend against each other will tend toward similarity values of -1.0.







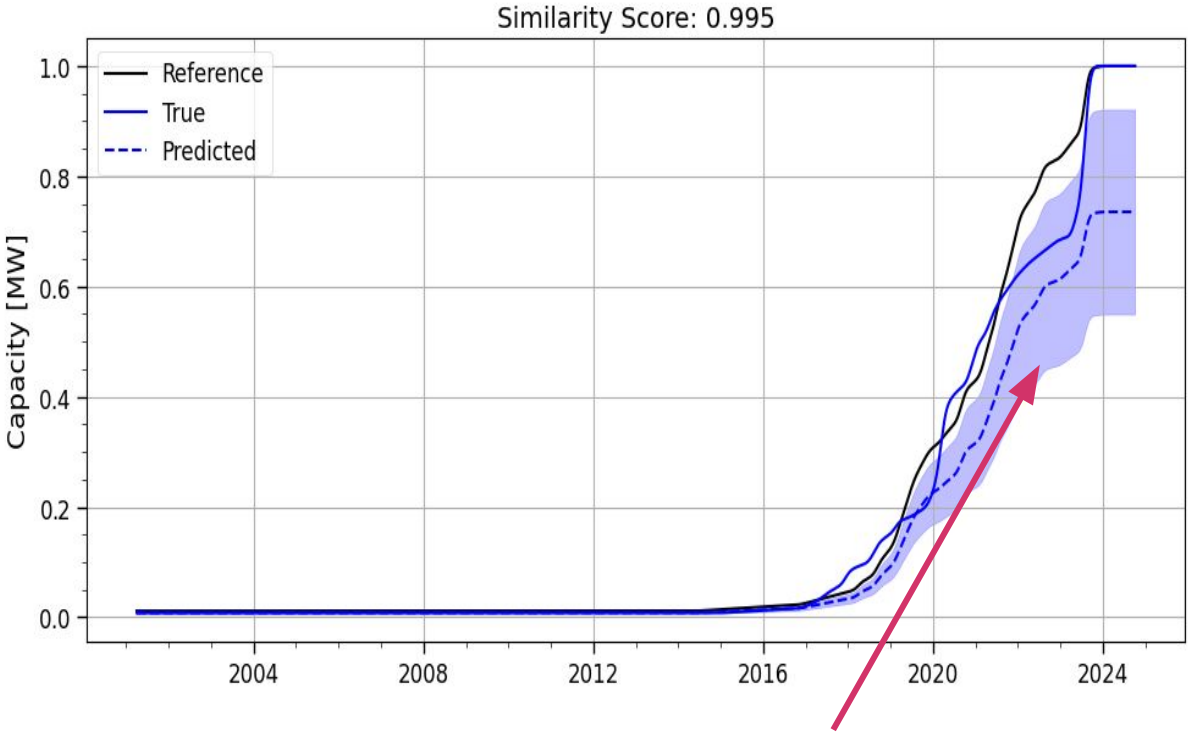
More Similarity Matrices

Predicting Capacity of Source

It will be an imperfect measure, but for any given pair of capacity groups (i.e. Solar by zip code) with a ***high similarity coefficient***, the general trend in the cumulative capacity of the unknown sample may be recovered.

In the figure to the right, a high similarity score was found for two zip codes with correlated Solar capacity growth. With ignorance to the back half of the data, the capacity can be ***roughly*** recovered.

NOTE: There is a known error with how the acceleration and error propagation is being backfilled into the cumulative capacity, driving the difference between the true and predicted capacity curves.



Expected error bounds grow with time.

Thank You