

# ML-Based Risk Assessment of Renewable Energy

---

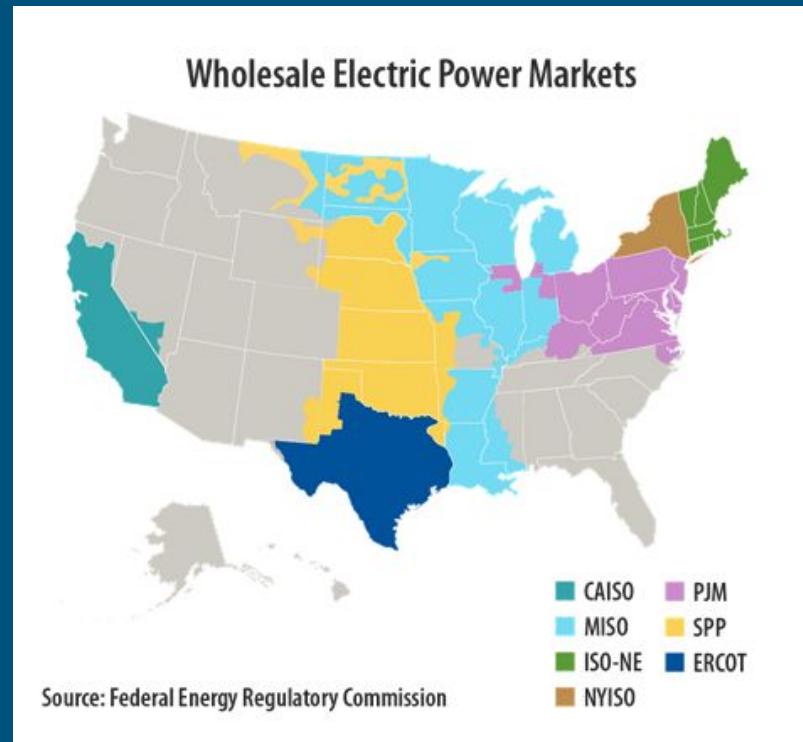
Kevin Bergdoll, Parker Fu, Muyuan He,  
Sam Joffe



# Electricity/Power Markets

---

- Power grids/markets
- Utilities
- Locational Marginal Pricing (LMP)
- Day-ahead market
- Real-time market
- Capacity market



# Risk

---

- Asset risk
- System risk
- How likely are you to deliver on what you promise?
- How can we measure the risk?

# Problem Statement

---

- Develop risk scores that address the asset and systemic risk of energy assets using ML techniques
- Calculated on a daily or hourly basis depending on the market

# Project Components

---

- Domain Research
- Data Collection
- Data Storage
- Machine Learning Algorithms

# Partners/Collaborators

---



# Agenda

---

- Introduction
- Weather
- Supply
- Demand
- Power Outages and Wind
- Data Storage
- Conclusion

# Weather

---

# Weather - what do we want?

---

- Historical and forecasted weather data
  - Hourly
  - Location
  - Specific date (historical)
  - Next day (forecasted)

coordinate	timestamp	temperature	dewpoint	windDirection	windSpeed	windGust	relativeHumidity
42.6526#-73.7562	2021-05-05T05:38:00+00:00	11.7	11.7	150	12.96	None	100
42.3130#-78.0195	2021-05-05T05:26:00+00:00	12.8	12.8	270	5.4	None	100
40.8448#-73.8648	2021-05-05T04:51:00+00:00	12.2	8.9	120	12.96	None	80.266011265195
42.1792#-75.8534	2021-05-05T05:42:00+00:00	15	14.4	340	12.96	None	96.204921152027
42.2318#-78.7476	2021-05-05T05:18:00+00:00	None	None	None	None	None	None
42.7655#-76.5488	2021-05-05T05:44:00+00:00	15	15	280	5.4	None	100
42.2313#-79.5603	2021-05-05T05:18:00+00:00	None	None	None	None	None	None
42.1362#-76.7798	2021-05-05T05:32:00+00:00	16.7	13.9	230	5.4	None	83.550482942249
42.4972#-75.6208	2021-05-05T05:42:00+00:00	15	14.4	340	12.96	None	96.204921152027
44.7904#-73.6006	2021-05-05T05:44:00+00:00	10.6	8.3	150	22.32	None	85.680481185637
42.3679#-73.5594	2021-05-05T05:35:00+00:00	10.6	10	0	0	None	96.074594670841

# Weather - Data Collection

---

API

Long time range

coordinate	timestamp	temperature	dewpoint	windDirection	windSpeed	windGust	relativeHumidity
42.6526#-73.7562	2021-05-04T14:51:00+00:00	9.4	8.9	10	5.4	None	96.688464288609
42.3130#-78.0195	2021-05-04T15:04:00+00:00	17.8	13.9	240	14.76	None	77.937149946939
40.8448#-73.8648	2021-05-04T15:38:00+00:00	16.1	11.7	300	5.4	None	75.162754589756
42.1792#-75.8534	2021-05-04T15:35:00+00:00	None	None	None	None	None	None
42.2318#-78.7476	2021-05-04T14:56:00+00:00	16.1	15	170	22.32	38.88	93.195965844652
42.7655#-76.5488	2021-05-04T13:56:00+00:00	15.6	14.4	0	0	None	92.570484068067
42.2313#-79.5603	2021-05-04T14:56:00+00:00	16.1	15	170	22.32	38.88	93.195965844652

coordinate	generatedAt	elevation	number1	starttime1	endtime1	isDaytime1	temperature1	temp unit1	windspeed1	winddirection1
42.6526#-73.7562	2021-05-04 12:00:44.396667	31.0896	1	2021-05-04T10:00:00-04:00	2021-05-04T18:00:00-04:00	TRUE	61	F	5 mph	NE
42.3130#-78.0195	2021-05-04 12:00:44.774553	438.912	1	2021-05-04T10:00:00-04:00	2021-05-04T18:00:00-04:00	TRUE	73	F	9 mph	SW



## Weather - New Sources

- Tomorrow.io
- Openweather.co
- Visualcrossing.com
- Worldweatheronline.com
- .....

Weather.com!!!!!!



The  
Weather  
Channel

An IBM Business

Hourly Weather - Bethlehem, PA

As of 4:42 pm EDT

Wednesday, October 13

4:45 pm **69°** Cloudy 1% WSW 7 mph

5:00 pm **69°** Mostly Cloudy 2% WSW 7 mph

6:00 pm **69°** Partly Cloudy 2% WSW 6 mph

7:00 pm **66°** Partly Cloudy 2% WSW 5 mph

Feels Like  
**66°**

Wind  
**WSW 5 mph**

Humidity  
**86%**

UV Index  
**0 of 10**

# Weather - Web Scraping

Time: everyday from 9:30am

Location: 62 counties in NY state

Interval: next day 24 hours

Attribute: (see the graph)

Data type: hourly forecast weather data

Albany County, NY	12 am	59°	Mostly Cloudy	59°	SE 5 mph	87%	0 of 10
Albany County, NY	1 am	57°	Mostly Cloudy	57°	SE 5 mph	87%	0 of 10
Albany County, NY	2 am	56°	Partly Cloudy	56°	SE 5 mph	88%	0 of 10
Albany County, NY	3 am	55°	Partly Cloudy	54°	SE 5 mph	90%	0 of 10
Albany County, NY	4 am	54°	Mostly Cloudy	54°	SE 4 mph	92%	0 of 10
Albany County, NY	5 am	55°	Mostly Cloudy	55°	SSE 3 mph	90%	0 of 10
Albany County, NY	6 am	54°	Mostly Cloudy	54°	SE 3 mph	92%	0 of 10
Albany County, NY	7 am	54°	Mostly Cloudy	54°	SSE 2 mph	92%	0 of 10
Albany County, NY	8 am	55°	Cloudy	55°	SE 2 mph	91%	0 of 10
Albany County, NY	9 am	56°	Cloudy	56°	SE 3 mph	88%	1 of 10
Albany County, NY	10 am	59°	Cloudy	59°	SE 4 mph	82%	1 of 10
Albany County, NY	11 am	61°	Cloudy	61°	SE 7 mph	76%	2 of 10
Albany County, NY	12 pm	63°	Cloudy	62°	SE 9 mph	70%	3 of 10
Albany County, NY	1 pm	63°	Cloudy	62°	SE 9 mph	69%	3 of 10
Albany County, NY	2 pm	64°	Cloudy	64°	SE 9 mph	67%	2 of 10
Albany County, NY	3 pm	64°	Cloudy	63°	SE 9 mph	67%	2 of 10
Albany County, NY	4 pm	64°	Cloudy	64°	SE 9 mph	65%	1 of 10
Albany County, NY	5 pm	63°	Cloudy	63°	SE 8 mph	68%	0 of 10
Albany County, NY	6 pm	61°	Cloudy	61°	SE 7 mph	72%	0 of 10
Albany County, NY	7 pm	60°	Cloudy	59°	SE 6 mph	76%	0 of 10
Albany County, NY	8 pm	58°	Cloudy	58°	SE 5 mph	77%	0 of 10
Albany County, NY	9 pm	58°	Cloudy	57°	SE 4 mph	80%	0 of 10
Albany County, NY	10 pm	57°	Cloudy	57°	SE 4 mph	82%	0 of 10
Albany County, NY	11 pm	57°	Cloudy	57°	SE 4 mph	84%	0 of 10

# Weather - Query

---

Provide query interface in Python:

Three functions:

Query by period (from date to date)

Query by day( Monday - Sunday)

Query by season (Spring - Winter)

```
which of the following operations you would like to perform on your dataset
1:-----query by season
      you could query the dataset by 4 seasons: Spring, Summer, Fall, Winter
      for example: you can get all Summer weather data in this dataset
2:-----query by day
      you could query the dataset by 7 days of the week: MONDAY, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday
      for example: you can get all Monday weather data in this dataset
3:-----query by period
      you could query the dataset by a period of time: from yyyy-mm-dd to yyyy-mm-dd
      for example: you can get all weather data that is from 2021-10-27 to 2021-10-28 in this dataset
4:-----query by detail
      you could query the dataset by all the detailed above combined together
      for example: you can get all the weather data that is on Monday from 2021-10-27 to 2021-10-28 in this dataset
                           or: you can get all the Summer weather data that is on Monday in this merged dataset
enter 1 or 2 or 3 or 4 as explained above to continue the query operation, enter 0 to quit the system
```

# Weather - database

---

Store data in the SQLite database after it's been scraping from the website

Convenient to organize

Easier for required query operation

The screenshot shows a screenshot of DB Browser for SQLite. The window title is "DB Browser for SQLite - /Users/rhym/Desktop/Chrome\_Driver/sqlite.db". The main area displays a table named "Weather" with the following columns: County, T\_date, Temperature, Weather, Feels, Wind\_Direction, Wind\_Speed, Humidity, and UV. The table contains 24 rows of data, each representing a weather record for a specific date and location. The data includes various weather conditions like Clear, Mostly Clear, Partly Cloudy, etc., with corresponding temperatures ranging from 26 to 44 degrees Celsius.

	County	T_date	Temperature	Weather	Feels	Wind_Direction	Wind_Speed	Humidity	UV
1	Dutchess Count...	2021-11-20 00...	30	Clear	30	W	2	4	0
2	Dutchess Count...	2021-11-20 01...	29	Clear	29	NNW	2	4	0
3	Dutchess Count...	2021-11-20 02...	28	Mostly Clear	28	NW	2	4	0
4	Dutchess Count...	2021-11-20 03...	28	Partly Cloudy	28	NNW	2	4	0
5	Dutchess Count...	2021-11-20 04...	27	Partly Cloudy	27	N	1	5	0
6	Dutchess Count...	2021-11-20 05...	27	Partly Cloudy	27	NNW	0	5	0
7	Dutchess Count...	2021-11-20 06...	26	Clear	26	NW	0	5	0
8	Dutchess Count...	2021-11-20 07...	25	Sunny	25	S	0	6	0
9	Dutchess Count...	2021-11-20 08...	29	Sunny	29	S	0	4	0
10	Dutchess Count...	2021-11-20 09...	34	Sunny	34	W	1	1	1
11	Dutchess Count...	2021-11-20 10...	37	Sunny	37	WSW	2	0	1
12	Dutchess Count...	2021-11-20 11...	40	Mostly Sunny	40	WSW	2	0	2
13	Dutchess Count...	2021-11-20 12...	42	Partly Cloudy	42	SSW	3	0	2
14	Dutchess Count...	2021-11-20 13...	44	Mostly Cloudy	44	SSW	6	0	1
15	Dutchess Count...	2021-11-20 14...	44	Mostly Cloudy	44	SSW	6	0	1
16	Dutchess Count...	2021-11-20 15...	45	Mostly Cloudy	45	SSW	6	0	0
17	Dutchess Count...	2021-11-20 16...	46	Cloudy	43	S	6	0	0
18	Dutchess Count...	2021-11-20 17...	41	Cloudy	41	S	5	0	0
19	Dutchess Count...	2021-11-20 18...	38	Cloudy	38	S	5	1	0
20	Dutchess Count...	2021-11-20 19...	37	Cloudy	37	S	5	1	0
21	Dutchess Count...	2021-11-20 20...	37	Cloudy	37	S	4	2	0
22	Dutchess Count...	2021-11-20 21...	36	Cloudy	36	S	4	2	0
23	Dutchess Count...	2021-11-20 22...	36	Cloudy	36	S	4	4	0
24	Dutchess Count...	2021-11-20 23...	35	Cloudy	35	S	4	5	0

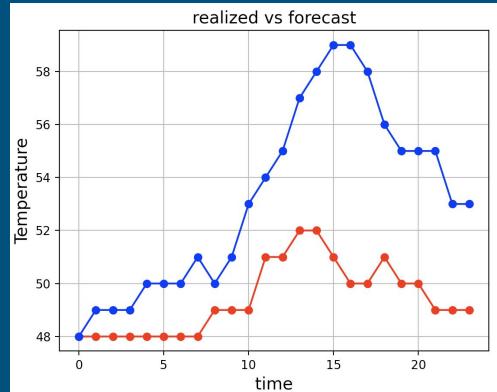
# Weather - Compare

---

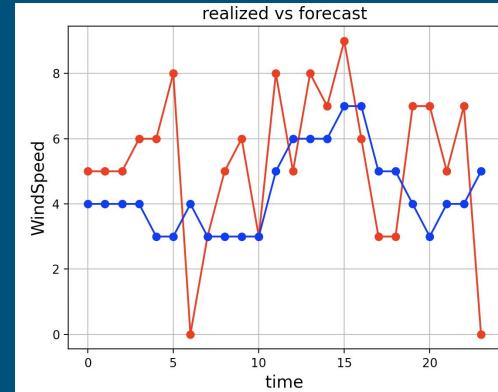
Compare the realized data and forecast data

Compute the rmse for Temperature and Wind Speed

Rmse: 4.48



Rmse: 2.57



# Energy Supply

---

# Data Sources

---

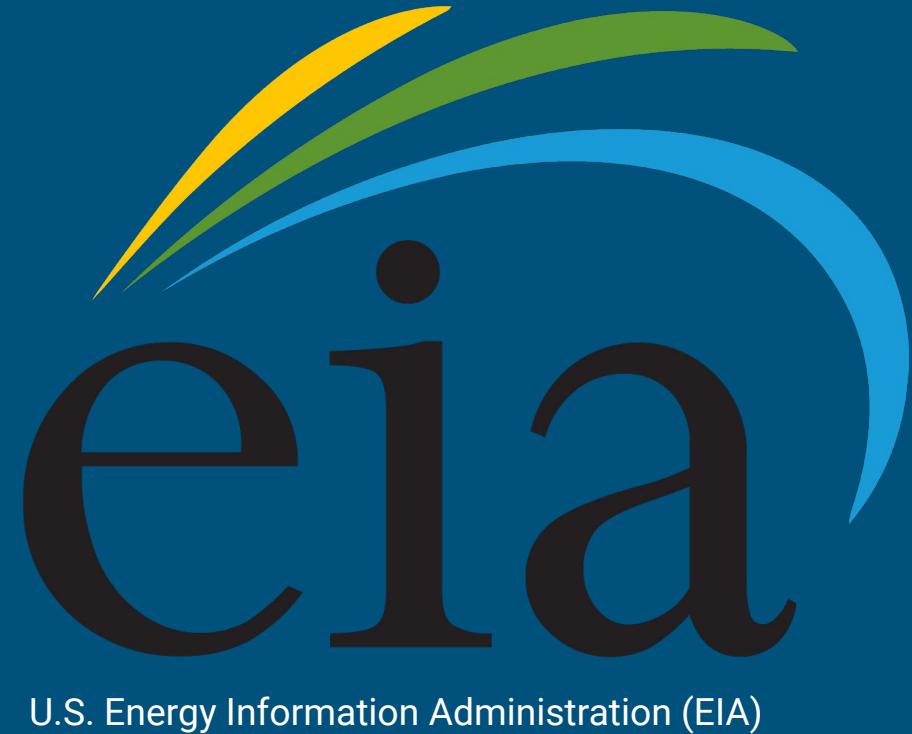
<https://www.epa.gov/egrid/download-data>

<https://www.eia.gov/electricity/data/eia861/>

<https://ampd.epa.gov/ampd/>



Environmental Protection Agency (EPA)



# Collected Data

State	Facility Name	Facility ID (Unit ID)	Associated Year	Date	Hour	Program(s)	SO2 (pounds)	Avg. NOx Rate (lb/MMBtu)	NOx (pounds)	CO2 (short tons)	Heat Input County	Input County	Owner	Operator	Operating Status	Facility Latitude	Facility Longitude	Utility Name	Utility ID	Utility Total Capacity (MW)	Utility Net Generation (MWh)
NY	Harlem Riv	7914.HR02		2018.5/14/2018	7	ARP	0	0.063	0	0	0.048.Bronx Cou	New York	E New York	I Operating	40.7989	-73.9147	New York	F	15296	#####	

- EPA datasets:

State Facility Name Facility ID (ORISPL) Unit ID Associated Stacks Year Date Hour Program(s)

SO2 (pounds) Avg. NOx Rate (lb/MMBtu) NOx (pounds) CO2 (short tons)

Heat Input (MMBtu) County Owner Operator Operating Status Facility Latitude  
Facility Longitude

- Egrid datasets:

Utility Name Utility ID

- EIA-861 datasets:

Utility Total Capacity (MW) Utility Net Generation (MWh)

# Difficulties Encountered

---

- No direct connection between the EIA and the EPA.
- Datasets after null rows are moved only have 7000 despite the fact that, for each term, half a million observations were collected.
- Collecting over a million datasets once will fail in EPA.

## Datasets before removing the null values

- Around 50,000 observations for each term, half of one year.
- Observations for one operator for one term expected is 4380 observations.

558189	NY	Vernon Bo	7909	VB02		2018	6/30/2018	19	ARP	0.242	0.008	3.2	23.9	402.9	Queens Co	New York	f	New York	f	Operating
558190	NY	Vernon Bo	7909	VB02		2018	6/30/2018	20	ARP	0.237	0.007	2.8	23.5	395.7	Queens Co	New York	f	New York	f	Operating
558191	NY	Vernon Bo	7909	VB02		2018	6/30/2018	21	ARP	0.005	0.052	0.408	0.472	7.912	Queens Co	New York	f	New York	f	Operating

# Matching Process Through Python

- Correlate EPA with egrid by connecting the Unit ID from the EPA and the Egrid and add the columns of Utility name and Utility ID into the EPA.
- Correlate EPA with EIA by connecting the Unit ID from the EPA with the utility number from the EIA and add the columns of Total Capacity and Utility Number into the EPA.
- Reformat the form of datasets. Drop the rows with null values.

```
# Matching IDs from base_file and add_file, respectively.  
leftIDs = [' Facility ID (ORISPL)', ' Unit ID']  
rightIDs = ['DOE/EIA ORIS plant or facility code', 'Unit ID']  
  
# Names of columns from add_file that need to be added to base_file.  
addCols = ['Utility name', 'Utility ID', 'DOE/EIA ORIS plant or facility code']
```

```
# Names of columns from add_file1 that need to be added to base_file1.  
addCols1 = ['Total Capacity', 'Utility Number']  
  
# Matching IDs from base_file1 and add_file1, respectively.  
leftIDs1 = ['Utility ID']  
rightIDs1 = ['Utility Number']
```

```
df = pd.read_csv(input_file, low_memory=False, index_col=False)  
df.drop('Associated Stacks', axis=1, inplace=True)  
df.dropna(how = 'any', inplace = True)
```

# Analysis of all the datasets

---

- temporal granularity of data sets: hourly.
- spatial granularity: County, Facility Longitude, and Facility Latitude.
- data integrity: comparing the number of observations expected (7000) with how many realized. And the possible explanations for it would be some operators are not really working.

True Observations of combined datasets for every half of a year:

- 2016\_1: 7,600
- 2016\_2: 7,300
- 2017\_1: 6,100
- 2017\_2: 7,400
- 2018\_1: 20,000
- 2018\_2: 6,300
- 2019\_1: 6,800
- 2019\_2: 7,700
- 2020\_1: 8,200
- 2020\_2: 7,800

Observations around 7,000

# Demand/Pricing

---



# New York ISO

## Independent System Operator

Load Forecast	CSV
NY Load Forecast	<a href="#">CSV</a>
Zonal Load Commitment	<a href="#">CSV</a>
Load Forecast Weather Data	<a href="#">CSV</a>

Actual Load	CSV	Last Updated
November 30, 2021	<a href="#">CSV</a>	12/01/21 00:15 EST
November 29, 2021	<a href="#">CSV</a>	11/30/21 00:22 EST
November 28, 2021	<a href="#">CSV</a>	11/29/21 00:15 EST
November 27, 2021	<a href="#">CSV</a>	11/28/21 00:11 EST
November 26, 2021	<a href="#">CSV</a>	11/27/21 00:12 EST
November 25, 2021	<a href="#">CSV</a>	11/26/21 00:13 EST
November 24, 2021	<a href="#">CSV</a>	11/25/21 00:12 EST

[Archive](#)[Custom Report](#)

### Integrated Real-Time

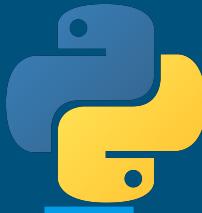


### Real-Time Actual Load

CSV Files	Last Updated
<a href="#">12-03-2021</a>	12/03/21 02:47 EST
<a href="#">12-02-2021</a>	12/03/21 00:18 EST
<a href="#">12-01-2021</a>	12/02/21 00:13 EST
<a href="#">11-30-2021</a>	12/01/21 00:13 EST
<a href="#">11-29-2021</a>	11/30/21 00:22 EST
<a href="#">11-28-2021</a>	11/29/21 00:15 EST
<a href="#">11-27-2021</a>	11/28/21 00:11 EST
<a href="#">11-26-2021</a>	11/27/21 00:12 EST
<a href="#">11-25-2021</a>	11/26/21 00:13 EST
<a href="#">11-24-2021</a>	11/25/21 00:12 EST
<a href="#">11-23-2021</a>	11/24/21 00:17 EST

### Archived Files (zip format)

CSV Files	Last Updated
<a href="#">12-2021</a>	12/03/21 02:47 EST
<a href="#">11-2021</a>	12/01/21 00:13 EST
<a href="#">10-2021</a>	11/01/21 00:08 EDT
<a href="#">09-2021</a>	10/01/21 00:03 EDT
<a href="#">08-2021</a>	09/01/21 00:11 EDT
<a href="#">07-2021</a>	08/01/21 00:05 EDT
<a href="#">06-2021</a>	07/01/21 00:11 EDT
<a href="#">05-2021</a>	06/01/21 00:06 EDT
<a href="#">04-2021</a>	05/01/21 00:07 EDT
<a href="#">03-2021</a>	04/01/21 00:02 EDT
<a href="#">02-2021</a>	03/01/21 00:04 EST
<a href="#">01-2021</a>	02/01/21 00:18 EST
<a href="#">12-2020</a>	01/01/21 00:06 EST
<a href="#">11-2020</a>	12/01/20 00:02 EST
<a href="#">10-2020</a>	11/02/20 16:01 EST
<a href="#">09-2020</a>	09/30/20 23:58 EDT
<a href="#">08-2020</a>	08/31/20 23:57 EDT
<a href="#">07-2020</a>	07/31/20 23:57 EDT
<a href="#">06-2020</a>	06/30/20 23:57 EDT
<a href="#">05-2020</a>	05/31/20 23:57 EDT
<a href="#">04-2020</a>	04/30/20 23:57 EDT
<a href="#">03-2020</a>	03/31/20 23:57 EDT
<a href="#">02-2020</a>	02/29/20 23:57 EDT



# Python

In [4]: load20170105

Out[4]:

	Time Stamp	Time Zone	Name	PTID	Load
0	01/05/2017 00:00:00	EST	CAPITL	61757	1232.6
1	01/05/2017 00:00:00	EST	CENTRL	61754	1801.2
2	01/05/2017 00:00:00	EST	DUNWOD	61760	680.1
3	01/05/2017 00:00:00	In [5]:	loadF20170105		
4	01/05/2017 00:00:00	Out[5]:			
...	...				
3163	01/05/2017 23:55:00				
3164	01/05/2017 23:55:00				
3165	01/05/2017 23:55:00				
3166	01/05/2017 23:55:00				
3167	01/05/2017 23:55:00				
3168 rows × 5 columns					

	Time Stamp	Capitol	Centrl	Dunwod	Genese	H
0	01/05/2017 00:00:00	1194	1650	563	1005	9
1	01/05/2017 01:00	1168	1610	549	978	9
2	01/05/2017 02:00	1157	1586	543	964	9
3	01/05/2017 03:00	1159	1585	540	964	9
4	01/05/2017 04:00	1183	1622	549	987	9
...	...	...	...	...	...	...
139	01/10/2017 19:00	1655	2116	824	1287	13
140	01/10/2017 20:00	1606	2051	799	1247	13
141	01/10/2017 21:00	1525	1941	761	1182	1269
142	01/10/2017 22:00	1425	1809	702	1100	1181
143	01/10/2017 23:00	1341	1703	647	1034	1101
144 rows × 13 columns						

In [6]: price20170105

Out[6]:

	Time Stamp	Name	PTID	LBMP (\$/MWhr)	Marginal Cost Losses (\$/MWhr)	Marginal Cost Congestion (\$/MWhr)
0	01/05/2017 00:05:00	CAPITL	61757	26.95	1.57	-4.21
1	01/05/2017 00:05:00	CENTRL	61754	21.87	0.19	-0.51
2	01/05/2017 00:05:00	DUNWOD	61760	26.25	1.97	-3.11
3	01/05/2017 00:05:00	GENESE	61753			

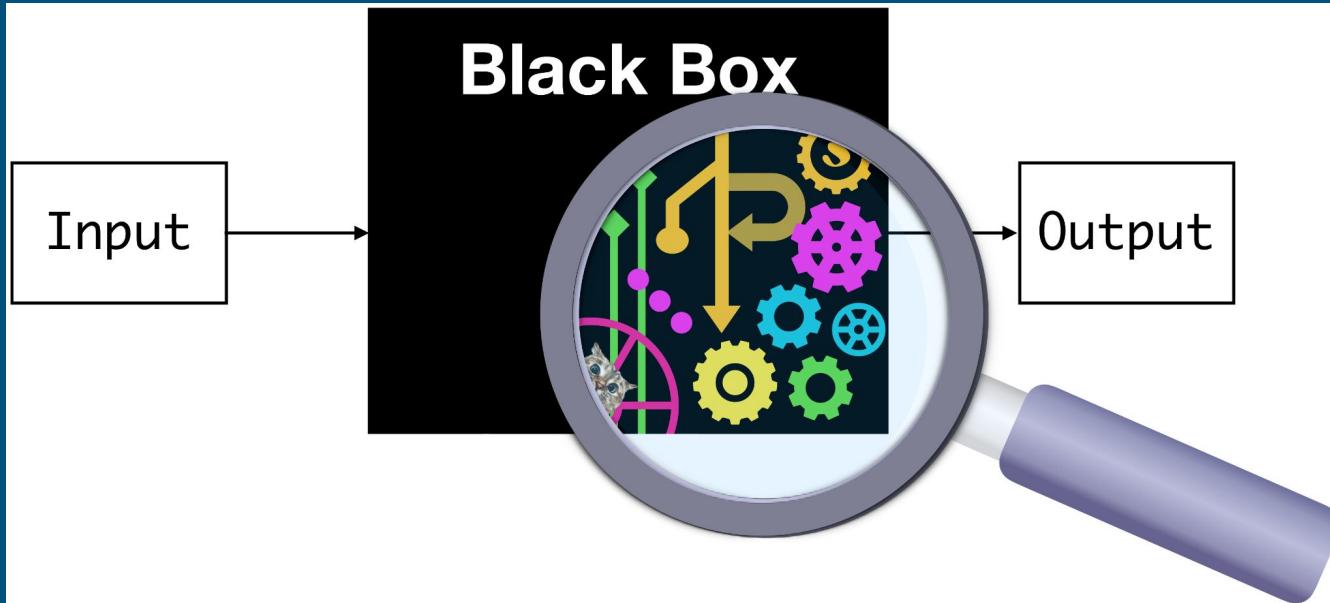
In [7]: priceF20170105

Out[7]:

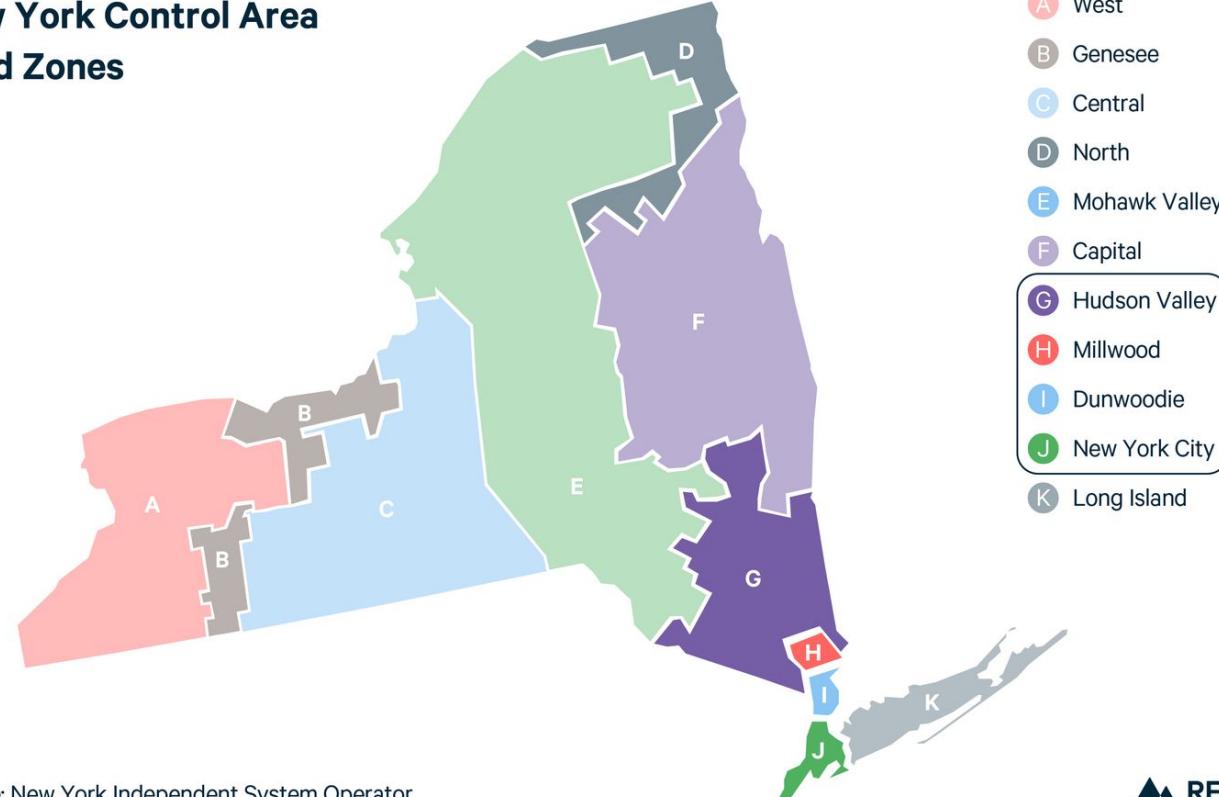
	Time Stamp	Name	PTID	LBMP (\$/MWhr)	Marginal Cost Losses (\$/MWhr)	Marginal Cost Congestion (\$/MWhr)
0	01/05/2017 00:00:00	CAPITL	61757	39.51	1.41	-16.11
1	01/05/2017 00:00:00	CENTRL	61754	24.18	0.51	-1.69
2	01/05/2017 00:00:00	DUNWOD	61760	35.17	2.05	-11.13
3	01/05/2017 00:00:00	GENESE	61753	23.58	0.29	-1.30
4	01/05/2017 00:00:00	H Q	61844	20.30	-0.81	0.88
...	...	...	...	...	...	...
4315	01/06/2017 00:00:00	NORTH	61755			
4316	01/06/2017 00:00:00	NPX	61845			
4317	01/06/2017 00:00:00	O H	61846			
4318	01/06/2017 00:00:00	PJM	61847			
4319	01/06/2017 00:00:00	WEST	61752			
355	01/05/2017 23:00	NORTH	61755	25.08	-1.07	0.00
356	01/05/2017 23:00	NPX	61845	41.03	1.91	-12.97
357	01/05/2017 23:00	O H	61846	26.79	-0.99	-1.63
358	01/05/2017 23:00	PJM	61847	32.39	0.60	-5.64
359	01/05/2017 23:00	WEST	61752	27.45	-0.29	-1.59

360 rows × 6 columns

# Forecasts and Risk Scores

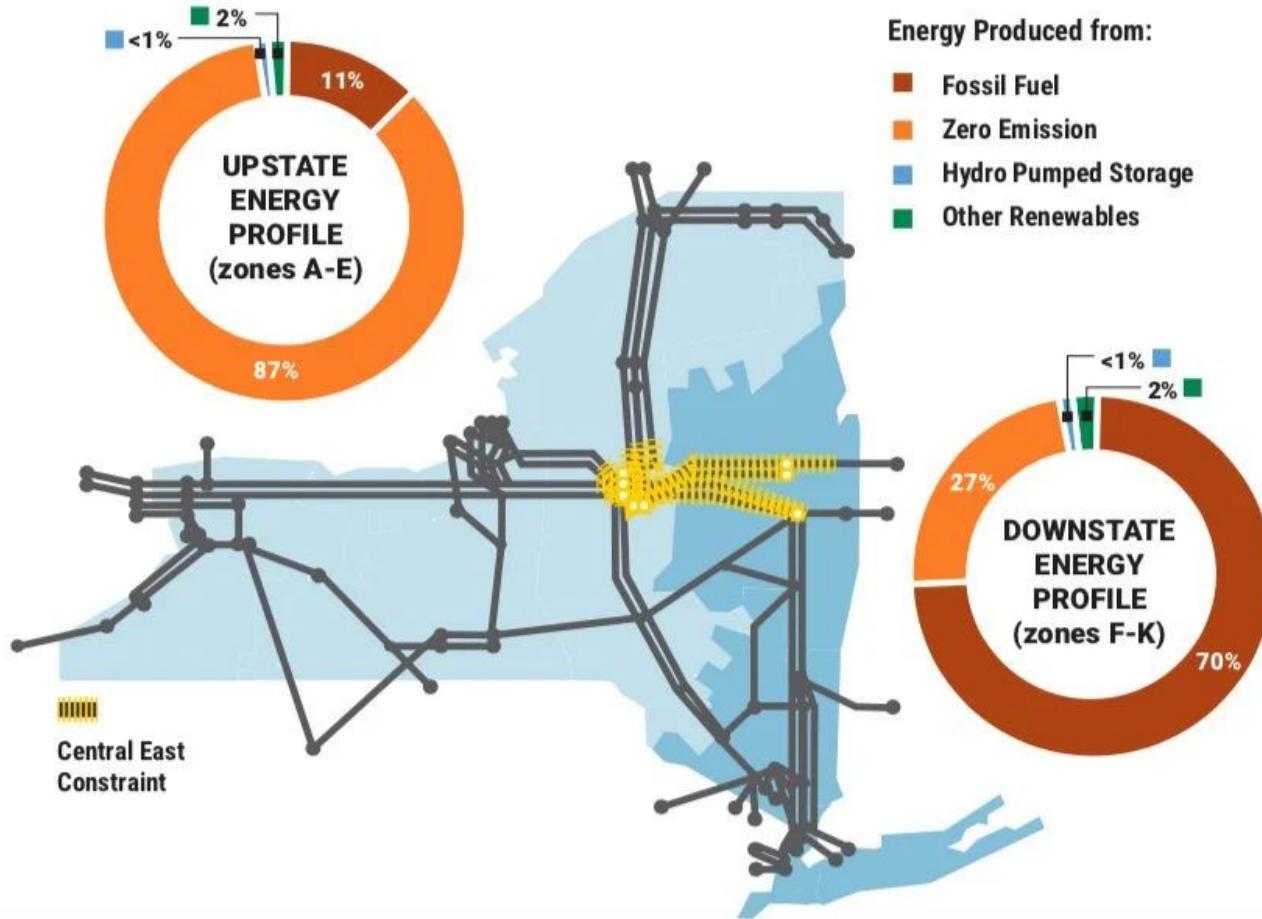


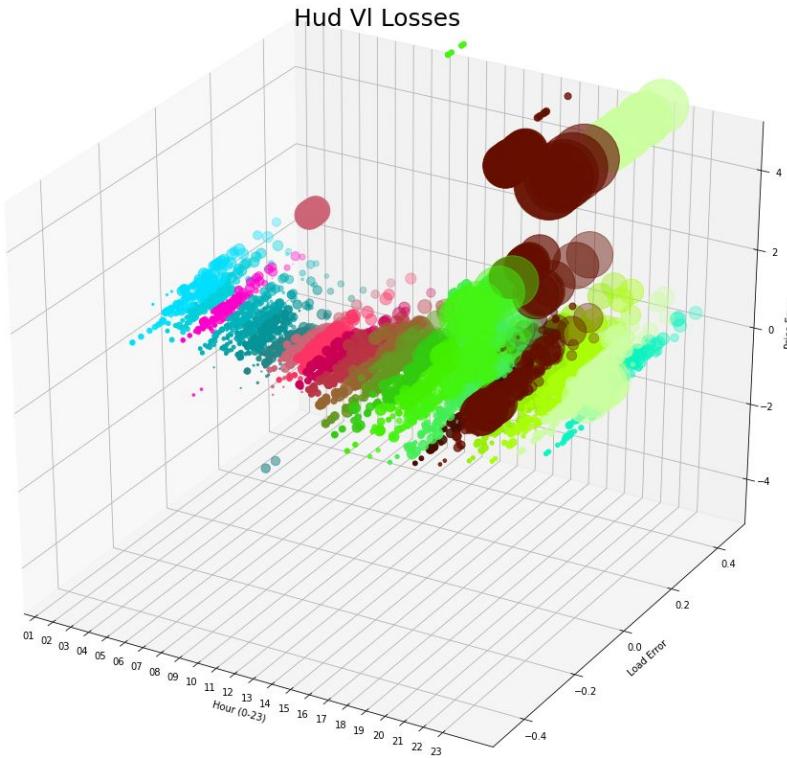
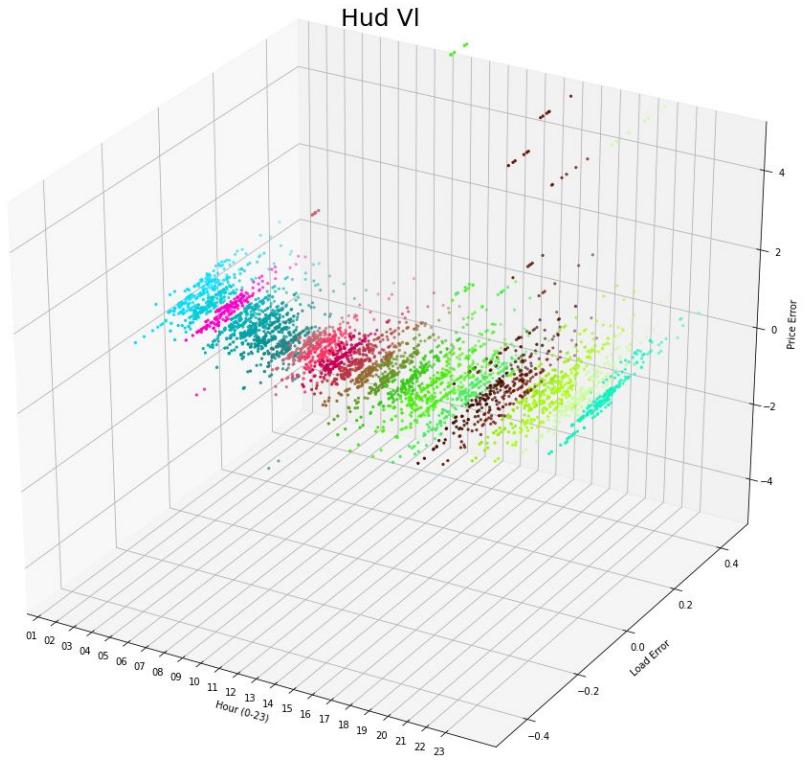
## New York Control Area Load Zones

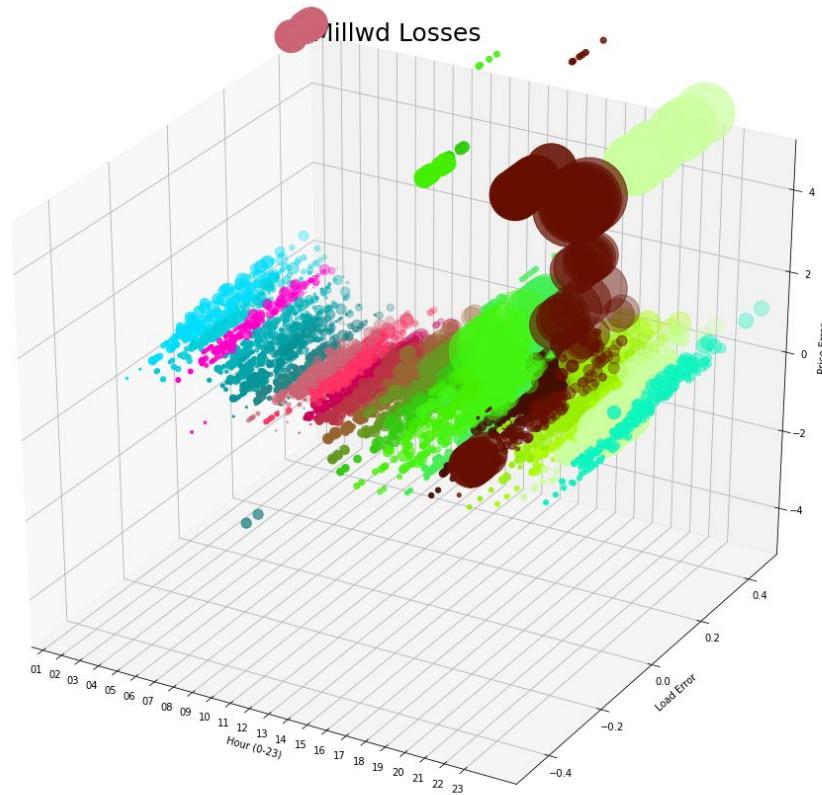
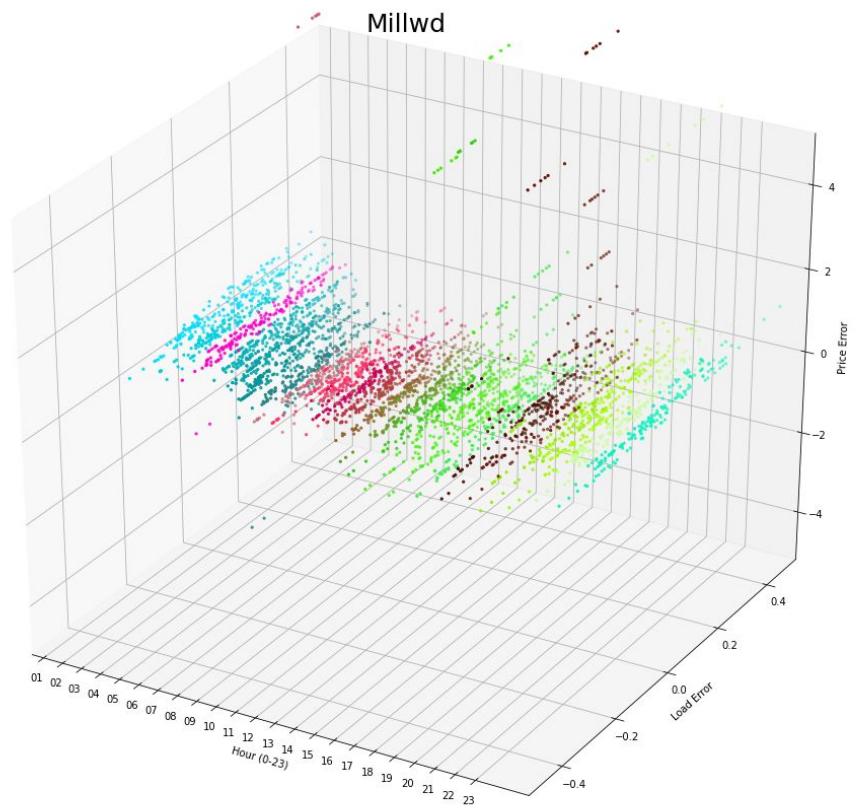


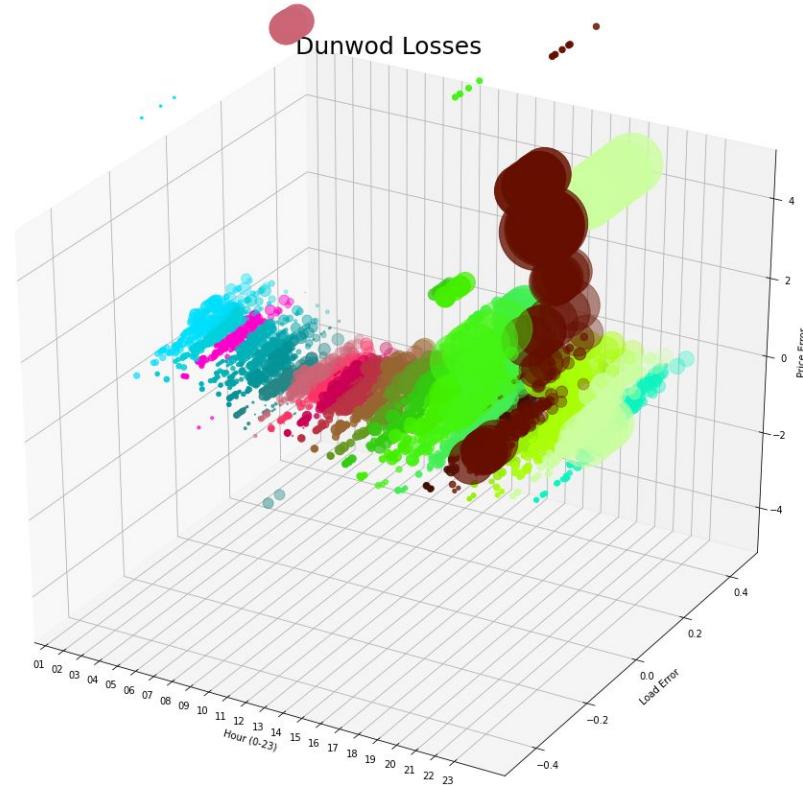
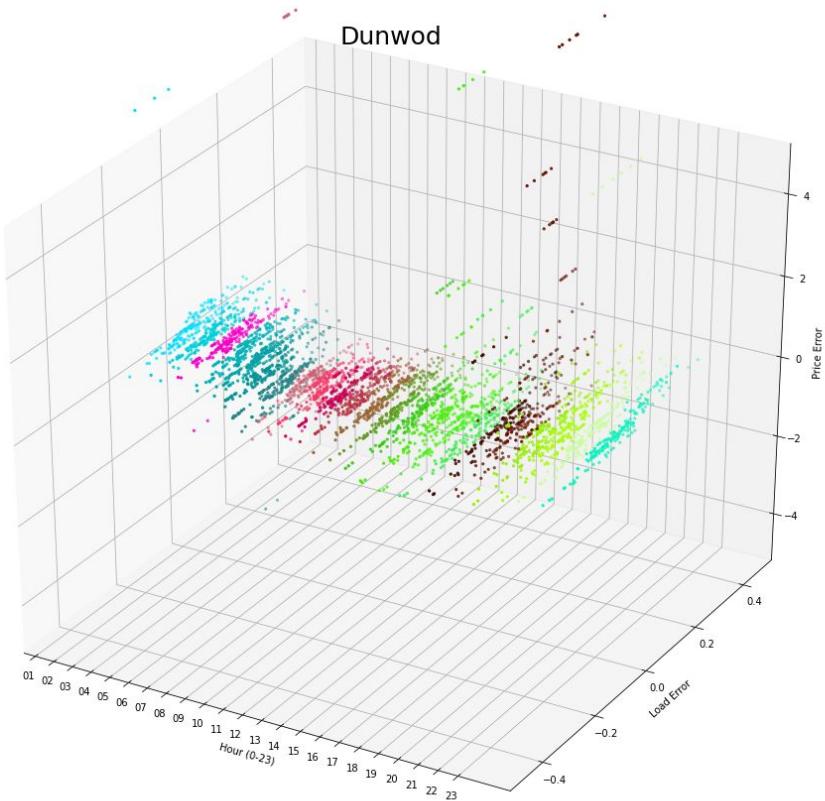
Source: New York Independent System Operator

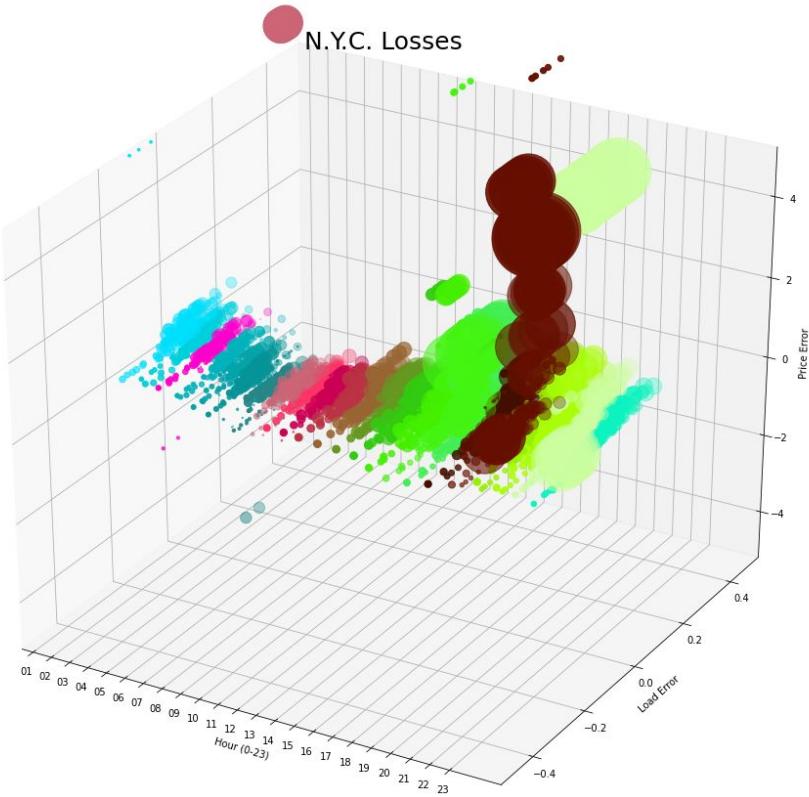
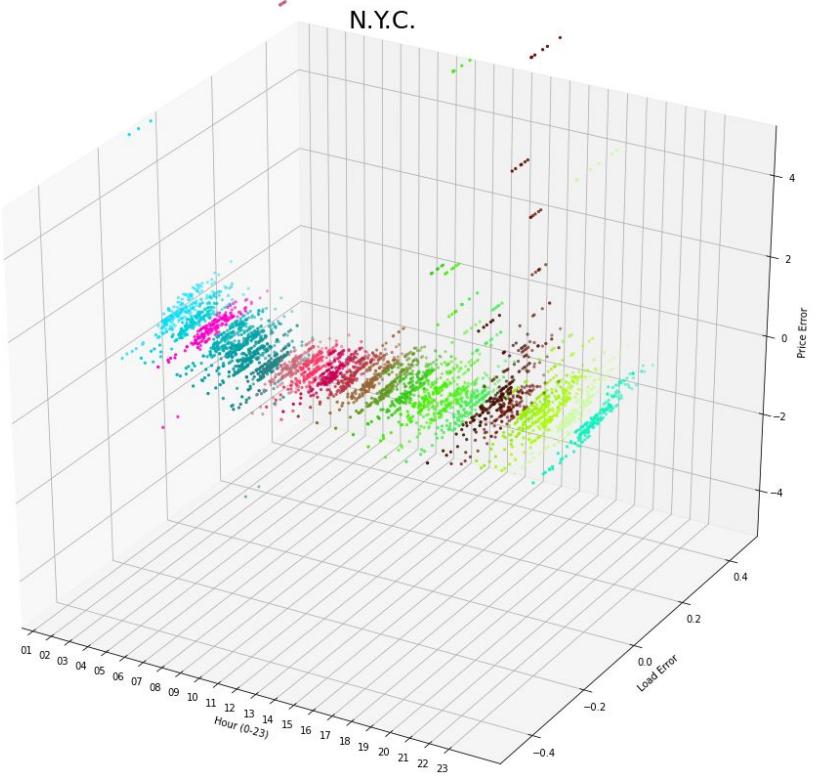
## ► Tale of Two Grids

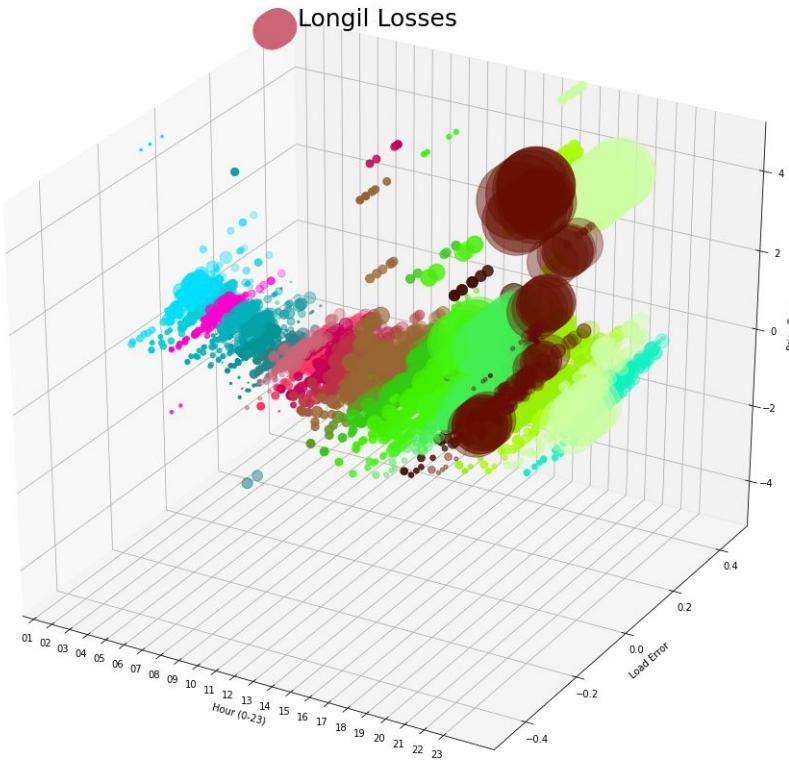
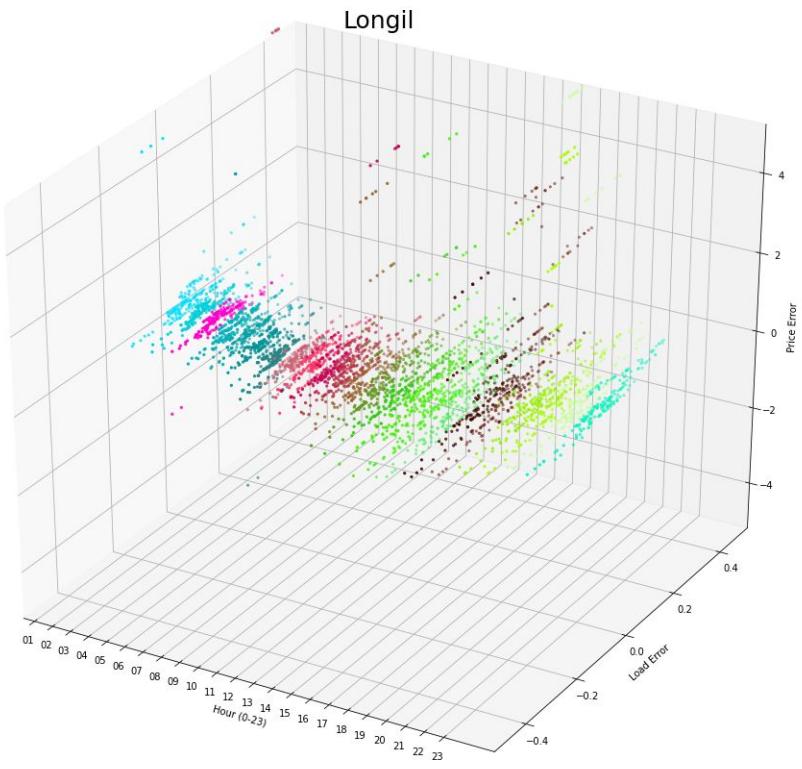




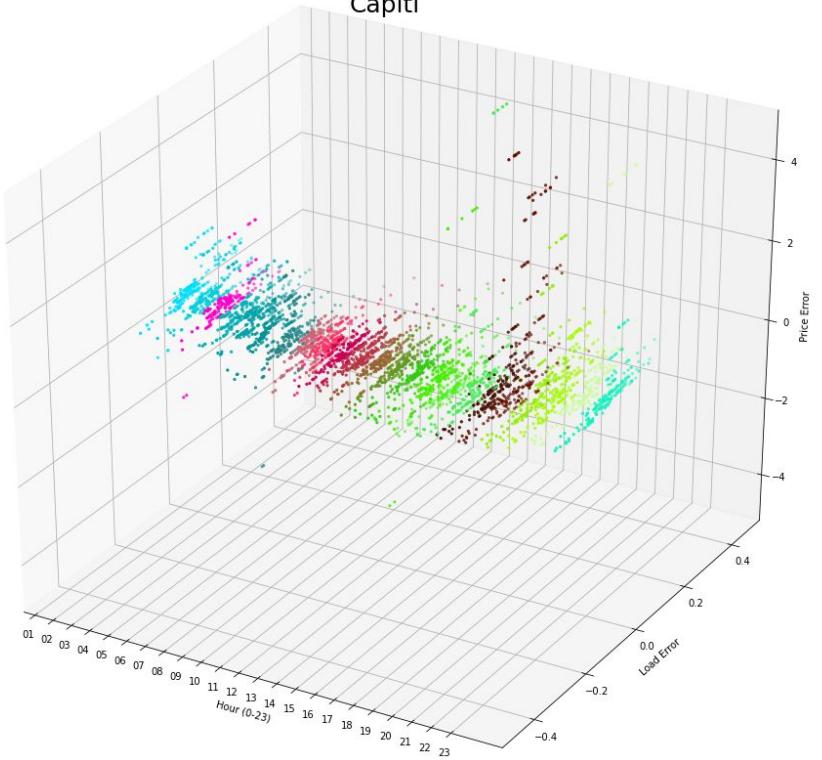




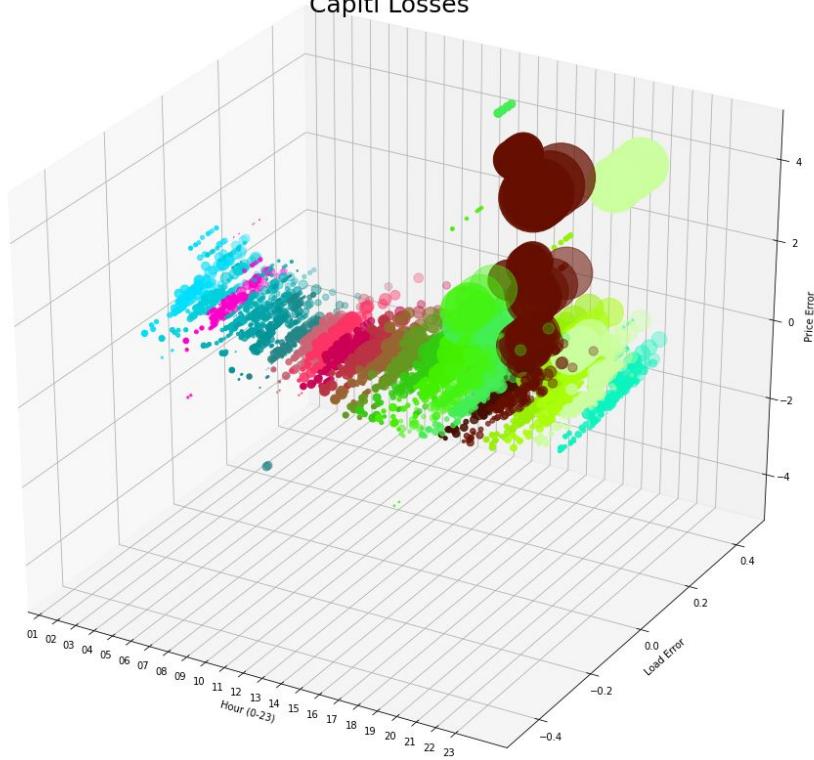




Capitl

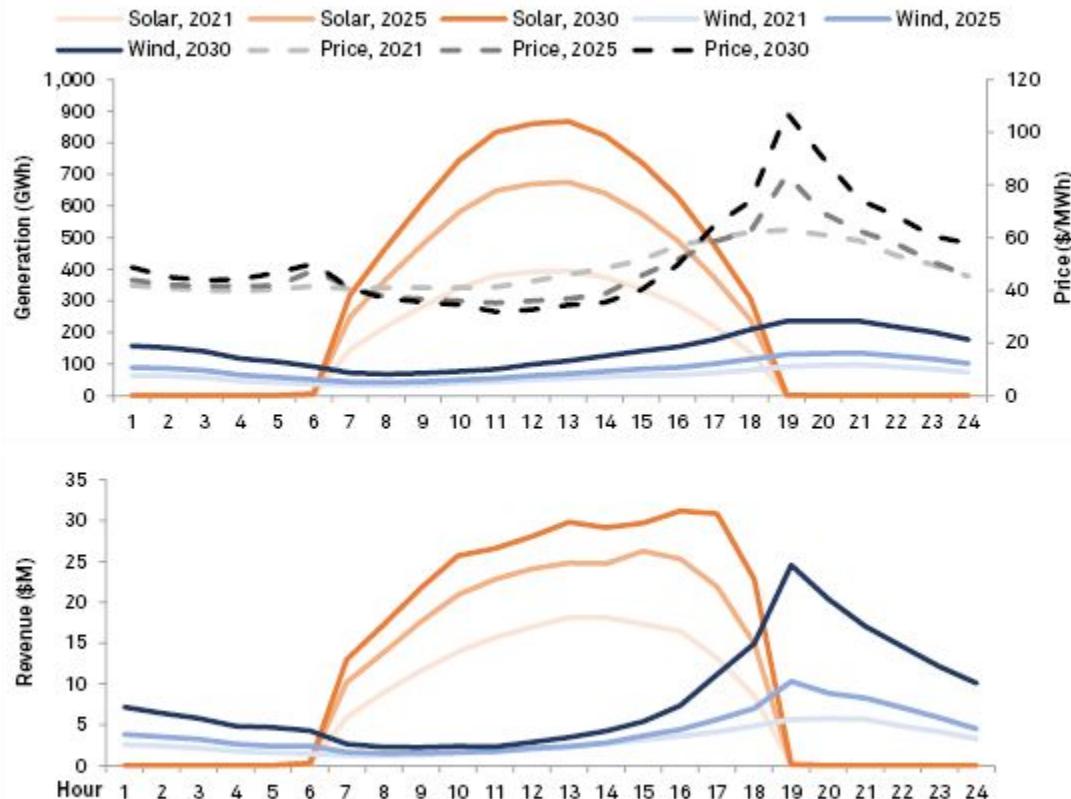


Capitl Losses



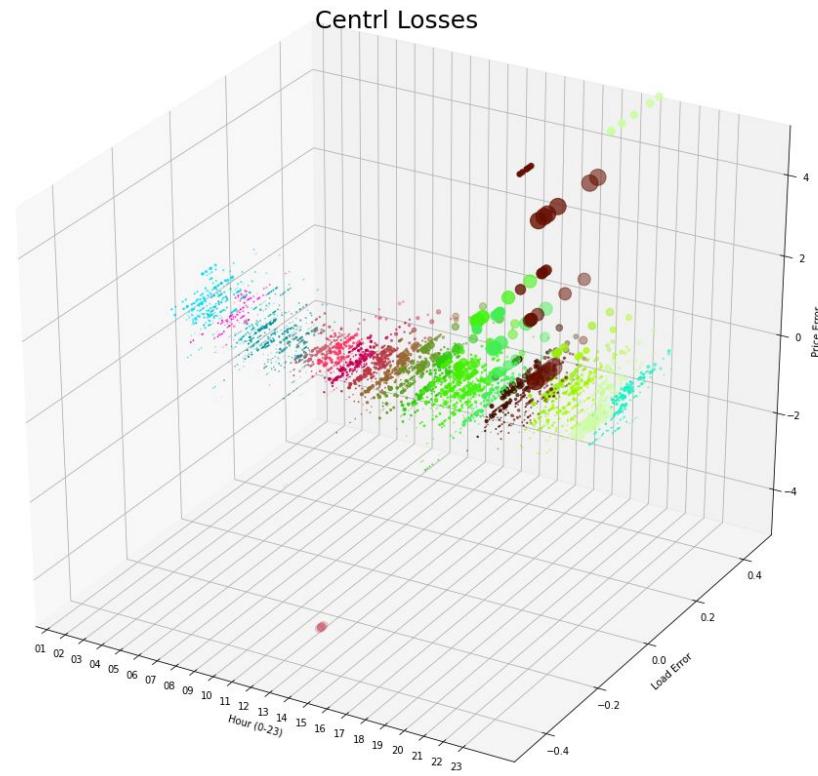
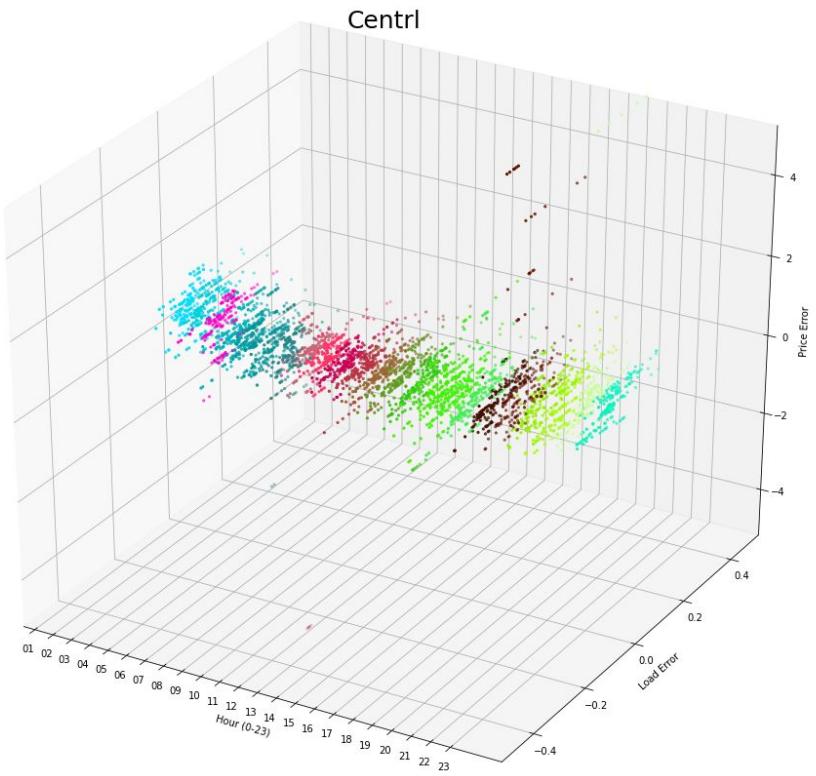
## California

Top: CAISO August solar and wind generation (GWh), SP-15 price (\$/MWh)  
Bottom: Revenue (\$M) of CAISO wind and solar generation

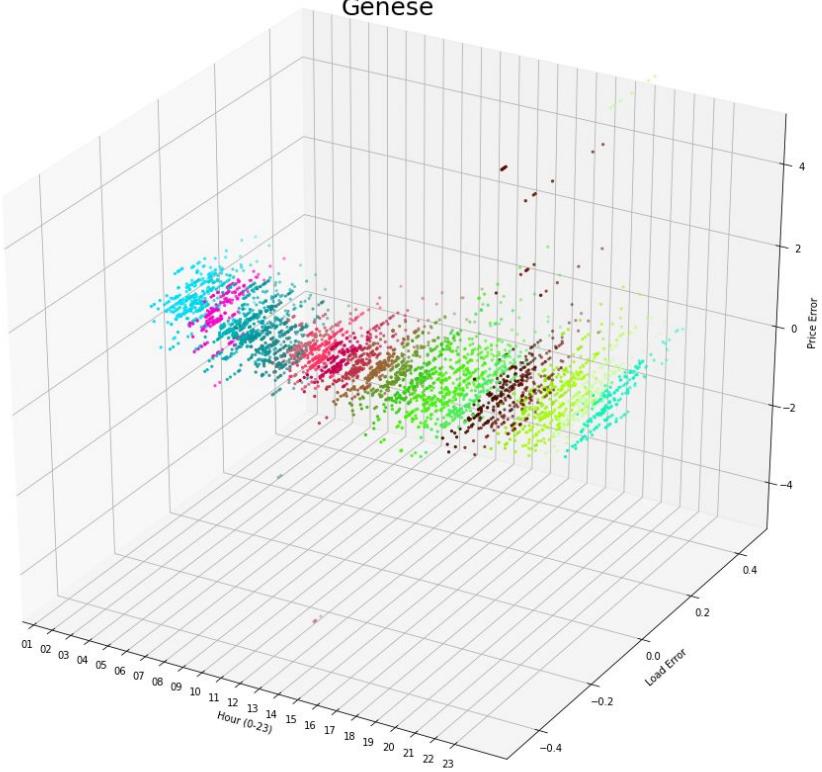


As of: Sept 30, 2020.

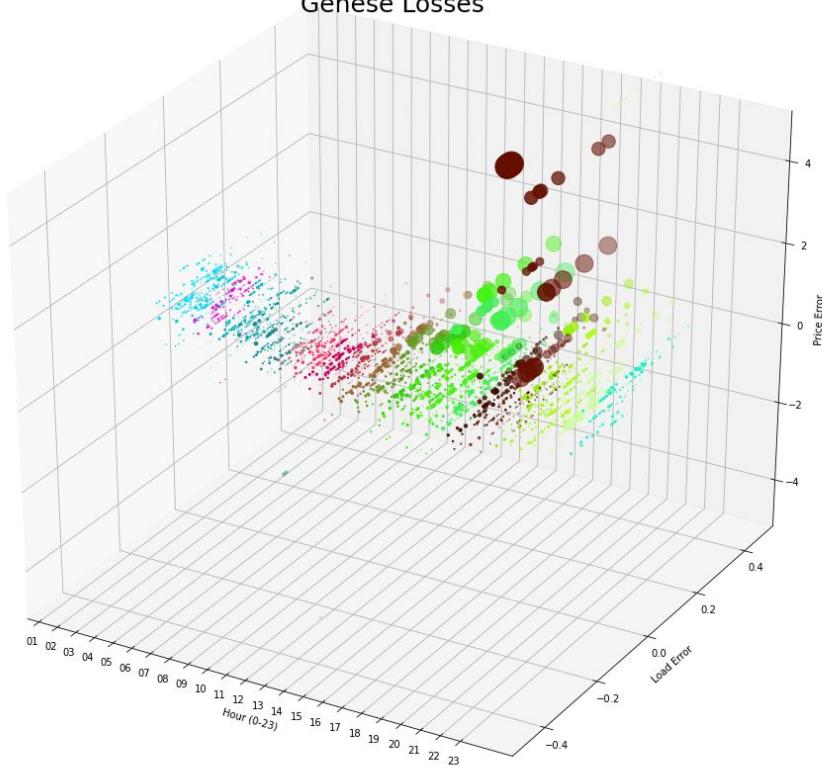
Source: S&P Global Market Intelligence

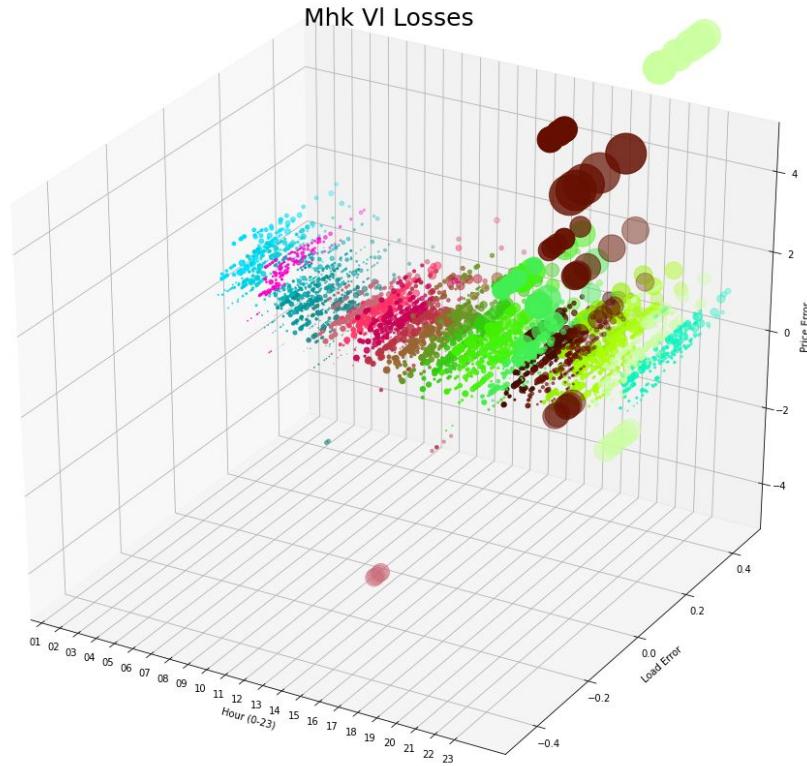
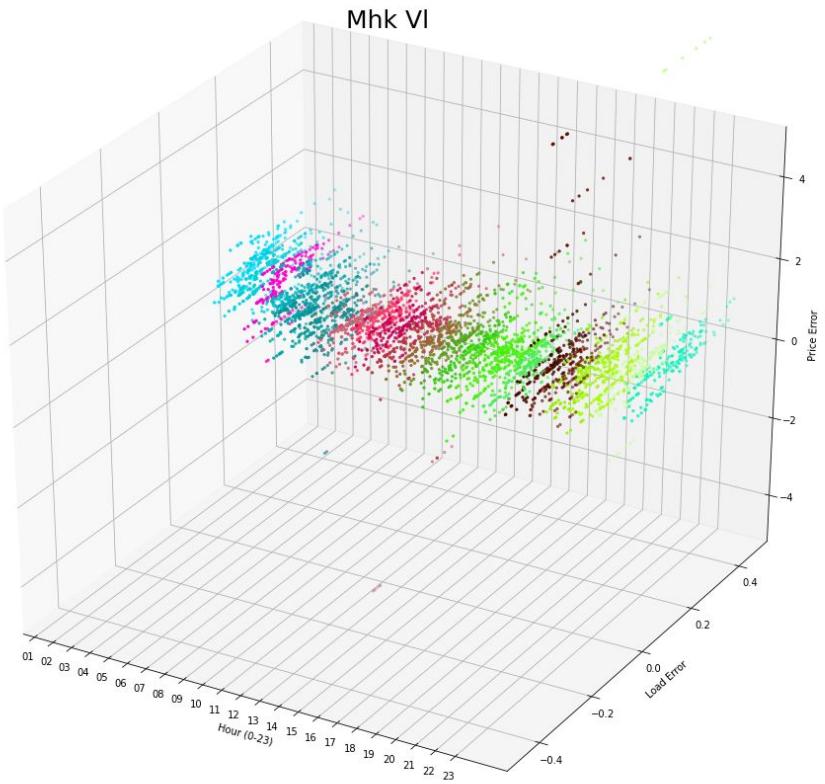


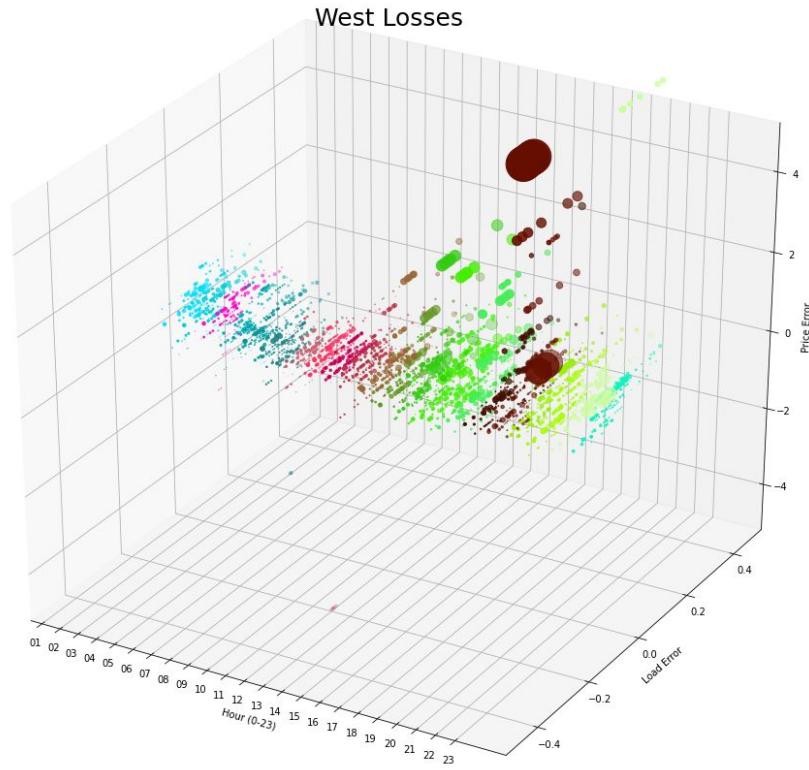
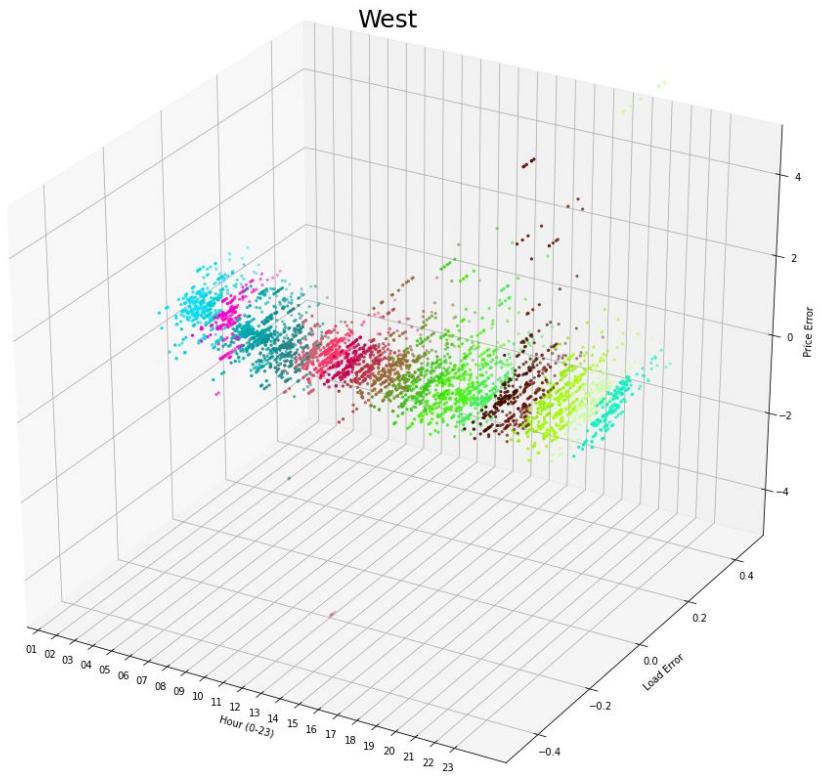
Genese



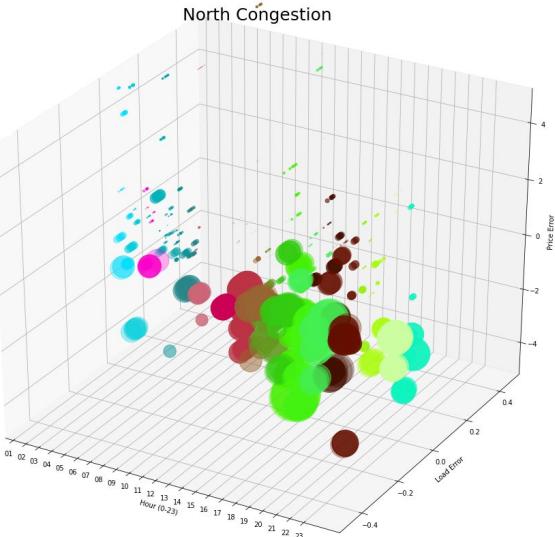
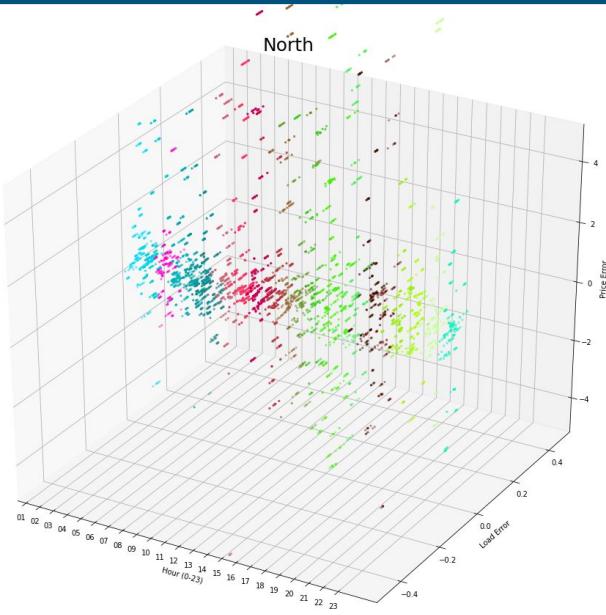
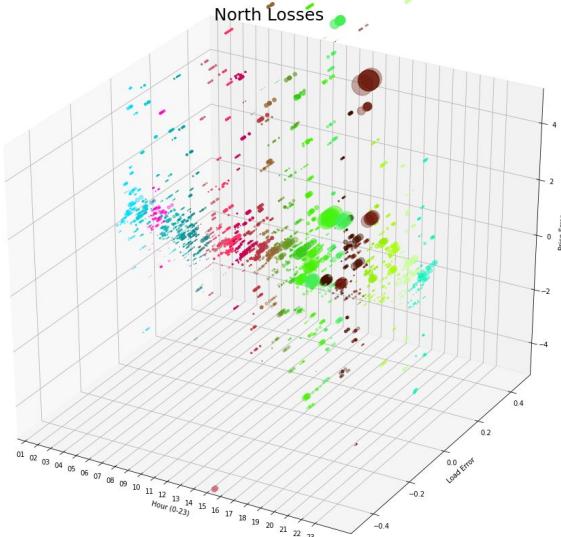
Genese Losses

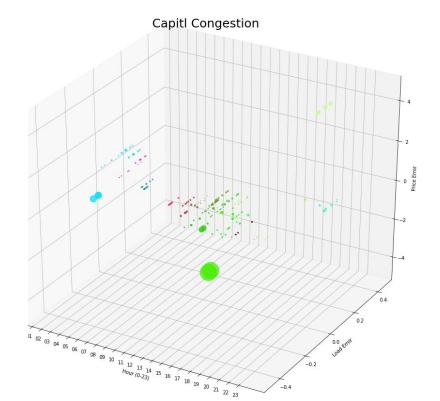
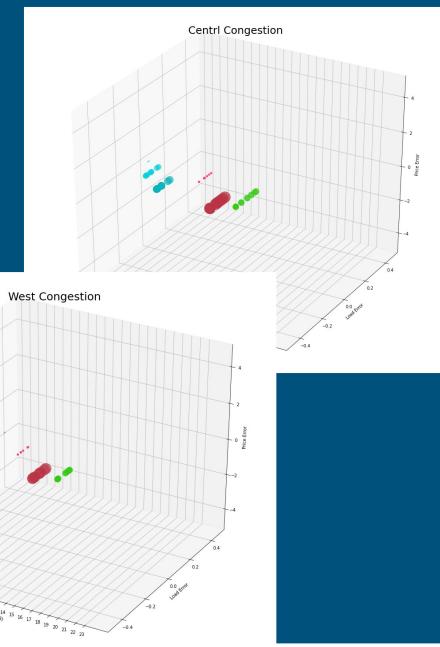
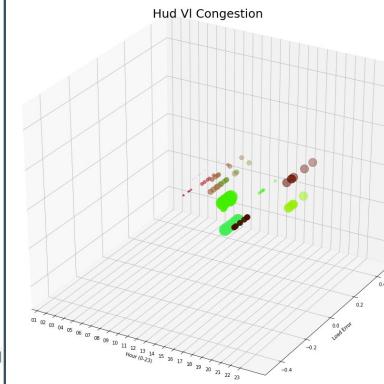
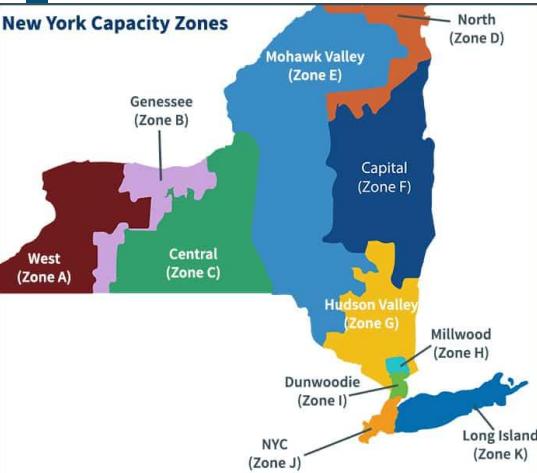
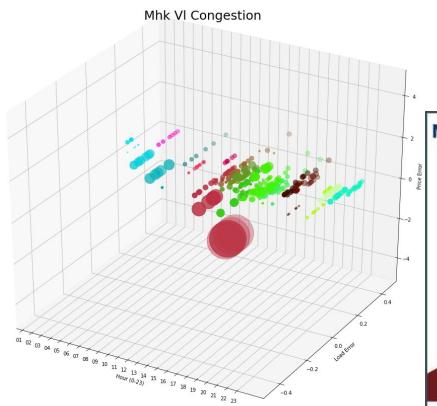
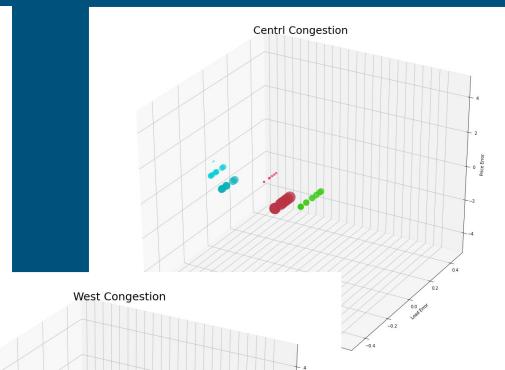
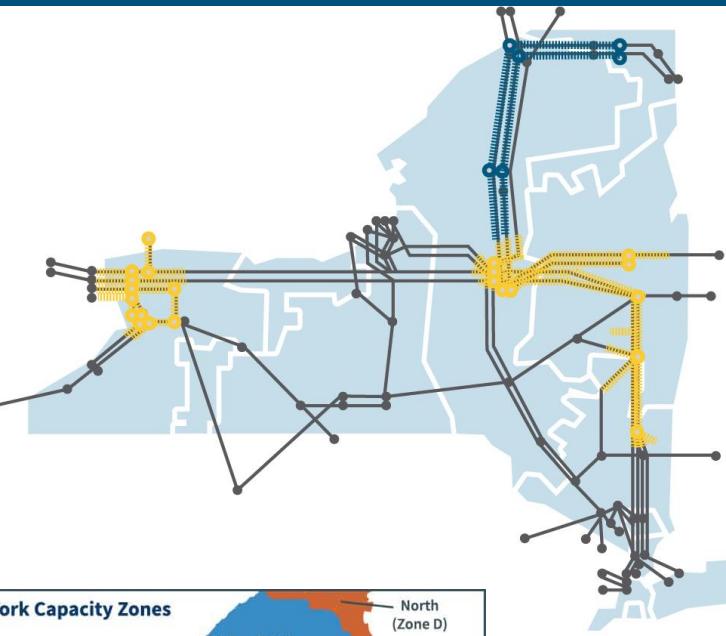
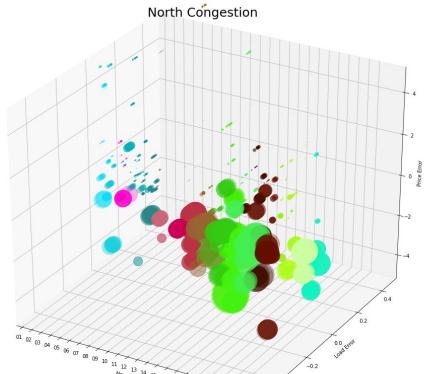






# Unpredictable price, Canada



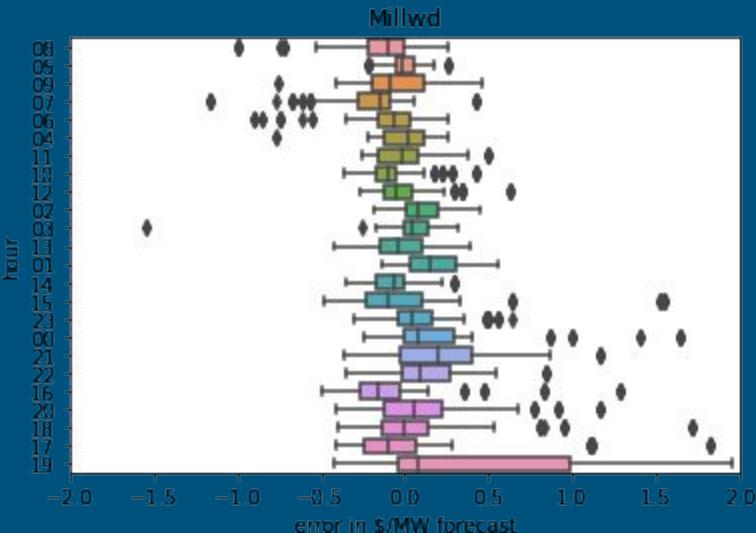
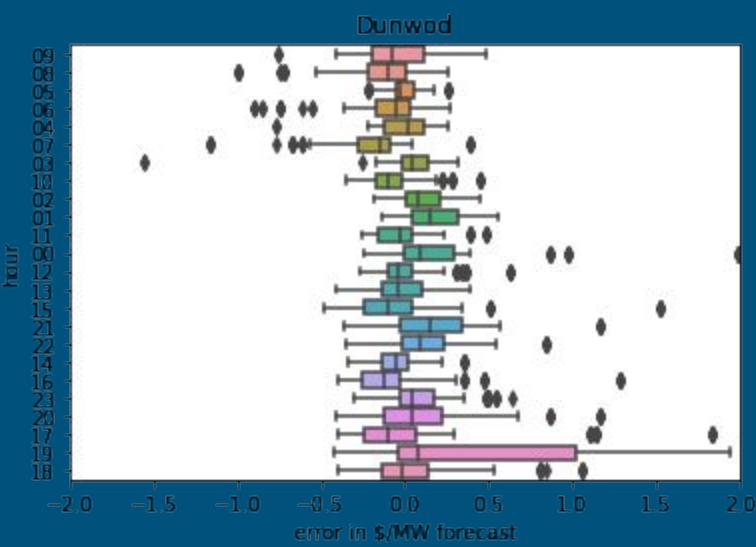


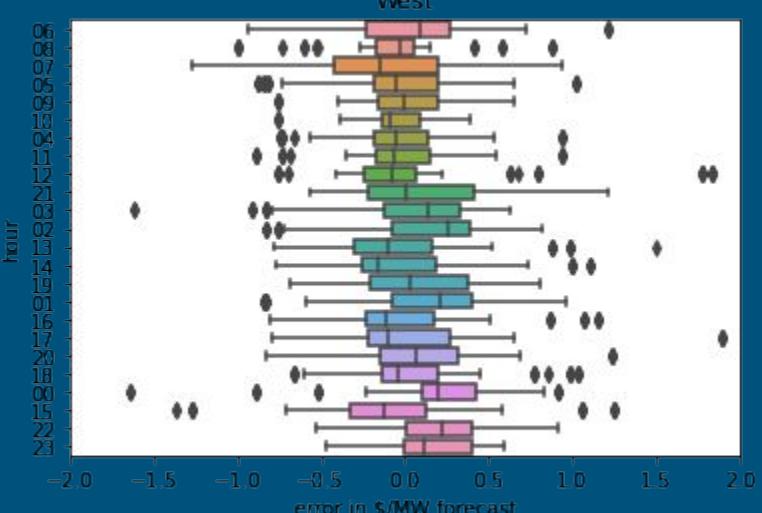
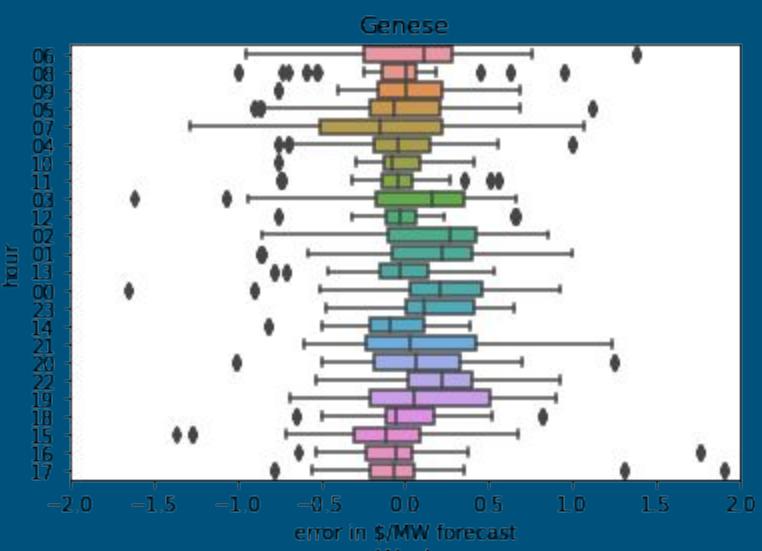
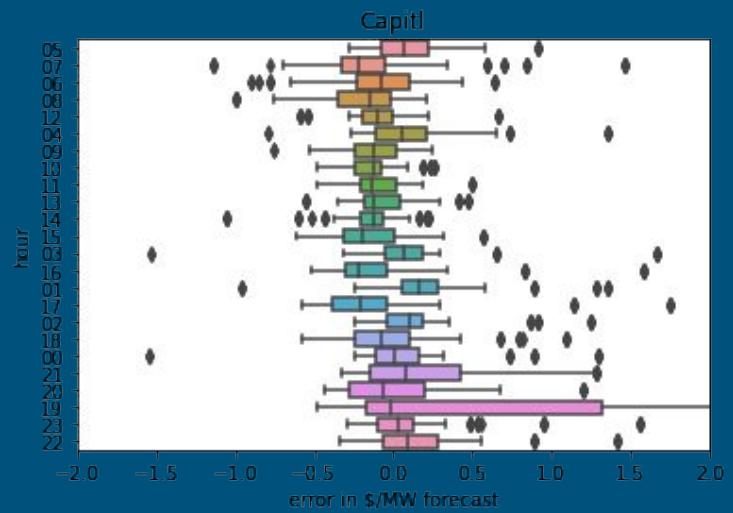
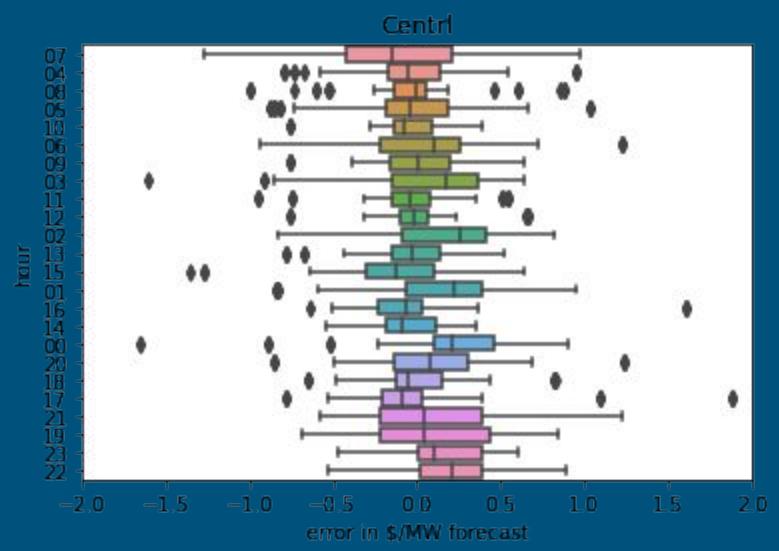
Dunwood, Millwd,  
Longil, N.Y.C. are N/A

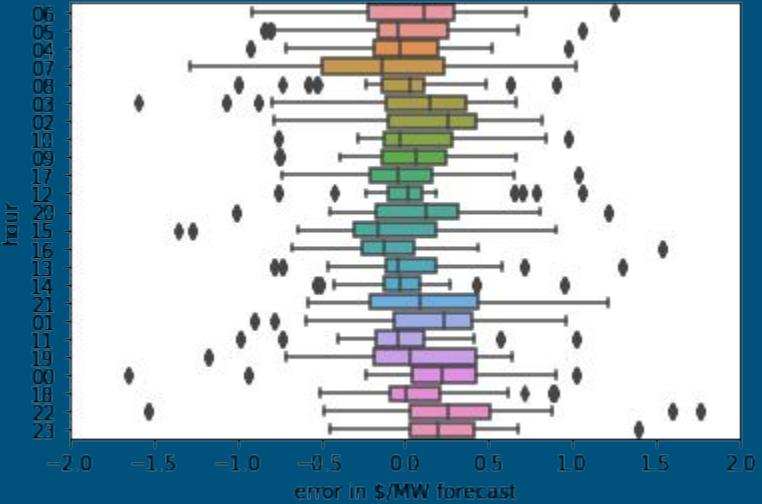
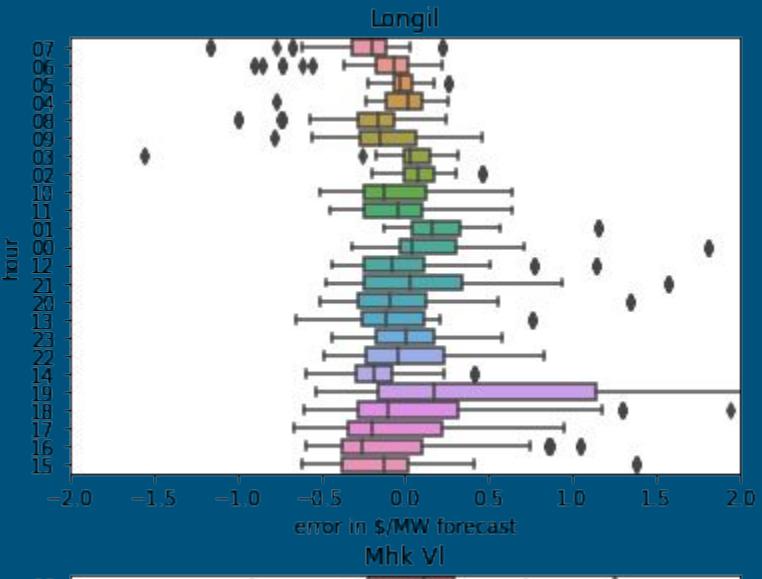
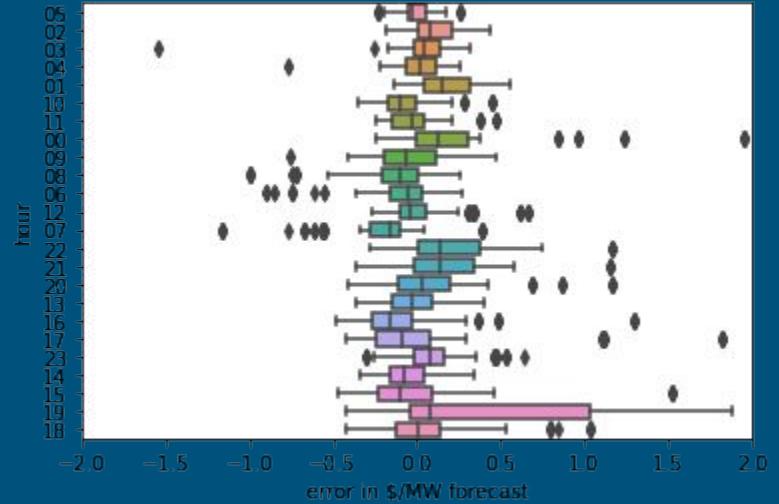
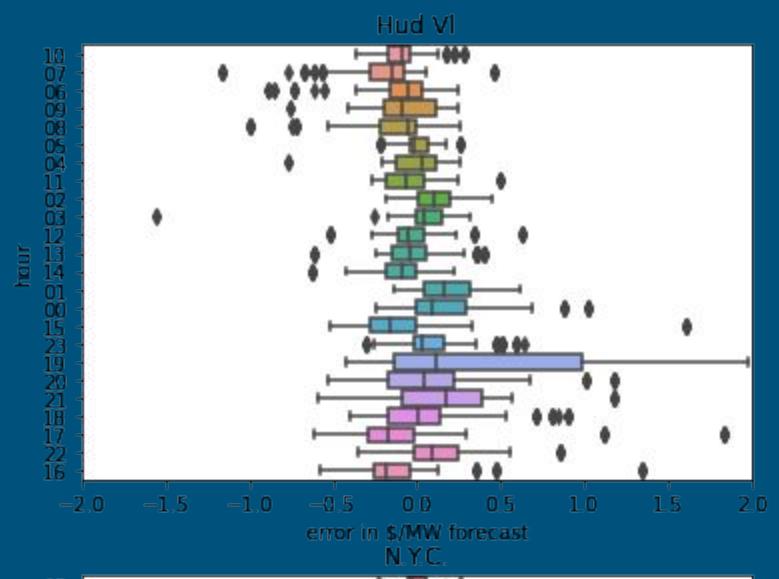
# Box plots

---

- Summer 2020
- Hours are sorted by ascending order of median % error in load forecast







# First Friday of Jan (2017-2021)

---

hour	avgPriceError	avgLoadError	avgTempError	tp corr	tl corr	pl corr	tpF	tlF	plF	tpr	tlr	plr
0	0.011	0.042	4.221	0.600	0.326	0.868	(201.30313988109216, 2.220446049250313e-15)	(23552.246153846158, 1.1102230246251565e-16)	(116.99890109890111, 1.624256285026604e-13)	0.60	0.33	0.87
1	-0.063	0.040	4.987	0.513	0.538	0.641	(1018.1287368150244, 1.1102230246251565e-16)	(54964.81111111109, 1.1102230246251565e-16)	(53.98611111111109, 1.0255907234579809e-09)	0.51	0.54	0.64
2	-0.153	0.030	5.695	0.275	0.511	-0.004	(316.44849294162543, 4.142575171783847e-12)	(98739.46428571429, 1.1102230246251565e-16)	(312.02380952380946, 4.474309811541843e-12)	0.27	0.51	0.00
3	0.328	0.015	5.837	-0.387	0.444	-0.482	(35.90234966533658, 5.086548293675719e-08)	(49861.406211180096, 1.1102230246251565e-16)	(1388.8062111801241, 1.1102230246251565e-16)	-0.39	0.44	-0.48
4	-118.284	-0.007	5.449	0.393	0.509	0.456	(0.0002713742845854594, 1.0)	(53960.71844660195, 1.1102230246251565e-16)	(198842416.21844658, 1.1102230246251565e-16)	0.39	0.51	0.46
6	-0.195	0.009	6.720	0.213	-0.094	0.679	(177.60852355584655, 1.1102230246251565e-16)	(46334.037267080756, 1.1102230246251565e-16)	(260.87732919254665, 1.1102230246251565e-16)	0.21	-0.09	0.68
7	-0.191	0.051	6.375	-0.273	-0.259	0.898	(315.97343586158684, 5.551115123125783e-16)	(67328.6991062562, 1.1102230246251565e-16)	(213.0834160873883, 1.0769163338864018e-14)	-0.27	-0.26	0.90
8	-0.158	0.046	6.496	-0.191	0.158	0.653	(554.2735631731282, 1.5654144647214707e-14)	(36992.11563731933, 1.1102230246251565e-16)	(66.73981603153744, 4.486842675177627e-09)	-0.19	0.16	0.65
9	-0.145	0.055	7.952	0.359	0.458	0.359	(1228.5504553899134, 1.1102230246251565e-16)	(92220.68027210882, 1.1102230246251565e-16)	(75.06462585034012, 2.2539742294824805e-09)	0.36	0.46	0.36
10	0.043	0.056	9.289	0.374	0.413	0.950	(706.1968377707225, 7.08988423525625e-13)	(31290.0278551532, 1.1102230246251565e-16)	(44.30779944289694, 6.127495444374631e-07)	0.37	0.41	0.95
11	0.143	0.068	9.333	0.551	0.412	0.945	(407.4335203366059, 1.038613639536834e-12)	(11470.882527147089, 1.1102230246251565e-16)	(28.153998025666336, 1.8550713741793956e-06)	0.55	0.41	0.94
12	0.092	0.073	9.333	0.401	0.398	0.937	(713.0209248934077, 4.829470157119431e-14)	(9881.82181818181, 1.1102230246251565e-16)	(13.859090909090899, 6.629043236783261e-05)	0.40	0.40	0.94
13	0.133	0.077	8.663	0.200	0.411	0.863	(505.4770033889742, 2.708944180085382e-14)	(14501.024154589371, 1.1102230246251565e-16)	(28.68780193236715, 5.82957312644794e-07)	0.20	0.41	0.86
14	0.055	0.073	8.396	0.328	0.098	0.841	(572.0899043834595, 2.0263790645458357e-12)	(22016.917562724015, 1.1102230246251565e-16)	(38.48506571087216, 1.205550257843413e-06)	0.33	0.10	0.84
15	-0.106	0.030	8.360	0.075	0.046	0.978	(166.26632954066582, 6.430270538260174e-09)	(37739.56521739131, 1.1102230246251565e-16)	(226.9826086956522, 1.6032369876128882e-09)	0.07	0.05	0.98
16	-0.221	0.025	7.273	0.028	0.037	0.899	(181.77843273793133, 6.064695412533183e-10)	(41344.38305709024, 1.1102230246251565e-16)	(227.4438305709025, 1.9958401598074715e-10)	0.03	0.04	0.90
17	0.190	0.060	6.500	-0.235	-0.099	0.765	(84.40179422038742, 5.4848757757852695e-09)	(19687.5, 1.1102230246251565e-16)	(233.25925925925927, 2.194233683638913e-11)	-0.24	-0.10	0.77
18	-0.072	0.066	6.447	0.445	0.150	0.877	(531.309106327788, 2.9297675396833256e-12)	(25060.91294117647, 1.1102230246251565e-16)	(47.168235294117636, 4.5318234320745887e-07)	0.44	0.15	0.88
19	-0.002	0.070	6.123	0.411	-0.352	0.491	(908.1155985839175, 1.27675647831893e-14)	(26185.891666666661, 1.1102230246251565e-16)	(28.835416666666664, 1.6387170546083851e-06)	0.41	-0.35	0.49
20	0.040	0.065	6.500	0.421	0.269	0.435	(457.9545454545453, 5.474509734426647e-13)	(85744.68085106382, 1.1102230246251565e-16)	(187.2340425531915, 7.27853331600556e-11)	0.42	0.27	0.44
21	-0.029	0.079	6.447	0.060	-0.214	0.726	(903.387447755057, 1.1102230246251565e-16)	(50387.79369627508, 1.1102230246251565e-16)	(55.77650429799429, 3.2372006186776048e-09)	0.06	-0.21	0.73
22	0.157	0.075	7.890	-0.305	-0.191	0.983	(525.331112793384, 2.580158309228864e-13)	(44916.660194174765, 1.1102230246251565e-16)	(85.50161812297735, 5.115056378457439e-09)	-0.30	-0.19	0.98

# Power Outages

---

# Power Outages

---

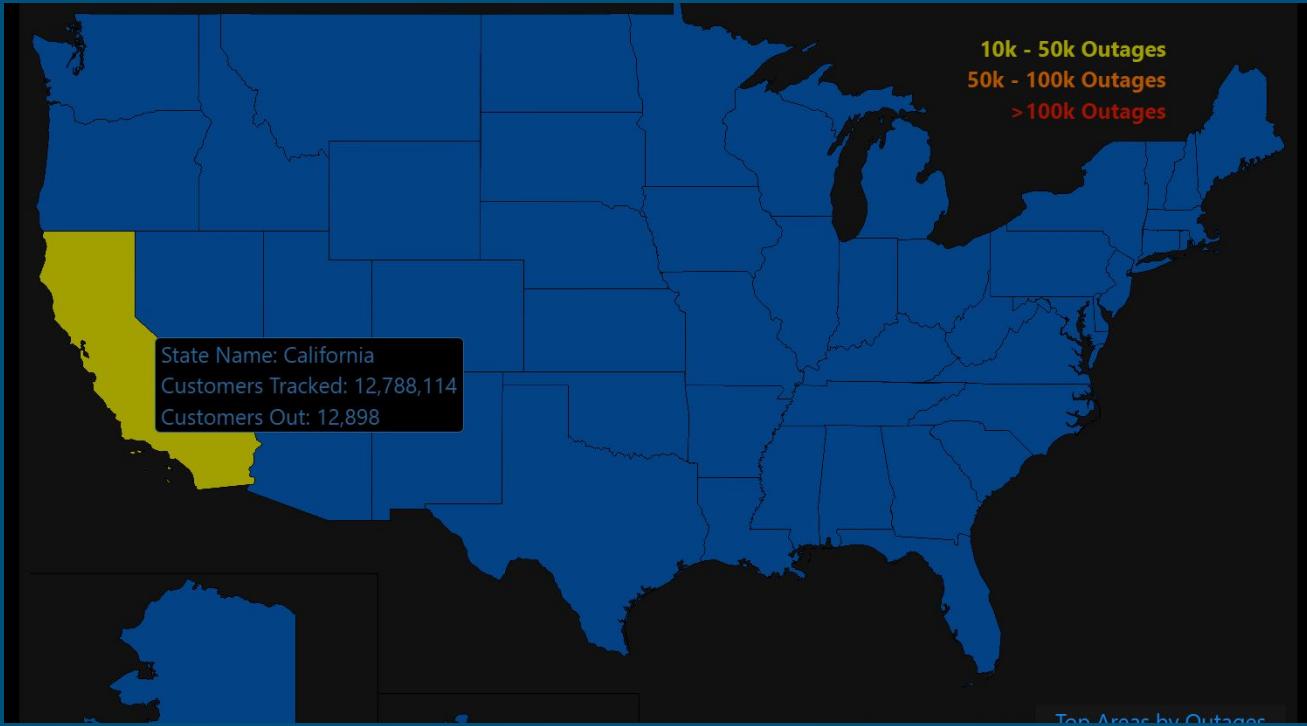
- Demand and supply side
- Planned vs unplanned

POWER_OUTAGES
<i>Start time</i>
<i>End time</i>
<i># Of Customers Affected</i>
<i>Demand Loss (megawatts)</i>
<i>Utility Company</i>
<i>State</i>
<i>County</i>



# PowerOutage.US

---



# PowerOutage.US API

## City Outage Information

JSON URL: [https://poweroutage.us/api/json\\_1\\_4/getcityoutageinfo?key=\[Key\]](https://poweroutage.us/api/json_1_4/getcityoutageinfo?key=[Key])

### Input Parameters

Parameter Name	Data Type	Description	Required
StateId	INT	Returns the Counties for the specified StateId	No
CountyId	INT	Returns the cities for the specified CountyId	No

Returns "CountyOutageInfo" records with the following definition:

Field Name	Data Type	Description
CityId	INT	Unique Identity Id
CountyId	INT	CountyId for reference to the CountyOutageInfo records
StateId	TINYINT	StateId for reference to the StateOutageInfo records
CityName	VARCHAR(100)	The Name of the City
TrackedCount	INT	The number of Customers Tracked
OutageCount	INT	The number of Customers without power
LastUpdatedDateTime	SMALLDATETIME	The Date and Time the Record was last updated. UTC time zone

### Example:

```
[  
  {  
    "CityId":8315,  
    "CountyId":1,  
    "StateId":1,  
    "CityName":"Auburn",  
    "TrackedCount":12484,  
    "OutageCount":0,  
    "LastUpdatedDateTime":"\Date(1536178828960)\\"  
  },  
  {  
    "CityId":572,  
    "CountyId":329,  
    "StateId":34,  
    "CityName":"Unknown",  
    "TrackedCount":60661,  
    "OutageCount":0,  
    "LastUpdatedDateTime":"\Date(1536521613890)\\"  
  }]
```

# PowerOutages.US Costs

---

- \$100/month for access to API
  - Only current outages
- Historical data by quote

# Wind

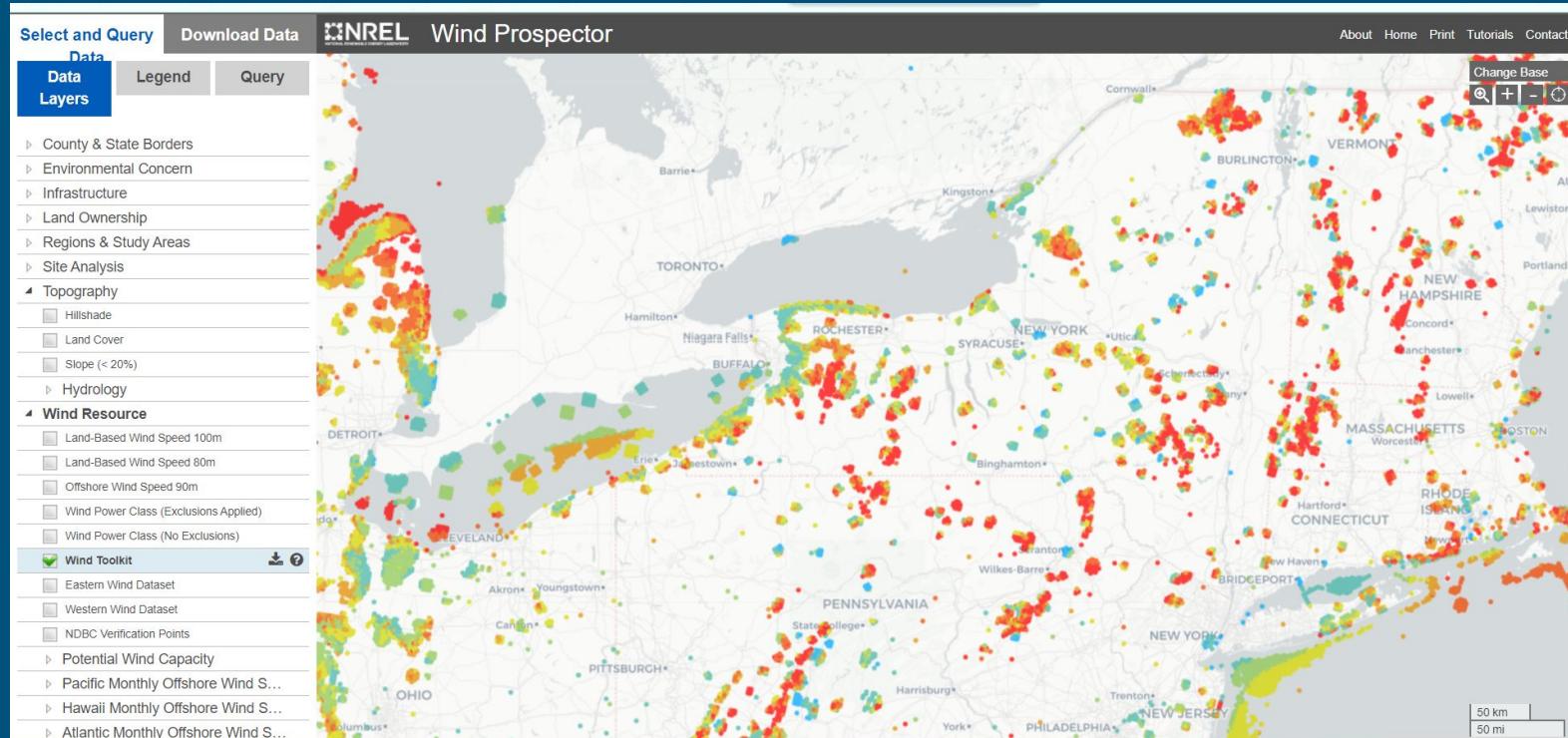
---

# Why Wind?

---

- Wind Turbines
- Dependent on Weather
- We have a good historical dataset

# WIND Toolkit



# WIND Toolkit API

---

GET

```
/api/wind-toolkit/v2/wind/wtk-download  
.csv?api_key={ {API_KEY} } &wkt=POINT (-78  
48.3215  
32.104233) &names=2009&utc=true&leap_da  
y=true&email=user@nrel.gov&reason=exam  
ple&affiliation=NREL
```

# Wind Variables

---

**Table 4**

WIND Toolkit data available via the online API or stored at NREL.

Data	Available through the API	Stored at NREL
Time (UTC)	Yes	Yes
Wind speed (m/s)	100 m	10 m, 40 m, 60 m, 80 m, 100 m, 120 m, 140 m, 160 m, 180 m, 200 m
Wind direction (°)	100 m	10 m, 40 m, 60 m, 80 m, 100 m, 120 m, 140 m, 160 m, 180 m, 200 m
Temperature (K)	2 m	10 m, 40 m, 60 m, 80 m, 100 m, 120 m, 140 m, 160 m, 180 m, 200 m
Relative humidity (%)	—	2 m, 10 m
Barometric pressure (Pa)	Surface	100 m, 200 m
Air density (kg/m <sup>3</sup> )	100 m	
Wind power (MW)	Every 5 min	
Wind power forecasts (MW)	At 1-h, 4-h, 6-h, and 24-h forecast horizons	
Precipitation rate	—	Surface
Solar radiation	—	Shortwave downward radiation, diffuse irradiance, direct normal irradiance, global horizontal shortwave irradiance
Atmospheric stability	—	1/L (Monin–Obukhov length)
Skin temperature	—	Surface
Upward heat flux	—	Surface
Boundary layer height	—	Yes
$u_*$ in similarity theory	—	Yes
Capacity factors	Yes	

1	SiteID	2297733		Site Timez		-5 Data Time		0 Longitude		-74.1201 Latitude		40.64046											
		Year	Month	Day	Hour	Minute	surface	air	air pressur	air pressur	relative hu	surface pr	wind speed	wind direc									
3	2009	1	1	1	0	30	101100	99800	98540	51.24	0	7.69	10.64	11.58	12.25	12.75	13.16	13.49	13.77	14.23	320.52	320.7	
4	2009	1	1	1	1	30	101230	99920	98650	49.2	0	7.33	10.22	11.24	12	12.55	13	13.37	13.69	14.21	317.88	318.3	
5	2009	1	1	1	2	30	101330	100020	98750	47.01	0	7.28	10.25	11.34	12.16	12.77	13.28	13.7	14.06	14.65	315.94	316.6	
6	2009	1	1	1	3	30	101370	100050	98780	48.6	0	8.23	11.32	12.39	13.16	13.73	14.2	14.59	14.93	15.48	316.41	317.0	
7	2009	1	1	1	4	30	101430	100110	98840	49.31	0	7.89	10.78	11.72	12.38	12.86	13.25	13.57	13.83	14.28	315.33	315.6	
8	2009	1	1	1	5	30	101520	100210	98930	48.77	0	5.48	8.02	9	9.73	10.3	10.77	11.16	11.51	12.09	313.78	314.4	
9	2009	1	1	1	6	30	101650	100330	99040	48.9	0	7.72	10.48	11.57	12.34	12.89	13.35	13.72	14.02	14.53	314.39	314.8	
10	2009	1	1	1	7	30	101750	100420	99130	49.74	0	8.84	11.55	12.56	13.21	13.66	14.03	14.31	14.56	14.99	310.81	311.5	
11	2009	1	1	1	8	30	101780	100460	99170	49.67	0	6.86	9.47	10.25	10.81	11.22	11.57	11.87	12.13	12.58	307.99	309.1	
12	2009	1	1	1	9	30	101870	100540	99250	49.03	0	7.05	9.85	10.64	11.21	11.64	11.98	12.28	12.53	12.99	309.65	310.9	
13	2009	1	1	1	10	30	102050	100720	99420	49.16	0	6.29	8.55	9.42	10.03	10.48	10.85	11.15	11.43	11.9	302.02	303.6	
14	2009	1	1	1	11	30	102110	100780	99480	49.94	0	6.23	8.24	8.8	9.13	9.29	9.4	9.46	9.5	9.53	292.55	294.4	
15	2009	1	1	1	12	30	102200	100870	99570	50.06	0	4.27	5.78	6.23	6.53	6.78	6.99	7.19	7.39	7.79	287.56	290.7	
16	2009	1	1	1	13	30	102320	100980	99680	48.49	0	4.35	5.96	6.51	6.87	7.14	7.36	7.54	7.71	8.01	305.7	308.6	
17	2009	1	1	1	14	30	102360	101030	99730	46.94	0	4.92	6.44	6.84	7.1	7.28	7.42	7.54	7.63	7.78	305.43	306.7	
18	2009	1	1	1	15	30	102370	101040	99740	45.91	0	5.14	6.69	7.05	7.28	7.42	7.53	7.61	7.68	7.78	299.3	300.5	
19	2009	1	1	1	16	30	102430	101100	99810	44.53	0	4.67	6.01	6.45	6.72	6.9	7.03	7.14	7.22	7.35	296.14	297.4	
20	2009	1	1	1	17	30	102370	101050	99760	42.1	0	4.94	6.45	6.87	7.14	7.31	7.45	7.55	7.64	7.77	300.81	302.0	
21	2009	1	1	1	18	30	102360	101040	99760	39.53	0	4.06	5.46	5.87	6.15	6.35	6.51	6.63	6.74	6.9	293.74	295.1	
22	2009	1	1	1	19	30	102360	101050	99770	38.6	0	4.27	5.79	6.27	6.56	6.75	6.9	7.01	7.09	7.22	299.33	300.2	
23	2009	1	1	1	20	30	102330	101020	99740	38.31	0	4.31	5.87	6.31	6.58	6.76	6.89	7	7.08	7.22	300.77	301.2	
24	2009	1	1	1	21	30	102360	101050	99770	40.12	0	3.74	5.09	5.5	5.75	5.95	6.11	6.25	6.35	6.52	297.34	297.8	
25	2009	1	1	1	22	30	102370	101060	99790	43.36	0	3.75	5.12	5.6	5.93	6.13	6.3	6.42	6.61	7.07	299.07	298.1	
26	2009	1	1	1	23	30	102380	101060	99790	46.02	0	2.76	4.17	4.84	5.37	5.79	6.15	6.49	6.7	6.9	284.37	284.2	
27	2009	1	1	2	0	30	102350	101030	99760	47.52	0	1.92	3.17	3.85	4.45	4.8	5.11	5.38	5.66	6.02	289.8	290.1	
28	2009	1	2	1	30	102340	101020	99740	45.64	0	2.82	4.16	4.78	5.28	5.65	5.99	6.35	6.71	7.44	298.06	299.4		
29	2009	1	2	2	30	102310	101000	99720	43.51	0	0.75	1.32	1.79	2.35	2.65	2.94	3.23	3.56	4.32	268.82	282.9		
30	2009	1	2	3	30	102250	100940	99660	42.66	0	0.66	1.07	1.31	1.53	1.88	2.25	2.66	3.04	3.88	262.38	274.2		
31	2009	1	2	4	30	102200	100890	99620	43.4	0	0.97	1.25	1.34	1.39	1.46	1.57	1.74	2.04	2.73	192.7	201		
32	2009	1	2	5	30	102120	100810	99540	48.18	0	1.24	1.53	1.59	1.62	1.88	2.15	2.47	2.8	3.45	177.12	185.5		
33	2009	1	2	6	30	102070	100760	99490	53.71	0	1.75	2.27	2.55	2.85	3.06	3.29	3.54	3.81	4.4	183.64	187.9		
34	2009	1	2	7	30	102010	100700	99420	52.21	0	1.52	2.05	2.22	2.51	2.84	3.15	3.87	4.28	5.09	182.80	190		

# Data Wrangling

---

Wind Turbine  
Location

Vincenty Distance

Electrical Bus

Electrical Bus

# Data Wrangling

---

	A	B	C	D	E
12	41.271	-73.953	Westchester	Millwood/Dunwood	87060
13	41.271	-73.953	Westchester	Millwood/Dunwood	87060
14	43.459	-76.532	Oswego	Central	106443
15	40.776	-73.911	Queens	New York City	79582
16	43.521	-76.408	Oswego	Central	106443
17	43.359	-78.605	Niagara	West	102598
18	43.521	-76.41	Oswego	Central	106443
19	41.204	-73.969	Rockland	Hudson Valley	86794
20	41.574	-73.966	Orange	Hudson Valley	90616
21	43.278	-77.31	Wayne	Genesee	103639
22	42.574	-73.859	Albany	Capital	103651
23	40.924	-73.342	Suffolk	Long Island	83099
24	40.592	-74.2	Richmond	New York City	79178
25	40.617	-73.648	Nassau	Long Island	83099
26	42.273	-73.849	Greene	Hudson Valley	99614
27	42.442	-74.448	Schoharie	Capital	100347
28	40.814	-72.94	Suffolk	Long Island	84349
29	40.728	-73.974	New York	New York City	79582
30	40.788	-73.913	Queens	New York City	79582
31	40.747	-73.499	Nassau	Long Island	83099
32	42.625	-73.75	Rensselaer	Capital	103844
33	40.750	-73.945	Q	New York City	79582

# Wind Modelling

---

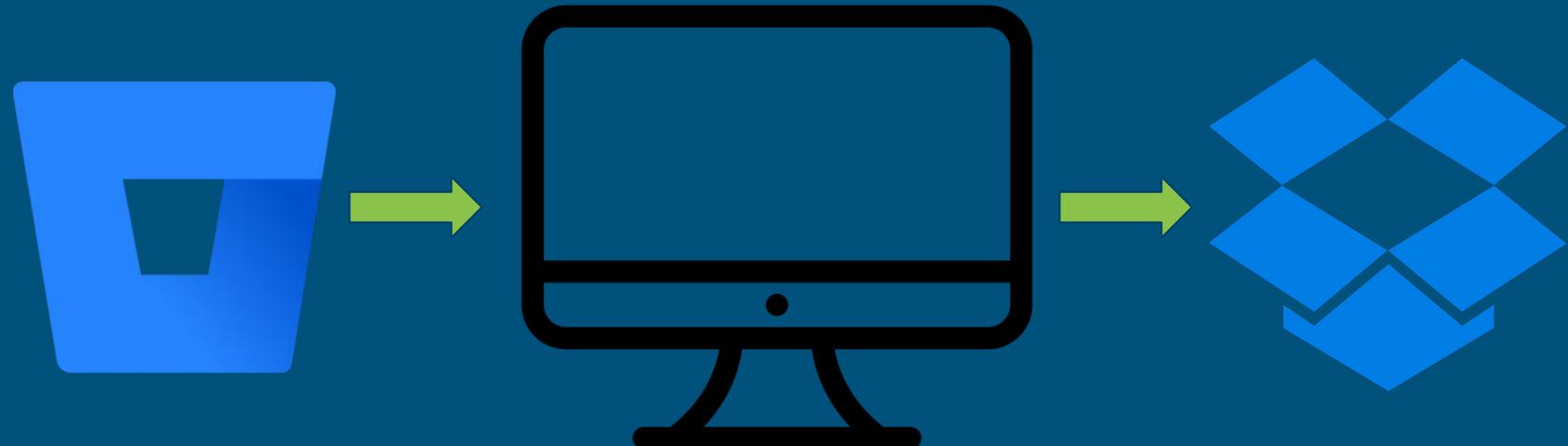
- WIND Toolkit created a very accurate model for predicting variables related to wind
- We may want to use it, or possibly generating our own wind model will be more useful and/or accurate

# Data Storage

---

# Data Storage

---



Bitbucket

HPCC + other Linux  
machines

Dropbox

# Future Steps

---

- Feature Analysis
- Time Series ML Model Research
- Expanding scripts to run on larger dataset (time, location)

# Thank You

---

Prof. Kishore

Prof. Lamadrid

Prof. Urban

Prof. Mikitka

David Morrisette

Jason Robinson

Brian Simboli

Mark Scott

Stuart Hinson

# Questions?

---