DEVELOPING PREDICTIVE MODELS FOR GRANULAR GRID LOCATIONS

Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

2025 AIEC-WLRA LOAD RESEARCH AND DATA ANALYTICS CONFERENCE

SEPTEMBER 22-24, SEATTLE, WA

OVERVIEW

- Study objectives
- Selecting locations for study
- Developing accurate granular predictive models
- Operationalizing forecasts into precise battery instructions
- Mid-study results
- Key takeaways

Pacific Gas & Electric

PG&E Territory

- 70,000 sq. mile service territory
- 26,000 employees
- Services 16+ million customers
- 5.5 million electric service accounts
- 4.5 million natural gas accounts
- 108,000 circuit miles of electric distribution lines
- 18,000 circuit miles of electric transmission lines
- 3,000+ feeders, 800+ substations, 16 distinct planning areas

Enabling Customer Solutions

- 1 in 7 of all electric vehicles in the US is in PG&E's service area
- 1 in 5 of all solar rooftops in the US is in PG&E's service area

STUDY OBJECTIVE AND VALUE DRIVERS

VALUE DRIVERS	OBJECTIVES	State of Project
Distribution infrastructure is sized to meet the local peak demand	Develop a VPP for distribution grid services	PG&E selected highly loaded circuits in our territory, with high potential for load management deferral (programs)
If flexible resources deliver load relief when local demand is nearing capacity limitations they may help avoid or defer growth-related distribution investments	Test ability to send different dispatch instructions to different locations	PG&E contracted 3 large aggregators, who receive our signals to meet needed load shapes
Resources must deliver the right magnitude of resources at the right locations for the right hours to deliver distribution load relief value	Assess ability to shape load based on distribution needs	PG&E ran beta tests from April-June, and operations from July-October

TIMING OF GRID VALUE VARIES BY LOCATION

Proactive Load Shaping for firmness of DERs

- Prove that BTM resources can achieve load shaping needs across D,T,G
- Sunrun, Everbright/Nextera, Tesla and SPAN are using first of its kind tech to orchestrate DERs
- Grid assets selected based on key criterion: challenging load shapes, assets in DACs, Goldilocks overloading
- Grid needs are dynamic, proactive load shaping works to meet changing needs

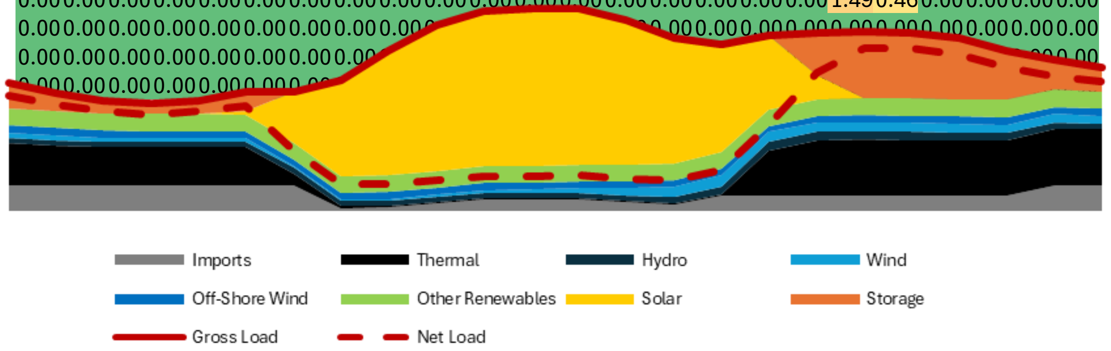
PG&E Grid 2030

16 Feeders for Distribution

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.10	0.67	0.63	0.93	0.80	0.64	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.78	0.30	0.76	0.91	0.81	0.78	0.65	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.39	0.41	0.55	0.78	0.55	0.50	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.20	0.00	0.00	0.00	0.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

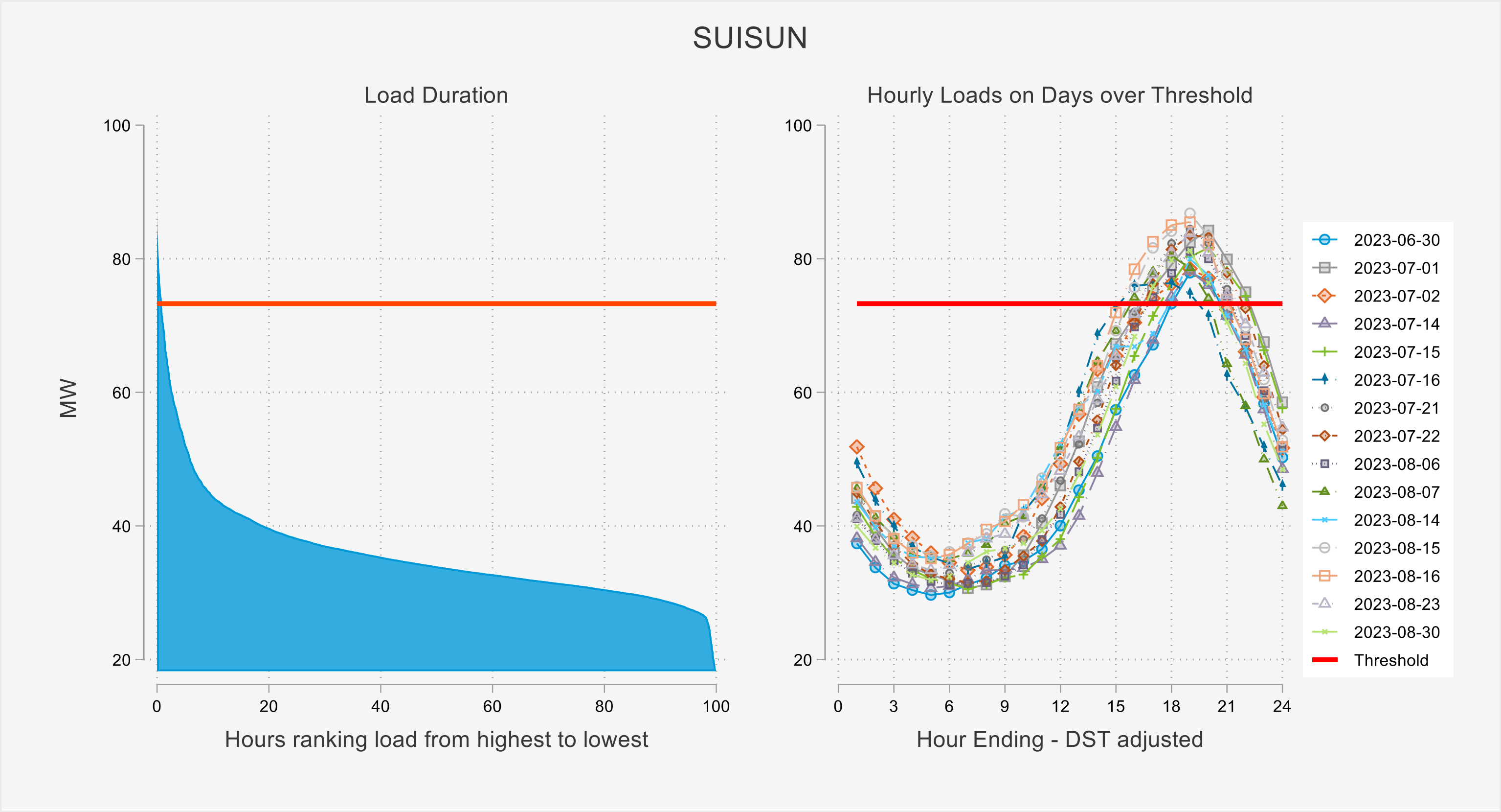
9 Substations

Tx	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.58	3.25	5.53	4.10	2.39	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.78	6.38	7.60	6.11	3.31	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	2.09	1.90	2.95	3.95	1.57	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.49	0.46	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



16 substation clusters simulating 2035 grid

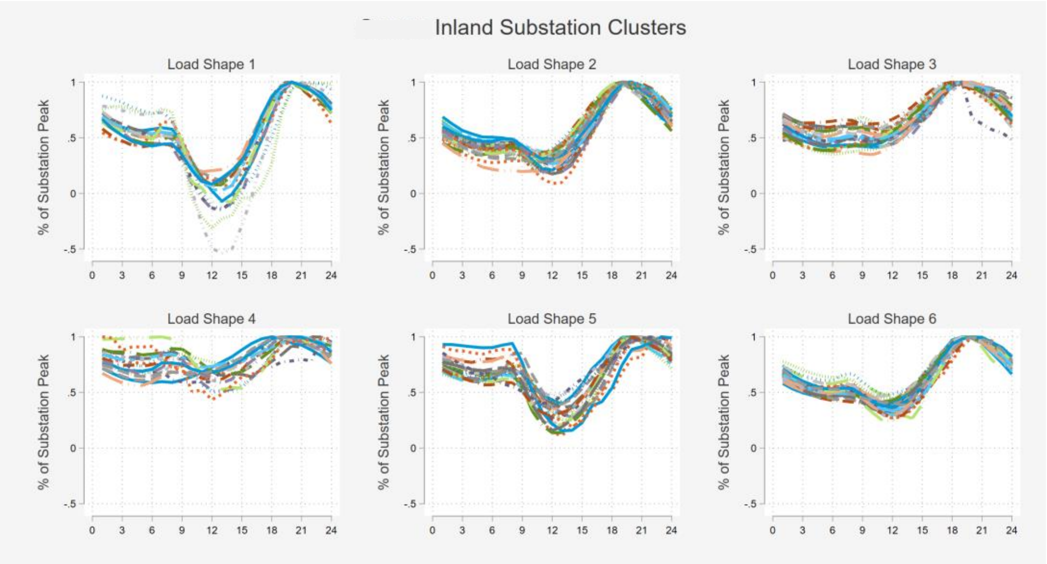
THE MAIN OBJECTIVE IS TO KEEP LOADS BELOW EACH LOCATIONS OPERATING LIMIT (WITH A BUFFER)

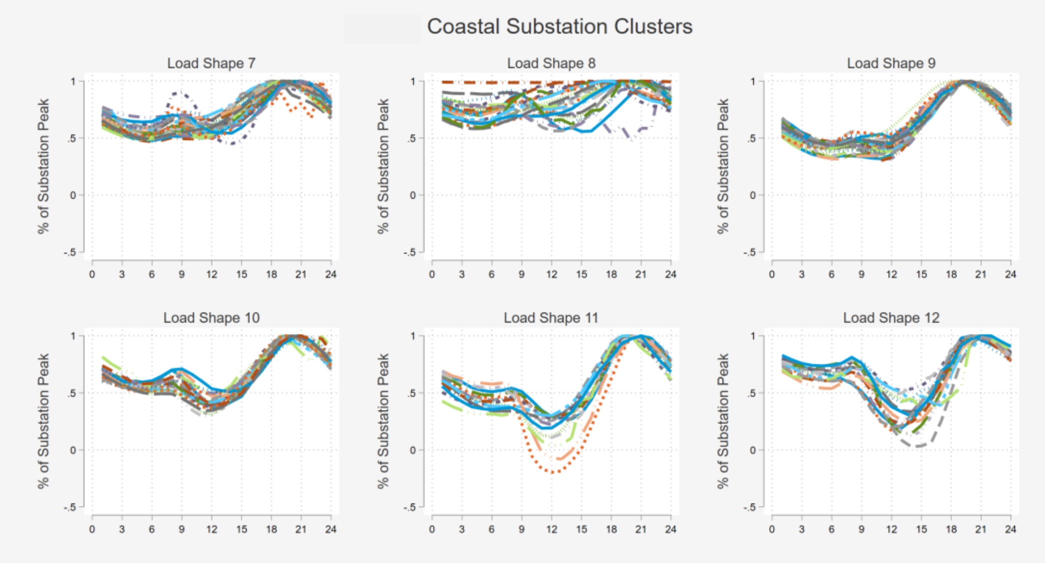


- For the field study, we:
  - ✓ Selected locations that are forecasted to be high loaded by 2030
  - ✓ Need to dispatch resources more often so we can learn (e.g., 75-100 hours)
  - ✓ Need variability in dispatch days and shapes.
  - ✓ Need meaningful resources to assess if a location specific portfolio can perform

SITE SELECTION METHODOLOGY

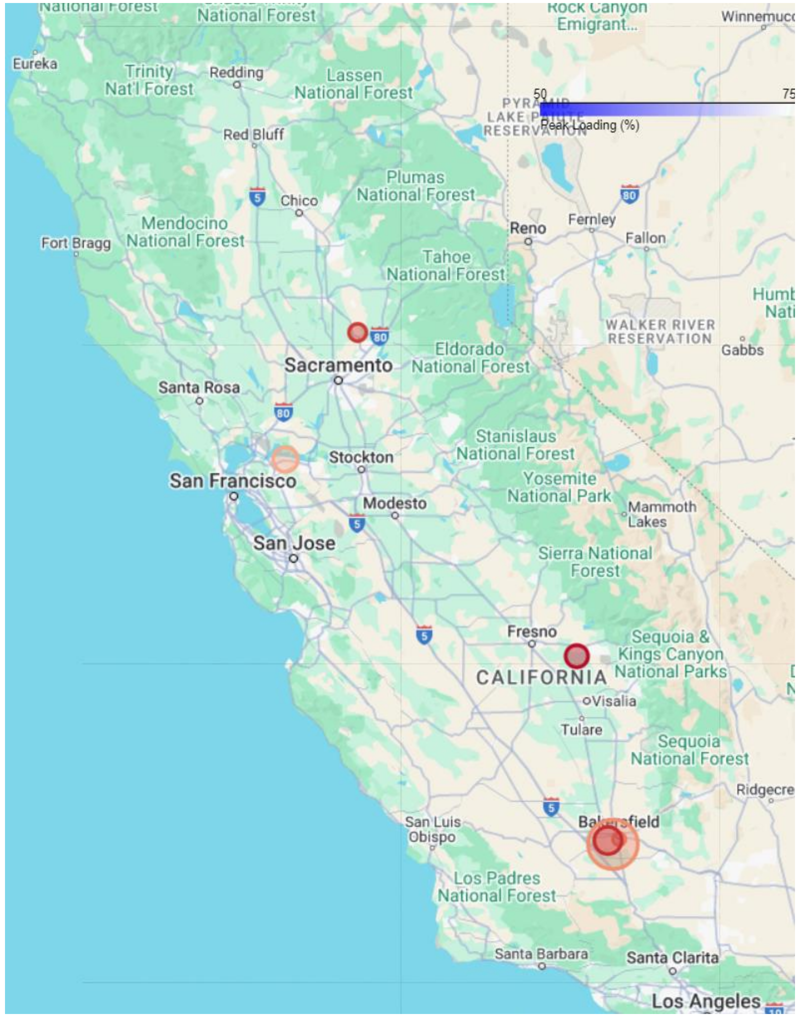
- Process Substation & Feeder Forecast Loading Data
- Merge in Battery Sites
- Retain feeders and substations which contain sizeable battery penetration and clean Pi Tag data
- From site candidates: keep the highly loaded locations (2030) until participation quota is reached
- Finalize site selection with PG&E Team

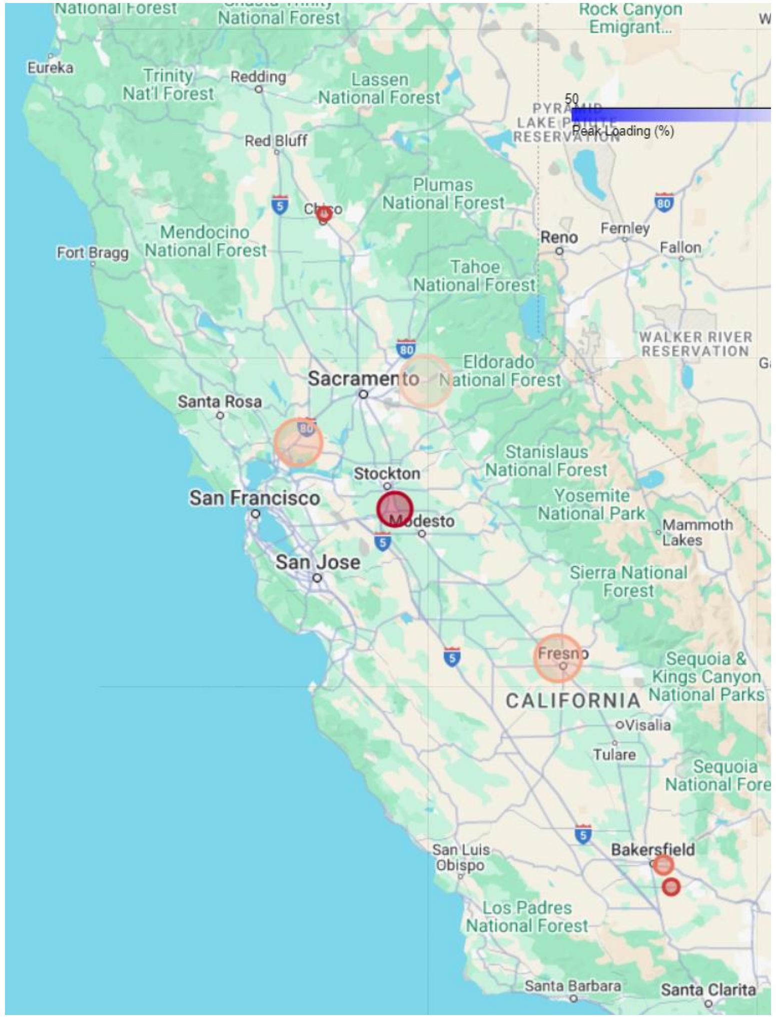




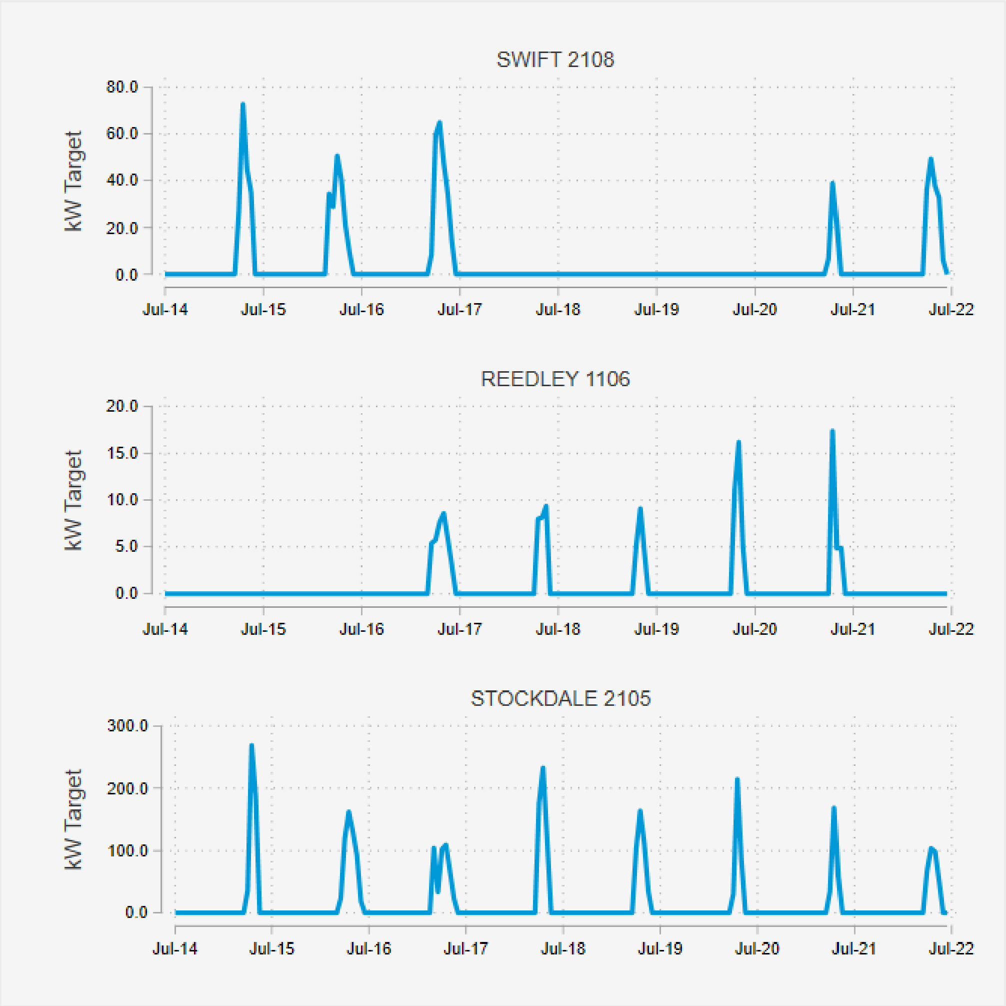
PARTICIPANT MIX

- 1,600+ batteries
- 120 smart panels
- 9 substations
- 18 feeders
- 16 future state substation clusters
  - ✓ Artificially added solar and EV loads projected for 2035 in IPER forecast





EACH LOCATION GETS DIFFERENT INSTRUCTIONS AND TARGET KW IS SCALED BASED ON GRID NEED AND RESOURCES AVAILABLE



HOW WAS THIS ACCOMPLISHED?

Develop accurate predictive models

- Machine learning accuracy tournament
- Each location gets its own mode

↓

Automate inputs needed for forecasts

- Forecasted weather
- CAISO System load forecasts

↓

Convert demand forecasts into dispatch instructions

- Same day target kW by location and hour
- Weekly outlook

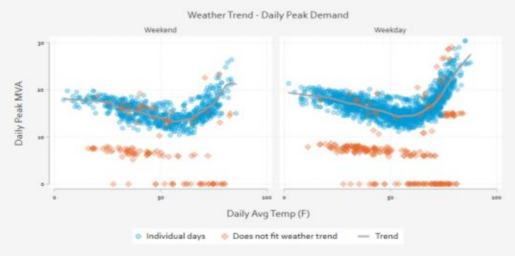
↓

Automate communications and dispatch

DEMAND MODEL DEVELOPMENT FOR EACH LOCATION

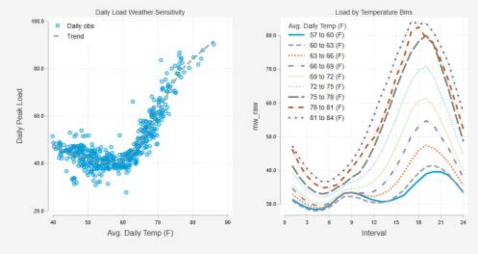
Data Prep Cleaning

- 2022-2024 SCADA data for each feeder and substation
- Merge in local weather a day-ahead CAISO forecasts
- Remove historical DR days
- Identify load transfers
- Remove outages and meter recording issues
- Avoid mixing data issues such as load transfers with load growth/declines



Exploratory Data Analysis

- Explore relationship of all possible predictive variables for DER adoption
  - ✓ Correlation
  - ✓ Plots
  - ✓ Bivariate regressions
- Identify the key predictors of adoption
- Identify non-linear patterns



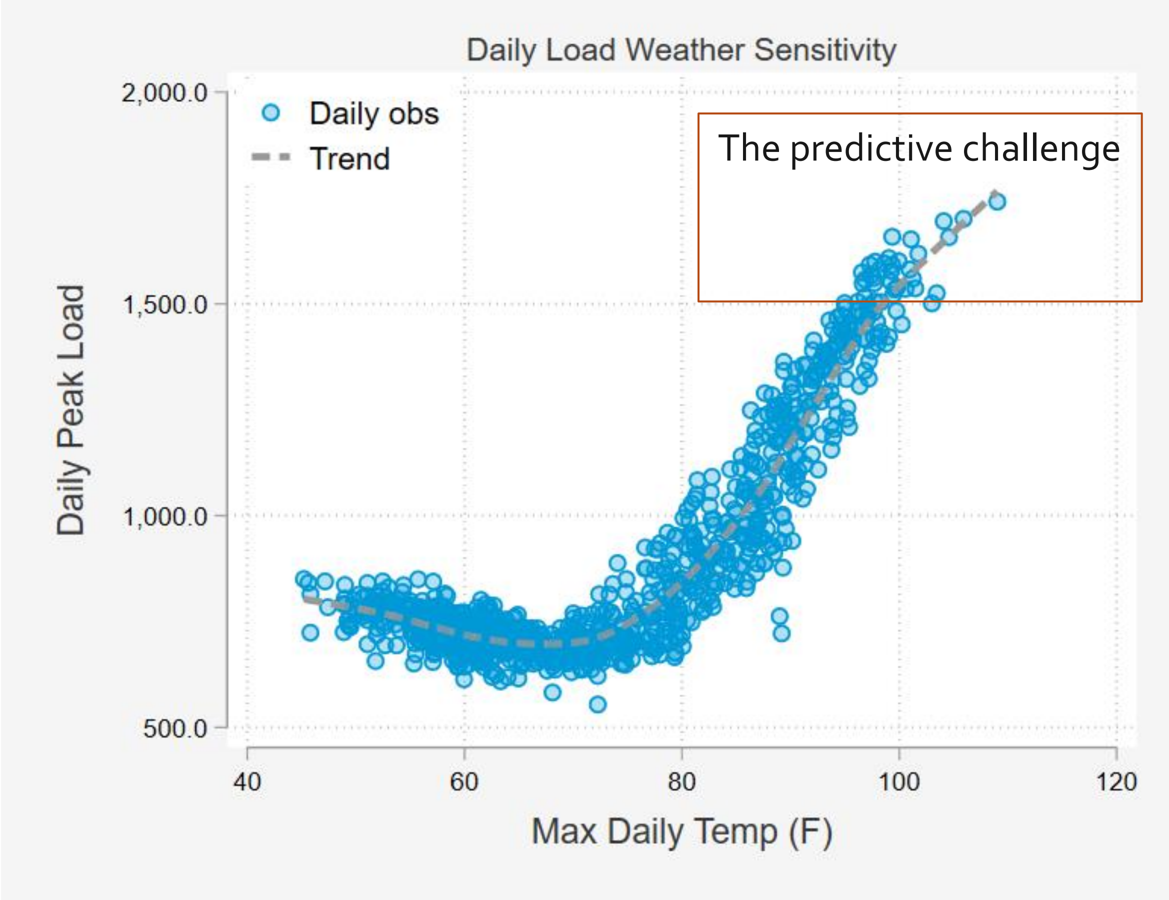
Model Tournament

- MORE DETAIL IN NEXT SLIDE

Apply model

- Run model using all historical data for each location
- Store model for each location
- Produce a function to forecast demand based on weather forecast and other explanatory variables:

MODEL TOURNAMENT



STEP 1 Identify the days when accurate predictions are needed most

STEP 2 Split data into testing and training data

STEP 3 Train models on predictive features

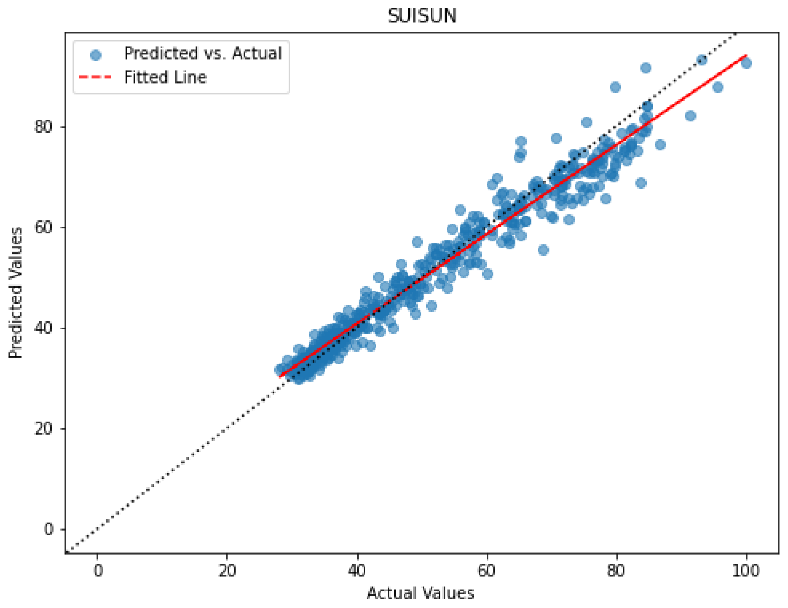
- Spline regression
- Random Forest Regressor
- Gradient Boost Regressor
- Light GBM
- Net Elastic Regressor
- XGBoost

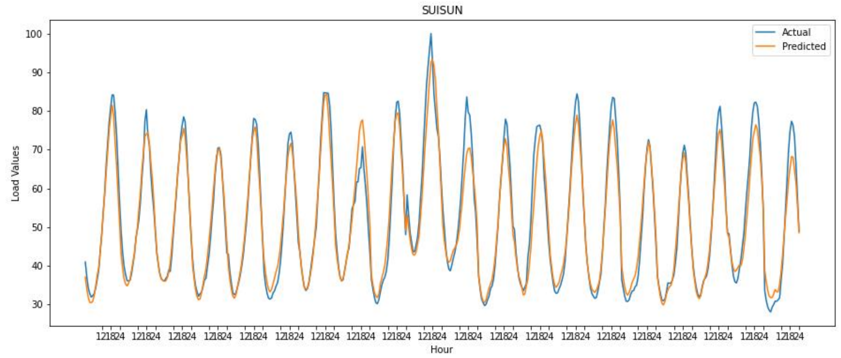
STEP 4 Assess accuracy using testing data by comparing out-of-sample predictions to actuals

STEP 5 Select best unbiased model

PERFORMANCE METRICS USING TESTING DATA (TOP 50 DAYS OVER 3 YEARS)

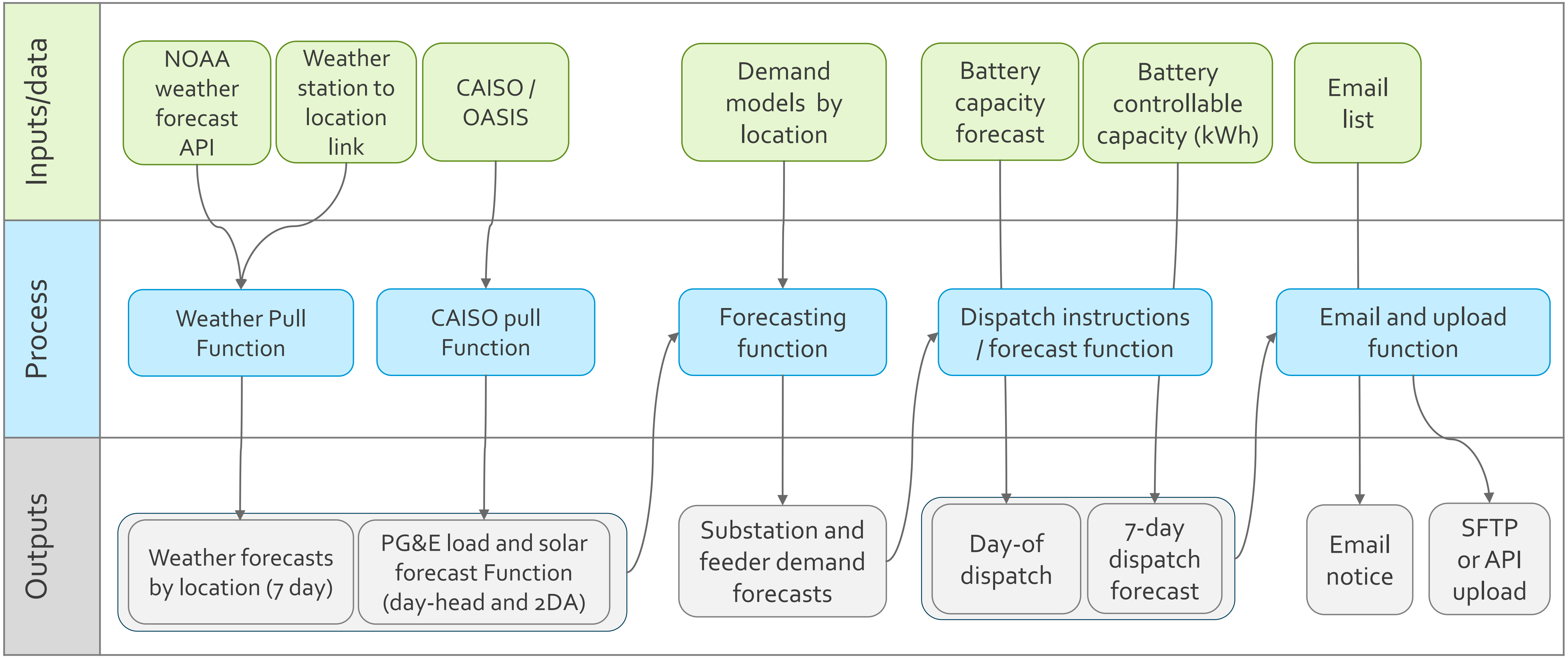
Location	Model	Avg. Actual	Avg. Predicted	RRMSE	RRMSE2	Weighted RMSE	R2	MAE	MAPE	% Bias
ASHLAN AVE	LightGBM	67.44	65.81	5.00	0.07	4.46	0.96	3.57	0.05	-2.41
BAKERSFIELD 2109	ElasticNet	9.74	9.80	0.63	0.06	0.60	0.95	0.50	0.06	0.64
HICKS 1116	RF	3.97	3.90	0.64	0.16	0.57	0.90	0.46	0.19	-1.77
HICKS 2101	RF	4.59	4.35	1.65	0.36	1.37	0.66	1.16	0.50	-5.30
LINCOLN 2108	ElasticNet	4.60	4.53	0.55	0.12	0.54	0.86	0.42	0.10	-1.70
MAGUNDEN	ElasticNet	39.79	39.96	3.30	0.08	3.25	0.91	2.70	0.08	0.42
MANTECA	LightGBM	70.08	69.67	2.29	0.03	1.82	0.98	1.31	0.02	-0.57
MC KEE 1103	LightGBM	3.54	3.42	0.58	0.16	0.50	0.84	0.34	0.10	-3.62
MC KEE 1107	LightGBM	4.66	4.53	0.48	0.10	0.42	0.94	0.34	0.08	-2.79
MORGAN HILL	RF	50.36	50.32	4.39	0.09	3.89	0.91	3.16	0.06	-0.10
MORGAN HILL 2108	ElasticNet	9.58	9.34	0.86	0.09	0.83	0.84	0.66	0.07	-2.43
MORGAN HILL 2109	RF	8.42	8.14	1.09	0.13	0.92	0.84	0.75	0.08	-3.28
MORGAN HILL 2111	ElasticNet	7.01	6.93	0.68	0.10	0.62	0.83	0.48	0.07	-1.08
REEDLEY 1106	ElasticNet	6.11	6.10	0.57	0.09	0.60	0.86	0.39	0.09	-0.27
SARATOGA 1107	ElasticNet	3.64	3.64	0.44	0.12	0.38	0.87	0.31	0.09	0.02
SHINGLE SPRINGS	RF	26.98	26.53	2.27	0.08	2.16	0.96	1.67	0.07	-1.64
SPAN Coastal 1	ElasticNet	1,959.04	1,958.80	83.28	0.04	78.02	0.97	65.04	0.03	-0.01
SPAN Coastal 2	XGBoost	2,176.36	2,126.47	166.41	0.08	161.59	0.54	130.38	0.06	-2.29
SPAN Inland 1	ElasticNet	1,408.71	1,392.41	56.74	0.04	55.16	0.93	44.82	0.03	-1.16
SPAN Inland 2	ElasticNet	2,615.37	2,577.72	159.50	0.06	139.32	0.97	105.71	0.04	-1.44
STELLING	LightGBM	36.09	35.28	3.51	0.10	3.40	0.82	2.60	0.07	-2.25
STOCKDALE 2105	RF	10.61	10.55	1.01	0.10	0.99	0.94	0.63	0.07	-0.54
SUISUN	ElasticNet	52.06	51.26	3.82	0.07	3.58	0.95	2.90	0.06	-1.55
SWIFT 2107	LightGBM	6.76	6.70	0.77	0.11	0.69	0.92	0.56	0.09	-0.88
SWIFT 2108	LightGBM	7.03	6.70	1.27	0.18	0.94	0.83	0.72	0.11	-4.75
SYCAMORE CREEK	RF	33.65	33.16	3.78	0.11	3.36	0.91	2.66	0.08	-1.44
Sunrun Coastal 1	ElasticNet	1,083.11	1,130.34	203.78	0.19	200.78	0.65	160.71	0.16	4.36
Sunrun Coastal 2	XGBoost	1,291.57	1,278.61	55.16	0.04	54.41	0.79	42.31	0.03	-1.00
Sunrun Coastal 3	XGBoost	985.37	979.03	42.43	0.04	39.78	0.99	31.86	0.04	-0.64
Sunrun Coastal 4	ElasticNet	462.41	447.70	44.23	0.10	39.20	0.91	29.36	0.06	-3.18
Sunrun Coastal 5	ElasticNet	124.36	125.47	8.91	0.07	8.35	0.99	6.72	0.24	0.89
Sunrun Coastal 6	ElasticNet	133.09	131.09	6.63	0.05	6.45	0.98	5.50	0.05	-1.50
Sunrun Inland 1	ElasticNet	184.22	181.49	19.72	0.11	19.22	0.98	15.76	0.37	-1.48
Sunrun Inland 2	XGBoost	656.02	654.25	41.05	0.06	37.64	0.99	31.19	0.07	-0.27
Sunrun Inland 3	ElasticNet	921.81	924.69	38.98	0.04	37.93	0.98	32.35	0.04	0.31
Sunrun Inland 4	ElasticNet	585.46	570.93	30.57	0.05	28.08	0.90	23.35	0.04	-2.48
Sunrun Inland 5	ElasticNet	353.11	342.55	19.36	0.05	18.89	0.98	15.64	0.05	-2.99
Sunrun Inland 6	ElasticNet	889.81	884.82	28.93	0.03	28.49	0.99	23.12	0.03	-0.56
TEVIS 2104	ElasticNet	6.24	6.38	0.89	0.14	0.89	0.96	0.66	1.15	2.24
TEVIS 2105	ElasticNet	11.03	11.00	0.73	0.07	0.69	0.96	0.52	0.05	-0.26
WEEDPATCH	ElasticNet	17.27	16.90	1.94	0.11	1.83	0.79	1.55	0.09	-2.13
WILLOW PASS 2108	LightGBM	6.34	6.09	1.07	0.17	0.95	0.84	0.78	0.13	-3.95





Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

PROCESS FOR FORECASTING AND DISPATCH INSTRUCTIONS



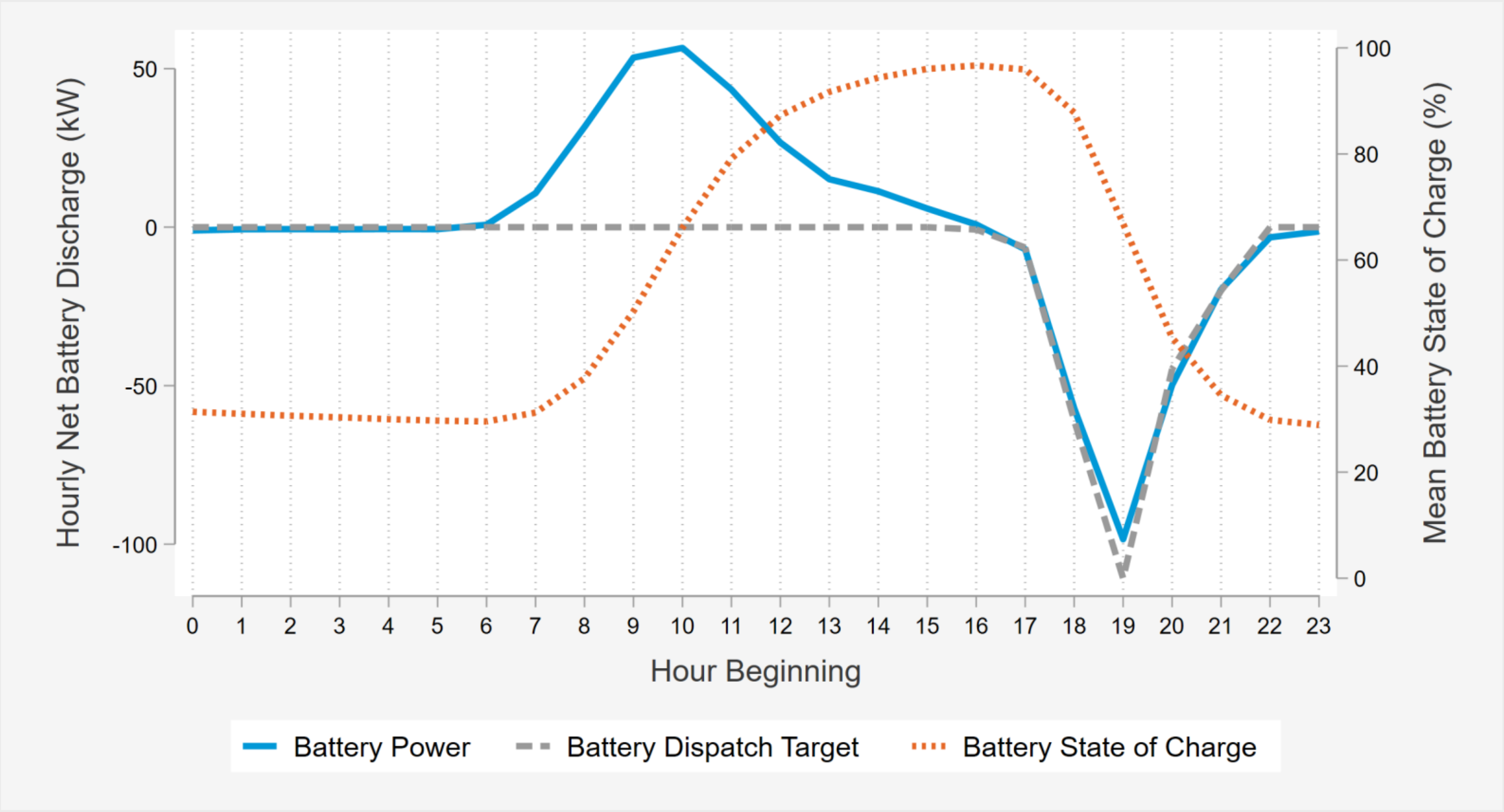
KEY QUESTIONS AND METRICS

1 Did the resource dispatch on the hours and locations requested?

3 Did the resource tend to over or under-deliver the target kW?

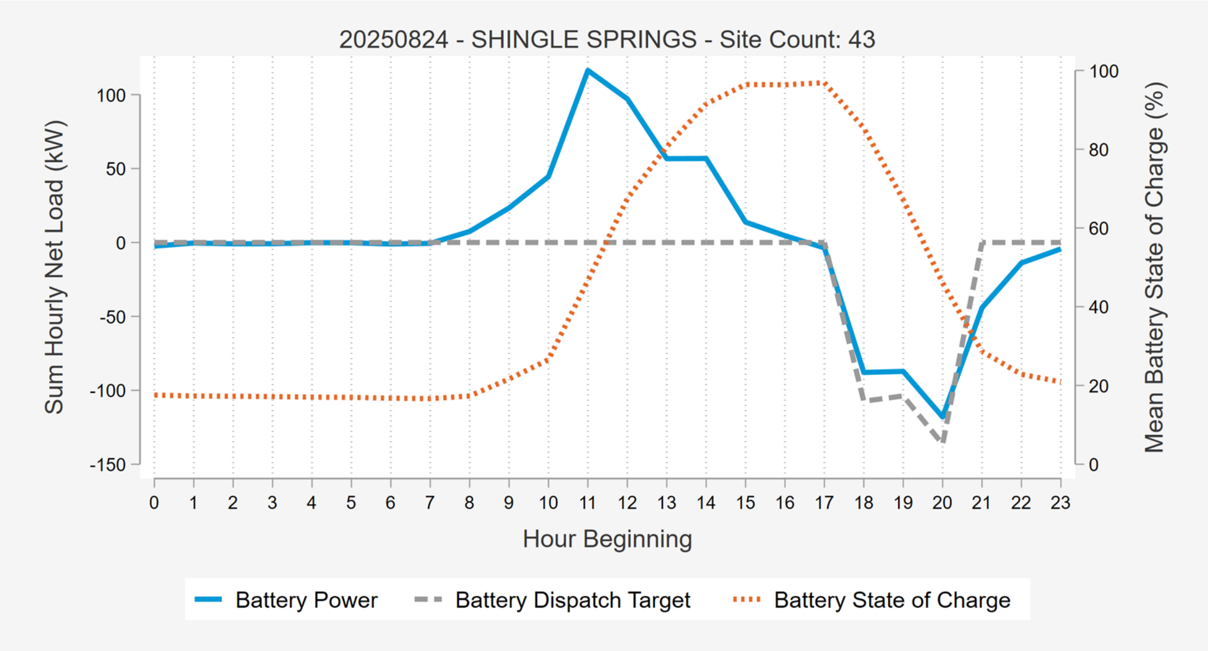
2 How closely did the resource follow the target kW instructions?

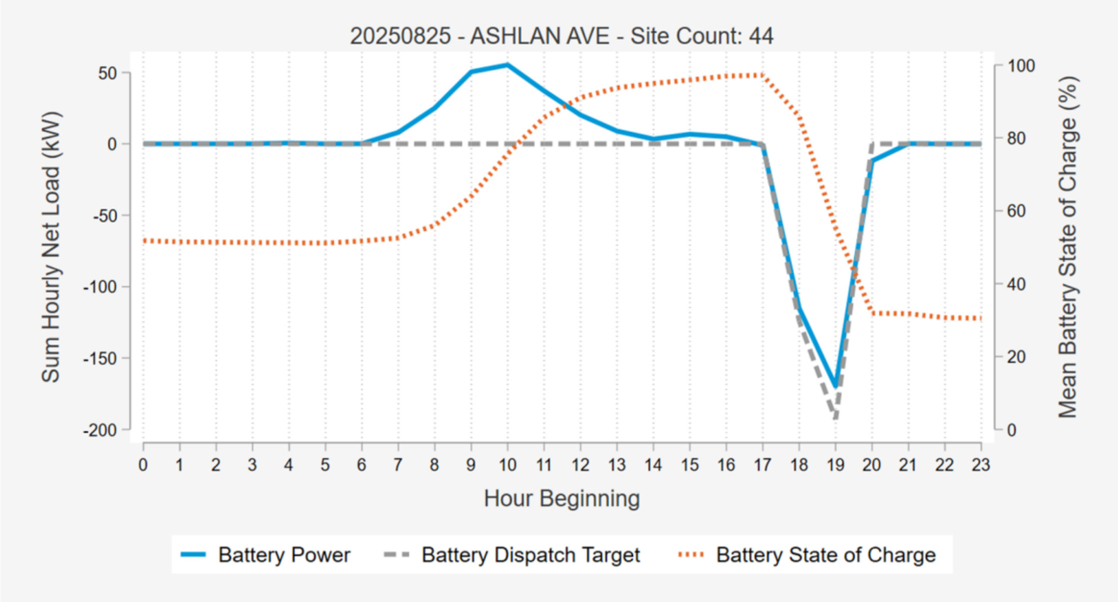
DID THE RESOURCE DISPATCH ON THE HOURS AND LOCATIONS REQUESTED?



- Results show the weighted average battery dispatch target and power readings at the group level across all event days.
- Battery groups are able to closely follow dispatch instructions in the hours and locations requested.
- On a typical event day, batteries are close to 100% charge and will draw down to 20-30%.

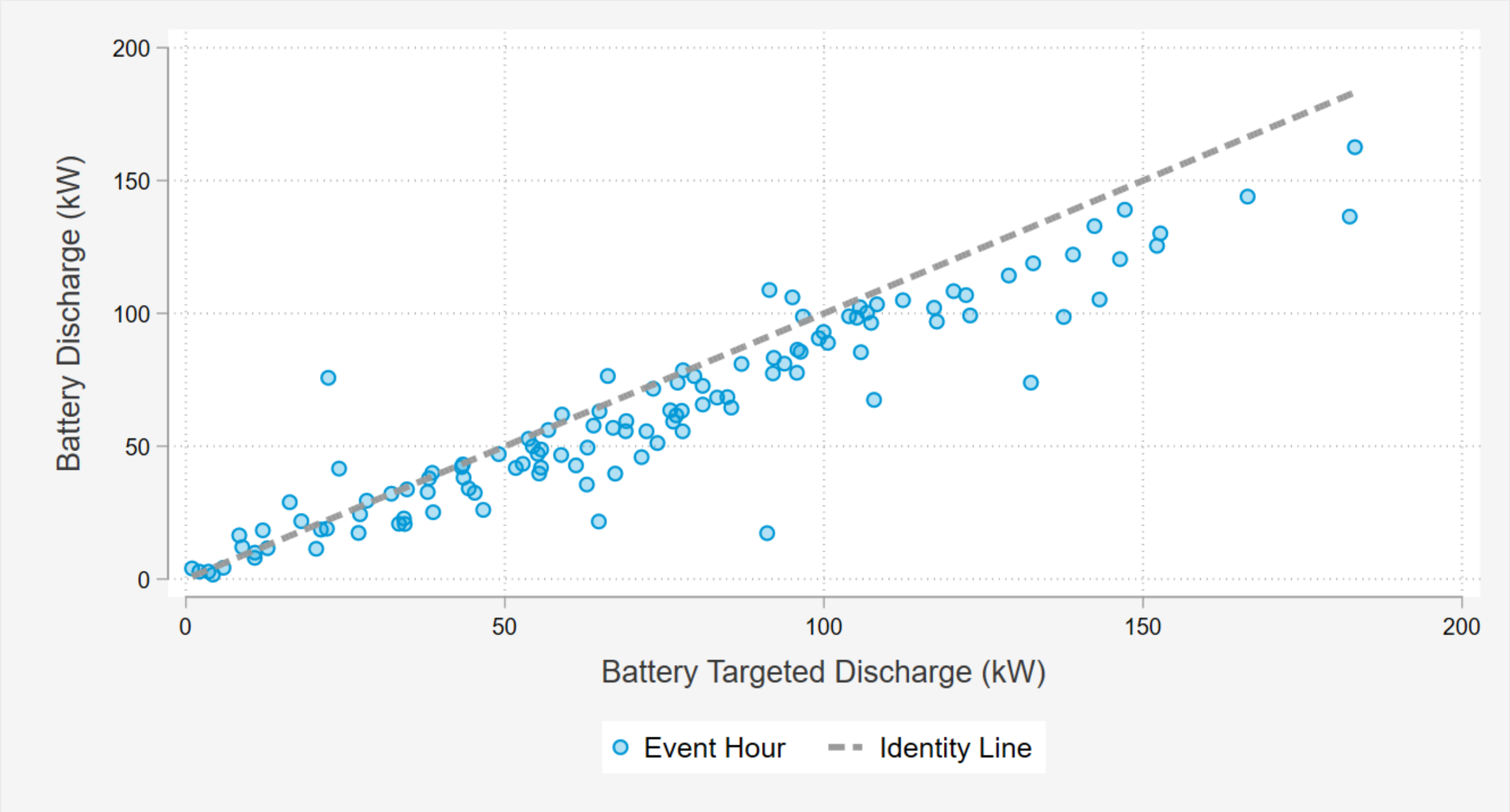
TWO ILLUSTRATE LOCATIONS, TWO DIFFERENT PROVIDERS





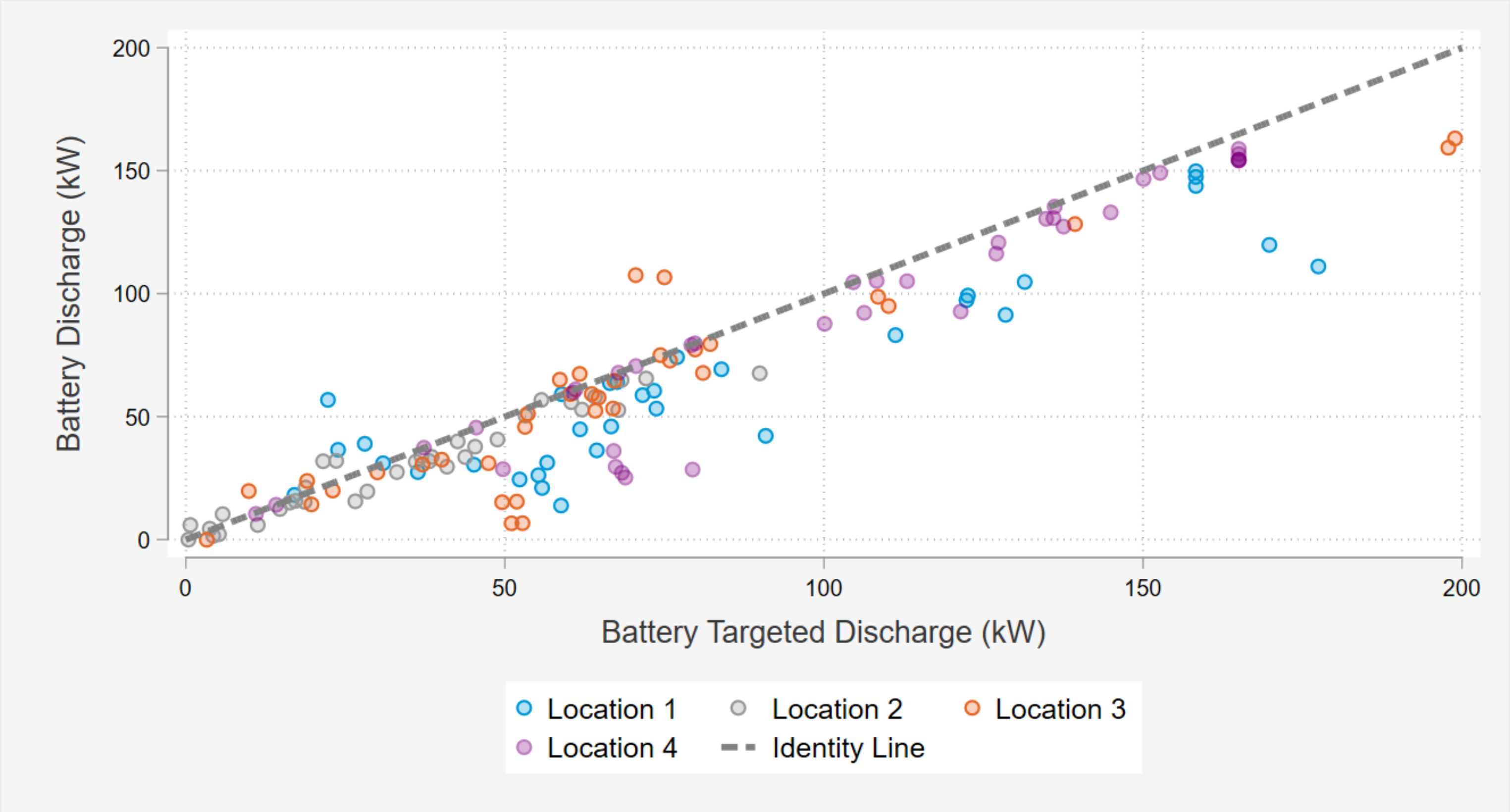
Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

HOW CLOSELY DID THE RESOURCE FOLLOW THE TARGET KW INSTRUCTIONS?



- Results show the weighted average battery dispatch target versus power readings at the group level across the all event days and hours.
- Battery Groups typically get close to the dispatch target but do not fully reach it.
- On a typical event day, batteries are close to 100% charge and will draw down to 20-30%.

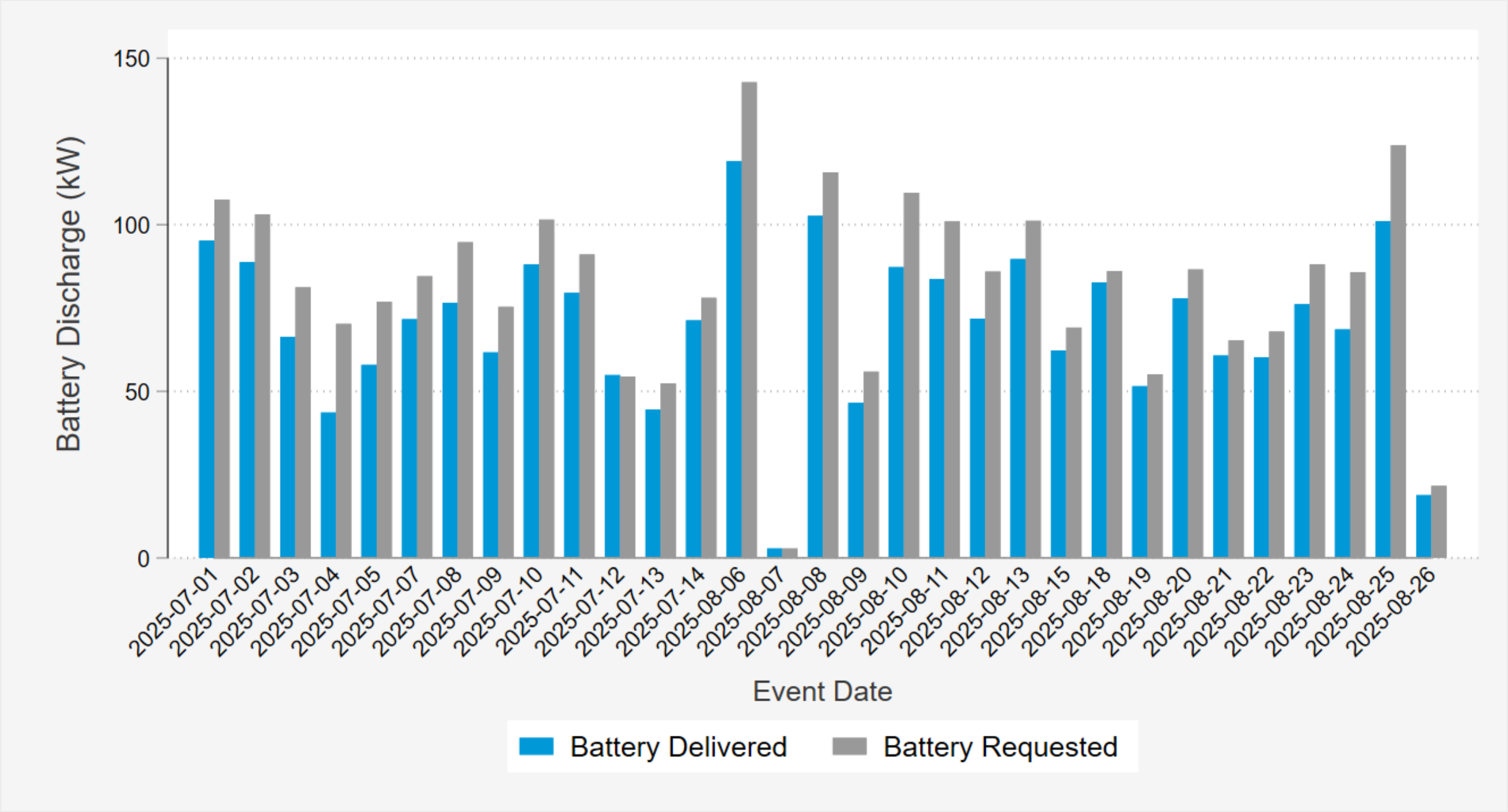
HOW CLOSELY DID THE RESOURCE FOLLOW THE TARGET KW INSTRUCTIONS AT SPECIFIC LOCATIONS?



- Results show the battery dispatch target versus power readings at the group level for 4 different group locations for all event hours.
- At each location, the groups generally followed the instructions
- Smaller battery groups had more variability

Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

DID THE RESOURCE TEND TO OVER OR UNDER-DELIVER THE TARGET KW?



- Results show the weighted average battery dispatch target and power readings at the group level by event date.
- No apparent trend in power delivered versus requested over the summer.
- The mix of locations dispatched on any given event date can have a big influence on resources requested.

KEY TAKEAWAYS

1. Accurate predictive models are critical to integration of battery storage into grid services
2. The new methods beat out the older predictive methods, but only slightly
3. Scaling and automating the granular forecasts requires understanding how battery providers ingest data via APIs
4. The timing and magnitude of load relief needed for distribution grid VPP varies by location
5. The batteries were able to deliver the requested kW at each location with high precision
6. The ability to deliver specific amounts of resources was tested with multiple vendors and battery systems

QUESTIONS?

Josh Bode
Partner
Demand Side Analytics
jbode@demandsideanalytics.com
415.786.0707

Trevor Udwin
Pacific Gas and Electric

APPENDIX

PROCESS FOR DEFINING LOAD LIMITING AT SITE

Identify the top 10 load hours for two years (20 total, location specific)

Calculate distribution peak load (DPLC) contribution for each SPAN premise

Provide DPLC to SPAN

SPAN determines the attainable load limit for each site

Payment is based on difference between the DPLC and observed loads on event days