April 29, 2026

End-Use Load Research Home Energy Metering Study: Data Standardization and Peak Load Analysis

Prepared For NEEA:
David Clement, Sr. Program Manager
Aaron James, Survey Data Analyst
Ben Spearing, Survey Program Manager

Prepared By:
Ethan Goldman, Principal
Resilient Edge
5 Pavilion Ave
South Burlington, VT 05403

and

AJ Howard, Principal
Linden Clean Energy
24 Brookside Ln
Portland, ME 04103

Northwest Energy Efficiency Alliance
PHONE
503-688-5400
EMAIL
info@neea.org

Table of Contents

Glossary of Acronyms

Acronym	Definition
BPA	Bonneville Power Authority
COP	Coefficient of Performance
CSV	Comma Separated Values
CVRMSE	Coefficient of Variation of the Root Mean Squared Error
ELCAP	End-Use Load and Consumer Assessment Program
ERV	Energy Recovery Ventilation
EV	Electric Vehicle
EVSE	Electric Vehicle Supply Equipment
HEMS	Home Energy Metering Study
HPWH	Heat Pump Water Heaters
HVAC	Heating Ventilation and Air Conditioning
HZ	Heating Zone
kW	Kilowatts
MST	Mountain Standard Time
MW	Megawatts
NEEA	Northwest Energy Efficiency Alliance
PPT	Pacific Prevailing Time
PST	Pacific Standard Time
RBSA	Residential Building Stock Assessment
TMYx	Typical Meteorological Year
USAF	United States Air Force
WMO	Meteorological Organization

Home Energy Metering Study Data Standardization and Electrification Peak Load Analysis

Section 1: Electrification Peak Load Analysis

1 Section 1 Introduction

1.1 Context

The rise of electrification and variable renewable generation will create challenges to meet peak load in the Northwest Region, which has not traditionally had significant capacity constraints given its relatively flat load and abundant hydropower resources. To address these challenges, utilities in the Region are investing in demand response and other low-cost solutions like strategic energy efficiency to help meet future peak loads. However, in many cases utilities lack the data needed to optimize planning assumptions around increased peak period loads in the future.

The Northwest Energy Efficiency Alliance (NEEA) Home Energy Metering Study (HEMS) offers the first large-scale, high-resolution metering Data for the Region since the End-Use Load and Consumer Assessment Program (ELCAP) study was completed in the 1980s. The HEMS dataset includes 7,690 points (metered circuits) across 418 residential single-family homes from the Northwest. The HEMS Dataset provides a great opportunity to analyze real field meter data and better understand the range of possible impacts of future residential electrification, particularly in terms of peak period loads.

1.2 Goals

This project aimed to demonstrate the usefulness of the HEMS data by creating planning scenarios around increases in peak load due to increased electrification in the residential sector. The project also strove to increase the usability of the HEMS Dataset for future researchers by performing additional data cleaning, restructuring, normalization and aggregation.

The project utilized the HEMS dataset and statistical methods to model ways that electrification may impact future residential loads in the region. Data analysis focused on heat pumps and heat pump water heaters (HPWHs), two rapidly growing, weather-dependent, electrified loads that can be identified in the HEMS data. The analysis developed load shapes for various replacement and new construction scenarios and considers what would happen as a higher percentage of homes adopt these devices using low, medium and high adoption forecasts developed by NEEA.

The project also increases access and usability of the HEMS Dataset by creating a cleaned dataset, a tool to provide easy access to aggregated data, and dynamic reporting tools for the HEMS team and funders. The Project also increases the transparency, reproducibility and updatability of data processing by using open-source tools and documented procedures for cleaning and summarizing data.

1.3 Research Objectives

The project aimed to achieve the following research objectives.

1. Increase data access and lower barriers to future use of HEMS data by making cleaned data and methods available.
2. Develop a weather normalized dataset and a standardized set of pre-aggregated data tables for use by other analysts and for reporting.
3. Perform data analysis to understand the range of possible impacts regarding how we can expect grid peak loads to change due to heat pumps and HPWH.
4. Provide static and dynamic reporting of the results that show key findings and methodology, as well as the ability to view/filter and download pre-aggregated data for further analysis.

1.4 Summary of Results

This project has successfully cleaned, organized, and aggregated the HEMS data to lower the barriers for future analysts. The project also normalized 88% of points based on weather and schedule using an open-sourced python library and methodology. It produced estimates of system-level impacts from increased heat pump and heat pump water heater adoption in the Northwest. These results showed that heat pumps and heat pump water heaters in the residential sector could lead to a modest increase in summer peak load and a modest decrease in winter morning peak load. We also estimated the potential impact of future shifts in temperature on weather-dependent electrical loads. Weather-sensitive loads like heat pumps (both for space conditioning and water heating) showed the strongest responses to colder temperatures in winter mornings and hotter temperatures in summer afternoons. Spring and fall load shapes showed more modest increases under both hotter and colder scenarios. Lastly, the project successfully developed a set of dynamic reporting dashboards for NEEA that were used to share findings and methodologies at every step in the project.

1.5 Structure of Report

This report includes three sections as follows

1. Electrification Peak Load Analysis – Provides an overview of the project and key findings with a focus on the electrification peak load analysis
2. Data Cleaning and Methodology – Provides a detailed review of data cleaning methodology and findings, and detailed methodology for the analysis of peak period loads.
3. Appendices – additional detail on inputs, outputs and findings from the analysis.

2 Statistical Analysis of Peak Load Scenarios

The team estimated the potential impact of two different types of peak period scenarios. First, we simulated the potential impact on the regional grid from electrification of residential space-conditioning and domestic hot water production. This analysis included the following steps:

1. Identifying the relevant peak periods in the region,
2. Weather normalizing the weather-dependent loads
3. Defining and estimating per-unit impacts of different technology adoption scenarios (e.g., ductless heat pump replacing resistive heating), and
4. Defining and estimating system-level impacts of high, medium and low adoption forecast scenarios.

Second, we used the OpenDSM EEmeter weather-regression models to estimate the changes in end-use load profiles predicted by hypothetical systematic changes in ambient dry-bulb temperature relative to the Typical Meteorological Year (TMYx) weather normals.

This section only provides a high-level overview of methodology. For more detailed information on the methodology and input assumptions for this analysis refer to Section 2, Data Cleaning and Methodology.

This analysis only includes residential single-family homes in the northwest. While illustrative of potential outcomes of residential electrification in the region, these analyses are intended to be used in combination with other forecasts, such as those for commercial buildings, solar PV and EV charging. This report also provides a list of recommendations for further analysis on this data set to examine other aspects of residential electrification.

2.1.1 Identification of Peak Periods

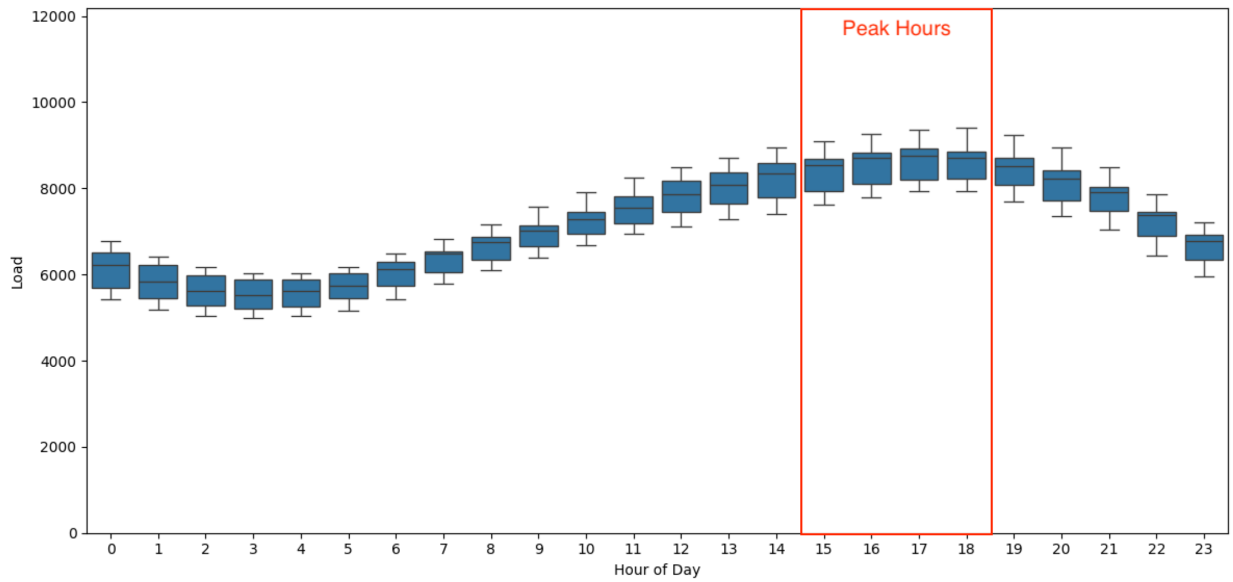
The team analyzed Bonneville Power Administration (BPA) system-level load data[1] to identify the peak period hours from each of the three peak periods: summer evening, winter evenings and winter mornings. Identifying peak period hours allows modeling the expected contribution to regional peak demand with scenarios of growing heat pump and HPWH loads. Based on analysis of the BPA system-level load data and reported information on regional usage[2], BPA usage accounts for roughly 30% of regional average yearly and peak load. We believe this is a reasonable proxy for estimating the peak grid load periods for the region.

Figure 1 shows the hourly load curves for the 25 peak days (five peak days from each of the last five years) selected for each peak period from the BPA data. As expected, the summer curve has a high evening peak and a much lower load in the morning, which is why we are not modeling a summer morning peak period. Winter, in contrast, has a clear dual peak both in morning and evening.

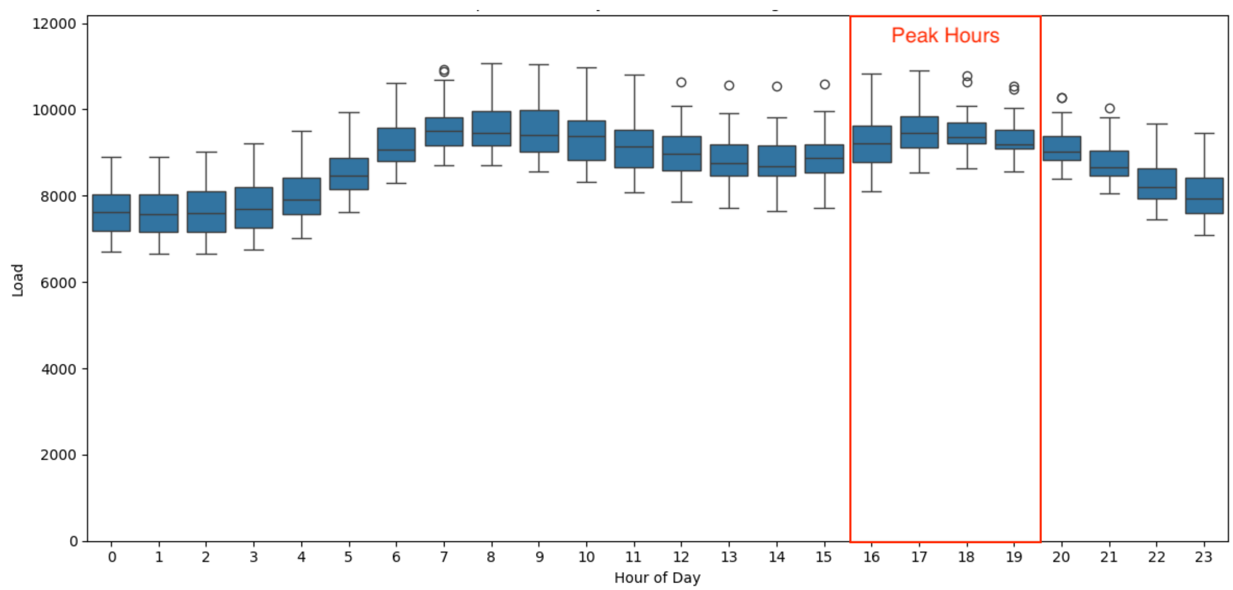
[1] https://transmission.bpa.gov/Business/Operations/FERC714/default.aspx
[2] https://www.nwcouncil.org/news/2025/05/02/pacific-northwest-load-forecast-2025/

Figure 1. Distribution of Loads by Hour for 25 Peak Days (2020 - 2025)

Summary Evenings



Winter Evenings



Winter Mornings

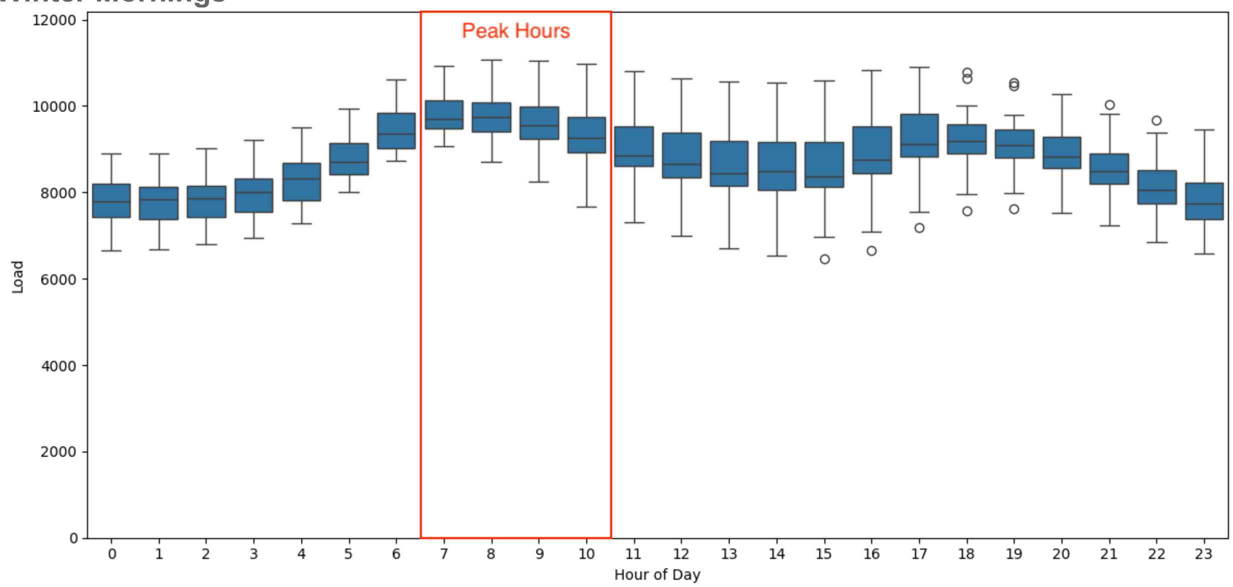
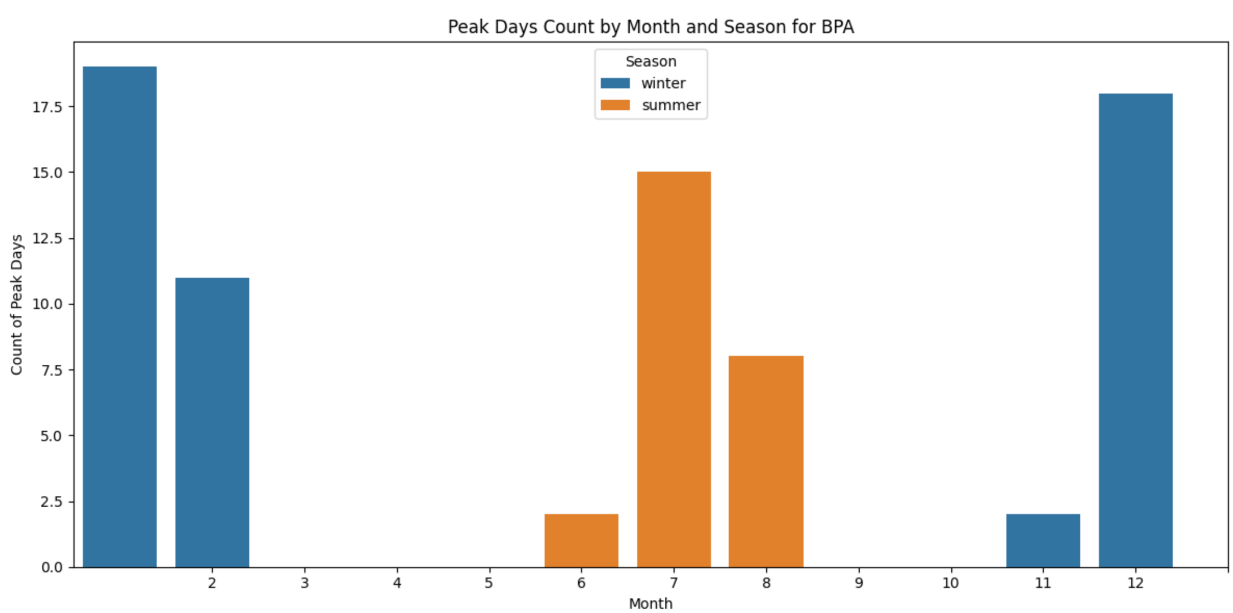


Figure 2 shows the count of peak days by month for summer and winter. The peak days for winter all fell in the months of November to February. The peak days for Summer were all in June to August.

Figure 2. Peak Day Counts by Month



To focus the analysis on the “peakiest” days when aggregating peak period load curves, the team decided to focus on the two highest months for the peak load analysis: July and August for summer and December and January for winter. Table 1 provides a summary of the selected peak period hours and months for the three peak periods. Note that an event running through the last hour starting 6:00 PM would end just before 7:00 PM, and the peak periods listed below are all four hours in duration

Table 1. Peak Period Hours by Season and Time of Day

	First Hour Start	Last Hour Start	First Month	Last Month
Summer Evenings	3:00 PM	6:00 PM	July	August
Winter Evenings	4:00 PM	7:00 PM	December	January
Winter Mornings	7:00 AM	10:00 AM	December	January

2.1.2 Weather Normalization of Weather Dependent Loads

To understand the impacts of electrification in a "normal" weather year, we used OpenDSM, which is an open-sourced tool for normalizing the load data based on outside temperature and schedule (month, weekday/weekend, hour of day). The weather normalization process involved the following steps:

1. Identifying a weather station to pull temperature actuals for each site, filtering for those which also have TMYx data available
2. Identifying a contiguous 12-month period for each site where each month has <10% missing power and temperature intervals. 6,327 of the 7,960 total points in the data set had sufficient data for normalization using this criterion.
3. Using those 12 months of data to create a weather model for each point.
4. Use the developed weather model to predict energy use using the most recently available TMYx weather normals for each site.

For more information on data cleaning and normalization, refer to volume 2, Data Cleaning and Methodology.

2.1.3 Technology Adoption Load Profiles

As homes in the region increasingly install ducted and ductless heat pumps for space heating and cooling, as well as heat pump water heaters, policy makers and grid planners want to estimate the impact of these technologies on the regional grid. The team used the HEMS data to address this question by modeling the net change in site-level electrical load for several different technology adoption scenarios for new installations and retrofit adoptions of electrified technologies. Retrofit adoptions were assessed against an assumed baseline technology (e.g., electric resistance or gas heating for heat pump adoptions). Note that HEMS only monitored the electrical consumption of fans, pumps, and other ancillary equipment for the gas heating systems, and this analysis is only estimating the change in these electrical loads, not the change in total heating energy use.

Heating

To more accurately classify the heating system type at each site, we first filtered out any space heating-type points that used less than 20W on average across all hours of the winter season, as points with loads below that level did not show any patterns of heating demand correlation. Table 2 provides a count of unused heating systems removed from the analysis.

Table 2. Unused Heating Points by Use Case and End-Use Type Label

Use Case	End-Use Type Label	Count
Heating	Ducted Heat Pump	17
Heating	Ductless Heat Pump	7
Heating	ERV	2
Heating	Electric Baseboard Heater	71
Heating	Electric Furnace	18
Heating	Other Zonal Heat	83

Next, we classified each site with a site-level heating type based on the combination of heating equipment used at the site. Table 3 provides a summary of site-level heating types and counts by site. Note that multiple combinations of end use types (shown in the middle column) can map to the same system type (left column).

Table 3. Site-Level Heating Types and Site Counts

Site Heating System Type	End Use Type Label List	Count
Ducted Heat Pump and Resistance Backup	Ducted Heat Pump, Electric Furnace	74
Ducted Heat Pump and Resistance Backup	Ducted Heat Pump	15
Ducted Heat Pump and Electric Resistance	Ducted Heat Pump, Electric Furnace, Other Zonal Heat	4
Ducted Heat Pump and Electric Resistance	Ducted Heat Pump, Other Zonal Heat	4
Ducted Heat Pump and Electric Resistance	Ducted Heat Pump, Electric Baseboard Heaters	2
Ducted Heat Pump and Electric Resistance	Ducted Heat Pump, Gas Furnace (Component), Other Zonal Heat	2
Ducted Heat Pump and Electric Resistance	Ducted Heat Pump, Electric Baseboard Heaters, Electric Furnace	1

Ducted Heat Pump and Gas Furnace	Ducted Heat Pump, Gas Furnace (Component)	11
Ductless Heat Pump	Ductless Heat Pump	31
Ductless Heat Pump and Electric Resistance	Ductless Heat Pump, Electric Baseboard Heaters	16
Ductless Heat Pump and Electric Resistance	Ductless Heat Pump, Other Zonal Heat	15
Ductless Heat Pump and Electric Resistance	Ductless Heat Pump, Electric Furnace	3
Ductless Heat Pump and Electric Resistance	Ductless Heat Pump, Electric Baseboard Heaters, Other Zonal Heat	1
Ductless Heat Pump and Electric Resistance	Ductless Heat Pump, Electric Furnace, Other Zonal Heat	1
Ductless Heat Pump and Gas Furnace	Ductless Heat Pump, Gas Furnace (Component)	2
Electric Resistance	Electric Furnace	43
Electric Resistance	Electric Baseboard Heaters	26
Electric Resistance	Other Zonal Heat	11
Electric Resistance	Electric Baseboard Heaters, Other Zonal Heat	9
Gas Furnace	Gas Furnace (Component)	97
Gas Furnace	ERV, Gas Furnace (Component)	3
N/A	Gas Furnace (Component), Other Zonal Heat	7
N/A	Electric Baseboard Heaters, Gas Furnace (Component)	3
N/A	Ducted Heat Pump, Ductless Heat Pump, Electric Furnace	2
N/A	Ductless Heat Pump, ERV, Gas Furnace (Component)	2
N/A	Ducted Heat Pump, Ductless Heat Pump	1
N/A	Ducted Heat Pump, Gas Furnace (Component), PTAC	1

Note that the heating type Ducted Heat Pump and Resistance Backup includes both heat pump-only sites and sites with both ducted heat pumps and resistive backup. This approach was taken because of the understanding that many ducted heat pumps are installed in combination with electric furnace backups or come with resistive backup. Instead of estimating the split between ducted heat pump only installations and those with resistive backup we assumed that the relative proportions of these systems in the HEMS population are the same as the proportions of system types for new installations and retrofits. Future analysis could possibly identify periods of resistive backup for heat pump installations through analysis of the 1-minute data.

Next, we aggregated the energy use of each end-use by site for each heating system type. The value for each site was divided by the conditioned square footage to get a per-square-foot measurement for each end-use and then those per-square-foot load shapes were averaged across the sites in each heating zone bin (HZ 1 and HZ 2-3).

We then took these profiles and multiplied them by the average square footage of each heating zone bin and then weighted them by the population of each bin to get a single region-wide population-weighted site-level peak load profile for each system type. Figure 3 shows the peak load profiles for each site-level heating type.

Figure 3. Peak Period Load Profiles for Site-Level Heating System Types

To translate these profiles into the adoption scenarios we took the retrofit profile and subtracted out the baseline profile to get an hourly trend of the changes in energy use by end use for each adoption scenario. Figure 4 shows the site-level heating profiles for each adoption scenario by system type for winter mornings and winter evenings.

Figure 4. Site Profiles for Heating Equipment Technology Scenarios

These show that while the electric resistance to heat pump scenarios result in a decrease in energy use during both winter mornings and winter evenings, the gas to heat pumps and new heat pumps result in an increase in energy use in the morning and evening. The decrease in the resistive to heat pump conversions is a result of the increased efficiency of heat pumps in comparison to electric resistance but is partially offset by electric resistance backup in the heat pump profile. Different forms of electric resistance are found in both the retrofit and baseline profiles and therefore this chart represents the net change in those end-uses due to the retrofit.

Cooling

To more accurately classify the cooling system type at each site, we first filtered out any space cooling-type points that used less than 5W on average across all hours of the summer season. Upon observation, the few points with seasonal average power levels below that threshold appear to either not be used or perhaps they were mislabeled in the breaker panel. Table 4 provides a count of unused cooling systems removed from the analysis.

Table 4. Unused Heating and Cooling points by Use Case and End-Use Type Label

Use Case	End-Use Type Label	Count
Cooling	Central AC	11
Cooling	Ducted Heat Pump	11
Cooling	Ductless Heat Pump	10
Cooling	ERV	2
Cooling	Room AC	3

Next, we classified each site with a site-level cooling type based on the combination of cooling equipment used at the site. Table 5 provides a summary of site-level cooling types and counts by site. Note that multiple combinations of end use types (shown in the middle column) can map to the same system type (left column).

Table 5. Site-Level Cooling Types and Site Counts

Site Cooling System Type	End Use Type Label List	Count
Ducted Heat Pump	Ducted Heat Pump	115
Central AC	Central AC	109
Ductless Heat Pump	Ductless Heat Pump	60
Central AC	Central AC, ERV	4
N/A	Central AC, Ductless Heat Pump	3
N/A	Ducted Heat Pump, Ductless Heat Pump	3
N/A	Central AC, Ductless Heat Pump, ERV	2
Room AC	Room AC	1
N/A	Central AC, Room AC	1

Next, we aggregated the energy use of each end-use for each site cooling system type. The value for each site was divided by the conditioned square footage to get a per square foot measure for each end-use and then those per-square-foot load shapes were averaged for each heating zone bin (HZ 1 and HZ 2-3).

We then took these profiles and multiplied them by the average square footage of each cooling zone bin and then weighted them by the population of each bin to get a single population-weighted site-level peak load profile for each site-level system type. Figure 5 shows the peak load profiles for each site-level cooling type.

Figure 5. Peak Period Load Profiles for Site-Level Cooling System Types

To translate these profiles into the adoption scenarios we took the retrofit profile and subtracted out the baseline profile to get a single hourly trend showing changes in energy use by end use for each adoption scenario.

Figure 6 shows the site-level cooling profiles for each cooling adoption scenario for summer evenings.

Figure 6. Per Site Cooling Profiles by Adoption Scenario

This shows a decrease of cooling load for central air conditioning to ducted heat pumps as the old cooling system is replaced by a more efficient heat pump unit. For room air conditioning to heat pump there is a small increase in energy use which may be the result of homes adding additional cooling capacity throughout the home and increasing comfort. This finding was based on a very small set of room air conditioners in the data set (it is likely that not all points using room air conditioning in the summer were identified since room air conditioning uses standard wall outlets and likely are not labeled on the panel) and is an area where additional research would be needed to establish the validity of this finding. New homes and homes going from no cooling system to heat pumps also result in increased cooling load on the system.

Water Heating

The process for calculating water heating impacts was similar to heating and cooling, with two exceptions:

1. It was assumed that each water heater retrofit was a one-to-one replacement of an existing water heater, where the single water heater served the entire site.
2. Since water heating is a year-round load, profiles were developed for all three peak periods.

Figure 7 shows the final technology adoption profiles for water heating for all three peak periods.

Figure 7. Per Site Water Heating Profiles by Site System Type

These results show savings for electric resistance to heat pump water heater conversions and increases in load from gas to heat pump water heater conversions and new heat pump water heater units. Both scenarios show a significant ramping of water heating load in the late evening which could be a result of increasing hot water usage over the evening that causes heat pumps to enter resistive mode to catch up with hot water demands.

2.1.4 System-Level Adoption Forecast Scenario Impacts

The site level adoption scenario impacts were scaled to the system level by multiplying them by the number of adopting homes in each of three defined adoption forecasts: low, medium, and high. Table 6 provides a summary of the number of new home adoptions by scenario.

Table 6 Count of Assumed Adopting Homes by Scenario

Scenario	Scenario Type	Low	Medium	High
Electric Resistance to Ducted Heat Pump	Heating	183,300	366,600	672,100
Gas Furnace to Ducted Heat Pump	Heating	98,700	197,400	361,900
Electric Resistance to Ductless Heat Pump	Heating	336,050	672,100	977,600
Gas Furnace to Ductless Heat Pump	Heating	180,950	361,900	526,400
Central Air Conditioner to Ducted Heat Pump	Cooling	112,800	225,600	413,600
No Cooling to Ducted Heat Pump	Cooling	169,200	338,400	620,400
Room Air Conditioner to Ductless Heat Pump	Cooling	336,050	672,100	977,600
No Cooling to Ductless Heat Pump	Cooling	98,700	197,400	361,900
Resistance to Heat Pump Water Heater	Water Heating	43,710	198,810	457,310
Gas to Heat Pump Water Heater	Water Heating	43,710	198,810	457,310
New Ducted Heat Pump	Heating	0	7,742	15,290
New Ductless Heat Pump	Heating	0	3,871	7,936
New Ducted Heat Pump	Cooling	0	7,742	15,290
New Water Heating	Water Heating	2,323	17,807	33,291

For more information about how the adoption forecasts were developed, see section 2.3 in Section 2: Data Cleaning and Methodology. Figure 8 shows the plots for the different adoption forecasts for the three peak periods.

Figure 8. System-Level Impacts for the Adoption Forecasts

These show that residential electrification could add from 175 to 642 MW to peak period loads during summer evenings[3], while winter evenings show a reduction of 161 - 609 MW and winter mornings show a reduction of 62 to 765 MW with the largest drop happening at the 7 o'clock hour. The 7am drop is a result of the transition from electric resistance to heat pumps for both ductless and ducted, as shown in Figure 4. When replacing inefficient equipment with more efficient equipment, the biggest impact will be in the hours of highest use. The savings from these reductions are overshadowing any other increases in the residential sector.

The reduced morning load is a direct result of the assumption of the ratio of electric resistive heating and gas heating that is being replaced by heat pumps, as shown in Table 6 Count of Assumed Adopting Homes by Scenario. We assumed that electric resistive heating would be replaced 65% of the time and gas would be replaced 35% of the time, given that the customer economics are better for electric resistance replacements. Note that if a higher percentage of the heat pumps were replacing gas heating systems, the net impact would increase, potentially even tipping the balance to positive net load impact for some periods. Again, note that these impacts only include residential single-family homes and the specific technologies listed; it is not a forecast for overall changes in regional grid demand.

Table 7 provides the net peak impacts by hour for each peak period and adoption forecast scenario.

---

[3] Based on the NW Power Council's summer peak load value of 33,000 MW, this represents 0.5 - 2% of regional peak load (https://www.nwcouncil.org/news/2025/05/02/pacific-northwest-load-forecast-2025/

Table 7. System Level Peak Period Impacts by Adoption Scenario in Megawatts (MW)

Summer Evenings				
Hour	**15**	**16**	**17**	**18**
Low	175	179	178	175
Medium	357	368	362	359
High	618	642	629	622
Winter Evenings				
Hour	**16**	**17**	**18**	**19**
Low	-177	-197	-210	-161
Medium	-332	-372	-389	-273
High	-527	-583	-609	-405
Winter Mornings				
Hour	**7**	**8**	**9**	**10**
Low	-277	-89	-62	-82
Medium	-529	-151	-98	-137
High	-765	-166	-113	-197

2.1.5 Temperature Scenarios by End-Use

In recent years, the Northwest region has experienced extreme temperature events, such as the "heat dome" of 2021, with increasing regularity. While it is notoriously difficult to model outlier events that are scarce in the training data, the OpenDSM models can nonetheless provide some insights into how weather-sensitive loads such as electric space heating and cooling are expected to change under different weather conditions.

The goal of this analysis is to simulate a year of hourly energy use for a range of hotter and colder temperatures. The team selected +15°F as the upper bound as that represents the deviation from the TMYx maximum temperature that was recorded during the 2021 heat dome event. To gauge the result from possible decreased temperatures, we mirrored that delta to extend to -15 °F for the lower bound.

We used the same OpenDSM models described earlier in this report to forecast the expected hourly energy consumption for six theoretical weather years at 5°F intervals between -15°F and

15°F (-15°F, -10°F, -5°F, +5°F, +10°F, +15°F). These weather data sets were constructed by adding each temperature offset to the dry bulb temperature for each hour. For example, in the -5°F data set, the temperature would be 5° lower in all hours; this includes both summer and winter days.

Figure 9 provides an example 24-hour seasonal load shape aggregated across all ducted heat pumps in the data set. Note how for summer, warmer temperatures cause a roughly 80% increase of load over the baseline in the 5 pm hour. Note that the model also shows that colder temperatures would result in increased morning heating load even during the summer, suggesting that there are already some cool summer mornings in this climate where heat pumps are being called for heat. However, since there are likely very few hours of data that OpenDSM can use to model this relationship, we would caution against reading too much into the portion of this analysis that extrapolates summer morning heating impacts.

In winter, the main impact is roughly 25% increase in the heat pump load shape during the 7 am hour which could drive the region to increased winter morning peaks.

Figure 9. Seasonal Daily Load Shapes for Ducted Heat Pumps Under Various Temperature Scenarios

Figure 10 shows another example with heat pump water heaters. In this case, increases in peak load occur around the 8pm hour in the colder scenarios, likely due to the increased temperature difference between ambient and the hot water temperature setpoint. These devices show a decrease in energy use due to warmer temperatures during evening peaks including in the summertime.

Figure 10. Seasonal Daily Load Shapes for Heat Pump Water Heaters Under Various Temperature Scenarios

Figure 11 shows the load shapes for “mains” (whole home) energy use.

Figure 11. Seasonal Daily Load Shapes for Mains Under Various Temperature Scenarios

As expected, these charts more closely resemble the heat pump charts as the predominate weather dependent load in the whole home is the heating and cooling systems. However, these charts do not show the pronounced increase in spring and fall loads under the highest temperatures, as these effects for ducted heat pumps must be counteracted by other decreased loads (e.g., water heating) or ductless heat pumps that show a much smaller increase in these periods.

We also looked at how average usage varied by temperature scenario offset across different seasons and day periods (morning, afternoon, evening, night). Figure 12 shows the relationship for ducted heat pumps. This view makes it easier to see the “slope” of kW per degree of temperature shift at specific temperatures and what the nature of the relationship. For ducted heat pumps, the minimum average kW occurs just over 50 degrees, with temperatures lower than this having a negative slope and temperatures above this having a positive slope as expected for a heat pump providing both heating and cooling.

Figure 12. Average Load by Temperature Offset for Ducted Heat Pumps

The figure shows that morning and evening loads tend to fall as temperatures rise except for summer where both increased temperatures and decreased temperatures lead to increased usage. Afternoons and evenings show decreased usage in winter as temperatures rise and increased usage in the summer as would be expected. Afternoons and evenings also show that both increased and decreased temperatures show increased usage in the shoulder seasons of spring and fall which is also intuitive as temperature changes could tip these shoulder seasons into net heating or cooling periods depending on which way the temperatures change.

Figure 13 shows the relationship for heat pump water heaters.

Figure 13. Average Load by Temperature Offset for Heat Pump Water Heaters

This end use shows a consistent increase in usage with lower temperatures and decrease for higher temperatures, presumably driven by inlet water and perhaps also by ambient temperature for those units in unconditioned space.

Figure 14 shows the relationship for the mains (whole home) power.

Figure 14. Average Load by Temperature Offset for Mains

As with the previous charts, these charts more closely resemble the heat pump charts as the predominate weather dependent load in the whole home is the heating and cooling systems.

3 Conclusions

The conclusions are organized in two sections:
1. Answers to research questions
2. Recommendations for additional research

3.1 Answers to Research Questions

3.1.1 Increase Data Access and Lower Barriers to Future Use

This project has successfully cleaned, organized, and aggregated the HEMS data to lower the barriers for future analysts. This work has resulted in the following improvements to the data set:
1. Standardization of data tables, including
    a. Creation of a consistent and easier-to-use point ID for labeling points and connecting data tables
    b. Merging of key elements from the Residential Building Stock Assessment (RBSA) into the site table for convenience
    c. Improved end-use labels and categories based on reclassification provided by NEEA
2. Selection of new weather stations and providing new actual and typical (TMYx) temperature data with improved data coverage
3. Creation of a cleaned and filled table that combines power and temperature data and fills missing intervals using open-sourced approach from OpenDSM

For more detailed information on the methodology and input assumptions for this analysis refer to Volume 2, Data Cleaning and Methodology.

3.1.2 Develop a Weather Normalized Dataset and Standardized Pre-Aggregated Data Tables

The project successfully normalized 88% of points based on weather and schedule using an open-sourced python library (OpenDSM) and standard-year TMYx temperature data. We also created a number of data aggregations with numerous statistical measures (e.g., mean, median, minimum, maximum, standard deviation, 25th percentile, 75th percentile), including
1. Daily, weekly, monthly and annual average power by point
2. Monthly and seasonal 24-hour load shapes for weekdays and weekends
3. Seasonal power ratios (summer to shoulder seasons, winter to shoulder season)

The normalization process surfaced a number of challenges, including:

1. OpenDSM has been primarily used for site-level data analysis and sub-meter data does not always have the same characteristics. For instance, while a whole-home electric meter will almost never have a zero-kWh reading for a 15-minute interval, itis common for a single circuit reading to remain at zero for long periods of time, depending on the loads that are connected to it. See section 3.8.1 in Section 2 for more discussion on how we addressed these issues.
2. The normalization process on circuit level metered data produced model error metrics (e.g., R squared, CVRMSE) are higher than is typically expected for well-fitting models for whole site loads. This is likely due to loads that have neither strong weekly schedule patterns nor strong weather-dependent trends. After reviewing modeling results for both points with high and low model errors the team determined modeling results still produced adequate results even when modeling errors were high. We based this decision on the comparison between modeled and actual data aggregations such as seasonal 24-hour load shapes. The team decided to force model creation for as many points as possible and provide model metrics for future analysts to filter out modeled points based on modeling error if deemed necessary.

For more detailed information on normalization and aggregations refer to Volume 2, Data Cleaning and Methodology.

3.1.3 Perform Data Analysis to Understand the Impacts from Heat Pumps and Heat Pump Hot Water Feeders on Grid Load During Peak Periods

The project successfully demonstrated the usefulness of this data set by producing estimates of system-level impacts of increased heat pump and heat pump water heater adoption in the Northwest. As a step in this process, we identified three peak periods for analysis based on BPA system-level data and created peak period load profiles by end-use for 16 adoption scenarios. These scenarios were then scaled into system level forecasts using three adoption scenarios (high, medium and low) that show potential net increases of summer evening peaks due to the addition of cooling load from heat pumps and a net reduction of winter peaks due to the increased efficiency of heat pumps in comparison to resistive heating. This latter finding is a result of a large, assumed number of resistive heat to heat pump conversions so that the savings from these installations outweigh the increase in load due to conversions from natural gas equipment to heat pump .

This result may seem counterintuitive, given the prevalence of gas-fired heating systems in the region, and how the site-level heating profiles calculated using HEMS data show that gas-to-heat pump conversions will lead to increased winter grid loads. However, the assumptions in the adoption forecast described a scenario where twice as many homes used heat pumps to replace electric resistance heating vs. gas heating. This assumption was driven by superior customer economics for the former case over the latter. However, there are many factors that will influence adoption of heat pumps over the next decade, including economic incentives, building codes and other policies, as well as popular attitudes towards fossil fuels and heat pumps, so this outcome is highly sensitive to factors outside the scope of this study.

Like any forecasting analysis with a 10-year time horizon, this is less intended as a definitive prediction of the future as it is a tool for exploring the relationships between patterns of peak grid demand, equipment adoption rates, and weather-driven performance of electrical loads. The analysis was performed using transparent, repeatable software code to facilitate future analysis with updated assumptions as the trajectory of regional electrification evolves. The analysis code also generated intermediate outputs with both device-level and site-level load shapes, allowing analysis and stakeholders to better understand the drivers behind the forecast's conclusions. These unit-level load shapes are stored as standardized, indexed files that can also facilitate additional scenario analysis without even editing or running any code.

3.1.4 Estimate Impact of Shifts in Temperature on Electrical Loads

The OpenDSM weather models were used to predict how the seasonal load shapes of different end-uses might change if the regional temperature norms continue to shift, as they have started to do in recent years. We created synthetic weather data sets for each weather station, shifting the outdoor dry bulb temperature both higher and lower, then modeled how each point would perform under these alternative weather scenarios. Predictably, weather-sensitive loads like heat pumps (both for space conditioning and water heating) showed the strongest responses to colder temperatures in winter mornings and hotter temperatures in summer afternoons. Spring and fall load shapes showed more modest increases under both hotter and colder scenarios. These results can also be useful for regional planners to understand how sensitive these types of loads would be to changing climate norms so that they can overlay that variability on the forecasts of grid impact due to electrification.

3.1.5 Provide Dynamic Reporting of the Results

This project successfully developed a set of dynamic reporting dashboards for NEEA that were used to share findings and methodologies at every step in the project. These dashboards provide useful tools for NEEA or other program funders to do further exploration on the datasets and to perform additional data cleaning and circuit-level end-use categorization review for any additional end-use labels that could be updated for increased accuracy. These dashboards have already proven useful for the analysis team and stakeholders to explore the results of this work through the development process, but there are far more insights to be mined by others who are looking to answer different questions with this rich data set.

3.2 Applicability of the Analysis

While this data set represents one of the largest and most long-running field studies of its kind, it is also attempting to address a wide array of research questions. It includes numerous types of end-use equipment in different configurations across multiple geographic regions with different climates and building stock trends. These variations cannot always be fully normalized when extrapolating to the larger population, because these estimates are based on the sample of

HEMS sites and the proportion of equipment types they have installed, particularly regarding system type definitions that incorporate different combinations of end use types. For example, “ducted heat pump” includes homes that also have load on circuits labeled "electric furnace” because the project organizers indicated that most of those loads are electric resistance backup heating elements installed as part of the ducted heat pump. If the blend of heat pump-only and heat pump plus resistance is different in the general population, that will affect the load shape for this system type, but we do not have enough information to correct for this bias. Future work could include a population-level load shape that is weighted based on the prevalence of different combinations of Heating Ventilation and Air Conditioning (HVAC) system types.

Some end-uses are also better suited for circuit-level metering, namely large loads like HVAC and DHW systems that are hard-wired to an individual breaker and that do not typically require circuit tracing. By contrast, plug load appliances are harder to correlate to a circuit metered in the breaker panel. For example, the data set only included four points labeled as Room AC and the energy use of these points was highly variable. Because room AC is a general load powered by standard wall sockets instead of dedicated circuits it is likely that many more points serviced room AC loads, but those labels were not captured in metering. In addition, for those labeled as room AC, there is the possibility of significant other loads on the circuits from other devices plugged into outlets on the same circuit in the household.

Also note that we defined system types based on the distinct set of end use types that exceed a minimum seasonal energy use threshold. Defining the system type of each site in this way means that the retrofit profiles for heat pump plus a backup system will exclude sites that are able to fully shut off the electric resistance or fossil fuel backup, as those will be considered “heat pump only” homes. We made this choice because it is easier to take the “clean” profiles that consist of sites that all have the same set of known loads and then later create weighted average profiles that incorporate the heat pump-only profile on a specified percentage of homes to represent the assumption that some homes fully displace their baseline systems. The one exception to this is “ducted heat pump” systems, as those were defined as including homes both with and without electric furnace loads, since that was believed to be standard backup component of these heat pump systems, even for new construction.

3.3 Recommendations for Additional Research

Throughout this project, the combined team recognized a number of areas of additional research that were out of scope for this project. These included:

1. Analysis of the ramping effect of heat pumps during and immediately preceding peak periods to assess how quickly additional power sources need to come online to meet increasing heating and cooling load from heat pumps.
2. Additional analysis on selected weather stations and missing data thresholds to increase the number of modeled points. Six sites were excluded from normalization though they had twelve months of usable power data and twelve months of usable temperature data but did not have twelve months usable for both. In addition, the project used a low threshold of 10% of missing intervals, while research from the OpenDSM team has

indicated that on site-level data up to 50% of intervals can be reliably filled using OpenDSM with good results. Further research could be done to see if a looser missing interval threshold would capture more modeled points

3. Additional analysis of Electric Vehicle Supply Equipment load shapes. These devices were excluded from the peak load analysis and determined to be out of scope for this project due to the low number of points (17) in the data set. It was determined that these were too few to meaningfully fill bins when using the original binning strategy based on heating zone, vintage and square footage. However, with the revised approach of binning power data normalized on square footage by heating zone only, these devices could be analyzed using similar strategies.
4. Additional analysis of the effect of room air conditioners to heat pump conversions. As noted above, only four room air conditioner circuits were identified in the data and it is likely that a significant number of circuits used for room air conditioning in the summer were missed since they are not typically labeled at the circuit level. In addition, circuits with room air conditioning could also have mixed loads since other uses such as lighting, televisions, or other appliances could be plugged into the same circuit and add significant non-HVAC base load. The results of this research could be approved by developing room air conditioning load profiles using other sources of data as applicable.
5. Develop a population-level load shape that is weighted based on the prevalence of different combinations of HVAC system types. Some of the nominal HVAC system types can include more than one type of HVAC equipment. For example, “Ducted heat pump” describes a system type that is applied to sites that may or may not include an electric resistance furnace as a backup heat source. These end-use load shapes were averaged to the site level, which implicitly incorporates the proportion of sites with and without electric resistance furnaces (in this example). If the sample of homes in the HEMS data set has over- or under-represented the proportion of centrally ducted heat pumps that use electric resistance heat versus those that operate purely off the heat pump, relative to the larger population of sites with centrally ducted heat pumps, that could misrepresent the population-wide average load shape for this heating system type. Population-level metrics of the prevalence of different combinations of HVAC systems could be used to weight these averages, but that was out of scope for the current study.
6. Estimate HP resistive backup reasonably well from HP circuit load by tracking 1-minute changes through time.

Home Energy Metering Study Data Standardization and Electrification Peak Load Analysis

Section 2: Data Cleaning and Methodology

1 Section 2 Introduction

This volume contains detailed methodology and results for the data cleaning of the HEMS data set and detailed methodology for the statistical analysis of peak loads.

2 Data Cleaning

2.1 Overview

The purpose of the data cleaning was to increase data access and lower barriers to future use of HEMS data by making cleaned data and methods available. The project resulted in a set of cleaned and formatted data tables for use in future HEMS analysis. The main output tables included:

1. Cleaned HEMS data
    a. **Point –** "Points" refer to metered circuits within the raw and final data schemas. This table uses a new point ID field to uniquely identify points and map them to power readings. This file also includes the site ID, the raw end use type label, and an updated end use type label. **Site -** an updated site table with new selected weather station information and merged with some data fields from the RBSA.
    b. **Power_reading_15m -** a standardized version of the raw POWER15 table, it contains average kW readings for each point in 15-minute intervals with the new point IDs added as a unique identifier. This file also has duplicates and low power points removed.
    c. **Combined_power_temperature_filled -** a new version of the power data in hourly intervals that has been combined with temperature data from EEweather and filled with interpolated intervals using OpenDSM.
2. Data aggregates
    a. **Annual_avg_power** - statistical metrics (e.g., mean, median, count, min, max) for each point by year.
    b. **Monthly_avg_power -** statistical metrics (e.g., mean, median, count, min, max) for each point by month.
    c. **Weekly_avg_power -** statistical metrics (e.g., mean, median, count, min, max) for each point by month.
    d. **Monthly_power_loadshape -** 24-hour average load shapes for each point for weekdays and weekends for each month.
    e. **Seasonal_power_loadshape -** 24-hour average load shapes for each point for weekdays and weekends for each season.
3. Typical Meteorological Year (TMYx)
    a. **Weather_station:** name and location of each weather station

b. **Tmyx_hourly_temperature:** weather data for each station, mapped to the standard reference output year.
c. **Temperature_scenario_predictions:** hourly power data predictions using various temperature offsets from the TMYx data (+/- 15 degrees Fahrenheit in 5°F increments).

Note that several additional descriptions of input assumptions, analysis outputs and data summary files are available for users that are interested in looking deeper into the data.

2.2 Technical Approach

This project used publicly available and open-sourced software for all analysis. The main analysis programming language was Python. The team utilized Kedro as a pipeline manager and created dynamic reporting tools using Evidence. The pipeline stored most interim files as parquet to minimize file size and preserve column data types where possible. For large tables (e.g., power readings, temperature) the data was stored using a local DuckDB database for faster performance. Weather and schedule normalization was performed using OpenDSM. "Actual year" temperature data from NOAA was used and was retrieved using EEweather, and TMYx data was used as normalized temperature data for weather modeling.

2.3 Data Retrieval and Cleaning

Cleaning of the HEMs data set included the following steps:

1. **Standardize -** establish base tables from raw data and align data table names and column data types to project specifications
2. **Restructure -** add new unique identifiers, perform conversions of data values and add additional columns from joining to other standardized tables
3. **Check -** look for missing, duplicate, or outlier values
4. **Clean -** filter out bad data, fill missing data with interpolated data
5. **Calculate Features -** summarize data to aid analysis and reporting

2.3.1 HEMS Data

The raw HEMS data was retrieved from the NEEA End-Use Load Research project data portal[4].

Table 8 shows the data inputs used for the project from the HEMS dataset.

[4] https://neea.org/northwest-end-use-load-research-project/

Table 8. HEMS Data Inputs

Data File	Description
POWER15.csv (partitioned CSV)	15-minute average power readings for each metered point (circuit). These readings were recorded with true-power meters that measured power factor as well as volts and amps, stored as 1-min kWh averages, then aggregated by NEEA into 15-min averages
HEMS SITES.csv	HEMS sites including location (zip code and time zone), RBSA site ID, heating zone, and original weather station ID.
HEMS POINTS ALT.csv	Individual metered circuits or “points” with end use, circuit name, original point ID, as well as a corrected end use for some circuits. Linked to HEMS SITES.csv with site ID.
HEMS_RBSA_BASE TABLE.csv	A RBSA data file with a row for each site which contains the site ID, RBSA site ID, RBSA version[5], and characteristics of the home such as vintage and square footage
HEMS Site Geographic Information.csv	Geographical information including zip code, latitude and longitude. Note: This information is not publicly available and was only used to increase accuracy of matching sites to weather stations.

2.3.2 Standardization

The first step in data processing is standardization, which includes the following transformations of the raw data:

1. Renaming columns with all lower-case letters separated by underscores (e.g., point_id)
2. Setting correct data types for table columns
3. Creating primary key identifiers for table (e.g., a new point_id was created as the original data did not have a point id for every point. The new point ID is a sequential series starting at 1 after sorting all unique points from both the raw point table and power tables by HEMS site ID and register name)
4. Selecting only necessary columns

2.3.3 Restructuring

Once each table is standardized, the tables are restructured as necessary. Restructuring primarily consists of merging columns from other tables, as necessary. For example:

1. The power table pulls in the point_id from the points table by merging on a combination of the hems_site_id and register_name. After the point_id is merged into

[5] The RBSA file includes data from both RBSA studies as well as data from an abridged survey with identical questions that was administered to HEMS participants who were not part of the RBSA study. For more details, see Table 10.

the power table as the foreign key for the points table, the register_name column is dropped.

2. A time_zone field is added to the site table mapping supplied time zone abbreviations to IANA time zone identifiers that can be used in Python/Pandas
    a. PST is mapped to ‘US/Pacific’
    b. MST is mapped to ‘US/Mountain’
3. The necessary RBSA data (e.g., square footage) is merged into the site table.

2.3.4 Defining End Use Types and Categories

The HEMS POINT ALT data includes two versions of the circuit type (end use type): 1) raw/original version, 2) a revised label based on investigation and analysis by NEEA. This project used this data to create a specific table for storing different versions of the circuit type label. Some circuit type labels were renamed for consistency and similar end use type labels were combined for consistency (e.g., ‘pump’ and ‘PUMP’). A table of updated circuit type labels is provided in Appendix 1.1.

The team also manually mapped end-use type labels to broader end use categories such as “HVAC,” “Appliance,” “Mains,” etc. For each end-use type label there were also three flags assigned to each:

1. HVAC is Heat – indicates HVAC end-uses that are used as heating
2. HVAC is Cool – Indicates HVAC end-uses that are used as cooling
3. HVAC is Non Load - indicates HVAC end-uses where only ancillary components were metered, specifically gas furnaces, so that those points could be excluded when executing the minimum load threshold rules.

The end use category mapping file showing the broader category and the flags described above is shown in Appendix 1.2.

2.3.5 Check

This step analyzed the data to identify the following data issues:

- Duplicate rows in the site data with the same HEMS site ID.
- Duplicate rows in the point data with the same point ID.
- Duplicate rows in the power data with the same interval start time and point id.
- Duplicate rows in the temperature data with the same interval start time and weather station USAF ID.
- “Low power” points where the absolute value of the power for the point never exceeds 1 watt.
- Points where no temperature data was present.

2.3.6 Clean

The cleaning step involved three sub steps, 1) data filtering, 2) missing interval analysis, and 3) data gap filling

In the filtering sub step, all values identified as having the data issues in the Check step: duplicates, low power points and points with no temperature data were removed as not being eligible for further analysis.

The HEMS power data and the EEweather temperature data were then resampled from 15 minutes to hourly and combined into one table which includes every possible interval from the first to the last month of power data. We then calculated which months had: 1) data for at least 90% of possible power intervals, 2) data for at least 90% of possible temperature intervals, and 3) data where at least 90% of hourly intervals have both power and temperature data. Months that met the third criteria were labeled as being usable for weather normalization.

The combined power and temperature data was then run through OpenDSM to create an hourly "baseline" object, which fills values for any missing intervals. OpenDSM fills gaps in the interval data by using interval data from the same hours of similar days that are nearby both before and after the missing data, which allows it to fill longer gaps while maintaining the load shape. To ensure values where the power equaled zero kW were retained, the OpenDSM flag of "is_electricity_data" was set to "False". This was necessary because OpenDSM was designed for whole-home data where electricity data intervals are almost never equal to zero, so the model assumes this is missing data, but gas data sometimes includes daily intervals with zero-value readings. The HEMS data is for individually metered circuits, so a value of zero is a valid value and should be retained. Note that the resulting "filled" data set retains markers added by OpenDSM for which intervals have been filled by their algorithm (interpolated_avg_kw, interpolated_temperature). It also includes markers for each interval for if that month has data usable for normalization (monthly_avg_kw_ok, monthly_temperature_ok).

2.4 Power Data Normalization

Normalizing hourly power data based on temperature and schedule (e.g., month, weekday/weekend) required the following steps described in this section:

1. Identifying weather stations for temperature data
2. Retrieving temperature data for selected weather stations
3. Selecting power and temperature data for normalization
4. Fitting a model of power versus temperature and schedule for each point
5. Predicting power using temperature normals

2.4.1 Identifying Weather Stations

The project identified weather stations using EEweather[6] for use in normalization based on the following methodology:

1. Filter available weather stations for those where:
    a. Distance between the site and station is less than or equal to 150 km

[6] https://eeweather.readthedocs.io/en/latest/api.html#isdstation-objects

b. Elevation difference between the site and station is less than or equal to 300 m
c. Weather data quality is designated as ‘high’
d. TMYx 2009 - 2023 data (the most recent set) is available
2. The closest weather station was then selected from the list of filtered weather stations

In developing the weather station selection methodology, the team analyzed the effect of tightening distance and/or elevation change on model accuracy, and it was determined that overall model accuracy was relatively insensitive to changes in these parameters. As such, the team agreed to a set of reasonable parameters that provided a weather station match for all sites.

For each site a single weather station was identified for retrieving temperature data. Each weather station is identified with a unique six-digit U.S. Air Force ID (weather_station_usaf_id).

2.4.2 Retrieving Temperature Data

For each unique weather stations identified for the HEMS sites, the following temperature data was retrieved:

1. Hourly “observed” temperatures for all years where power data was present
2. TMYx 2009-2023 data[7]

The TMYx data represents a “Typical Meteorological Year” and therefore does not represent a specific year. In order to consistently analyze and visualize the data, the TMYx timestamps were mapped to a single “reference” calendar year of 2025 for aggregation and analysis. The reference year is configurable in the Kedro pipeline parameters.

Note: TMYx weather normals were used instead of TMY3 because the TMYx normals are derived from 2009-2023 weather data for most stations, which is much more recent than the traditional NREL TMY3 data. The new “gridded” TMY from NREL does have weather normals derived from more recent weather data, but the interpolated locations do not directly correlate with weather stations to use for the actuals. Also, a software library like eeWeather for automatically downloading the data for a given weather station is not yet available, likely due to the more complicated asynchronous API required for the gridded TMY data.

2.4.3 Selecting Power and Temperature Data for Normalization

As previously noted, the data cleaning step included analysis to determine which months had enough data coverage for normalization (defined as having no more than 10% missing intervals for both power and temperature data). This analysis was leveraged to determine the time range to use for normalization, as the normalization process uses one year of data for an input. To this end, an analysis was completed to assess which points had one full year (twelve contiguous

[7] Source URLs for the TMYx file for each weather station were taken from the Region 4 source file, where the USAF ID matches to the World Meteorological Organization (WMO) column in the file https://climate.onebuilding.org/sources/Region4_USA_TMYx_EPW_Processing_locations.xlsx

months) of data usable for weather normalization. For points where at least twelve months of usable data could be identified, the team identified the oldest twelve-month period that met the criteria. The oldest set of months was selected because using data closest to when the end-use-type labels were set increases the likelihood that these labels remain accurate and are not affected by any upgrades or changes in the household over time.

2.4.4 Modeling Power by Point

Once the optimal normalization date range was determined for each point, the cleaned interval data (combined power and temperature data with missing intervals interpolated using OpenDSM) was filtered to the selected normalization dates. This input data was used to create an OpenDSM hourly baseline object for each point (as noted above, the flag ”is_electricity_data” was set to ”False” in this step so that OpenDSM did not interpret zeros as missing data).

The hourly model was then fit using the baseline data object. The team decided to create models for as many points as possible by setting the flag ‘ignore_disqualification’ to ‘True’ while fitting the data to force model creation for each point, including those with model error metrics that exceed the thresholds typically used when generating counterfactual baselines for demand-side impact measurements. We made this decision after comparing the seasonal weekday/weekend 24-hour profiles generated from the model’s TMYx forecasts to those generated from the raw data. The “poorly-fitting” models typically come from points that don’t have strong correlations with outdoor weather conditions, which is expected for many residential plug loads. In these cases, the model primarily reflects monthly hour-of-week patterns, which effectively captures the patterns of use that we believed most analysts would find useful. During the process, the analysis captured and documented all disqualifications, errors and warnings in the normalization process in case future analysts choose to exclude certain points and models based on model quality.

The OpenDSM EEmeter model uses an elastic net regression framework to model energy use against outdoor dry bulb temperature, as well as hour of day, day of week, and month of year. This model was developed through an open-source, stakeholder-driven process that ensures transparent, repeatable methods[8].

2.4.5 Predicting Power Using Temperature Normals

Lastly, each point’s model used the TMYx temperature data was used to create predicted power usage for a year of “typical” weather conditions. The TMYx data for each weather station was mapped to a single “reference” calendar year of 2025 so that all results could be aligned consistently during the aggregation analysis. Again, the flag ”ignore_disqualification” was set to ”True” to force prediction for each point and all errors and warnings were captured in this process for further analysis.

[8] More information about the modeling methods can be found here: https://opendsm.energy/documentation/eemeter/hourly_model/methodology/

2.5 Data Cleaning Findings

This section provides a summary of data quality for the HEMS data sets and is organized by the main data tables used in the analysis.

A key metric for data quality in the analysis is how many total points remain in various stages of the analysis pipeline. Out of the total 7,690 unique points in the data set, 6,327 (82%) were able to persist through the normalization step. Table 9 provides a summary of points that were retained for each step in the analysis.

Table 9. Point Count by Analysis Stage

Analysis Step	Remaining Points	Point Count Change	Percent Change	Total Change
Total Points	7,690		0.0%	0.0%
Points with Power Readings	7,645	-45	-0.6%	-0.6%
Cleaned & Filled Points	7,513	-132	-1.7%	-2.3%
Points with Adequate Data for Normalization	6,517	-996	-13.3%	-15.3%
Modeled/Predicted Points	6,327	-190	-2.9%	-17.7%

2.5.1 HEMS Site Data

The HEMS data was matched with the RBSA data to pull in home characteristics that were not available in the HEMS data itself. The main objective of NEEA's RBSA is to develop a current, robust, and representative characterization of the existing residential single-family and multi-family building stock in the Northwest as it relates to energy use; it is primarily based on site collected data and is updated approximately every 5 years. There were multiple "versions" of RBSA data that were drawn from, as described in Table 10.

Table 10. RBSA Versions and Site Counts for HEMs Analysis

RBSA Version	Description	Count
2017	Collected site data on 930 homes in 2017	222
2022	Collected site data on 1,737 homes in 2022	133
Mini RBSA	Collected limited data on 27 homes that were recruited outside of 2017 and 2022 RBSAs that did not agree to participating in a full RBSA site visit.	27
RBSA for HEMS	Collected full onsite data of 30 homes that also participated in the HEMS study using RBSA 2022 data collection protocols.	30
N/A	6 Homes that were recruited outside of the 2017 and 2022 RBSAs that declined any further data collection, so their data contains very little HOME/HVAC data.	6

One duplicate site was identified in the dataset. In this case, there were two separate HEMS site IDs that shared a single RBSA ID. It was determined that this was a home with a related outbuilding on a different meter and the two sites were combined for analysis purposes. This brought the total sites in the data set from 419 sites to 418 sites.

Six sites have no RBSA ID, as there is no RBSA data for these sites.

2.5.2 RBSA Site Data

As noted above, two sites (7531 and 7532) have the same RBSA Site ID (SITE_01317). These sites were combined for analysis. Six HEMS sites have no available RBSA data.

For square footage calculations, we used data from the RBSA. The RBSA data has two square footage values, total floor area, and conditioned area. We used conditioned square footage for analysis. The team attempted to compute unconditioned square footage by subtracting conditioned square footage from total floor area, but 181 sites (43%) had a negative calculated unconditioned square footage when subtracting conditioned from total. Twelve of the 418 RBSA (3%) sites are missing conditioned area data. 31 of 418 RBSA sites (7%) were also missing home vintage data, though this was not ultimately used for binning.

2.5.3 Point Data

We found that the points in the points and power tables did not match. There were both points that appeared in the points table but were missing from the power table, and points that were appeared in the power table, but not in the points table. Therefore, we combined all unique points from the points table (7,661) and unique points from the power table that were not present in the points table (29) for a total of 7,690 unique points.

The raw points table included a unique point ID, however 256 of 7,690 (3.33%) points were missing this identifier. Without a unique point identifier, matching the point data to the power data in the raw data requires matching on a combination of the HEMS site ID and the text field labeled 'regname' (register name). To simplify file joins and ensure each point had a unique identifier, a new numerical point ID was added to both the points and power tables for easier mapping between tables. This also reduces the power data file size by replacing the register name string column with the numerical point ID, which is more efficient to store and to query. The original point ID was retained in a point_id_original column.

We also filtered out "low power" points, which were points where the absolute value of the power readings never exceeds 1 Watt. In total, 132 of 7,690 points (1.7%) were filtered out for having consistently low power.

The team also incorporated cleaned end-use type labels that were provided in the raw HEMS Point Alt table. These updated labels were created by NEEA based on previous analysis and professional judgement. In total, 9% of total points had an end use type label change between the raw label (circuit_label_type_desc) and the cleaned label (CIRC_TYPE). These different end use type labels were mapped to two different versions of the end use type label (01_raw and 02_cleaned) in the point end use type table. 8% of total points also changed category (e.g., ‘HVAC’ to ‘Other’). A summary of all end-use type label changes is found in Appendix 1.1.

Table 11 shows the count of points for each end-use type label using the cleaned labels.

Table 11. Summary of Points by End Use Type Label

End Use Type Label	End Use Category	Point Count	Percent of Total
Other	Other	3,471	45.1%
Stove/Oven/Range	Appliance	429	5.6%
Clothes Dryer	Appliance	399	5.2%
Mains	Mains	375	4.9%
Clothes Washer	Appliance	324	4.2%
Dishwasher	Appliance	317	4.1%
Refrigerator/Freezer	Appliance	295	3.8%
Electric Resistance Storage Water Heater	DHW	229	3.0%
Electric Baseboard Heater	HVAC	199	2.6%
Microwave	Appliance	195	2.5%
Garbage Disposal	Appliance	178	2.3%
Electric Furnace	HVAC	178	2.3%
Other Zonal Heat	HVAC	170	2.2%
Ducted Heat Pump	HVAC	149	1.9%
Central Air Conditioner	HVAC	139	1.8%
Gas Furnace Component	HVAC	135	1.8%
Ductless Heat Pump	HVAC	95	1.2%
Hot Tub	Other-Large	77	1.0%
Pump	Other-large	65	0.8%
Heat Pump Water Heater	DHW	65	0.8%
Solar	PV	31	0.4%
End Use Load Research Box	Null	30	0.4%
Mains With Solar	Mains	28	0.4%
Sub Panel	Other	27	0.4%
Central Vacuum	Appliance	25	0.3%
Electric Vehicle Charger	EVSE	17	0.2%

Electric Instantaneous Water Heater	DHW	15	0.2%
Energy Recovery Ventilator	HVAC	8	0.1%
Ignitor	Other	8	0.1%
Room Air Conditioner	HVAC	6	0.1%
Gas Instantaneous Water Heater	DHW	4	0.1%
Other Large Load	Other-Large	4	0.1%
Other With Solar	Other	2	0.0%
Packaged Terminal Air Conditioner	HVAC	1	0.0%

Only a little more than half of labeled points were assigned a specific end-use, with 45% of points in the NEEA cleaned data have an end use type label of ‘Other’. This makes sense as a significant portion of home electrical circuits feed wall outlets where various end uses could be plugged in.

2.5.4 Power Data

To simplify table joins, the new point ID was added to the power table. The original point ID was retained in the point data for mapping back to the raw HEMS data.

The primary cleaning of the power data was the removal of duplicates. In total, 150,078 duplicate power readings were identified based on point_id and start_time. All duplicates were either identical power values or rounding errors. All duplicates were removed by keeping the first instance of the duplicate in the data.

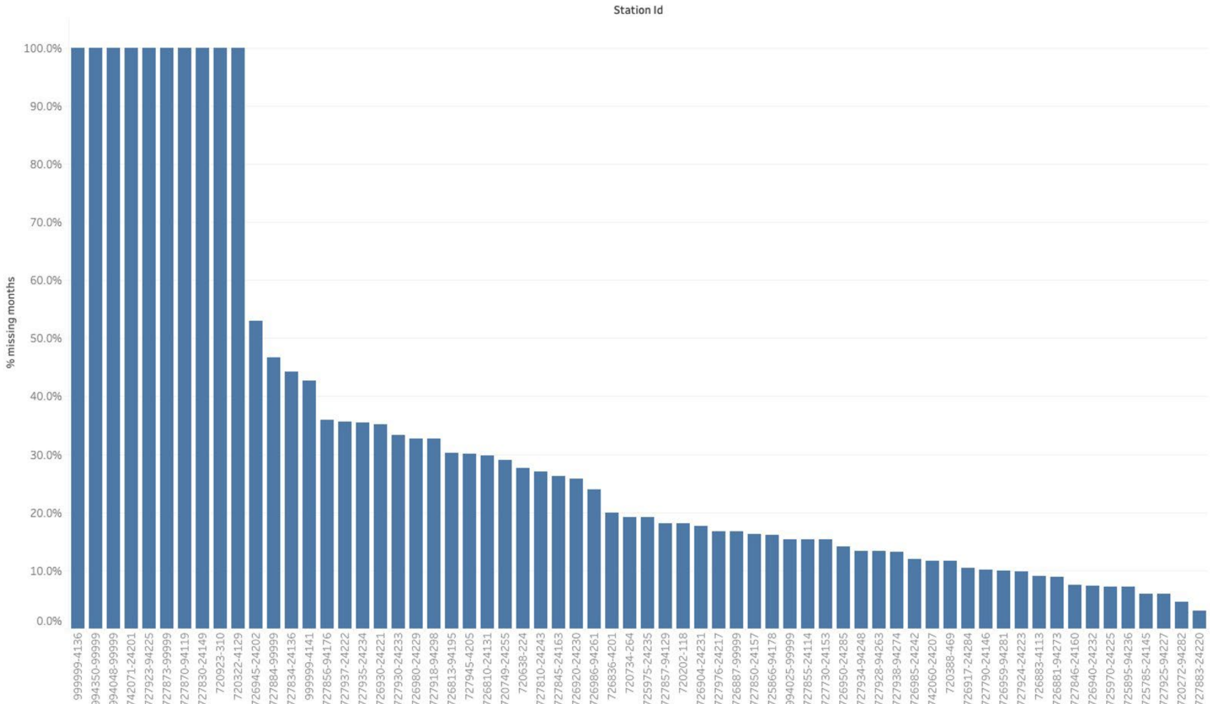
2.5.5 Weather Data

As noted above, the team pulled new temperature data for the weather normalization process. The primary reason for this was that we found significant gaps in the HEMS weather table.

1. For 10 of 64 stations (16%) all months were missing more than 10% of intervals, which was the cutoff for considering the data usable for weather normalization
2. For 30 of 64 stations (47%) more than 20% had data unusable for weather normalization
3. September 2024 was missing ~90% of expected values across all stations
4. The data also contained duplicate values

Figure 15 shows the % of months missing more than 10% of temperature intervals for each weather station in the original HEMS data.

Figure 15. Percentage of Months with Insufficient Data from HEMS Weather Table



2.5.6 Site Residuals

As a tool to potentially help identify significant possible missing points, the team calculated energy residuals based on the difference between mains power readings and sum of all points. A snapshot of this analysis from the project dashboards is shown Figure 16. This chart has the points with highest residuals sorted to the top, so shows the worst-case examples with the highest residuals. This figure uses the 12 months of data used for normalization, when available. For points with insufficient data for normalization, all available months are included.

Figure 16. Power Residuals Between Mains and the Sum of All Points

There were 394 sites that had mains metering data that allowed us to calculate residual values. Of those, 227 had a <10% weighted average difference between the sum of channels and the mains. There are 17 sites that have >100% difference; of these, 6 had a sum of mains power less than 2,000 kWh, which is likely an error. Of the remaining 11 of those sites, 7 actually had a larger sum of circuit-level energy than mains energy. The 150 sites with between 10% and

100% residuals are shown in the following two plots. 8 shows that most of this remaining set have between 10% and 50% residuals; these likely have some circuits that were not metered. Those with negative residuals, meaning that the sum of circuit-level energy is less than the mains energy, seem suspect. In 9, we can see that the most extreme cases are those with either very low total energy in the mains channel or very high amounts of mains energy. Such cases would be best addressed with follow-up field work to determine if the values are accurate or not.

Figure 17: Histogram of Site Residuals

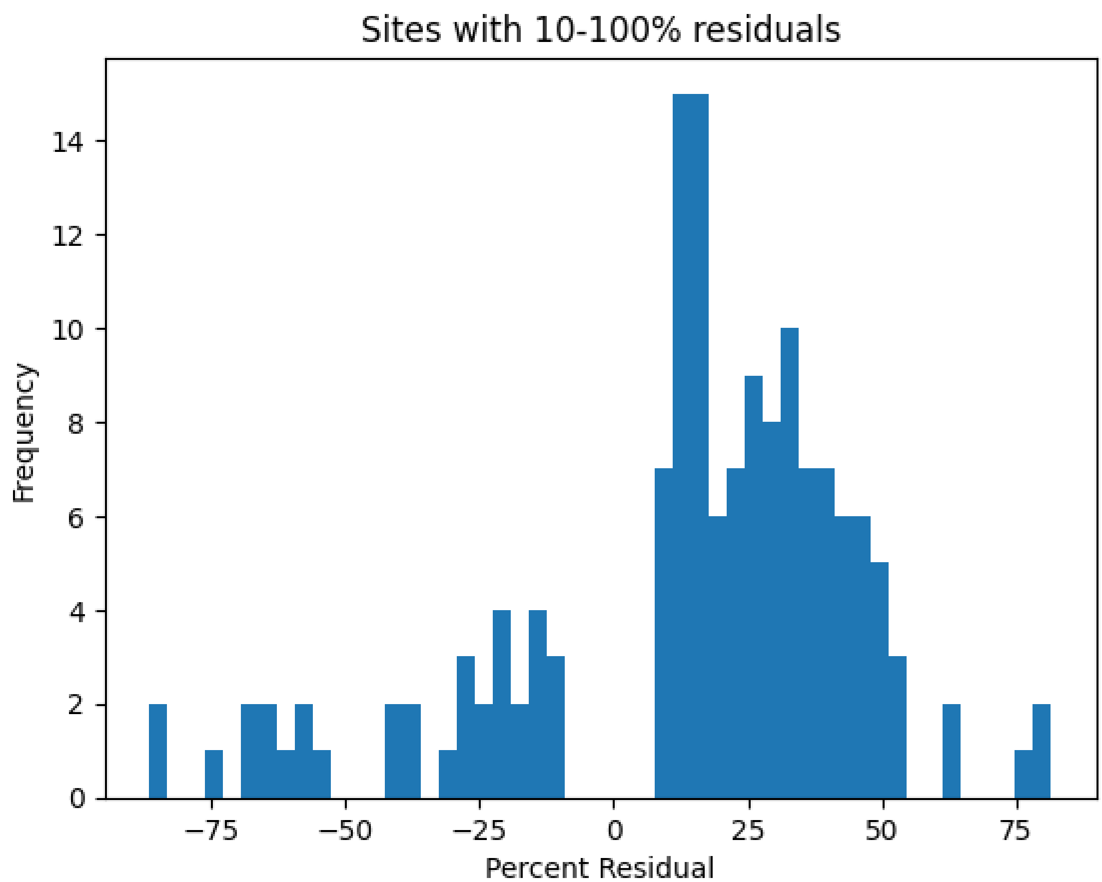
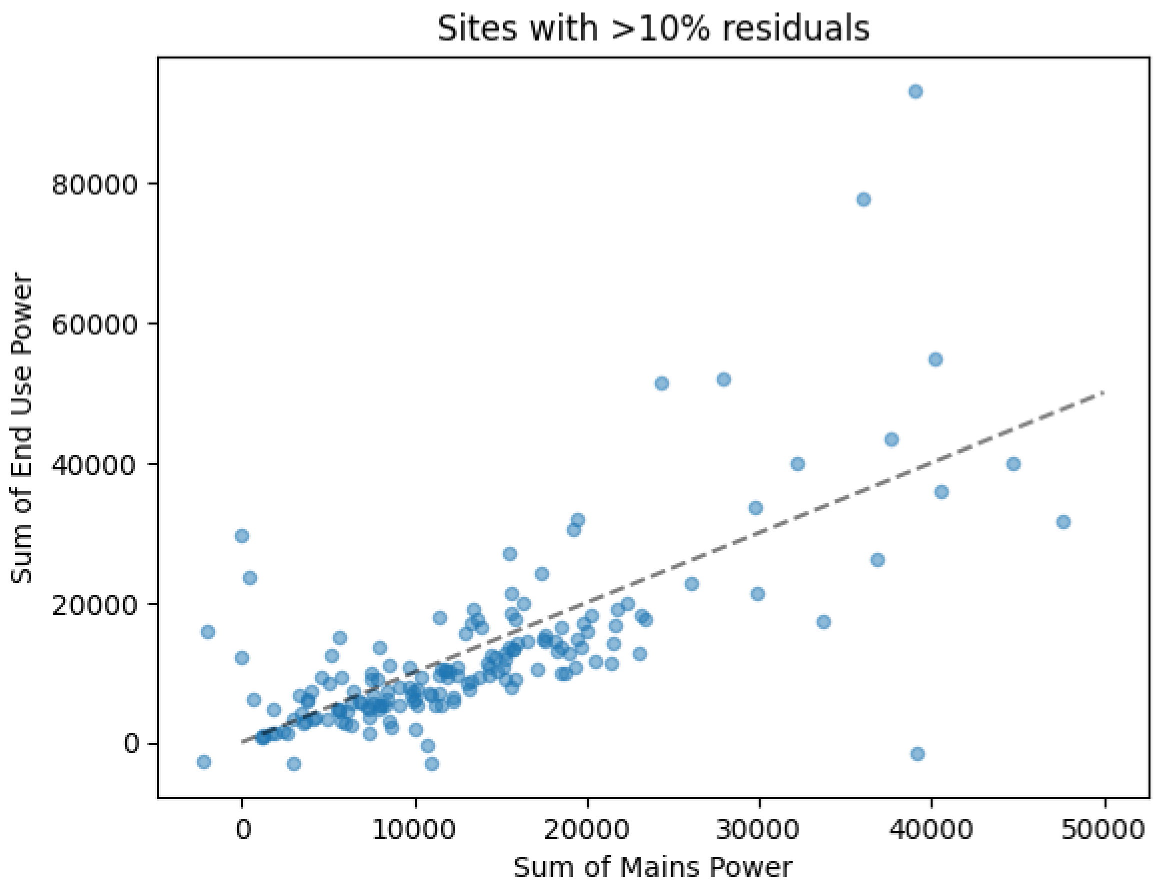


Figure 18: Scatter Plot of Site Residuals



3 Statistical Analysis of Electrification & Peak Loads

This section provides detailed methodology for the statistical analysis of electrification and peak loads. The overview of findings from this work are described in Section 1.

The statistical analysis of electrification and peak loads required the following steps described in this section:

1. Identifying regional peak periods
2. Development of future temperature scenarios
3. Development of future electrification scenarios
4. Development of bins for weighting
5. Development of peak period profiles
6. Peak load impact calculations

3.1 Identifying Regional Peak Periods

To support the statistical analysis of peak loads, the team needed to identify relevant peak load periods. Through discussions with NEEA, it was determined that there were three peak load periods of interest:

1. Summer Evenings
2. Winter Mornings
3. Winter Evenings

To determine the specific hour ranges for these periods, the team took system load data from BPA's annual power system reports for FERC[9]. This data includes hourly intervals of system load for the BPA service territory. While not a perfect fit to the NEEA territory, this data was determined to be the most representative data for the region to determine peak periods.

The analysis included the following steps:

1. Five years of BPA data was taken from 2020 - 2024 and combined in a standard format.
2. Time intervals were converted to Pacific Prevailing Time (PPT)
3. Each interval datetime was labeled with the following initial definitions. These are the initial ranges where the peak periods were expected to land.
    a. Season
        i. Summer: June - September
        ii. Winter: October - May
    b. Day Periods
        i. Morning: 12:00 am - 11:00 am
        ii. Evening: 12:00 pm - 11:00 pm
4. Day type (weekday/weekend) and holidays were filtered from the data[10].
5. For each year, the five days with highest peak load for the three relevant peak load periods (Summer Evenings, Winter Mornings, Winter Evenings) were determined.
6. Hourly data from the 25 peak days (five days for each of the five years) was averaged to get an average hourly load curve for these peak days.
7. The four consecutive hours with the highest average load from the peak days were taken to be the peak load period hours for that period.

3.2 Development of Future Temperature Scenarios

To estimate how climate trends could impact the expected loads of different end uses in the study, the team used the OpenDSM EEmeter models to forecast the expected energy use for hypothetical "typical" weather data that was constructed by taking the TMYx data and shifting every reading hotter or colder by a consistent offset. For future temperature sensitivity modeling, the team took the 2021 heat dome as a reference point for extreme high temperatures. By

[9] https://transmission.bpa.gov/Business/Operations/FERC714/default.aspx

[10] Federal holidays were taken from the Office of Personnel and Management (https://www.opm.gov/policy-data-oversight/pay-leave/federal-holidays/#url=Overview), but inauguration day was excluded for 2021.

comparing regional temperature highs from 2021 to the normalized TMYx data, it was determined that the 2021 peak was roughly 15 degrees Fahrenheit above average. To capture this variability, the team decided to do temperature scenarios for +/- 15 degrees Fahrenheit relative to the TMYx data set. To limit the total number of scenarios, the team decided to do steps of +/- five degrees, so the final temperature scenarios are -15, -10, -5, +5, +10, +15.

3.3 Development of Future Electrification Scenarios

Electrification scenarios were developed to analyze the impact on peak-period regional grid loads from the future adoption of electrified weather-sensitive equipment (heating, cooling, water heating). These scenarios are designed to estimate the likely net hourly impact that would be caused by different adoption rates of ducted heat pumps, ductless heat pumps and heat pump water heaters by 2035.

Electric vehicle charging is also an important driver of new load in the Northwest, but Electric Vehicle Supply Equipment (EVSE) were excluded from the peak load analysis and determined to be out of scope for this project due to the low number of points (17) in the data set. It was determined that these were too few to meaningfully fill bins when using the original binning strategy based on heating zone, vintage and square footage. However, with the revised approach of binning power data normalized on square footage by heating zone only, these devices could potentially be analyzed using similar strategies. This data set still represents an extremely small sample, given the expected diversity of EV driving and charging behaviors, vehicle charging capacities and charging speeds. There are also multiple other larger studies of EV charging impacts on the grid, so the decision was made to remain focused on heat pumps.

The scenarios were split into retrofit replacements and new construction. Impacts from retrofits are calculated as the difference between the baseline load shape and the replacement equipment load shape. New construction additions were considered new loads and did not require any baseline condition. Note that this analysis is not intended to reflect any sort of attribution to incentive programs or assumptions about any counterfactual replacement equipment. The baseline and retrofit load profiles were developed by analyzing the HEMS data set and binning each site into different heating, cooling or water heating profiles based on the blend of end-use loads present in the HEMS data.

3.3.1 Heating Retrofit Scenarios

For retrofit replacements of heating, individual homes often have a mix of different heating equipment because of modification to the home over time. To address this complexity, the team binned homes from the HEMS data into “profiles” based on the different combinations of baseline heating equipment.

The first step was filtering out points with low power usage to account for the fact that some heating systems components seem to be unused. For example, heat pumps may be installed to replace electric resistance, but the original resistive heaters are left in as a backup but rarely if ever used. To account for this, the team filtered out any heating equipment where the average seasonal heat usage for winter was less than 20 watts.

After filtering these points, the HEMS homes were binned into simplified categories based on the heating end-uses present at the site. This also allowed us to group various combinations of electric resistance heat into a single profile when appropriate. For this analysis, categories containing heat pumps were segmented by ducted and ducted systems. Descriptions of the heating equipment profiles are provided in Table 12.

Each retrofit scenario describes the transition from a baseline profile to a retrofit profile. For the electric resistance to ductless heat pump scenario, we used the ductless heat pump with electric resistance profile for the retrofit, which excludes all the sites where only ductless heat pump loads were detected. This generally produces a conservative result that may overestimate the impacts since it assumes the baseline "backup" system will always be used at least some of the time. As a future exercise, the site-level profiles for "ductless heat pumps *only"* could be blended with the profile for "ductless heat pump plus electric resistance" to show how the net load could be lower if some of the retrofitting sites were able to fully eliminate all resistance backup heat.

Table 12. Heating Profile Descriptions

Heating Equipment Profile	**Site Heating Equipment Description**
Ducted Heat Pump and Resistance Backup	Ducted heat pumps only and ducted heat pump plus electric furnace
Ductless Heat Pump	Ductless heat pump only
Gas Furnace	Gas furnace only and gas furnace plus Energy Recovery Ventilation (ERV)*
Electric Resistance	Electric resistance heat (electric furnace, baseboard electric, or other zonal heat) only
Ducted Heat Pump and Gas Furnace	Ducted heat pumps and gas furnace**
Ducted Heat Pump and Electric Resistance***	Ducted heat pumps and electric resistance
Ductless Heat Pump and Electric Resistance	Ductless heat pumps and electric resistance

* The baseline electricity use of fossil-fueled heat is from "furnace components" such as blower fans and air handlers

** No corresponding profile for ductless heat pump and Gas Furnace was created due to having only four sites in the HEMS data sample and missing data amongst those points

*** This profile was not used for analysis

Table 13 includes the baseline scenarios used for heating equipment.

Table 13. Heating Baseline and Retrofit Profiles

Scenario	Baseline	Retrofit Profile
Electric Resistance to Ducted Heat Pump	Electric Resistance	Ducted Heat Pump and Resistance Backup
Gas Furnace to Ducted Heat Pump	Gas Furnace	Ducted Heat Pump and Gas Furnace
Electric Resistance to Ductless Heat Pump	Electric Resistance	Ductless Heat Pump and Electric Resistance
Gas Furnace to Ductless Heat Pump	Gas Furnace	Ductless Heat Pump*

*Ductless Heat Pump and Gas Furnace was not used as the retrofit profile due to a low prevalence of sites in the HEMS data.

For each scenario, NEEA developed low, medium, and high estimates of the adoption rates for each replacement scenario. Each scenario is based off a set of assumptions. The estimated current market role was provided by the load forecasting team at the Northwest Power and Conservation Council. Adoption rates, replacement rates, and future market penetration rates were estimated using respective technology data from the RBSA when possible. When data was not available, professional opinions of market trends from NEEA staff were utilized. The scenario estimates are found in Table 14. Note that the "subcategory row" percentages representing, to take the first section of Table 14 as an example, what portion of homes are installing ducted heat pumps to replace electric resistance, are a percentage of the population that installed ducted heat pumps in that scenario. Therefore, those subcategories should always sum to 100%. Using the same example, the "top-level" percentages of households that installed ducted heat pumps are applied to the entire population of residential sites.

Table 14. Heating Adoption Rate Scenarios

Type	Adoption Scenario	Current	Low	Medium	High
Ducted Heat Pump Replacements	% of households that installed ducted heat pumps	18%	24%	30%	40%
	- Replacing electric resistance	65%	65%	65%	65%
	- Replacing gas	35%	35%	35%	35%
Ductless Heat Pump Replacements	% of households that installed ductless heat pumps	8%	19%	30%	40%
	- Replacing electric resistance	65%	65%	65%	65%
	- Replacing gas	35%	35%	35%	35%

3.3.2 Cooling Retrofit Scenarios

The cooling methodology was the same as for heating, but the team filtered out any cooling equipment where the average seasonal cooling usage for summer was less than 5 watts. There were four cooling equipment profiles developed, as found in Table 15.

Table 15. Cooling Profile Descriptions

Cooling Equipment Scenario	Site Cooling Equipment Description
Central Air Conditioner	Central Air Conditioner and ERV
Room Air Conditioner	Room Air Conditioner only
Ducted Heat Pump	Ducted heat pumps only
Ductless Heat Pump	Ductless heat pumps only

For replacements, ducted heat pumps were expected to replace only legacy central A/C systems at a certain percentage, while ductless heat pumps were expected to only replace room A/C. In addition, some homes are assumed to start with no cooling, and so the addition of a heat pump represents new load. Table 16 summarizes the baseline scenarios used for cooling equipment.

Table 16. Cooling Baseline and Retrofit Profiles

Scenario	Baseline	Retrofit Profile
Central Air Conditioner to Ducted Heat Pump	Central Air Conditioner	Ducted Heat Pumps
No Cooling to Ducted Heat Pump	None	Ducted Heat Pumps
Room Air Conditioner to Ductless Heat Pump	Room Air Conditioner	Ductless Heat Pumps
No Cooling to Ductless Heat Pump	None	Ductless Heat Pumps

For each scenario, NEEA developed estimates of both the adoption rates and replacement scenarios for low adoption, medium adoption and high adoption scenarios. These are found in Table 17.

Table 17. Cooling Retrofit Adoption Rate Scenarios

Type	Adoption Scenario	Current	Low	Medium	High

Ducted Heat Pump Replacement	% of households that installed ducted heat pumps	18%	24%	30%	40%
	- Replace legacy central air conditioner	40%	40%	40%	40%
	- No baseline air conditioning	60%	60%	60%	60%
Ductless Heat Pump Replacements	% of households that installed ductless heat pumps	8%	19%	30%	40%
	- Replacing window air conditioner	65%	65%	65%	65%
	- No baseline air conditioning	35%	35%	35%	35%

3.3.3 Water Heating Retrofit Profiles

For water heating, there were three equipment scenarios considered, as found in Table 18. Instantaneous (e.g., “tankless”) water heaters were excluded from this analysis due to a low prevalence of units in the HEMS data.

Table 18. Water Heating Profile Descriptions

Water Heating Equipment Scenario	Site Water Heating Equipment Description
Electric Resistance Storage Water Heater	Electric resistance storage water heating only
Heat Pump Water Heater	Heat pump water heating only
No Electric Load	Water heating with no electric load (e.g., gas)**

** The HEMS dataset did not include enough data to determine baseline usage for non-electric water heating, so it is assumed to be zero, based on our understanding of the technology.

For heat pump water heaters, a retrofit replacement was assumed to be a one for one replacement of an existing water heater. Table 19 summarizes the baseline scenarios used for water heating equipment.

Table 19. Water Heating Baseline and Retrofit Profiles

Scenario	Baseline	Retrofit Profile
Resistive to Heat Pump Water Heater	Electric Resistance Storage Water Heater	Heat Pump Water Heater
Gas to Heat Pump Water Heater	No electric load	Heat Pump Water Heater

For each scenario, NEEA developed estimates of both the adoption rates and replacement scenarios for low adoption, medium adoption and high adoption scenarios. These are found in Table 20. Replacements were assumed to be evenly split between fossil-fueled water heaters (baseline energy usage of zero) and resistive water heaters.

Table 20. Water Heating Adoption Rate Scenarios

Adoption Scenario	Current	Low	Medium	High
% of existing households with Heat Pump Water Heater	3%	10%	30%	50%
- Replace resistance water heater	50%	50%	50%	50%
- Replace fossil-fueled water heater	50%	50%	50%	50%

3.3.4 New Construction Scenarios

For new construction, there were three equipment scenarios considered, as found in Table 21. Note that new construction scenarios for heat pumps provide both new heating and cooling load and so will contribute both to summer and winter peak analysis

Table 21. New Construction Profile Descriptions

New Construction Equipment Scenario	Energy Use Profile
New Ducted Heat Pump	Ducted heat pump and resistance backup
New Ductless Heat Pump	Ductless heat pumps only
New Heat Pump Water Heater	Heat pump water heating only

For new construction, there is no baseline scenario, as these are considered purely new load added to the system. Table 22 summarizes the baseline scenarios used for new construction equipment.

Table 22. New Construction Baseline and Retrofit Profiles

Scenario	Baseline	Retrofit Profile
New Ducted Heat Pump	None	New Ducted Heat Pump
New Ductless Heat Pump	None	New Ductless Heat Pump
New Heat Pump Water Heating	None	Heat Pump Water Heater

For each scenario, NEEA developed estimates of the number of new construction homes and the adoption rates for low adoption, medium adoption and high adoption scenarios. These are found in Table 23.

Table 23. New Construction Adoption Rate Scenarios

Adoption Scenario	Current	Low	Medium	High
Number of new construction households	0	77,420	77,420	77,420
% installing ducted heat pump	42%	42%	52%	62%
% installing ductless heat pump	23%	23%	28%	33%
% installing heat pump water heater	7%	10%	30%	50%

Note: for new construction, values represent the percent of new construction homes installing the technology

3.4 Development of Bins for Weighting

We explored a number of possible binning attributes from the RBSA population weighting table provided by NEEA, including vintage, home size, and heating zone (HZ). Figure 19 Shows the different heating zones in the Northwest by zip code with mostly coastal regions being in heating zone one and further inland regions mixed between heating zones 2 and 3.

Figure 19. Northwest Heating Zones by Zip Code

Source: RTF Climate Calculator https://rtf.nwcouncil.org/work-products/supporting-documents/climate-files/

While the HEMS study population contains a robust sample across all of the bin combinations, when defined by the three binning attributes, many of the site-level HVAC system types would not be well-represented in all the bins. Figure 20 provides a Histogram of homes per bin showing that over 20 bins would have less than 10 homes in the bin.

Figure 20. Distribution of Bin Counts

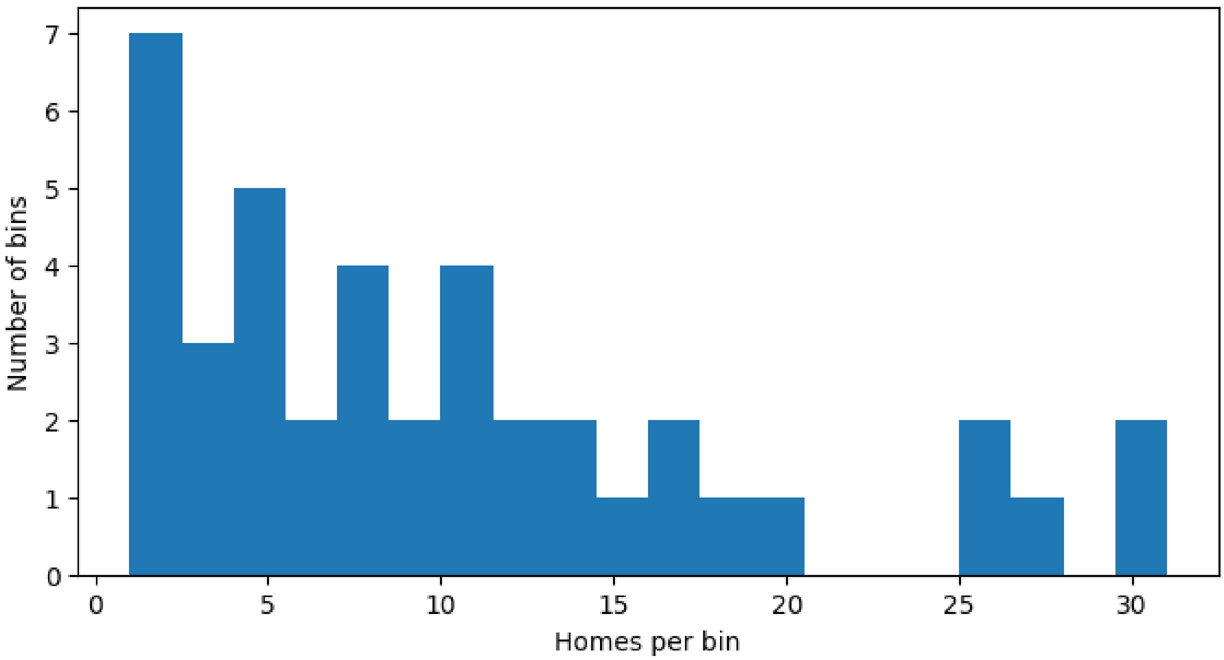
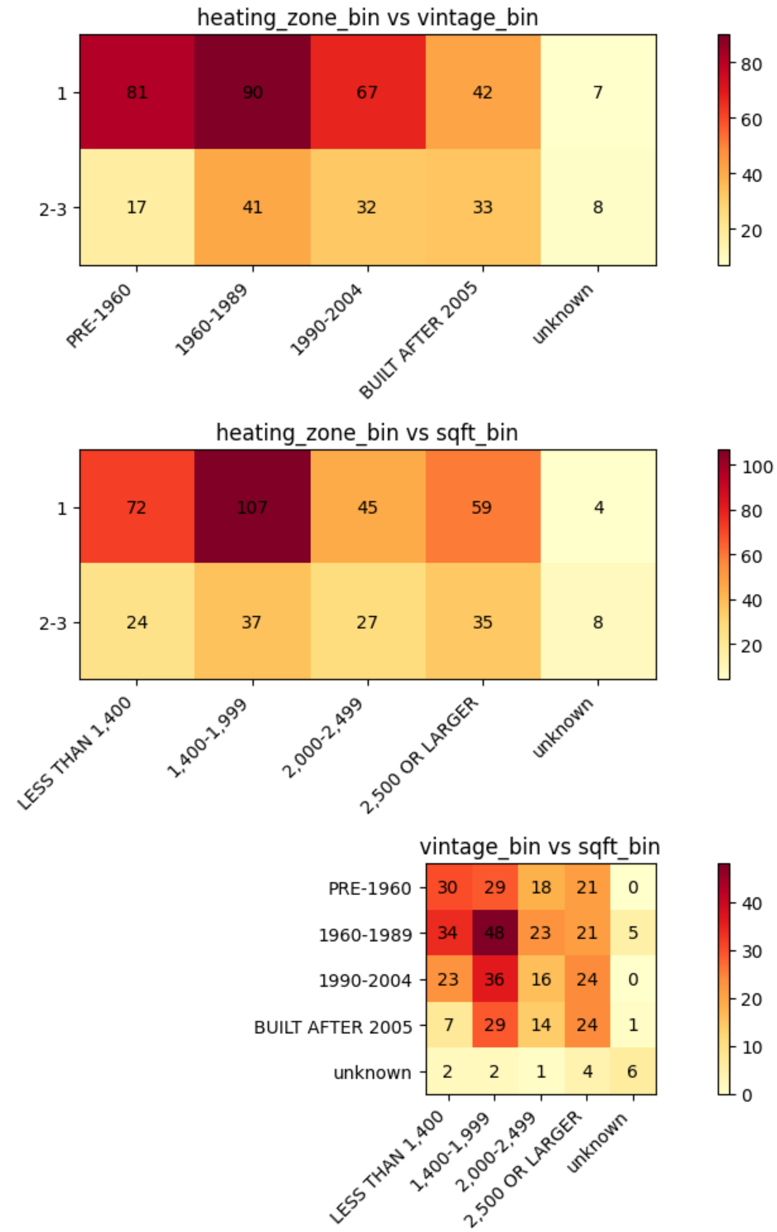


Figure 21 shows the bin counts by attribute. Each graphic shows the count of sites when looking at two of the attributes against each other. While some counts here seem potentially adequate, the actual bin counts would be less when sliced by the third attribute. Also, even 20 homes per bin is often too few to get good results when looking at the representation of end-uses within those homes. Most importantly, any given end use type will only be found in a small subset of sites, so these counts represent an upper bound for the feature of interest.

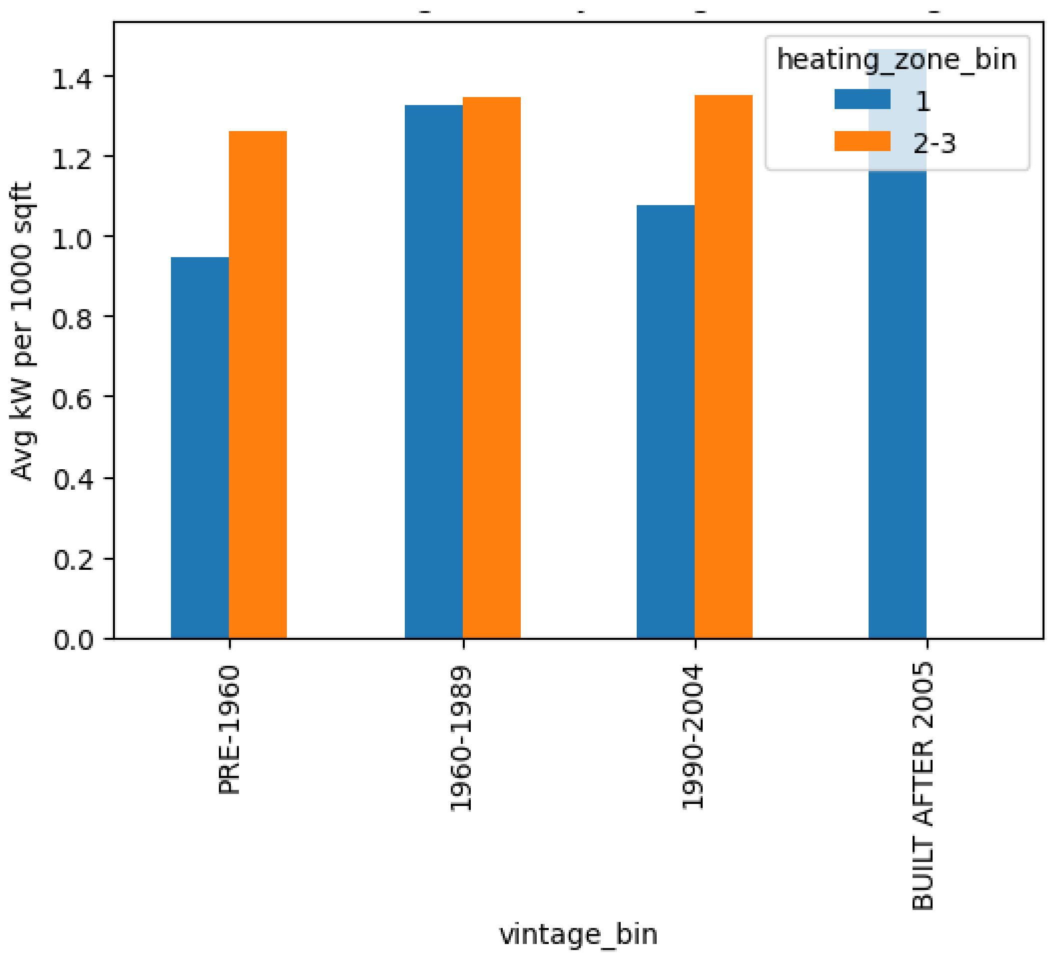
Figure 21. Bin Counts by Bin Attributes



We analyzed the correlation between each attribute and site HVAC load, using the site-level weather-normalized energy use per square foot measurements for the sites with electric resistance heating systems only as a proxy for the heating energy demand. While we expected to see a difference between newer and older homes, there was not a meaningful correlation in this data set, as shown in Figure 22. Also, 31 sites (7% of the total sites) were missing vintage, so those sites would have been dropped or added to an “unknown” category. Weighting by the square footage bins was unnecessary, since we are already normalizing by that metric. Note that 12 sites were missing square footage data and were excluded from this analysis. We did see a small correlation between heating zone and heating load, and we combined HZ 2 and HZ

3 into a single bin due to the small number of sites in each of those bins. We understand this approach has also been taken with other NEEA projects, so it should still be possible to cross-reference the results of this study with prior NEEA research.

Figure 22. Winter Electric Heating Loads by Vintage and Heating Zone Bin



3.5 Development of Peak Period Profiles

Peak period load profiles were calculated to determine the load contribution of each end use during the peak period. These profiles were calculated for each site level end-use equipment scenario. This was done separately for heating and cooling.

First, site-level system type was determined by creating flags that indicate what type of heating and cooling systems are present for each site. Then each unique site-level combination of heating and cooling end uses was mapped to a site-level system type. Points with low usage (average power for the season was <= 20w for heating or <= 5w for cooling, as points below these thresholds were observed to have effectively no activity consistent with HVAC equipment) were excluded from the determination of site-level system type, as this reduced the possible combinations and these low power points would have little impact on the energy use profile.

Some rare combinations that were not easily classified were marked NA and not included in subsequent analysis.

Next, the individual circuit level loads for the selected heating and cooling systems were aggregated at a site level during the peak periods and divided by the site square footage to get the heating and cooling profile for each site and end use type calculated per square foot. These site level profiles were then multiplied by the average square footage of sites in each heating zone to get a per-site profile by heating zone. To get a final, single, profile for the region, the heating zone profiles were averaged together, weighted by the total number of residences in each heating zone. Table 24 provides the values used for this site weighting

Table 24. Site Weighting Data by Heating Zone

Heating Zone	Site Count	Average Square Footage
1	3,438,363	1,889
2-3	1,264,748	2,018

Finally, a variety of statistics are calculated at the site level for the total loads by end use type, for each hour in peak load profile, including:

a. Mean
b. Median
c. Standard Deviation
d. Quartiles (25^{th} and 75^{th} percentiles)

3.6 Peak Load Impact Calculations

The peak load impact calculation involves first computing the total number of households currently in the region. This was found to be approximately 4,700,000 by summing the population weights from the NEEA HEMS weights data. Retrofits are considered changes to the existing installed base, so this number was used as the basis for the retrofits completed by 2035 (i.e. it was assumed that new construction homes built between now and 2035 were assumed not to be upgraded in this time).

To get the number of households completing upgrades, the team took the three adoption scenarios (low, medium and high) for the percent of homes that adopt this equipment and subtracted the current percentage to get the percent of total homes doing the retrofit between now and 2035 for each of the adoption scenarios.

We then multiplied these percentages by the total number of homes to get the number of households for each combination of equipment scenarios (e.g., Electric Resistance to Ductless Heat Pump) and adoption scenarios (high, medium, low).

For each equipment scenario, the peak period profile for the baseline site-level equipment type was subtracted from the peak period profile for the retrofit site-level equipment type and then multiplied by the total number of homes in that equipment scenario to get the total contribution of that equipment scenario to changes in load during the peak period. This calculation is performed for the three different assumed adoption rates to get a low, medium and high impact curve. The curves for each adoption scenario can then be added to get the estimated total contribution to future load change from these technologies.

3.7 Weather Station Selection

After filtering for 'high' quality weather stations with TMYx 2009-2023 data available, the team performed an analysis to understand the difference between tightening distance and/or elevation differences between the site and the selected weather station. The team picked two scenarios to test as indicated Table 25.

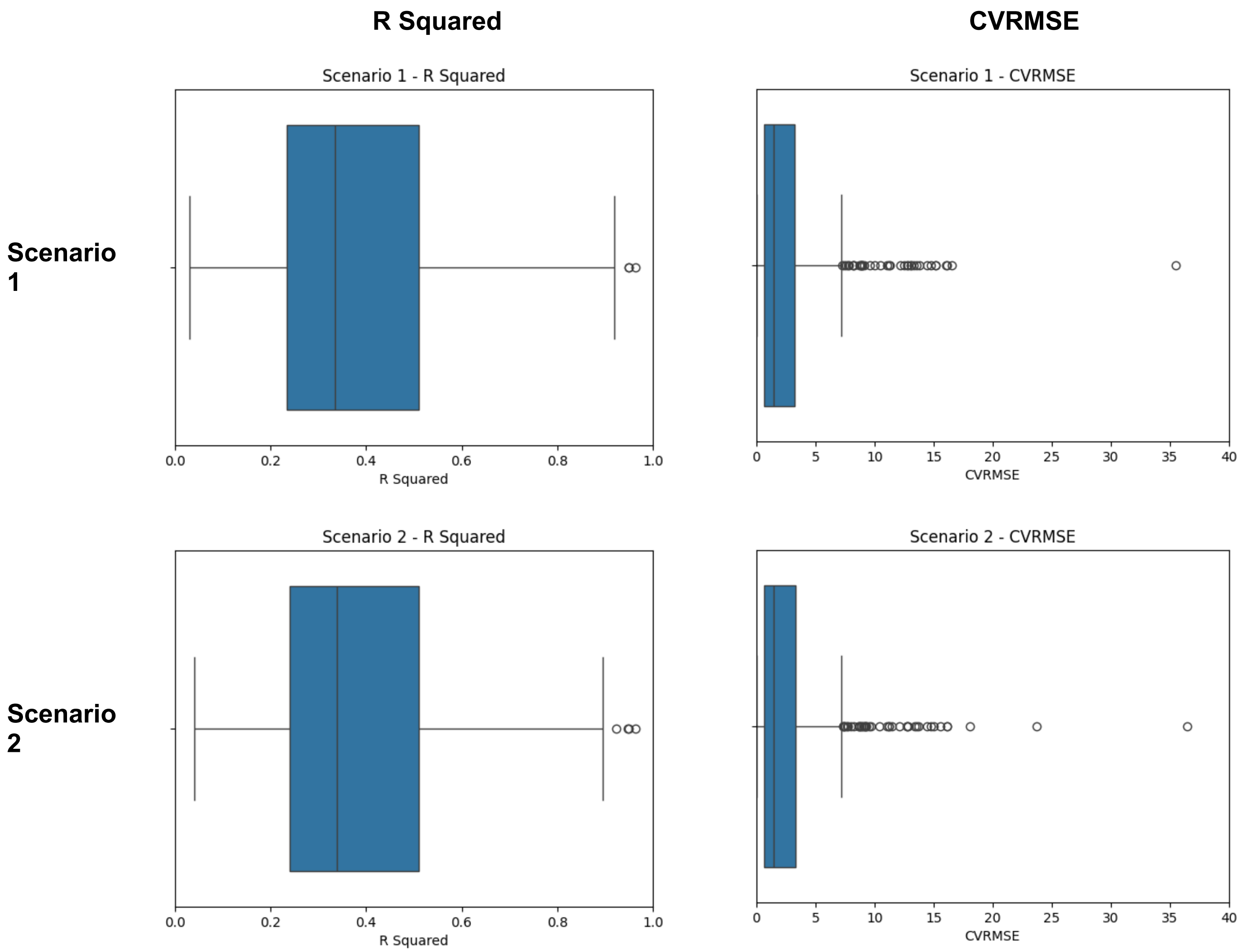
Table 25. Weather Station Selection Analysis Scenarios

	Scenario 1	**Scenario 2**
Distance Cutoff (km)	100	200
Elevation Cutoff (m)	300	100

After running the full pipeline for both scenarios, the team found 1,194 points from 64 HEMS sites where the selected weather station changed between the two scenarios. Of the 1,194 points, 850 points from scenario 1 and 872 points from scenario 2 had adequate power and temperature data coverage for modeling. This difference in modeled points was due to one site that did not have sufficient coverage of power and temperature data for a 12-month period when using the weather station selected in scenario 1. Note that this analysis was only performed on data through the end of 2024, as it was performed before 2025 data was available.

The team compared model accuracy across the two scenarios and found the median model accuracy using R-squared and CVRMSE metrics to be closely aligned. The median R-squared for Scenario 1 was 0.383 for scenario 1 and 0.385. The median CVRMSE for Scenario 1 was 2.38 for scenario 1 and 2.41 for scenario 2. Figure 23 shows the distribution of results for the two metrics across both scenarios. Note that the median value of the CVRMSE, represented by the line in the middle of the blue square in the figure, is much lower than the mean since the large outliers bring up the mean substantially.

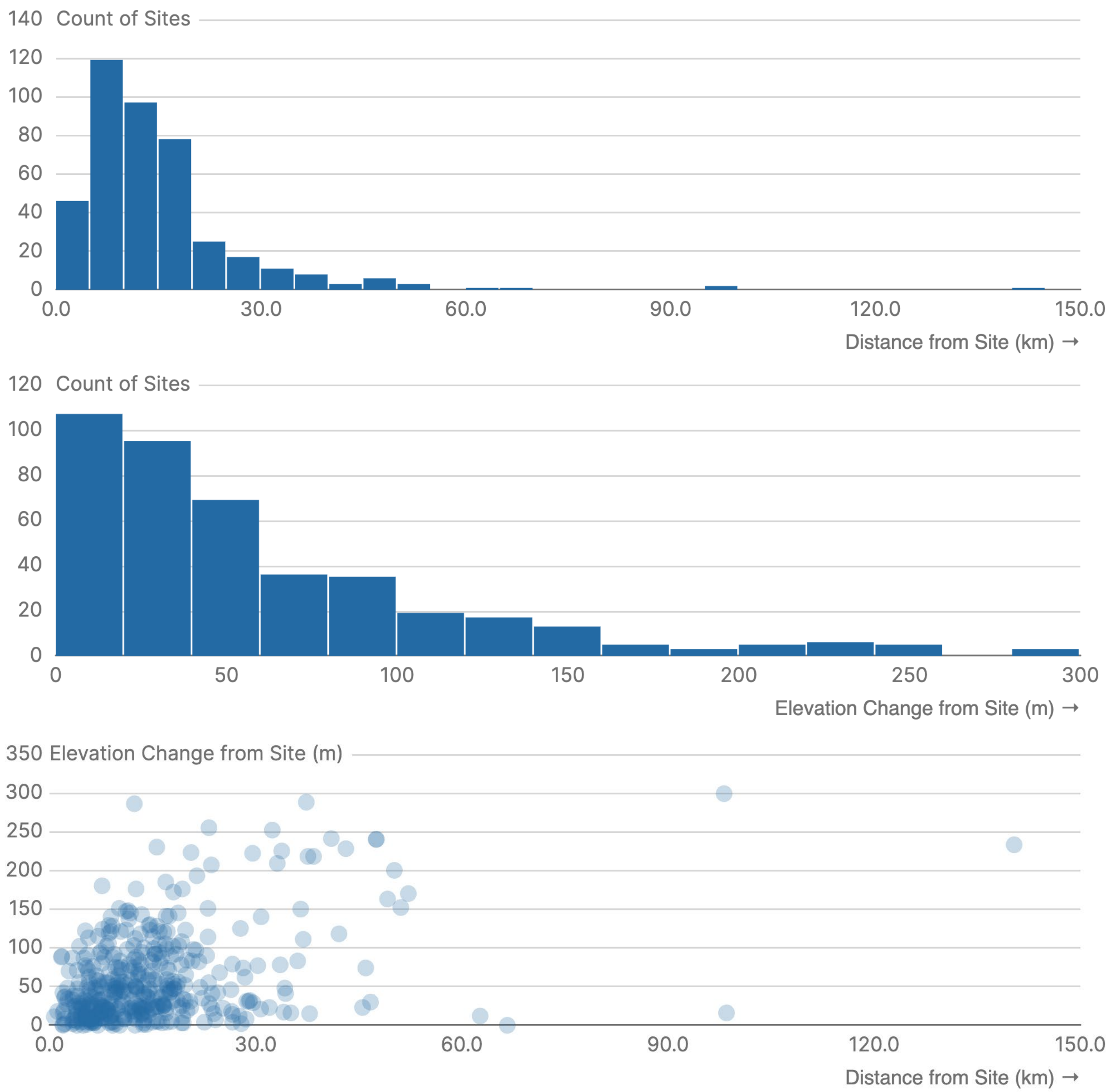
Figure 23. Model Accuracy Results for Different Weather Station Selection Criteria



As a result, the team determined that there was not a significant difference in model quality between tightening elevation difference and distance in weather station selection for weather modeling. The team then decided to use the base logic used by EEweather and set an elevation limit where all stations would have a match (<= 300m) and then select the closest weather station that met all the criteria, thereby prioritizing distance over elevation.

A breakdown of final weather stations that were selected for analysis is found in Figure 24.

Figure 24. Weather Station Distance and Elevation Change from Site



This selection technique resulted in:

1. Only 5 of 418 sites (1.1%) matched to stations over 60 km from the site
2. Only 30 of 418 sites (7.2%) matched to stations over 150 m elevation difference from the site.
3. 368 of 418 sites (88%) matched to stations that were within 150 m elevation difference and 30 km distance from the site.

3.8 Normalization and Modeling

3.8.1 Modeled Points

The first step in normalization was to determine if the data coverage of the power and temperature data was sufficient for each month. If more than 10% of intervals were missing for a month the team considered that month inadequate for modeling. This requirement is harmonized with the definition used by OpenDSM to trigger a warning for data completeness in its models.

In general, the data coverage was high with 90% of total months across the cleaned and filled points having <10% missing data for power and 95% of total months across the cleaned and filled points having <10% missing data for temperature. A total of 88% of total months across all points had adequate data for both power and temperature.

After attempting to identify a 12-month period that had sufficient data quality for both power and temperature, 996 of 7,690 points (13%) were filtered out before normalization because they did not meet the appropriate criteria. From these 996 points, there are 117 points from 6 sites that we lost despite having 12 months of contiguous power data and 12 months contiguous temp data, but not 12 contiguous points for both. While a different selection strategy could capture these points (such as searching for other weather stations with sufficient data), this was not considered for this analysis due to the low number of points and the added complexity of automating the analysis.

For the points with sufficient data for modeling, the team captured a variety of errors and warnings during the normalization process. In total, 190 of the remaining 6,517 points (2.9%) were not able to be modeled due to errors. 57 of 6,517 points (0.9%) had baseline disqualifications due to negative values in the baseline power reading data. These points all had the following end-use types: Solar, Mains, Mains With Solar, and Other With Solar. Negative power values are flagged by OpenDSM because the 'is_electricity' flag was set to False in order to interpret zeros as real data and OpenDSM does not expect negative readings for non-electricity data. While these points did have baseline disqualifications, they were successfully modeled anyway by setting the flag 'ignore_disqualification' to 'True' during modeling.

3.8.2 Modeling Accuracy

The team looked at two measures of model accuracy, the coefficient of determination (R-squared) and the Coefficient of Variation of the Root Mean Squared Error (CVRMSE). These two statistical metrics are commonly used to describe the model's "goodness of fit" and evaluate if a model is adequately capturing the relationship between independent and dependent variables. The R-squared is a value from 0 to 1.0 that is described as quantifying how much of the variation in the data is captured in the model. For example, if a weather-regression model predicted a linear fit between daily average temperature and energy use, and all the data points fell exactly along the model forecast line on a temperature-power scatter plot, that would represent an R-squared of 1.0 because the model fully describes the variation in the daily energy use data. R-squared is typically used to gauge explanatory power of a model. CVRMSE measures how far the model's predictions fall from the actual data, with larger penalties as the predictions get further from the right answer, which makes it very sensitive to outliers. A CVRMSE of 0.0 would be a perfect fit, but there is no limit to how large a CVRMSE can grow, though the raw RMSE score is normalized to the mean magnitude of the measurements. This measurement of the predictive error magnitude is useful for validating forecasting models, as it describes overall accuracy and bias.

For this analysis, we used the unadjusted R-squared and CVRMSE values. These are not adjusted for the number of samples and the number of predictors used in the model. While such adjustments may be beneficial for model selection, the unadjusted values are more transparent, as they report the raw error metrics in a transparent and interpretable fashion.

While there are a few R-squared and CVRMSE thresholds that are commonly used when evaluating models for use as counterfactual baselines, we did not directly apply any of them as a filtering criterion for the models we generated on these sub metered data sets. One reason is that the typical thresholds were developed for whole-home data, where the data represents the aggregate consumption of many appliances and the variations in less-predictable individual appliance use tends to be balanced out with more-predictable HVAC and always-on loads, forming a smoother overall trend with less variation relative to the total load. Note that Figure 25's R-squared visualization (the upper plot) indicates the 0.50 threshold for R-squared, which is commonly used as a quality metric for modeling. Note that we did not use this as a disqualification criterion.

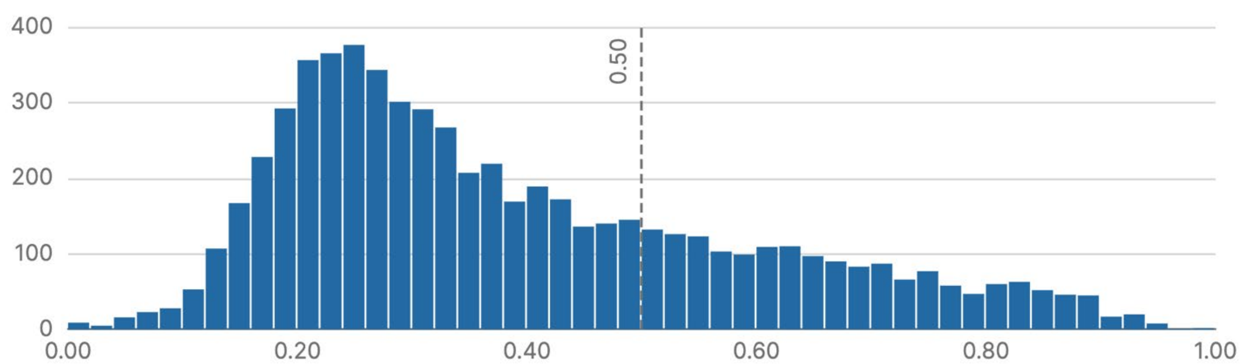
The average accuracies for the 6,327 modeled points were:

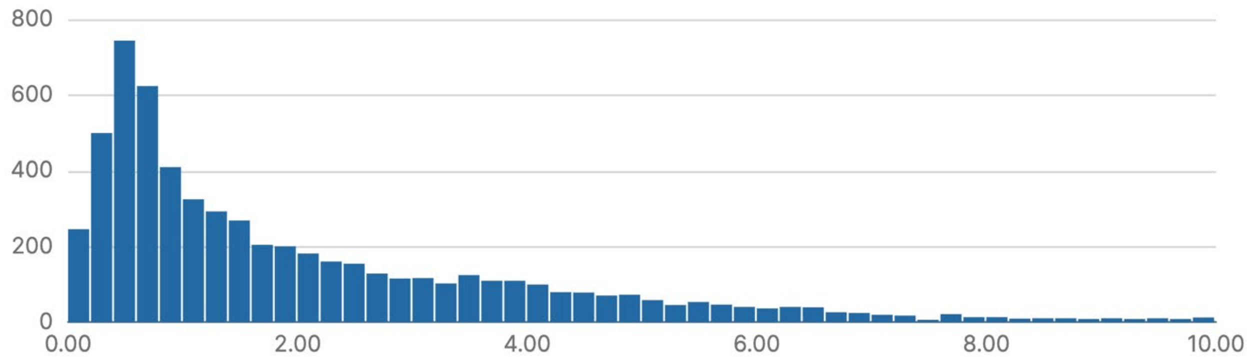
R-squared: 0.39

CVRMSE: 2.41

Figure 25 provides histograms of R-squared and CVRMSE for modeled points.

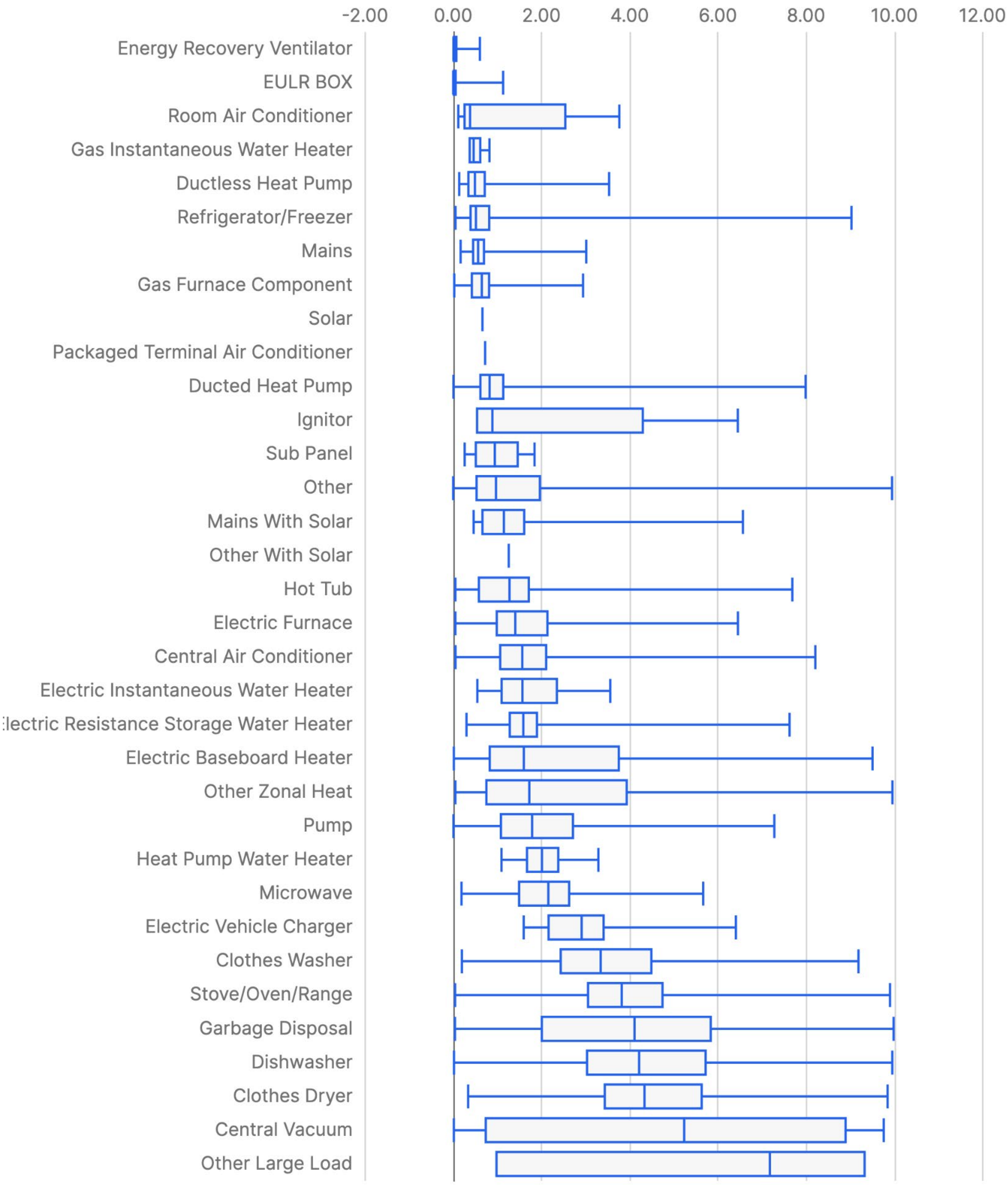
Figure 25. Count of Modeled Points by R-Squared and CVRMSE Value





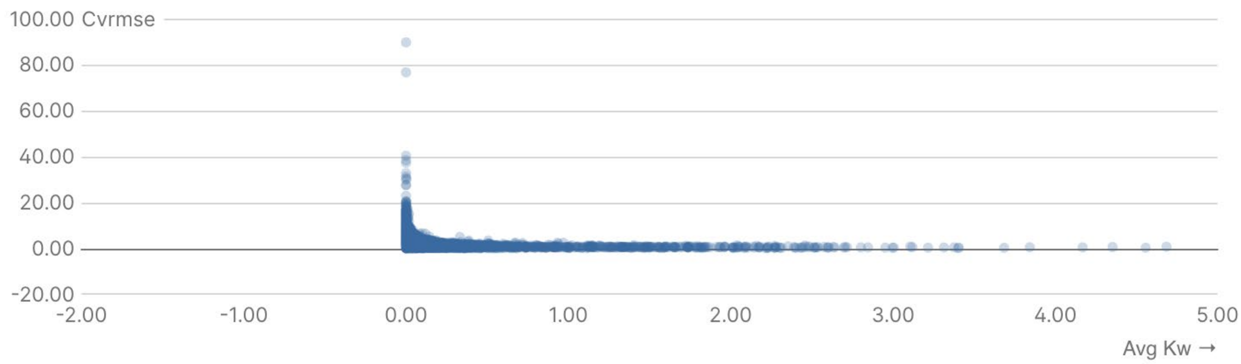
As shown in Figure 26, the distribution of CVRMSE scores varied widely by end use type of the point. The plot is sorted by median CVRMSE, and we can see many of the HVAC and always-on appliances falling near the top of the chart with the lowest median CVRMSE for those end use types. Some appliances, such as stoves, dishwashers, dryers, and vacuums would not be expected to exhibit predictable patterns driven by outdoor temperature and hour-of-week schedules. Note that 152 of the 6,325 models with CVRMSE greater than 10 were excluded from the plot to preserve scale.

Figure 26. Distribution of CVRMSE Values, by End Use



These poorly modeled points are all on very low-use circuits, almost all of which have <10W annual average consumption, as shown in Figure 27. Because the average use is so low, the CVRMSE tends to be high because it is calculated as error divided by average use.

Figure 27. CVRMSE vs Avg Kw by Point



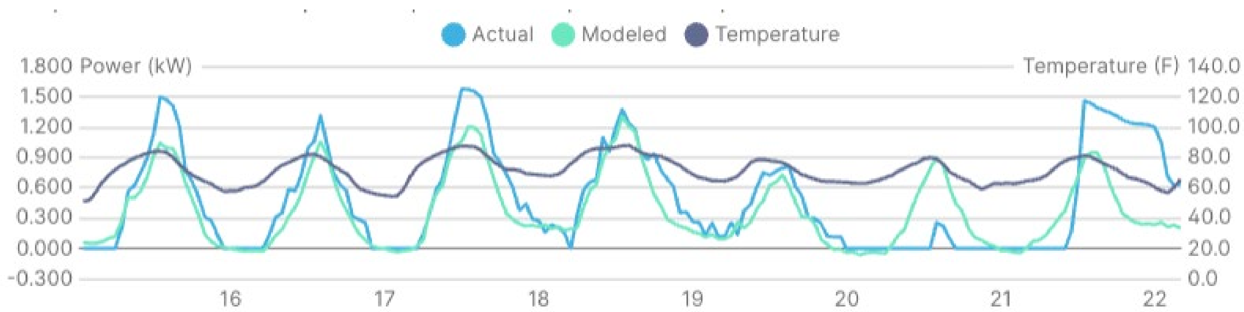
When we investigated several of the points with the highest CVRMSE values, we found that the energy use patterns in those points were indeed highly variable and the models did not do a great job predicting individual hourly energy use. However, the model's fall-back behavior is to predict primarily with time-of-day, day-of-week, and month-of-year patterns, so the monthly and seasonal weekday/weekend 24-hr load shapes produced by the model are very similar to the load shapes produced from the raw energy data. These load shapes and other aggregate trends like monthly average energy use are the most common use cases for this data set, so we ultimately determined that it was still beneficial to use the model outputs to generate a normalized 8,760-hr annual energy use profile for all points with adequate data, since that guarantees that there are consistent time-series data sets with the same weather and day-of-week values for each interval.

3.8.3 Considerations

One effect of weather regression modeling that is exacerbated when working with data from individual circuits is that the largest peak intervals tend to be under-predicted, as the model is trying to minimize error across many similar hours.

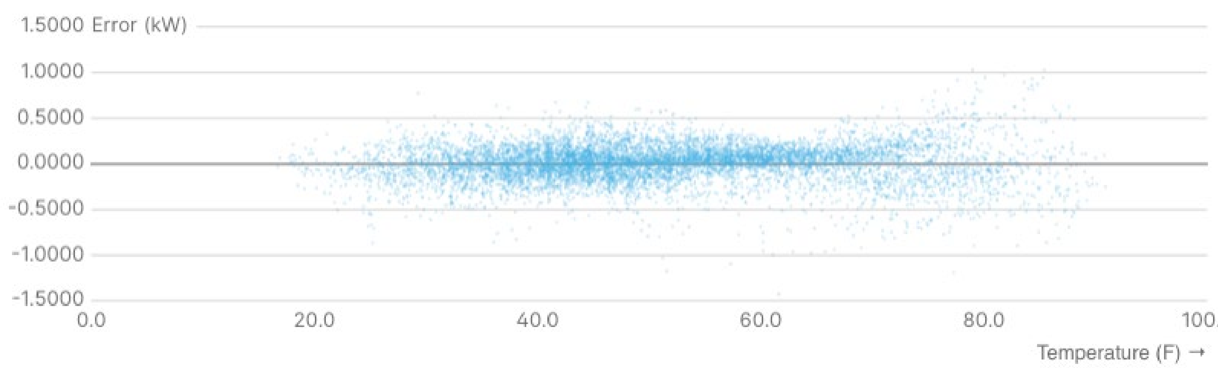
Figure 28 provides an example ducted heat pump (point_id = 6352) with an R-squared of 0.81 and a CVRMSE of 0.38, so this represents a best-case scenario where we have a well-fitting model. The figure shows one of the highest-demand periods of the year for this point ID. The "modeled" trend represents the prediction of the model on the same weather data used to train the model, so we are just looking at how well the model fits the actual data. During the highest period in the "Actual" data trend on August 17th, we see that the "Modeled" underestimates the peak by roughly 25%.

Figure 28. Actual vs Modeled Power and Temperature by Hour for a Week in August



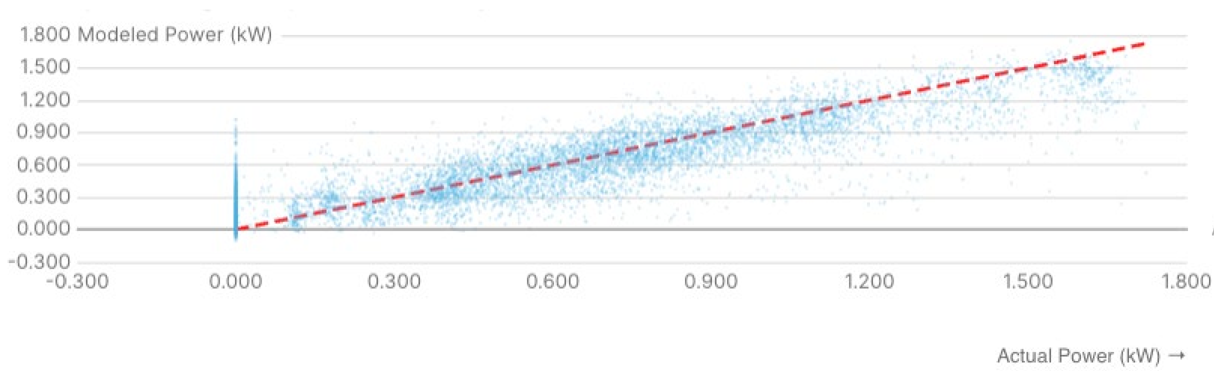
Looking at a plot of the same hourly error (actual minus modeled) versus outdoor temperature for the same point in Figure 29, we see that the hottest periods (above ~85°F) have a higher degree of error in both directions, as the model is trying to minimize overall error by balancing over- and under-estimation.

Figure 29. Error vs Temperature



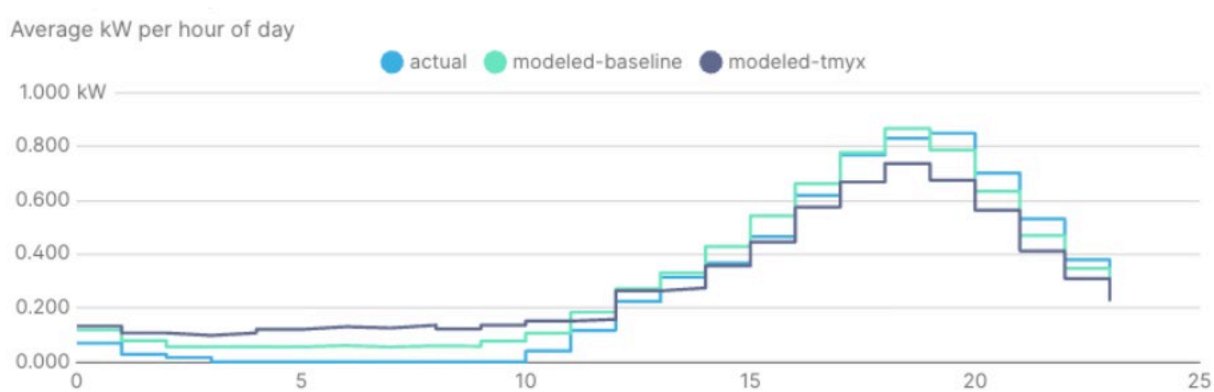
When we view the error for this point on a scatter plot of modeled power versus actual power in Figure 30, we can see that the highest-power intervals on the far right are systematically under-predicted, as shown by the majority of those markers that fall below the red 1:1 line. This may be due to the model failing to fully capture the non-linear behavior of decreased Coefficient of Performance (COP) at higher temperatures, or it may just be due to the periods of highest demand representing exogenous increases in demand that are not balanced by points that are in a similar actual power band but with a larger modeled power. That is, if there is some random noise in actual power around the modeled power prediction for a given set of weather and schedule conditions, those random variations cancel out everywhere but at the highest actual power conditions.

Figure 30. Modeled vs Actual Power



However, when the individual intervals for this point are aggregated into average load shapes, such as the August weekday profile shown in Figure 31, we see that the modeled values closely resemble the actual values in all hours for this example. While the under-estimation of actual peak hours remains an issue for estimating the peak behavior of individual loads, the seasonal average load shapes that we use to estimate grid impacts of electrification appear to be robust. Note that the "modeled TMYX" trend is somewhat different because the weather data in that forecast is somewhat different than the actual weather data that corresponds to the actual and modeled baseline trends.

Figure 31. Average Load shape for July



4 Aggregations

4.1 Average Loads by Time Period

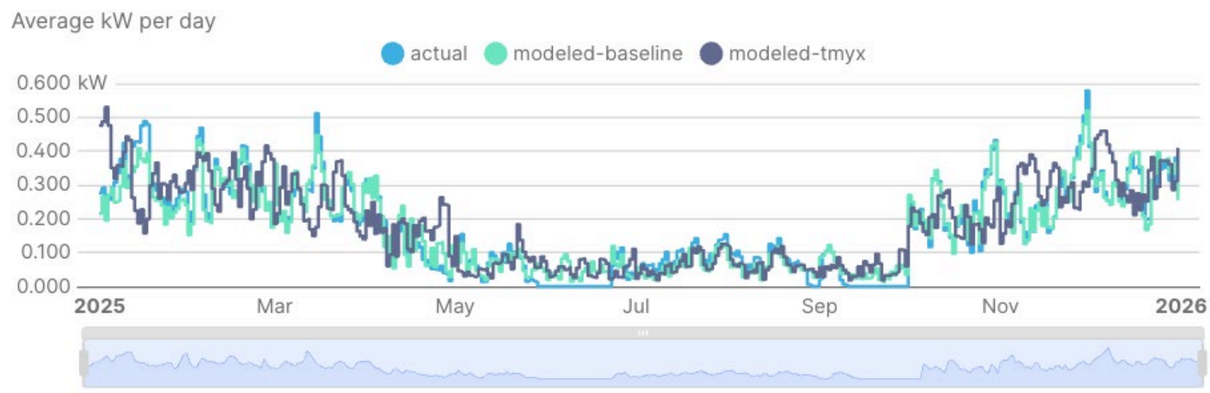
The project developed profiles for each point of the average load for a set of different time periods: daily, weekly and monthly. These profiles are visualized in the point details page for each point in the reporting dashboards. Example aggregations are shown below for a ductless heat pump. The charts show three different profiles:

1. **Actual:** raw data
2. **Model Baseline:** predicted data using the normalization models using the actual temperatures as the input for prediction.
3. **Model TMYx:** predicted data using the normalization models using the TMYx temperatures as the input for prediction.

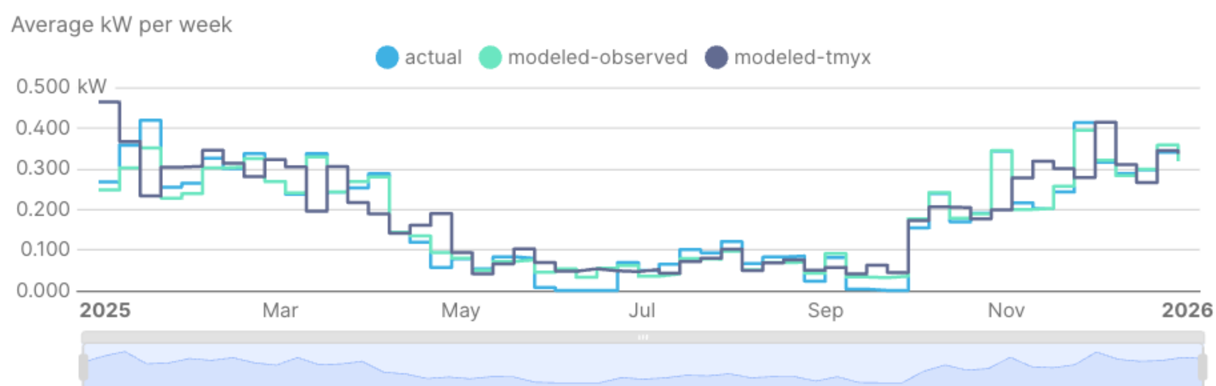
Figure 32 provides examples of these average load aggregations:

Figure 32. Example Average Load aggregations for a Ductless Heat Pump (point_id = 7)

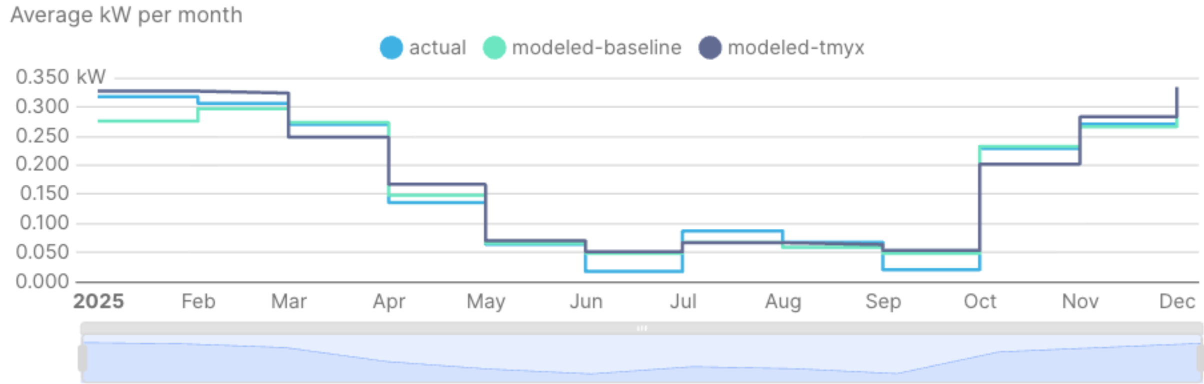
Daily



Weekly



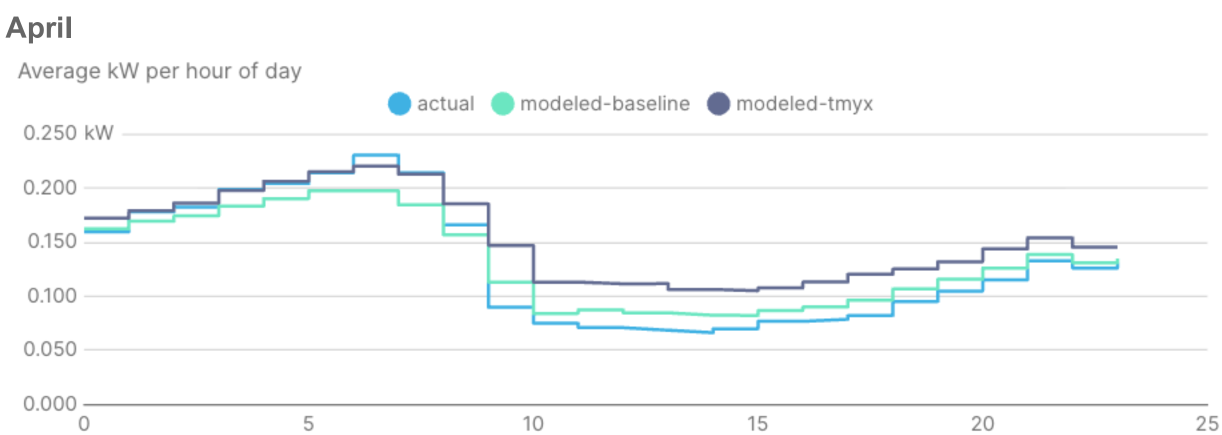
Monthly



4.2 Daily Average Load Shapes

The project outputs daily average load profiles for each point for a set of different time period aggregations: monthly and seasonally. The aggregations are also split into weekday and weekend profiles. These profiles are visualized in the point details page for each point in the reporting dashboards. Example aggregations are shown below for a ductless heat pump. The charts show the same three profiles as the time period aggregations. Figure 33 shows example monthly weekday profiles for a Ductless heat pump in April and August.

Figure 33. Monthly Weekday Average Daily Load Profile in kW for Ductless Heat Pump (point_id = 7)



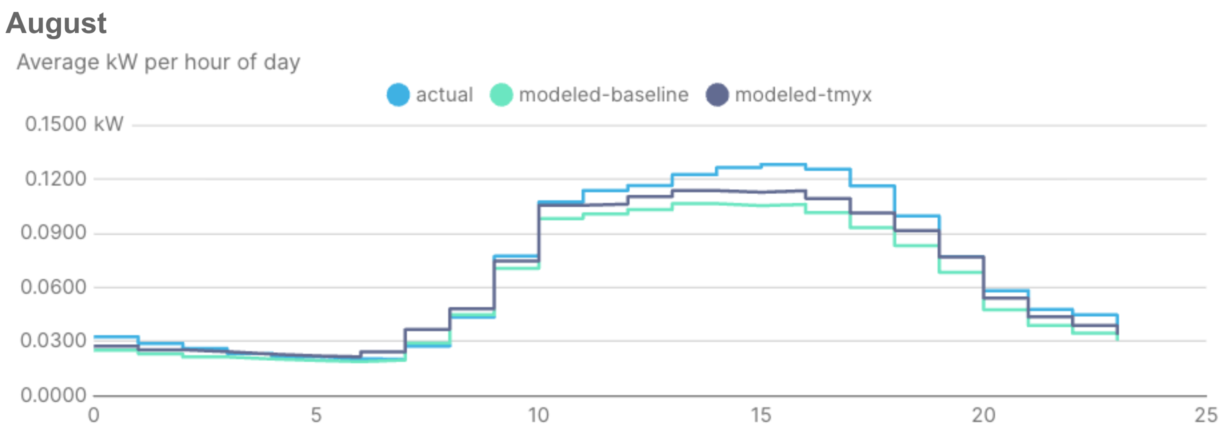
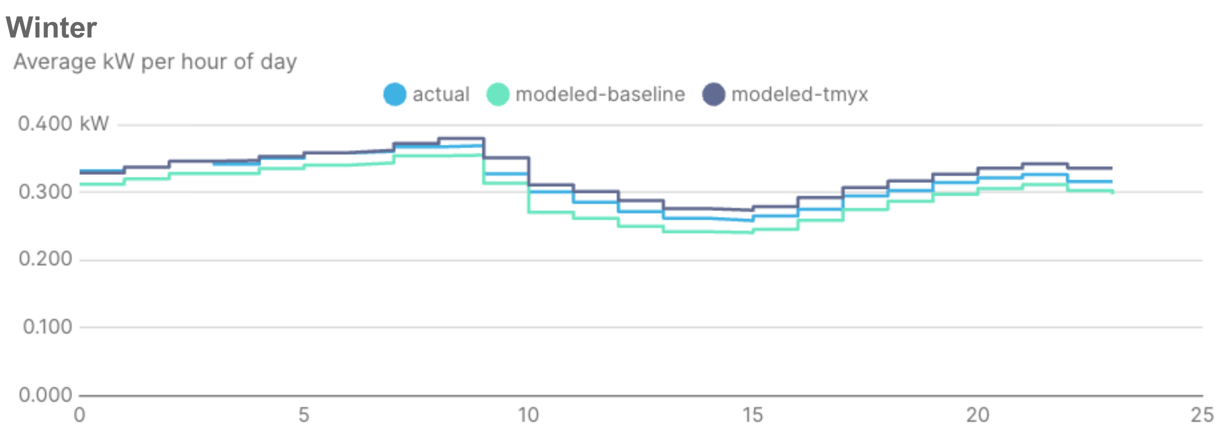
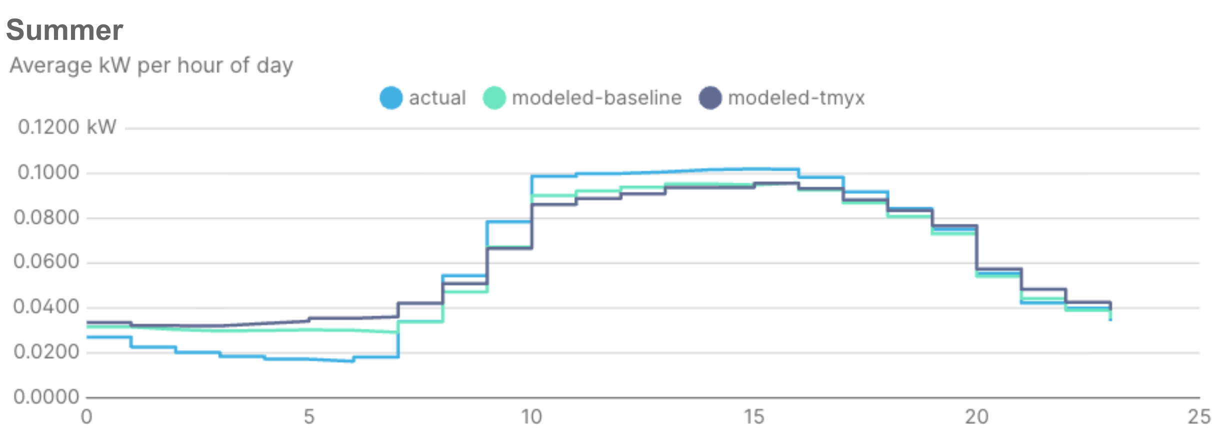


Figure 34 shows example seasonal weekday profiles for a Ductless heat pump in winter and summer.

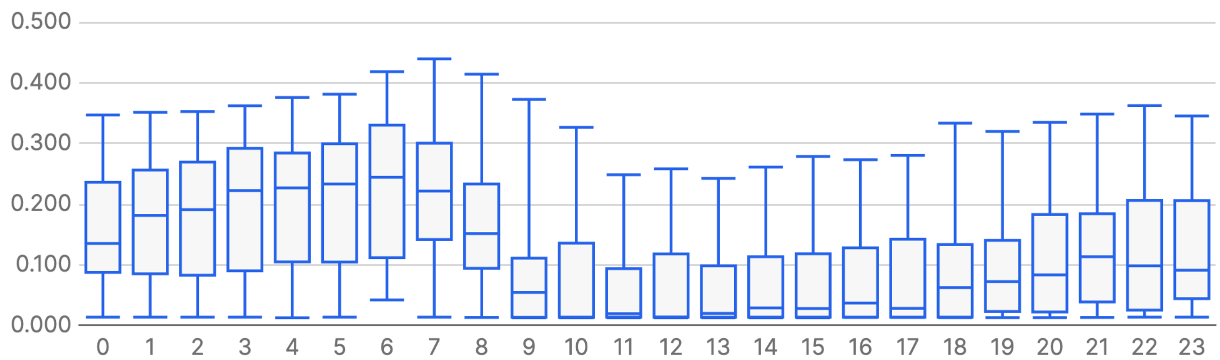
Figure 34. Seasonal Weekday Average Daily Load Profile in kW for Ductless Heat Pump (point_id = 7)





These aggregations are also available in box and whisker plots that show variation amongst each point in the aggregated profile. An example is shown in Figure 35 below.

Figure 35. Box Plot of Monthly Weekday Average Daily Actual Load Profile in kW for a Ductless Heat Pump in April (point_id = 7)



Note that when comparing to the Figure 33 average plot above for April, you can see the median point (represented by the middle line in the box) in hours 10 -13 is extremely low and near zero, while the range of values in those hours extend up to roughly 0.25 kW. This indicates that the average of 0.075 kW is the result of a smaller number of high usage hours during the month, rather than sustained daily usage at that level.

	First Hour Start	Last Hour Start	First Month	Last Month
Summer Evenings	3:00 PM	6:00 PM	July	August
Winter Evenings	4:00 PM	7:00 PM	December	January
Winter Mornings	7:00 AM	10:00 AM	December	January

Home Energy Metering Study Data Standardization and Electrification Peak Load Analysis

Section 3: Appendices

1 Appendices

1.1 Updated Circuit Type Labels

The following table includes a count of points where the raw circuit type label was modified into a cleaned version by NEEA. The table includes the original (raw) label, the new modified label from NEEA (cleaned) and the count of points that had that specific change.

Raw Label	NEEA Cleaned Label	Number of Points
Dishwasher	Refrigerator/Freezer	1
Ducted Heat Pump	Pump	1
Electric Resistance Storage Water Heater	Other Zonal Heat	1
Garbage Disposal	Refrigerator/Freezer	1
Mains With Solar	Mains	10
Needs Review	Central Vacuum	1
Needs Review	Other	8
Needs Review	Stove/Oven/Range	1
Other With Solar	Other	6
Other With Solar	Other Zonal Heat	1
Other With Solar	Pump	1
Other Zonal Heat	Electric Baseboard Heater	1
Other	Central Air Conditioner	2
Other	Central Vacuum	24
Other	Clothes Dryer	9
Other	Clothes Washer	26
Other	Dishwasher	20
Other	Ducted Heat Pump	7
Other	Electric Baseboard Heater	2
Other	Electric Furnace	9
Other	Electric Resistance Storage Water Heater	1
Other	Electric Vehicle Charger	3
Other	End Use Load Research Metering Box	30
Other	Energy Recovery Ventilator	8

Other	Garbage Disposal	15
Other	Gas Furnace Component	2
Other	Heat Pump Water Heater	1
Other	Hot Tub	32
Other	Ignitor	8
Other	Microwave	17
Other	Other Large Load	4
Other	Other Zonal Heat	50
Other	Packaged Terminal Air Conditioner	1
Other	Pump	63
Other	Refrigerator/Freezer	78
Other	Room Air Conditioner	3
Other	Stove/Oven/Range	13
Other	Sub Panel	27
Stove/Oven/Range	Clothes Dryer	1
Stove/Oven/Range	Other	1

1.2 End Use Category Mapping

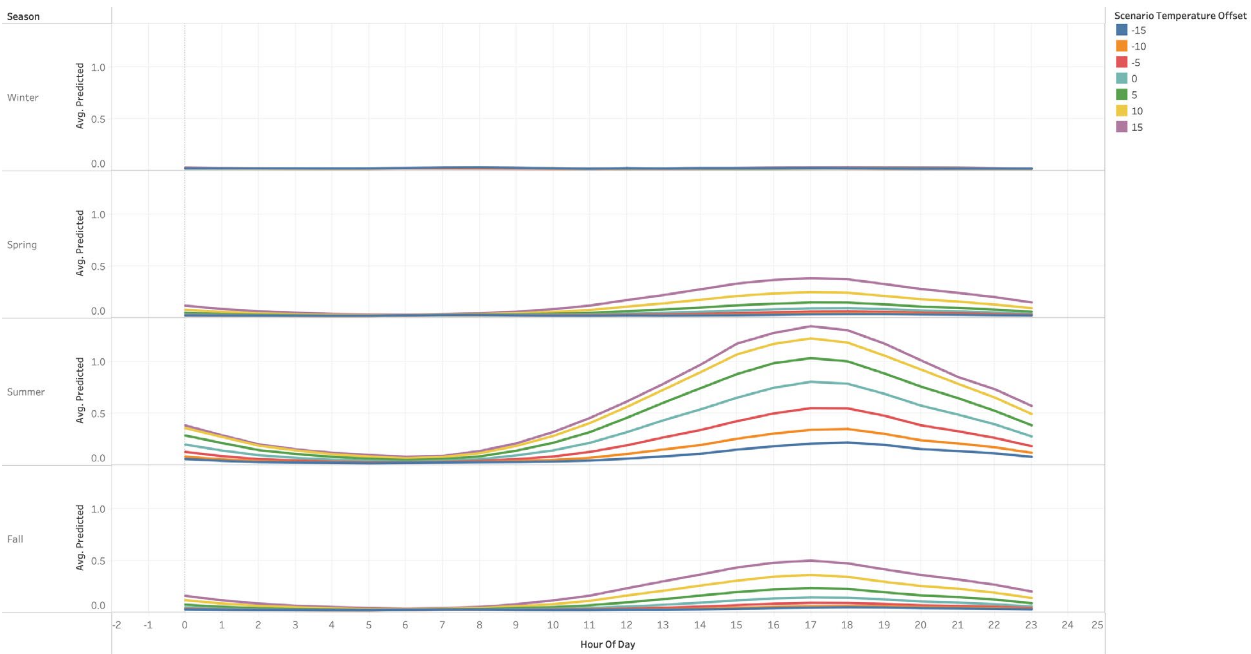
The following table includes a mapping of end use type labels to end use categories and the HVAC flags used to identify HVAC loads.

End Use Type Label	End Use Category	HVAC is Heat	HVAC is Cool	HVAC is Non-Load
Central Air Conditioner	HVAC		X	
Central Vacuum	Appliance			
Clothes Dryer	Appliance			
Clothes Washer	Appliance			
Dishwasher	Appliance			
Ducted Heat Pump	HVAC	X	X	
Ductless Heat Pump	HVAC	X	X	
Electric Baseboard Heater	HVAC	X		
Electric Furnace	HVAC	X		

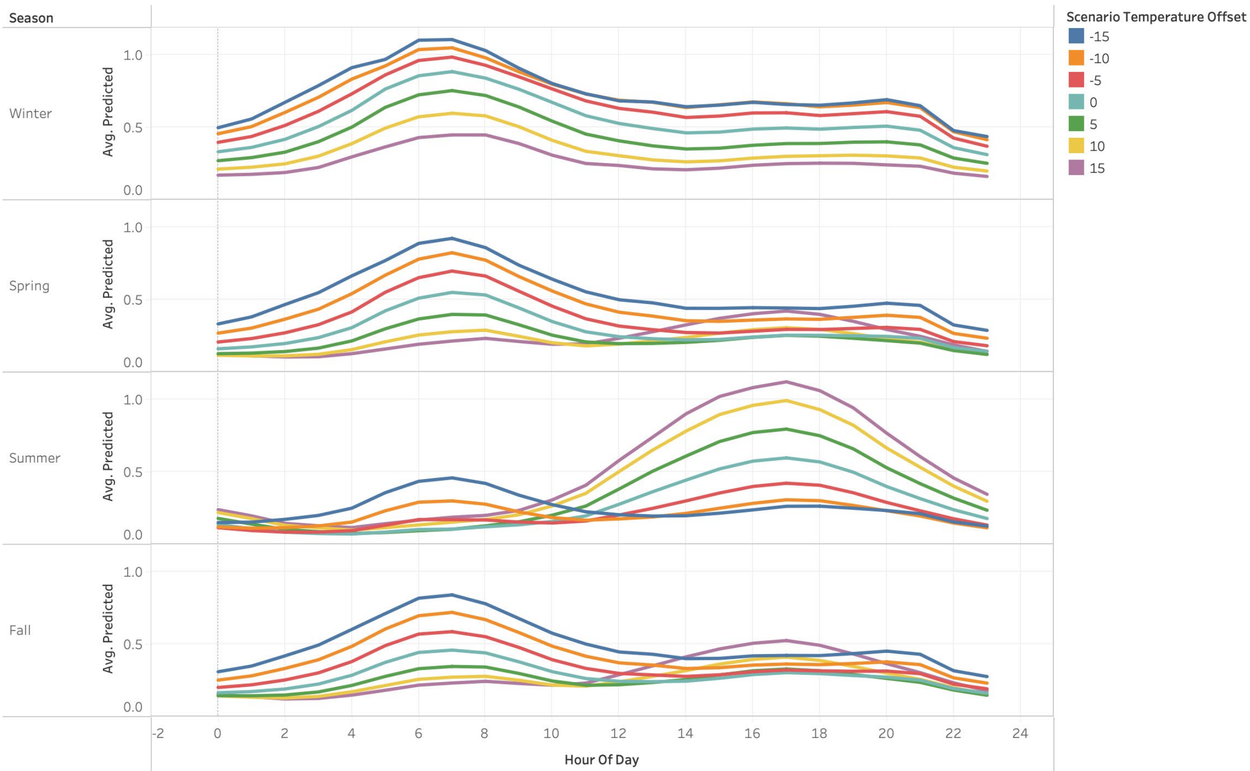
Electric Resistance Storage Water Heater	DHW			
Energy Recovery Ventilator	HVAC	X	X	
End Use Load Research Metering Box	Other			
Electric Vehicle Charger	EVSE			
Garbage Disposal	Appliance			
Gas Furnace Component	HVAC	X		X
Heat Pump Water Heater	DHW			
Hot Tub	Other-Large			
Ignitor	Other			
Electric Instantaneous Water Heater	DHW			
Gas Instantaneous Water Heater	DHW			
Mains	Mains			
Mains With Solar	Mains			
Microwave	Appliance			
Needs Review	Other			
Other	Other			
Other Large Load	Other-Large			
Other With Solar	Other			
Other Zonal Heat	HVAC	X		
Packaged Terminal Air Conditioner	HVAC	X		
Pump	Other-large			
Refrigerator/Freezer	Appliance			
Room Air Conditioner	HVAC		X	
Solar	PV			
Stove/Oven/Range	Appliance			
Sub Panel	Other			

1.3 Temperature Scenario Charts for Weather Sensitive End Uses

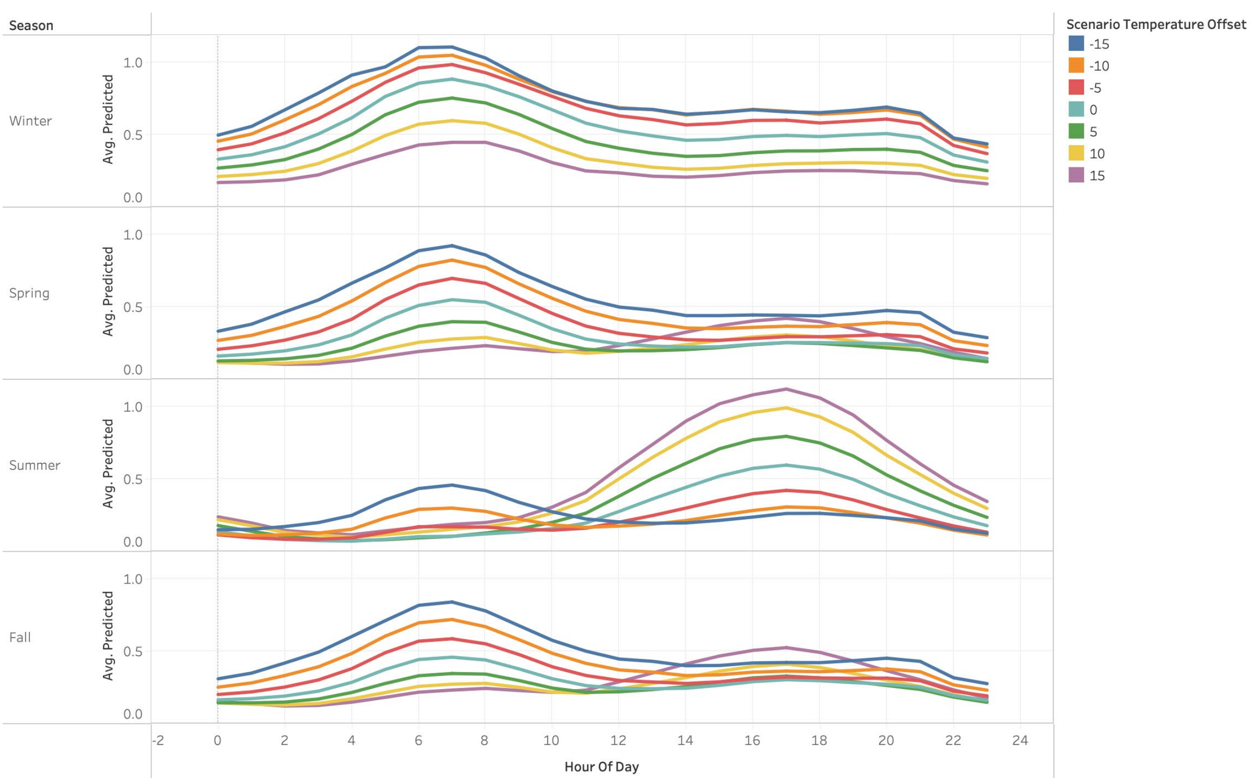
Central Air Conditioning



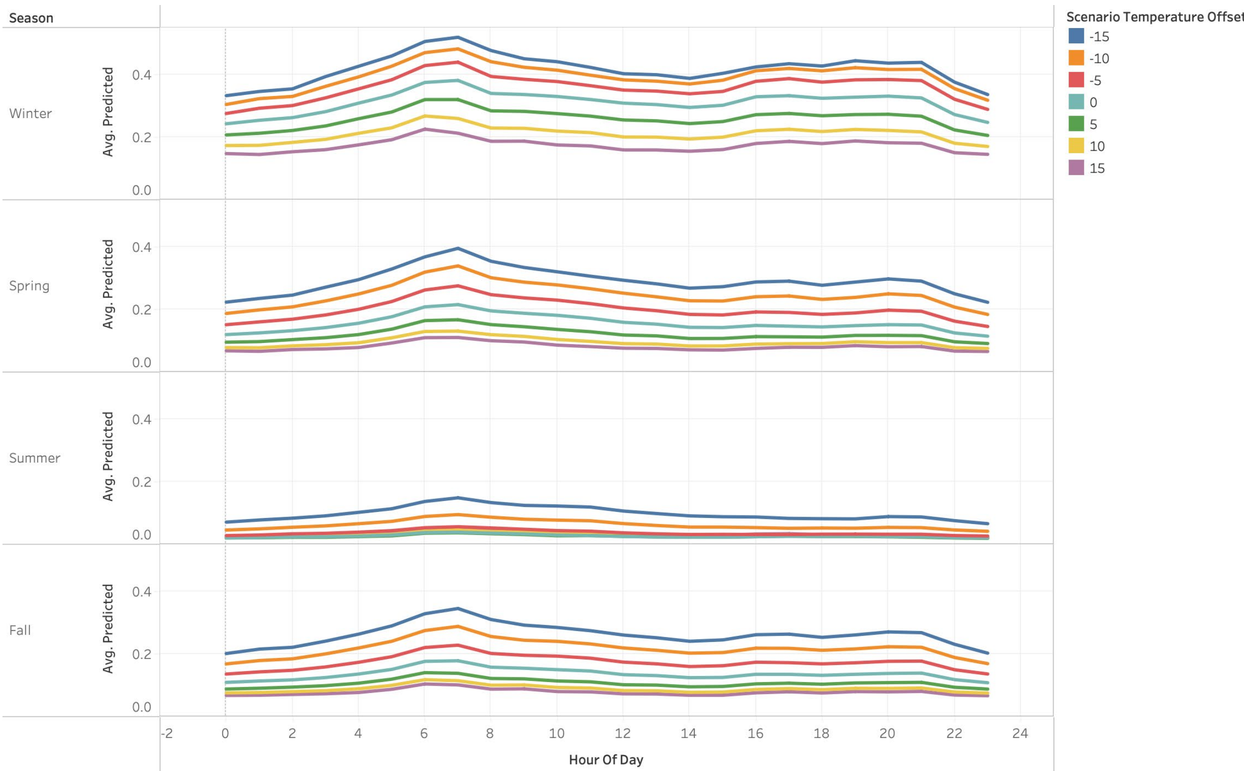
Ducted Heat Pump



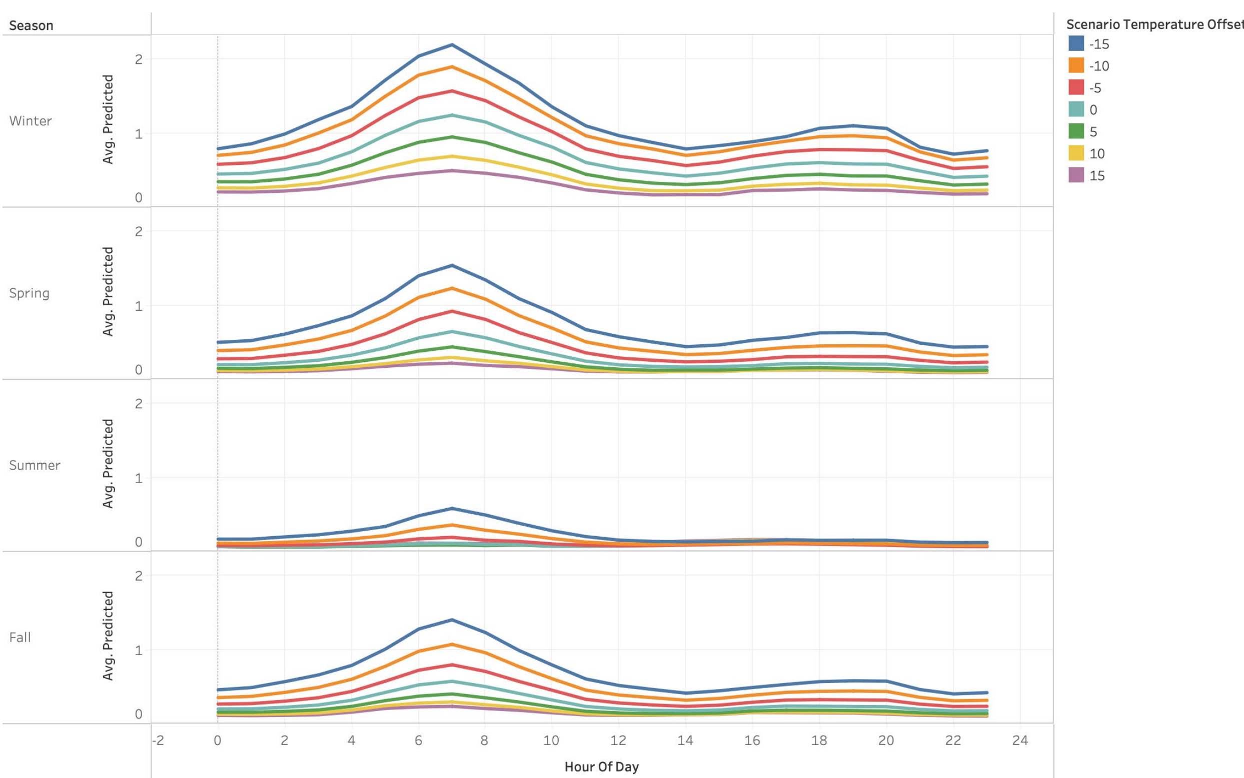
Ductless Heat Pump



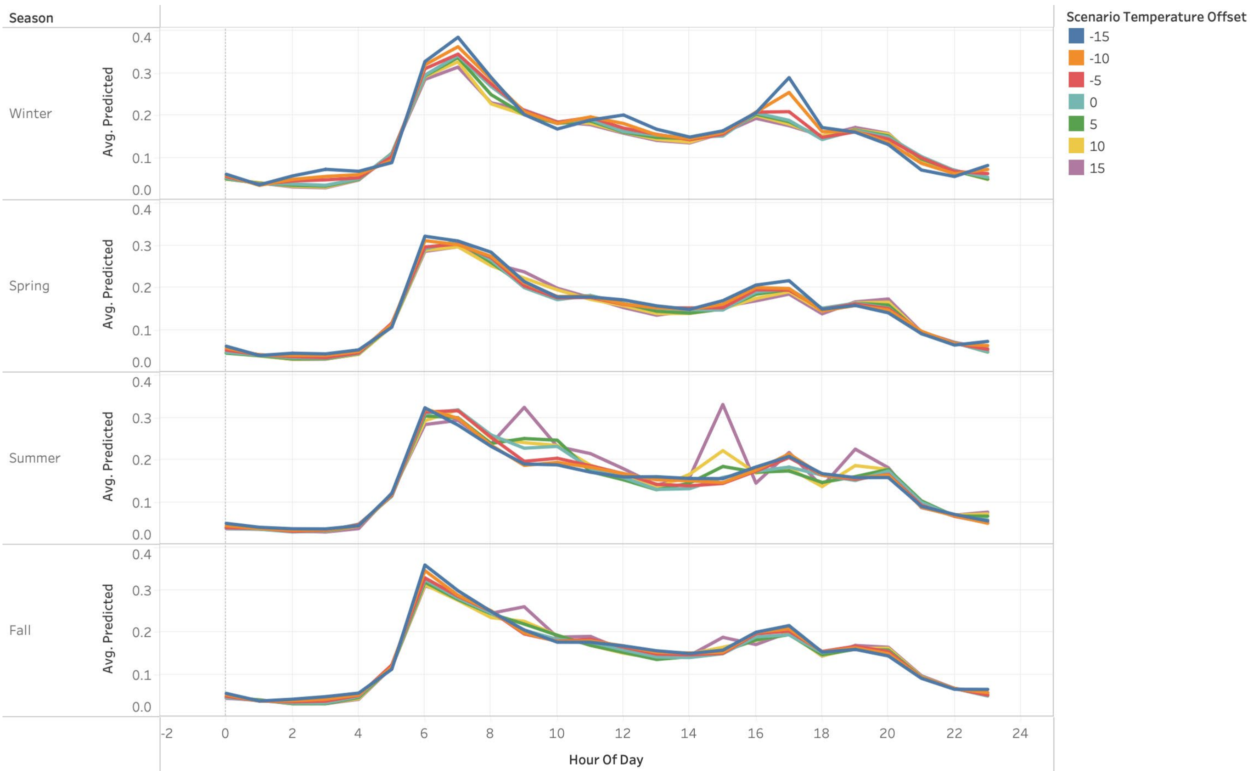
Electric Baseboard Heater



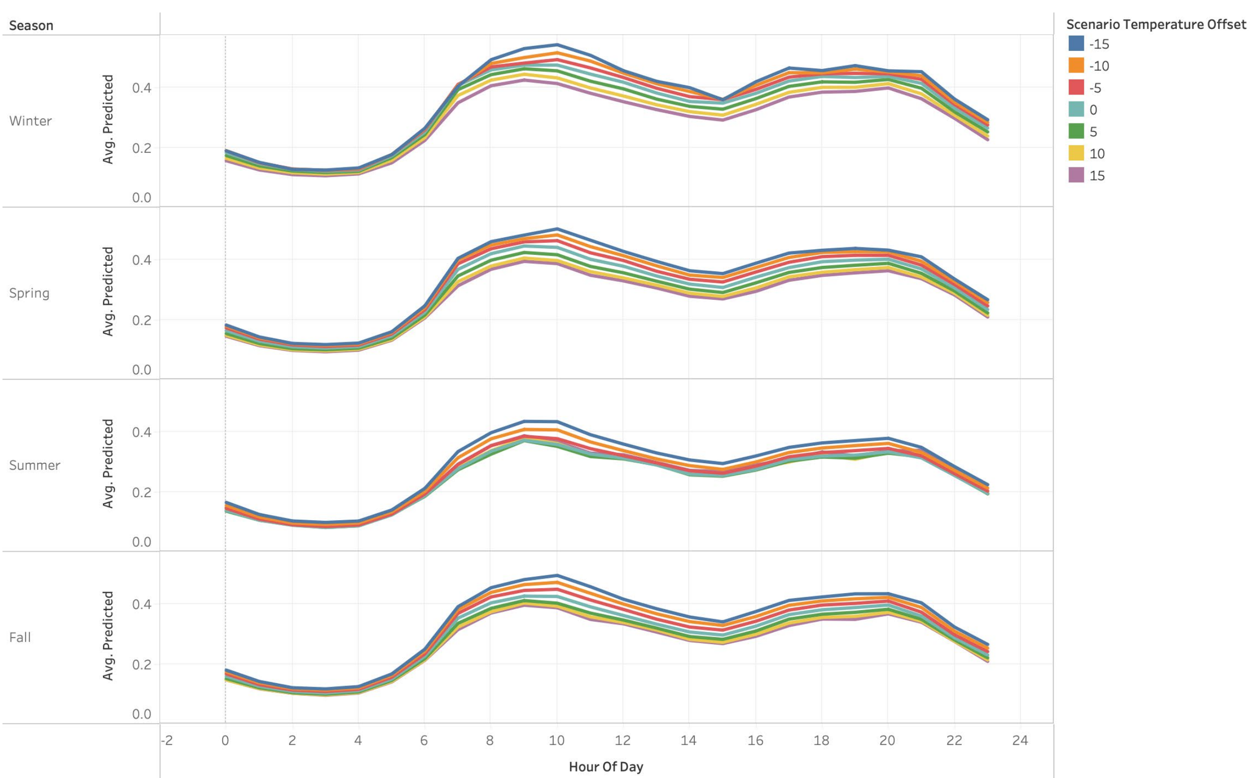
Electric Furnace



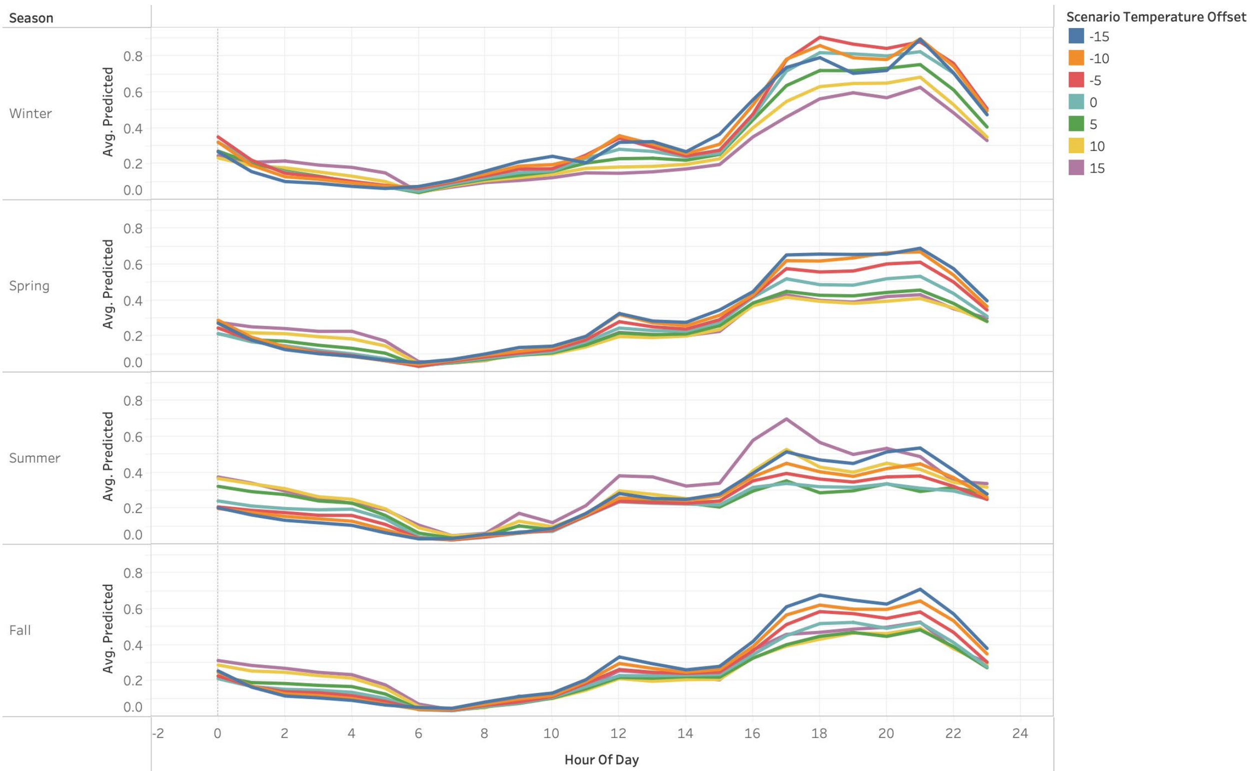
Electric Instantaneous Water Heater



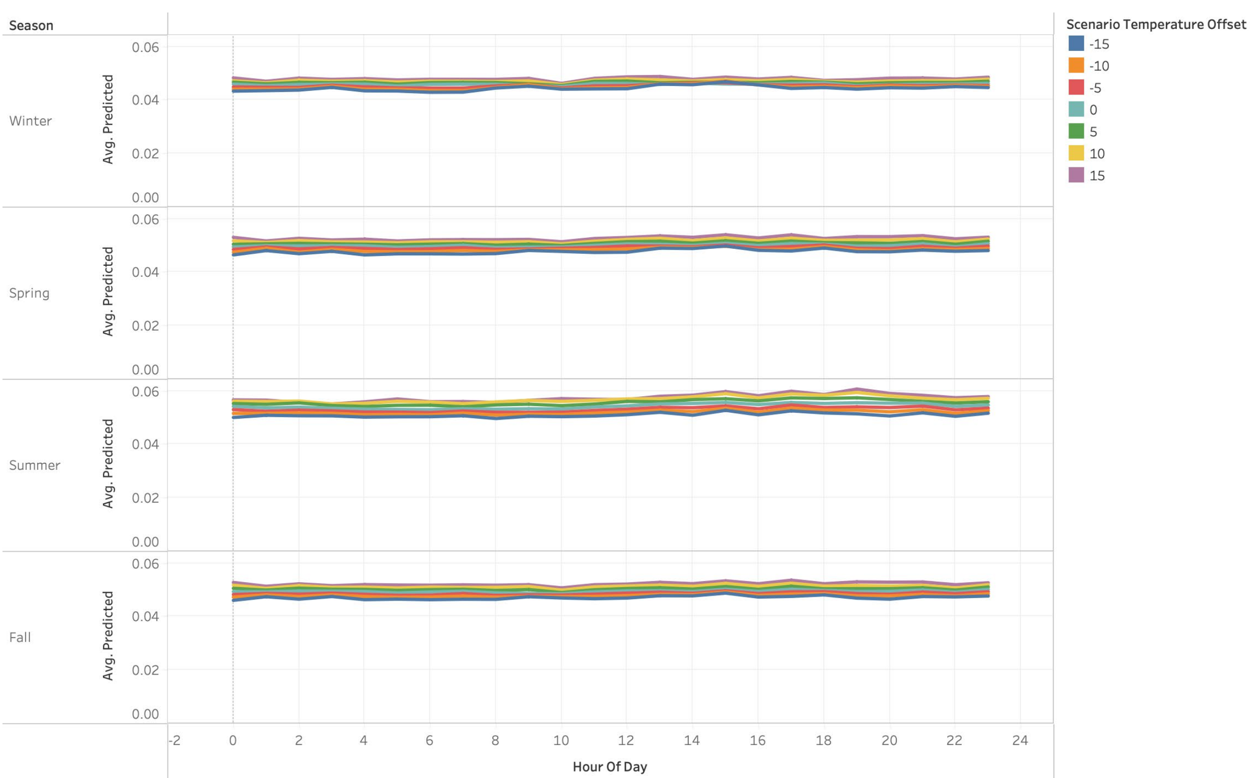
Electric Resistance Storage Water Heater



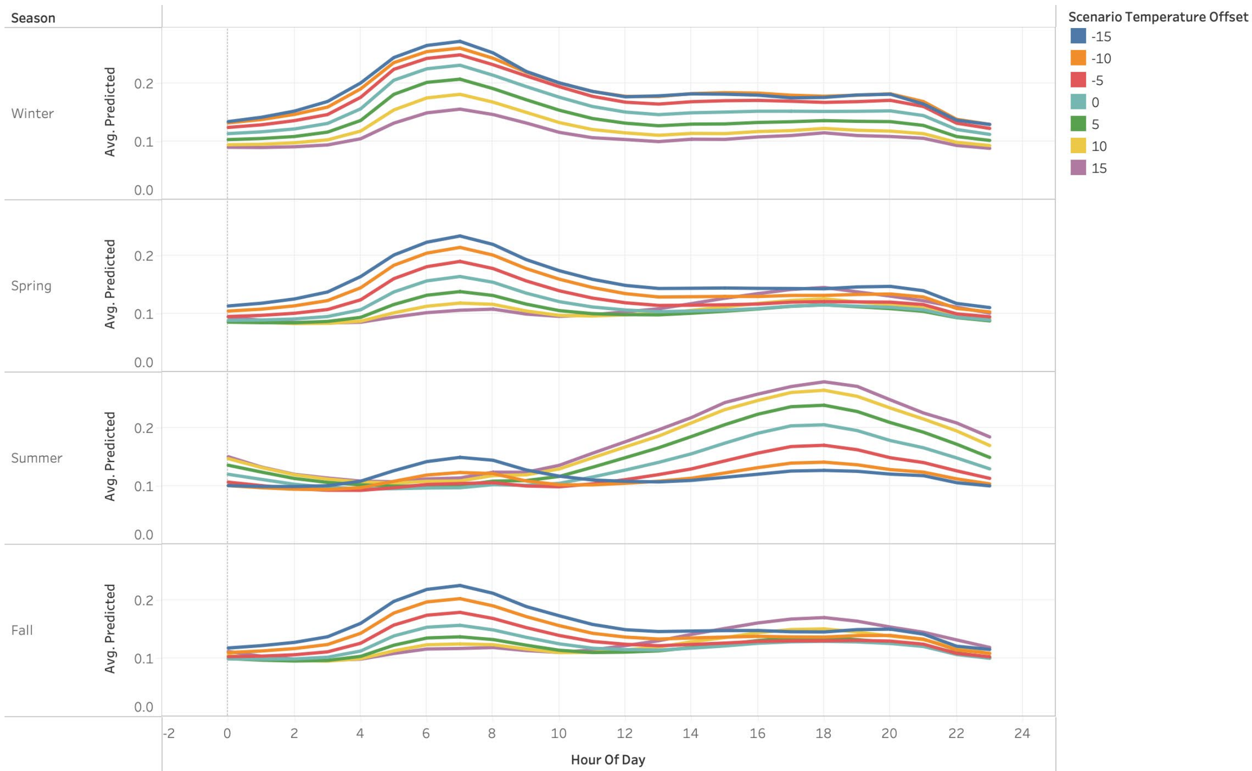
Electric Vehicle Charger



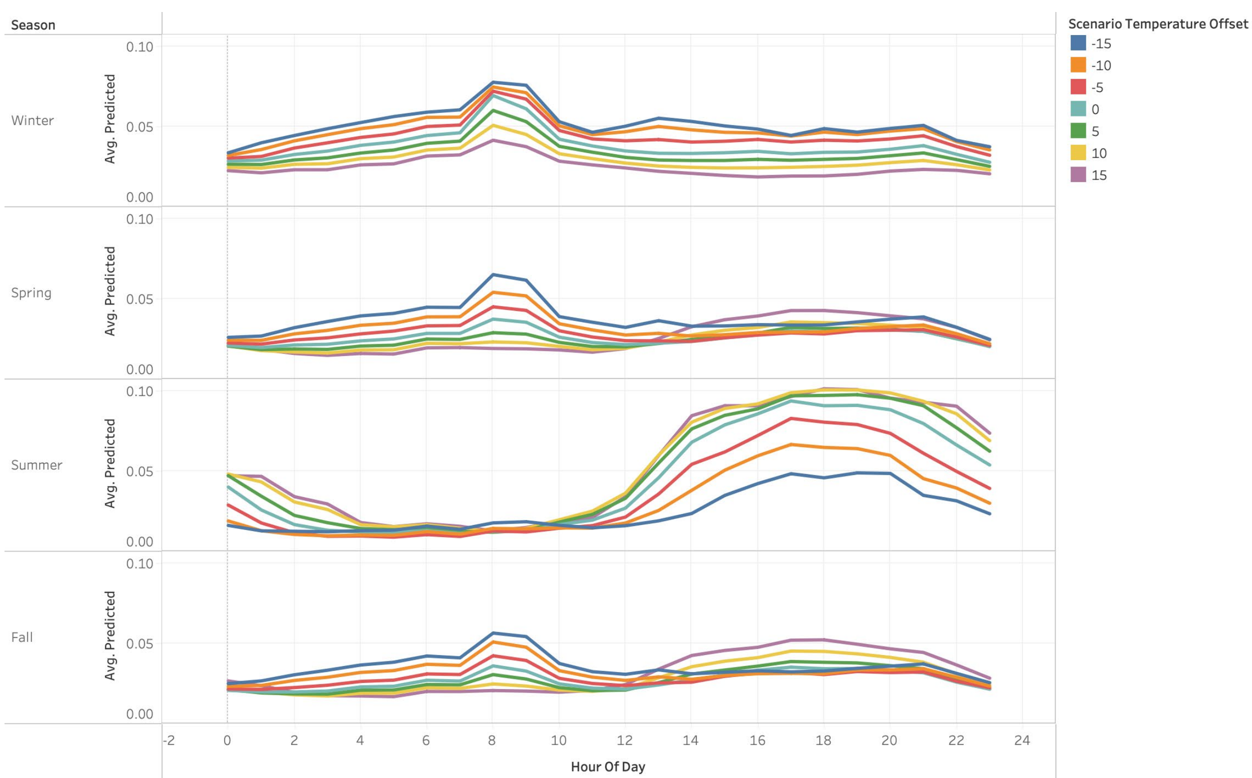
Energy Recovery Ventilator



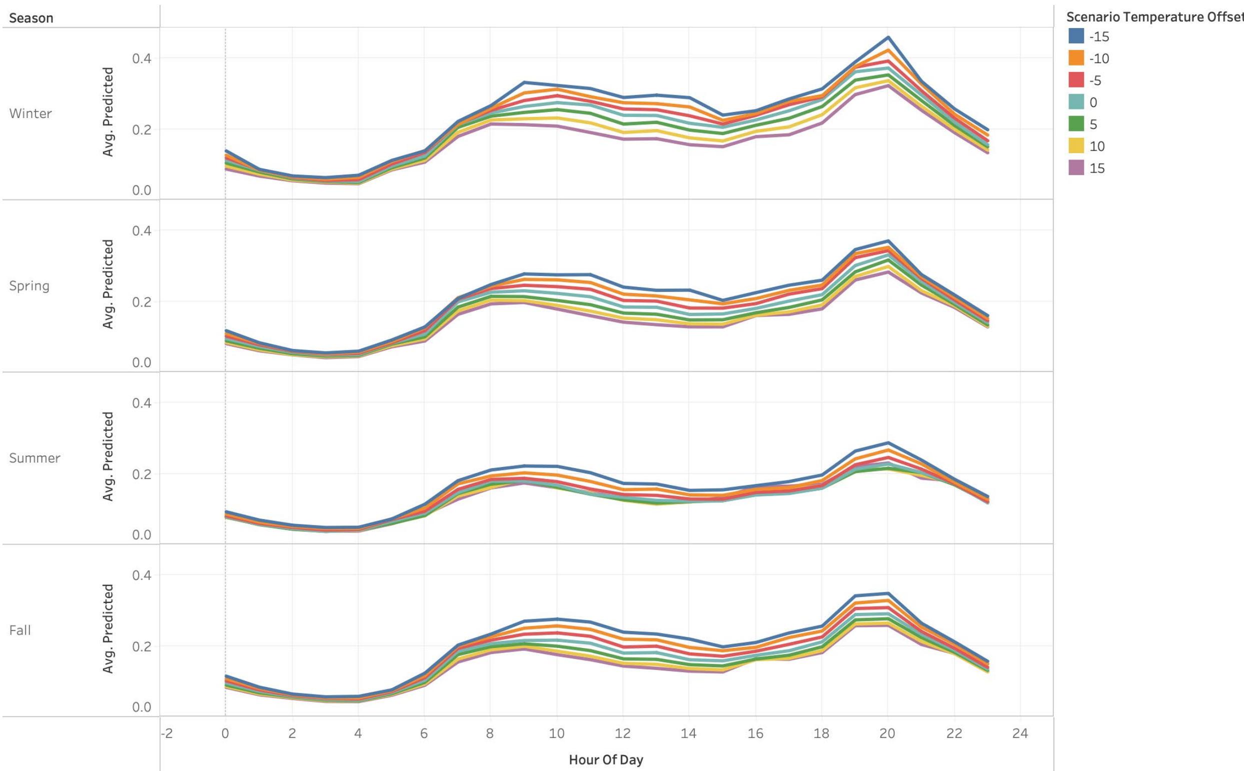
Gas Furnace Component



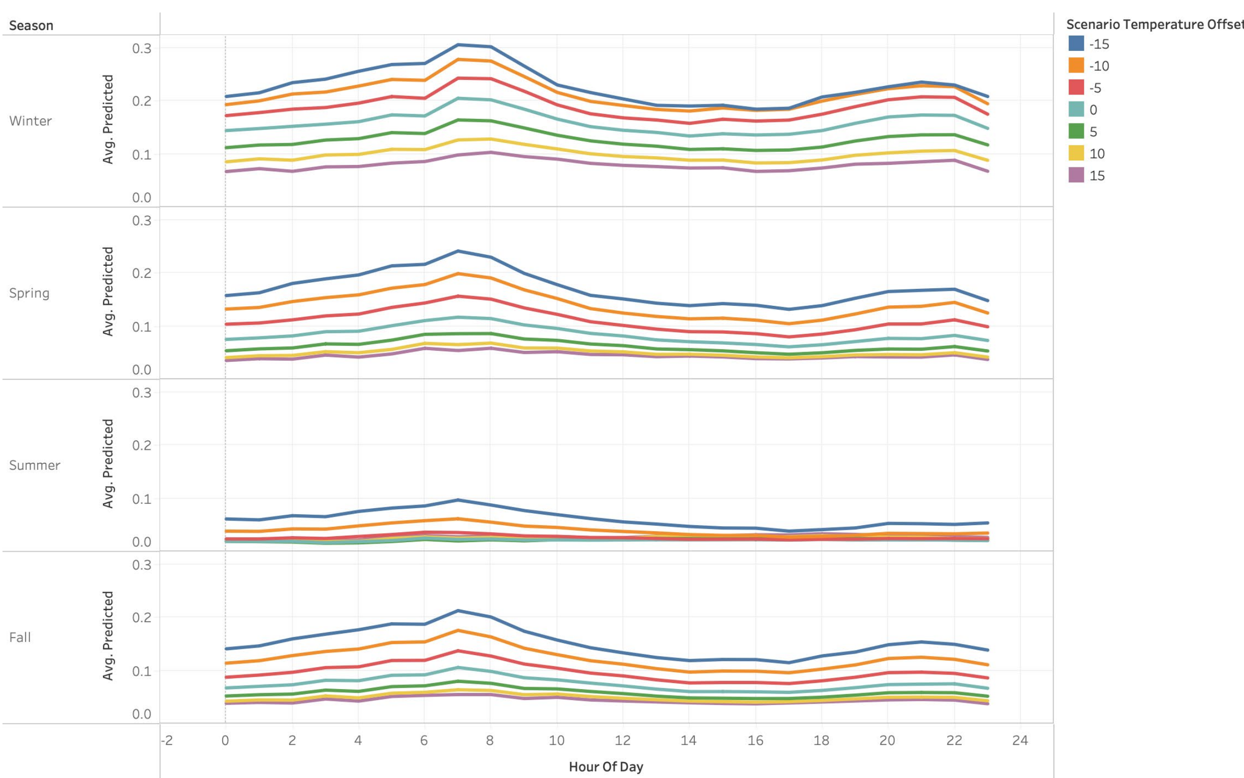
Gas Instantaneous Water Heater



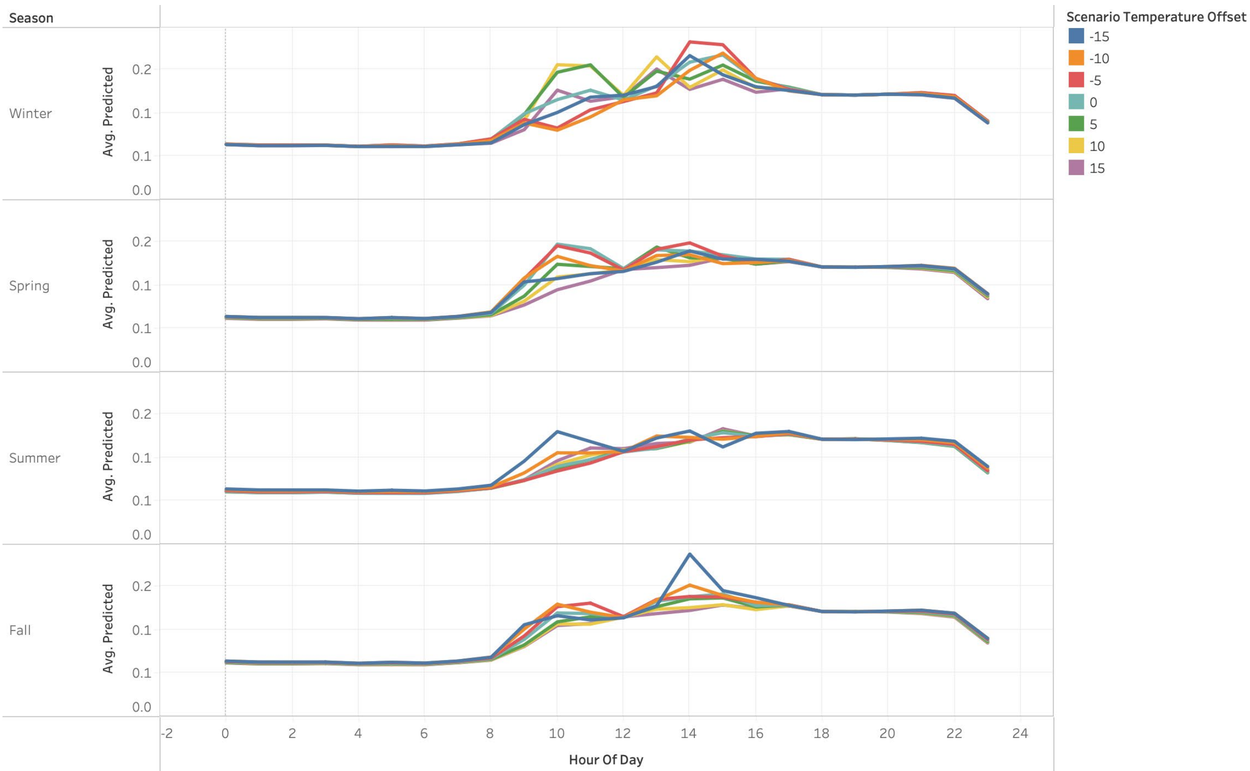
Heat Pump Water Heater



Other Zonal Heat



Packaged Terminal Air Conditioner



Room Air Conditioner