Demand Side Analytics
DATA DRIVEN RESEARCH AND INSIGHTS

Report

HOURLY IMPACT OF ENERGY EFFICIENCY EVALUATION PILOT

Prepared for
Vermont Department of Public Service

By Demand Side Analytics, LLC
Jesse Smith
Stephanie Bieler
Savannah Horner

November 2021

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY 3

2 INTRODUCTION 10

2.1 ADVANCED M&V OVERVIEW 10

2.2 PROJECT OBJECTIVES 11

3 METHODOLOGY 13

3.1 DATA SOURCES 13

DSM Tracking Data 13

Customer Characteristics 14

AMI Data 14

Weather Data 15

Avoided Costs and Emissions Rates 16

Efficiency Vermont Technical Reference Manual 19

3.2 SELECTING MEASURES 19

Selecting Final Prescriptive Sites 23

Selecting Final Custom Projects 24

3.3 PRE POST ANALYSIS 26

3.4 MATCHING ANALYSIS 28

3.5 ESTIMATION 31

3.6 ACCURACY ASSESSMENT 34

3.7 UNCERTAINTY 35

Prescriptive Measures 36

Custom Projects 39

4 RESULTS 40

4.1 ACCURACY ASSESSMENT 40

Residential 41

Commercial 42

Standard Errors 44

4.2 PRESCRIPTIVE RESIDENTIAL 45

Advanced Thermostat 45

BLPM Boiler Motor – Residential 48

ENERGY STAR Dehumidifiers 50

ENERGY STAR Clothes Dryer 52
ENERGY STAR Clothes Washer 54
Home Performance with ENERGY STAR 56
Heat Pump Water Heater 59
LED Indoor Fixture 63
LED Screw-Base Lamp 65
High Efficiency Pool Pump 68
Refrigerator 70
Variable Speed Cold Climate Heat Pump 72
Overview of Prescriptive Residential Results 76
4.3 PRESCRIPTIVE COMMERCIAL 77
BLPM Circulator Pump – Commercial 77
Commercial Refrigerator 79
Compressed Air 81
LED Agricultural Interior Fixtures 82
LED Exterior Fixtures 84
LED Interior Lighting 86
Overview of Prescriptive Commercial Results 91
4.4 CUSTOM 91
Goodness of Fit 92
Aggregate Impact 93
Aggregate Impact with COVID 95
Understanding Outliers 95
4.5 TIME-VALUE OF EFFICIENCY 97
5 COMPARISON WITH TRADITIONAL M&V METHODS 101
5.1 CUSTOM 101
5.2 PRESCRIPTIVE 104
6 SOFTWARE OPTIONS 111
7 CONCLUSIONS 113
8 TECHNICAL APPENDIX 115

1 EXECUTIVE SUMMARY

Demand Side Analytics (DSA) executed a contract on January 11, 2021, with the Vermont Department of Public Service (PSD) to aid in the investigation of advanced measurement and verification (M&V) practices for energy efficiency evaluation. DSA worked closely with Efficiency Vermont (EVT) to gain access to four years of program tracking data and 15-minute advanced metering infrastructure (AMI) data for all Green Mountain Power (GMP) customers.

Program administrators, regulators, and system planners seek to quantify the energy and peak demand impacts of conservation programs for a variety of technical and policy reasons. Readers of this report should keep in mind that these impacts cannot be directly measured because it is not possible to observe what consumption would have been absent program intervention (known as the counterfactual). While any attempt to quantify energy savings is inherently an estimate, methods have strengths and weaknesses and there are statistical metrics we can use to assess the uncertainty associated with our estimates.

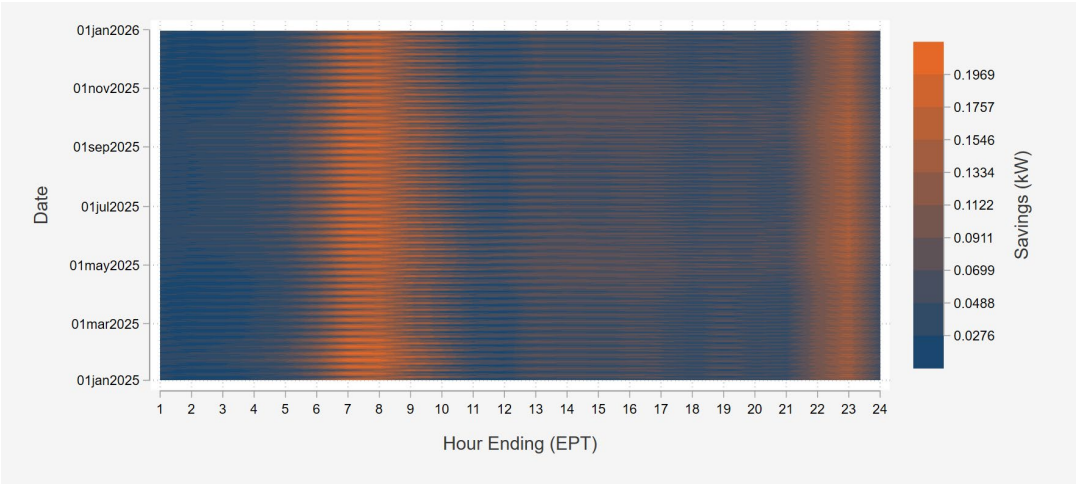
In this study, we implement Advanced M&V procedures that build upon the International Performance Measurement and Verification Protocol (IPMVP) Option C Whole Facility[1] approach to energy savings estimation. We do this through a regression model that follows Lawrence Berkeley National Laboratory's (LBNL) Time-of-Week Temperature (TOWT) Model[2], where the dependent variable is hourly electric consumption from the GMP meter and the independent variables contain information about the weather, day of week, and time of day. We model energy consumption separately in the pre-installation and the post-installation periods and compare the two periods in an attempt to understand how energy consumption changes among participants after installation of an energy efficiency measure.

This methodology estimates efficiency impacts in each hour of the year. Granular results provide insight into the distribution of energy savings across the year. Figure 1 shows a heat map of the average energy savings from replacing an electric resistance water heater with a high efficiency heat pump water heater. The heat pump water heater measure's energy savings are concentrated in the morning and evening (orange regions), presumably because this is when the homes analyzed use the most hot water. The striations in the morning hours show the different profile for weekdays and weekends, with savings occurring later on weekend mornings. Combining this type of output with granular avoided cost assumptions allows us to conceptualize a "time-value" of efficiency for our various measures of analysis.

[1] https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp Option C is whole facility.

[2] https://eta-publications.lbl.gov/sites/default/files/LBNL-4944E.pdf

Figure 1: HPWH with Electric Baseline Heat Map



In addition to implementing these methods and reporting the results, a core research objective for this study is to compare and contrast the Advanced M&V methods with Traditional M&V practices. Traditional M&V is a broad term. For purposes of this report, we discuss Traditional M&V in two categories.

- **Prescriptive.** Generalized measure characterizations housed in the Technical Reference Manual (TRM). Prescriptive measures seek to estimate the average capacity, efficiency, operating hours, and energy savings across program participants.
- **Custom.** Found most commonly in the commercial and industrial sectors – typically for larger projects. Program staff or independent evaluation contractors collect site-specific operating parameters and use actual equipment characteristics to estimate savings for a given project. Often these project analyses follow IPMVP Option A or B.[3]

Our analysis segments energy efficiency projects into these two broad categories using EVT's designation of project "track". We classify projects from five tracks as custom; 6012CNIR, 6013CUST, 6014CUST, 6017CUST, and 6020CUST, while the rest of the tracks are classified as prescriptive. Table 1 shows the total claimed electric savings from the EVT tracking database, across the two project types from 2017-2020. Custom projects are less common but tend to produce more savings per project. Approximately 75% of Efficiency Vermont's claimed electric savings come from prescriptive projects and the other 25% comes from custom projects.

[3] https://evo-world.org/en/products-services-mainmenu-en/protocols/ipmvp Option A is retrofit isolation - key parameter measurement. Option B is retrofit isolation – all parameter measurement.

Table 1: Distribution of EVT Electric Impacts by Project Type (2017-2020)

Project Type	Projects	MWh	Summer kW	Winter kW
Prescriptive	33,950	332,165	40,587	60,175
Custom	3,134	141,223	17,567	17,856
Total	**37,084**	**473,388**	**58,154**	**78,032**

This study applies Advanced M&V techniques to both project types. For prescriptive measures, we estimate the average performance across a pool of sites using the TOWT model with and without a matched comparison group (or control group). We run the method with matched controls once without COVID-19 period data[4] in the post-installation period and once with COVID-19 data in the post-installation period. We do not run the pre-post TOWT with COVID-19 period data in the post-installation period because this method does not control for the exogenous effect of the COVID-19 shutdowns. Section 4.1 discusses the presence of bias during the COVID-19 shutdowns and shows that the direction of bias varies by customer class. For custom projects, we estimate savings individually for each project using the pre-post TOWT.

Table 2 lists the dimensions for our comparison of Traditional and Advanced M&V methods and summarizes the key findings after implementing the Advanced M&V techniques across different project types. Generally, Advanced M&V is a powerful tool – but it is not the right tool for every job. Section 5 of the report explores each comparison area in detail.

[4] Data from March 1, 2020 onward is referred to data from the COVID-19 period in this report

Table 2: Comparing Advanced and Traditional M&V Practices

Metric	Considerations	Findings
Cost	What are the costs of integrating Advanced M&V methods into program delivery and evaluation? What savings in traditional M&V costs might be expected?	Integration costs can include both software and labor to train staff. These costs can be modest if open source tooling is used. Current levels of end-use metering could likely be cut in half and save upwards of $50,000 per year. Additional savings may be possible from streamlining or reducing dependency on the Technical Reference Manual (TRM) for prescriptive measures, as detailed in Section 5.
Uncertainty and Precision	For what applications do Advanced M&V methods produce more precise estimates of energy and demand savings? For what applications are traditional methods better suited?	Uncertainty is a function of sample size, effect size, and the ability of models to explain variation. Measures with high uptake will have larger available sample sizes. Non-residential sites with standard operating schedules whose electric loads are sensitive to weather will typically exhibit the best goodness-of-fit statistics. Advanced M&V methods generally work best for retrofit measures, but can also provide valuable insights for market opportunity measures such as hours of operation and load shape.
Timing	Are Advanced M&V methods able to provide impacts more quickly than traditional methods?	Given the robust meter data sharing procedures in place between GMP and EVT, impacts can be estimated one day into the post-installation period. While this fast feedback is a useful check, post-retrofit loads should be observed over a full range of conditions before projecting the impacts to an annual basis. Since most cases will require close to a full year of post installation data, Advanced M&V methods are likely no faster or take longer to produce than Traditional M&V.
Integration with Planning Activities	Granular impacts are one of the exciting features of Advanced M&V methods. How much effort will be required to integrate granular impacts into planning activities such as ISO-NE FCM certification and benefit-cost modeling?	For this study, we model avoided electric energy and CO_2 emissions using 8760 impacts. While more computationally intense, the calculations are still relatively straightforward because the Avoided Energy Supply Costs in New England study provides 8760 benefit streams. The current ISO-NE summer peak demand definition is 1-5pm on non-holiday weekdays in June, July, and August. When that definition shifts later in the afternoon as solar penetration increases, the summer coincidence factors used to compute summer demand impacts will need to be updated. Having granular impact load shapes could make that transition far simpler.
Customer Settlement	What are the implications to using Advanced M&V methods to determine rebate amounts (e.g., an Energy Savings Guarantee or Pay-for-Performance program)? How does it change the risk profile for the program administrator and participant?	The guaranteed savings program model is attractive because it encourages program administrators to look beyond average performance and pursue customers who can save the most energy. When measured savings determine the incentive amount, the cost of acquisition is also known with certainty in advance. Even for sites that are well suited to Advanced M&V methods, rebate amounts will remain a mixture of performance and random noise. This creates a risk for the participants as some participants will be over-compensated and others will be under-compensated. It will be important for EVT to clearly communicate the modeling procedures and payment risk to participants up front.

Based on the results of our analysis and understanding of the regulatory environment in Vermont, we offer the following key findings:

- **For 14 of the 20 prescriptive measures analyzed, our meter-based estimates return smaller energy and peak demand impacts than the reported values in the Efficiency Vermont tracking data for prescriptive measures.** PSD and Efficiency Vermont will need to consider carefully whether this indicates that perhaps the TRM characterizations for these measures are overstated, or that meter-based methods are not the right M&V approach for the measure. Demand Side Analytics believes that the appropriate conclusion varies by measure.
    - ✓ It is important to note the energy and demand savings stored in the EVT tracking data are not the final estimates of program and portfolio savings. EVT and PSD work with third-party evaluation contractors to verify and independently estimate impacts for ISO-NE compliance and other reporting functions. We do not attempt to compare the results of those efforts to the outputs of this study. In cases where our meter-based results diverge from the EVT tracking database claims, a logical next step would be to consider the results of other Vermont evaluation efforts for additional context.
- **Meter-based estimates are difficult to compare with prescriptive savings assumptions for market opportunity measures**. By definition, pre-post regression models assume the existing condition of the home or business as the baseline. Many measures in the EVT portfolio, and eight of the twenty (40%) prescriptive measures we analyze, assume a "market opportunity" perspective where the replaced equipment is assumed to have reached the end of its useful life so the baseline is a new code-minimum piece of equipment. Because this hypothetical piece of code-minimum equipment was never installed, household consumption for the baseline condition cannot be directly estimated with utility meter data.
    - ✓ For some measures we theorize that the replaced equipment is less efficient, on average, than the code-minimum assumed in the TRM baseline. For other measures, such as residential LED lighting, we believe the replaced equipment may be more efficient on average than the TRM baseline. For other measures, like dehumidifiers, we suspect of fraction of homes may not have contained the appliance at all prior to participation.
    - ✓ For market opportunity measures, code-minimum baseline is industry standard practice. What participants would have purchased absent program incentives is an attribution question and outside the scope of this study.
- **The analysis of custom projects returns noisy results at the customer level.** This evens out with aggregation, but in a Guaranteed Saving Product program design – where the participant incentive is determined via meter-based methods – the noisiness of individual customer regressions creates settlement risk. We believe it would be important for customers to see and understand the mathematical model of their energy use that will be used to measure savings and the associated uncertainty to avoid potentially contentious situations over rebate levels.
    - ✓ Adjusting for non-routine events could help reduce the noisiness and settlement risk, but this works against the automated advantages of meter-based methods.
- **Vermont is not a large state, and we are using a method where suitability is a function of sample size.** The limited number of measures or program offerings that achieve both large electric impacts and high participation may impose a ceiling on the applicability of Advanced M&V methods for prescriptive measures.

- **The four-period weighted averages and the granular 8760 valuation profiles tend to produce similar lifetime avoided energy and emissions benefits, unless the shape of the savings differs greatly from the shape of the system load.** These fluctuations are due to differences between the shape of savings and the shape of the system load. Figure 2 compares the four-period and 8760 estimates for all prescriptive measures.
  - ✓ To better capture the time value of efficiency, Vermont could move to more granular load shapes or develop separate four-period avoided costs by measure using more granular load shapes.
  - ✓ By far the most important drivers of lifetime avoided energy benefits are the amount of energy saved annually and the measure life, so these metrics should always be top priority.

Figure 2: Lifetime Avoided Energy Benefits and Lifetime Avoided CO_2 Emissions

Measure	Lifetime Avoided Energy Benefits ($2021)			Lifetime Avoided CO_2 Emissions (Short Tons)		
	8760	4-Period	Percent Difference	8760	4-Period	Percent Difference
BLPM Circulator Pump - Commercial	-$4,987	-$4,398	12%	-31.74	-30.94	3%
Variable Speed Cold Climate Heat Pump	-$1,352	-$1,253	7%	-9.98	-9.92	1%
Heat Pump Water Heater (Fossil Fuel Baseline)	-$515	-$497	4%	-4.20	-4.12	2%
ENERGY STAR Dehumidifier	-$278	-$290	-5%	-2.70	-2.78	-3%
Prescriptive LED Interior Lighting	-$26	-$33	-29%	-0.19	-0.28	-47%
Advanced Thermostat	$15	$8	45%	0.07	0.03	55%
ENERGY STAR Clothes Washer	$16	$12	26%	-0.02	-0.04	-136%
LED Screw Base Lamp	$35	$30	16%	0.23	0.17	27%
Upstream LED Interior Lighting	$51	$53	-3%	0.45	0.46	-3%
LED Indoor Fixture	$76	$65	15%	0.60	0.53	11%
Home Performance with ENERGY STAR	$78	$59	24%	0.34	0.30	13%
High Efficiency Pool Pump	$115	$128	-11%	1.19	1.28	-7%
BLPM Boiler Motor - Residential	$208	$205	1%	1.68	1.58	6%
ENERGY STAR Clothes Dryer	$215	$211	2%	1.93	1.79	7%
Refrigerator	$337	$341	-1%	2.78	2.75	1%
Heat Pump Water Heater (Electric Baseline)	$395	$410	-4%	3.58	3.58	0%
LED Exterior Fixtures	$608	$627	-3%	5.68	5.44	4%
Commercial Refrigerator	$5,022	$5,206	-4%	44.81	44.67	0%
LED Agricultural Interior Fixtures	$19,272	$18,779	3%	152.77	152.10	0%

- **More detailed load shapes can only improve the accuracy of capacity benefits and planning considerations around peak loads.** Our analysis results suggest the need to review the current peak demand impacts for residential ductless heat pumps. Our meter-based estimates of annual kWh consumption and winter peak load increases from cold-climate ductless heat pumps were stable across methods, precise, and around 50-60% of the assumptions in the EVT TRM. Table 3 summarizes the electric penalties for two common size ductless heat pumps from the residential retrofit and market opportunity TRM entries.[5] Our pre-post TOWT results suggest a penalty of 0.35 kW during the winter peak and approximately 1,500 kWh annually, which are lower than the assumed winter peak load and energy increases characterized in the

[5] In the EVT TRM, ductless heat pump installations are claimed in two parts. The retrofit measure characterizes the switch from fossil fuel heating to a code minimum heat pump and the impacts are claimed by Vermont Distribution Utilities' Tier III programs. The market opportunity measure is claimed by EVT and represents the electric efficiency savings between a code minimum and high efficiency heat pump. Impacts must be summed across the two measures to arrive at the TRM's estimate of the impact of the ductless heat pump on the grid.

EVT TRM. This result suggests that growth in winter peak demand and electric energy consumption from aggressive electrification efforts in Vermont may be more gradual than current planning assumptions suggest.

Table 3: Retrofit Variable Speed Cold Climate Heat Pump TRM Characterization

Type	Capacity	kWh Penalty	Summer kW Penalty	Winter kW Penalty
Single Zone	18,000	2,310	0.10	0.64
Single Zone	24,000	2,961	0.13	0.81
Multi Zone	18,000	2,193	0.10	0.61
Multi Zone	24,000	2,754	0.13	0.76
TRM Average		**2,555**	**0.12**	**0.70**
Pre-Post TOWT Modeling Result		**1,539**	**0.11**	**0.35**

2 INTRODUCTION

2.1 ADVANCED M&V OVERVIEW

The advantages of an Advanced M&V analysis framework are conceptually simple. In a jurisdiction with high AMI penetration, most program participants have revenue grade meters measuring electricity consumption at hourly, or sub-hourly, intervals. Using this metering infrastructure along with statistical procedures to control for the effects of weather and other factors offers the following potential advantages:

- **Eliminates the need for sampling.** Analyzing a census of program participants using traditional M&V methods is cost-prohibitive so evaluation contractors use statistical sampling to develop estimates of the parameters of interest. With Advanced M&V methods, there is very little incremental cost associated with analyzing additional premises.

- **Reduces the M&V burden on participants.** Sending technicians to homes and businesses to install and remove meters requires time and coordination. If M&V can be conducted remotely using the revenue meter, it creates less of an "ask" on program participants.

- **Enables program administrators to look beyond the average customer.** The use of average savings assumptions to set rebate levels and claim savings is standard industry practice but can create inefficiencies and misaligned incentives. In practice, some customers save more and some save less given the same measure. Additionally, the timing of savings can vary considerably which has important implications for valuation. An M&V approach that can identify these differences and segment customers creates powerful opportunities for targeting high-value customers and 'pay for performance' program designs.

- **The ability to embed M&V into program delivery.** Once the data transfer protocols, cleaning procedures, and methods are established, Advanced M&V methods can be conducted on a rolling basis. Savings performance still needs to be observed over a range of conditions to develop a reliable estimate of normalized savings, but the analysis and reporting can be largely automated.

- **Granular impact estimates.** Time-differentiated savings estimates are useful for a wide range of planning and valuation functions.

While Advanced M&V methods offer a number of advantages, it is important to remain mindful of the limitations or tradeoffs. The ability to measure impacts is a function of four key factors:

1) **The effect or signal size.** The effect size is easiest to understand as the percent change. It is easier to detect large changes than it is to detect small ones. This difficulty is immediately present when trying to disaggregate a small change within the large and variable load at a specific site.
2) **Data volatility or noise.** The more volatile the load, the more difficult it is to detect changes in energy use.

3) **The ability to explain variation and filter out noise.** Statistical regression models, control groups and other techniques, no matter how simple or complex, are tools to filter out noise (or explain variation) allowing the effect to be more easily detected. The use of hourly load and weather data allows us to build better models that filter out more noise than daily or monthly models. However, there are sites where the loads are not weather dependent, and these sites may not perform as well under this Advanced M&V methodology.

4) **Population size.** Aggregation of impacts across homes and businesses improves the precision of estimates and allows us to measure smaller impacts. However, it is important to understand and quantify uncertainty at the individual customer level if site-specific results are being considered for settlement.

There are various industry guidelines regarding acceptable use of Option C techniques. One common guideline is that the expected impact is at least 10% of premise load. This guideline deals with item #1 in the list above but ignores item #2. Ultimately, the "signal to noise" ratio is what matters. DSA calculates the relative precision, or fractional savings uncertainty (FSU), for each premise as part of our standard procedure. This metric uses the uncertainty (margin of error) as the numerator and the impact as the denominator. Any value greater than 100% indicates that the impact estimate is not statistically significant. California took a reasonable early position in requiring FSU values of less than or equal to 50% when Advanced M&V methods are used.

Advanced M&V methods produce estimates of the change in energy consumption at the meter. This is most appropriate for measures that utilize an early replacement or retrofit path where the baseline is the existing conditions. For market opportunity measures (or "replace-on-burnout" measures), the baseline is a code minimum piece of equipment that was never installed in the home or business. At the project kickoff meeting, we discussed the fact that market opportunity measures are a large share of EVT's measure catalog. Rather than exclude this entire category of measures from study up front, we decided to include some market opportunity measures in the analysis. In most cases, Advanced M&V methods should overstate the impact for these measures because the replaced equipment was less efficient than the code minimum baseline equipment. However, our analysis results show this was not always the case for market opportunity measures. Section 5.2 includes a dedicated discussion of market opportunity results.

2.2 PROJECT OBJECTIVES

The study included the following analysis components. Each element was included to answer one or more of the PSD's Advanced M&V research objectives:

- Implementation of the LBNL Time of Week and Temperature (TOWT) piecewise linear regression model as the base pre-post modeling framework. This approach returns impact estimates for each of the 8760 hours in a year and these granular estimates are useful for a variety of planning and valuation applications.

- Comparison of lifetime avoided energy and emissions impacts using granular 8760 impacts to the four-period definition currently used in Vermont. Avoided cost assumptions come from Avoided Energy Supply Costs in New England 2021 study (AESC 2021).[6]

- Testing of panel regression models to estimate the average impact across a pool of homes or businesses versus aggregation of results from individual-customer regression models.

- Estimating results using a matched control group for comparison with pre-post modeling results. A pre-post modeling framework assumes that the only change within a home or business between the baseline and reporting periods is the energy efficiency installation and weather. The current COVID-19 pandemic provided a salient reminder of the importance of controlling for exogenous factors.

- Clear reporting of uncertainty at multiple levels of aggregation. Modeling outputs from high frequency AMI data must be adjusted to account for autocorrelation. Aggregation of errors improves the precision of savings estimates, but it is important to understand the ability of Advanced M&V methods to measure impacts at the site-level, especially if Advanced M&V is being considered as a settlement mechanism with commercial customers in a "guaranteed savings product".

- Comparison of Advanced M&V practices with traditional practices. This tasks seeks to understand how widely applicable Advanced M&V methods are in Vermont, which types of measures or programs work well, which ones don't, and seeks to quantify potential cost savings from meter-based methods.

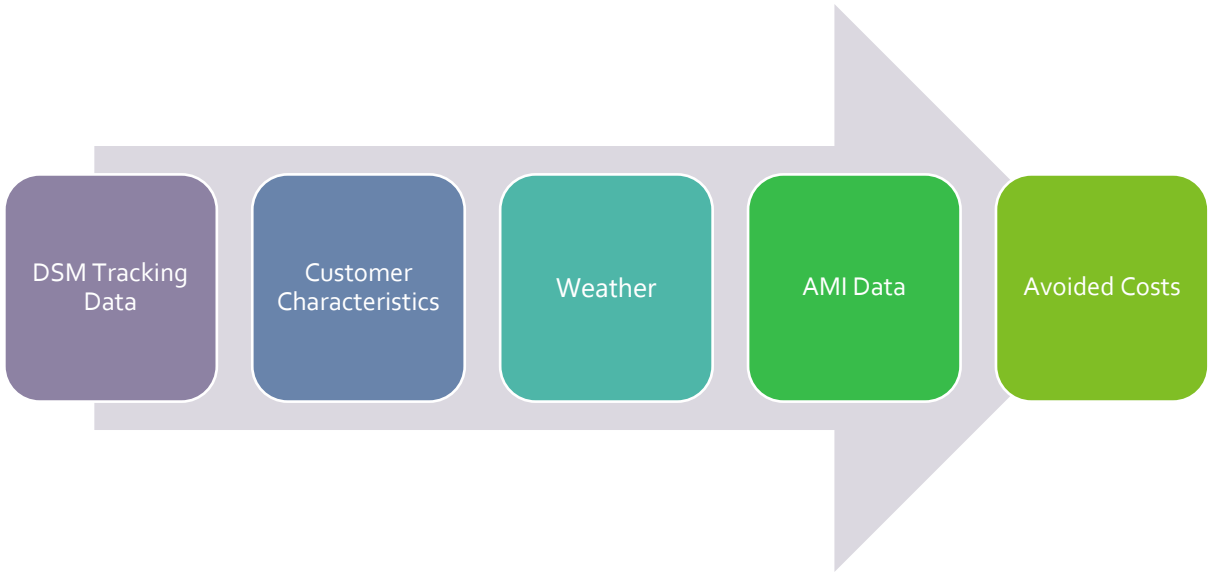
[6] https://www.synapse-energy.com/project/aesc-2021-materials

3 METHODOLOGY

3.1 DATA SOURCES

Figure 3 lists the five primary data streams used to complete the analysis. We discuss each item in detail below the figure.

Figure 3: Primary Data Requirements



DSM Tracking Data

DSM tracking data is the foundation of the analysis. DSA used EVT tracking data to identify homes and businesses for inclusion in the analysis. EVT provided tracking databases for 2017, 2018, 2019, and 2020. Fundamentally, we used the DSM tracking data to determine:

- Participants in an EVT program and what measure(s) they installed. For this analysis, we looked for sites that installed a measure of interest. Additionally, we noted whether the site installed additional measures. The installation of additional measures may confound the analysis of the measure(s) of interest.
- The claimed savings, including the expected annual energy, summer peak demand, and winter peak demand impacts from the installation.
  - ✓ **Expected Annual Energy (kWh/year):** Aggregate energy savings across the year.
  - ✓ **Summer Peak Demand (Summer kW):** Average demand reduction from 1PM – 5PM on non-holiday, weekdays in June through August.
  - ✓ **Winter Peak Demand (Winter kW):** Average demand reduction from 5PM – 7PM on non-holiday, weekdays in December through January.

- In-service date. Used to separate the baseline, or "pre" period from the reporting, or "post" period.

- Unique identifiers. Once we selected the customers of interest in the DSM tracking data, the site ID linked the customer to their source ID(s) in the customer characteristics file.

Customer Characteristics

EVT provided some basic information about the approximately 250,000 service points in the AMI dataset based on an extract from Green Mountain Power's customer information system. This information ultimately informed the matching model specification. Key fields from this dataset included:

- **Source ID**. Uniquely identifies the sites' metering to the utility.

- **Rate Code**. This is the tariff[7] under which electric usage from the meter is billed and any supplemental riders in place for the account. Knowing the rate code of each service point is useful when developing matched control groups, as it allows us to match residential accounts with other residential accounts and commercial accounts with other commercial accounts and potentially control for factors like solar PV and time-varying pricing.

- **Usage Type**. A categorical variable that defines whether the site is recorded as commercial or residential, as well as the level of usage for the type of site (e.g. Non-High Use Residential, Commercial 500k to 1Mil, Large Commercial or Industrial).

- **Service start and end date**. For this type of analysis, we want to analyze the performance of energy efficiency measures with the occupant held constant. If there is a change in ownership/tenancy during the baseline or reporting period, it is hard to separate changes in energy use attributable to energy efficiency from behavior changes in the occupants.

- **Location**. The service address, including county, city and, zip code of the meter.

- **Generation Description**. This flags sites with net metering.

- **Firmographics**. For commercial accounts, we received the primary business activity in the form of SIC codes.

AMI Data

The third component shown in Figure 3 is the AMI data itself. EVT provided access to a cloud-based SQL database with meter reads dating back to January 1, 2017. DSA aggregated the 15-minute interval data from Green Mountain Power electric meters into hourly interval data for analysis. EVT also supplied a mapping file that associates sites (from the tracking data) to service points (in the AMI data), but the mapping of participants to AMI data is incomplete. The incomplete mapping is not unexpected

[7] GMP tariff sheet. https://greenmountainpower.com/wp-content/uploads/2019/03/Rate-Descriptions-1.pdf

with a non-utility program administrator, but it does create challenges for control group development. Specifically:

- We received site ID to service point ID mapping information for around 50% of the GMP site IDs who participated in DSM between 2017 and 2020. While we know which service points do not have *mapped* participation, this does not mean that these sites do not have *unmapped* participation. Consequently, our control group matching pool is not a pure list of non-participants.
- GMP runs their own fuel savings and transformation programs that include electrification measures, such as heat pumps, EVs, and battery storage. This participation is not captured in the DSM tracking data we received from EVT. Identification of these participants and removing them from the control pool would improve matching procedures.
- EVT also operates midstream and retail buy-down offerings that do not include site information.

Weather Data

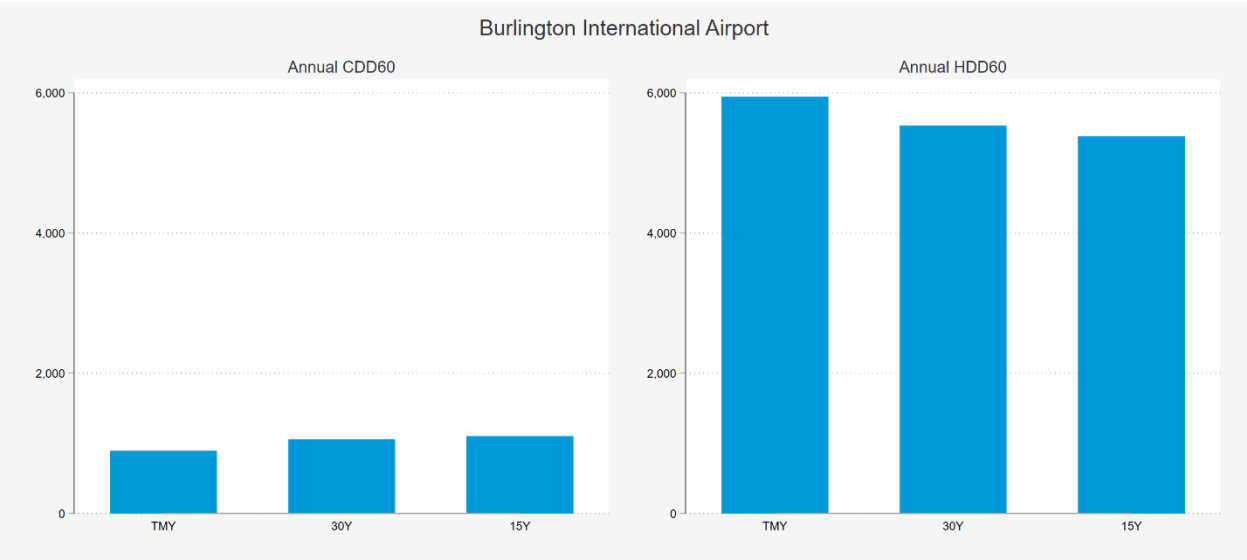
Two types of weather data are required for this analysis: historic and normal. All historic and normal weather data came from the only Class 1 weather station in the state, Burlington International Airport. Historic data is the actual hourly dry bulb temperature value recorded at the weather station and is used to estimate the regression models. Climate normal weather data is a projection of hourly weather data for the area based on long-run trends. For this project, we tested three potential definitions of "normal" weather: TMY3 weather data, 30-year climate normals, and 15-year climate normals. TMY data is provided by the National Renewable Energy Laboratory. TMY, which stands for "typical meteorological year", is meant to represent the median weather conditions for each hour in a calendar year at a given location.[8] TMY3 is the latest iteration of TMY data, which covers the years 1991-2005. The TMY3 data is the most popular data source for normal weather.

US climate normals also provide information on standard climate conditions at weather stations throughout the US[9]. The National Centers for Environmental Information develop 30-year and 15-year weather normals every ten years, with the most recent update completed in May 2021. The 30-year normals cover the years 1991-2020, while the 15-year normals cover 2006-2020. Figure 4 provides a comparison of the three normal weather data sources, using annual cooling and heating degree days (base 60 degrees F). TMY3 has the fewest cooling degree days and most heating degree days, followed by 30-year climate normals, and then 15-year normals with the most cooling days and least heating days. This pattern also follows the time range of each dataset, TMY3 data is based on older years while 15-year normals are the most recent.

[8] https://www.nrel.gov/docs/fy08osti/43156.pdf

[9] https://www.ncei.noaa.gov/products/us-climate-normals

Figure 4: Weather Normals Comparison



After consultation with PSD staff, we elected to use 30-year normals (1991-2020) for all results presented in this report. We use the climate normal data for predictions. After running the regression models using actual weather data, we store the regression coefficients and predict energy consumption under typical weather conditions. This approach ensures that the baseline and efficient case rely on identical weather assumptions.

Avoided Costs and Emissions Rates

One of the key drivers for this study was a desire to better understand the time value of efficiency. Wholesale energy costs and marginal emission rates can vary significantly from hour to hour and day to day depending on system load and other factors. The current four-period valuation framework relies on system load-weighted averages that have the potential to undervalue some measures and overvalue others. Our modeling approach returns 8760 impact estimates, so we leverage the 8760 Vermont avoided cost assumptions from AESC 2021 for this task. AESC 2021 presents multiple scenarios, or counterfactuals. We use the outputs of Counterfactual #3, which assumes a future in which program administrators install no new energy efficiency resources in 2021 or later years. To align with current practices in Vermont, we use "retail" avoided costs which include:

- Renewable Energy Credit (REC) cost to load
- Line losses of 9%
- Wholesale Risk Premium (WRP) of 4% during summer months and 0% during winter months

Where:

$$Retail\ Avoided\ Cost = (Wholesale\ Avoided\ Cost + REC) * (1 + Losses) * (1 + WRP)$$

AESC 2021 also provides a 15-year forecast (2021-2035) of 8760 marginal emission rates for region (short tons of CO_2 per MWh). The value of avoided carbon dioxide emissions is an often-discussed

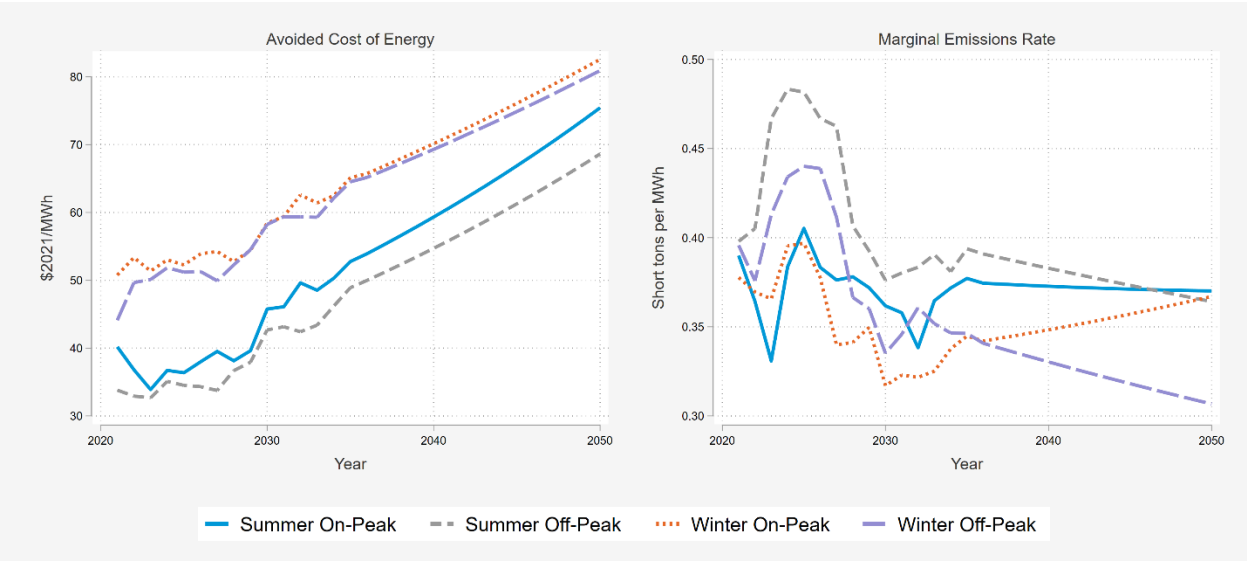
policy issue and AESC 2021 provides avoided costs using multiple methods. For this study, we focus on the quantity of emissions and do not monetize the values. Figure 5 shows the average marginal emission rate for hour ending 11 by month over a 15-year horizon. During shoulder months, we see a clear reduction in the project emissions rate, presumably due to increased penetration of solar photovoltaic generation in the region over time.

Figure 5: Average Marginal Emission Rate by Year and Month: Hour Ending 11

Year	January	February	March	April	May	June	July	August	September	October	November	December
2021	0.40	0.40	0.39	0.28	0.34	0.41	0.40	0.42	0.39	0.38	0.34	0.50
2022	0.42	0.38	0.37	0.26	0.38	0.35	0.40	0.40	0.37	0.32	0.38	0.38
2023	0.40	0.40	0.38	0.37	0.38	0.38	0.40	0.40	0.38	0.30	0.39	0.39
2024	0.43	0.37	0.28	0.18	0.36	0.36	0.46	0.41	0.41	0.40	0.44	0.46
2025	0.41	0.30	0.18	0.17	0.37	0.33	0.43	0.39	0.38	0.25	0.45	0.47
2026	0.36	0.25	0.14	0.17	0.23	0.24	0.38	0.36	0.26	0.24	0.41	0.45
2027	0.33	0.24	0.14	0.03	0.20	0.22	0.35	0.35	0.24	0.34	0.33	0.41
2028	0.29	0.18	0.11	0.03	0.18	0.18	0.32	0.31	0.18	0.17	0.31	0.38
2029	0.24	0.16	0.10	0.09	0.14	0.18	0.29	0.30	0.18	0.19	0.36	0.39
2030	0.23	0.16	0.10	0.02	0.13	0.17	0.28	0.27	0.17	0.27	0.28	0.36
2031	0.22	0.16	0.10	0.02	0.15	0.17	0.29	0.26	0.15	0.16	0.26	0.35
2032	0.22	0.15	0.10	0.10	0.13	0.16	0.29	0.28	0.15	0.19	0.30	0.35
2033	0.24	0.15	0.11	0.02	0.14	0.16	0.31	0.30	0.18	0.29	0.28	0.37
2034	0.24	0.15	0.10	0.03	0.17	0.18	0.30	0.30	0.18	0.17	0.31	0.37
2035	0.24	0.16	0.10	0.06	0.16	0.19	0.33	0.31	0.23	0.27	0.39	0.40

To calculate the lifetime avoided cost of energy and avoided emissions, our valuation streams must extend far enough in the future to cover the full effective useful life of each measure. Consistent with AESC 2021, we use the compound annual growth rate from 2030 to 2035 to project benefit streams through 2050. Figure 6 shows the four-period profiles for Vermont used for this analysis.

Figure 6: 30-Year Valuation Streams by Costing Period



The four costing periods shown in Figure 6 are defined as:

- **Summer On-Peak**: 7AM to 11PM on weekdays June-September
- **Summer Off-Peak**: 11PM to 7AM on weekdays June-September. All weekend hours June-September.
- **Winter On-Peak**: 7AM to 11PM on weekdays October-May
- **Winter Off-Peak**: 11PM to 7AM on weekdays October-May. All weekend hours October-May.

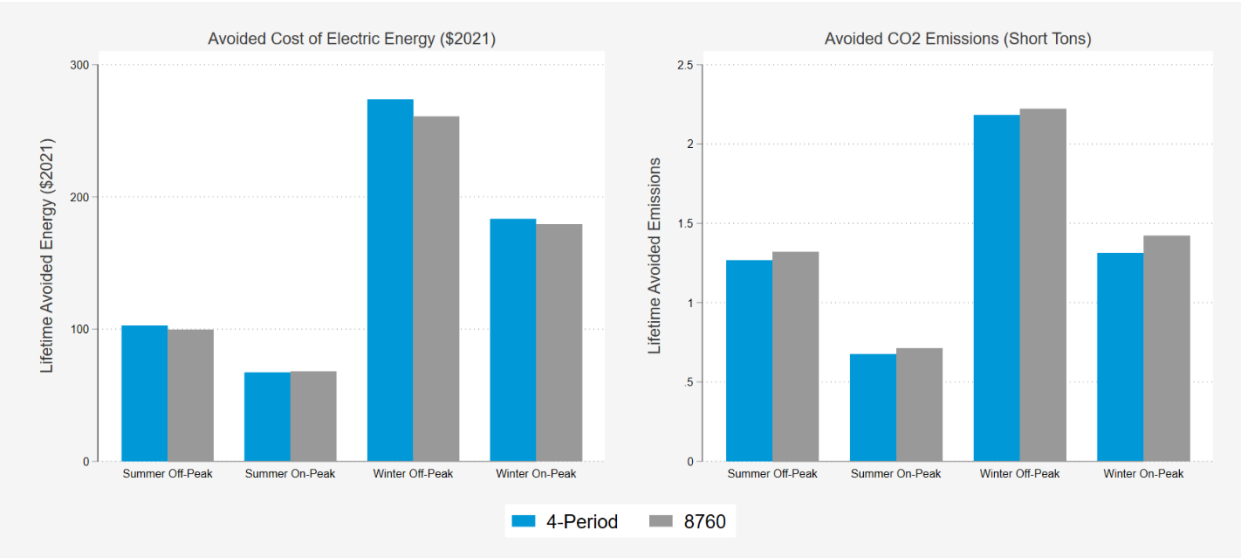
For each measure analyzed in Section 4, we compute the net present value of avoided energy costs and the lifetime avoided CO_2 emissions using our 8760 impact load shapes and the 8760 valuation profiles from AESC 2021. For comparison, we collapse the 8760 impact load shape into a four-period load shape, compute the NPV avoided energy costs, and lifetime avoided CO_2 emissions using the four-period valuation profiles from AESC 2021. Table 4 shows the results for the commercial LED exterior fixture measure. This measure exhibits a clear "dusk to dawn" savings profile, which is quite different from the shape of Vermont's system load. For this measure, use of 4-period valuation profiles based on system load returns estimates of avoided costs that are 3.1% higher and CO_2 emissions that are 4.2% lower than the granular 8760 profiles.

Table 4: Valuation Comparison for LED Exterior Fixtures (EUL = 15 years)

Costing Period	Annual kWh Savings	4-period Avoided Costs	8760 Avoided Costs	% Difference Avoided Energy	4-Period CO2 Emissions Lifetime	8760 CO2 Emissions Lifetime	% Difference Avoided Emissions
Summer Off	174	$103	$100	-3.1%	1.3	1.3	4.0%
Summer On	104	$67	$68	1.1%	0.7	0.7	5.4%
Winter Off	330	$274	$261	-4.9%	2.2	2.2	1.8%
Winter On	214	$183	$179	-2.2%	1.3	1.4	7.6%
Total	**823**	**$627**	**$608**	**-3.1%**	**5.4**	**5.7**	**4.2%**

In Section 4, we present this comparison in a graphical format. Figure 7 shows the results for the commercial LED exterior fixtures measure.

Figure 7: Valuation Comparison for LED Exterior Fixtures



EFFICIENCY VERMONT TECHNICAL REFERENCE MANUAL

For prescriptive measures, the energy and peak demand savings values stored in the EVT tracking database are generally based on measure characterizations in the Technical Reference Manual (TRM). The TRM contains a variety of assumptions regarding equipment capacity, efficiency, baseline, load shape, and measure life in an intuitive online format. TRM access allowed the analysis team to confirm our understanding of the measures of interest and identify changes to measure characterizations over time. There were significant changes to the residential ductless heat pump and heat pump water heater measure characterizations during the period of analysis. Understanding these changes is important context as readers interpret our results and compare our regression-based estimates with the energy impacts stored in the EVT tracking database.

3.2 SELECTING MEASURES

After receiving and processing the EVT Tracking Data, DSA took the original list of 21 measures from Appendix 1 of the RFP and ran them through five distinct stages in order to identify if the measure of interest had adequate data for modeling.

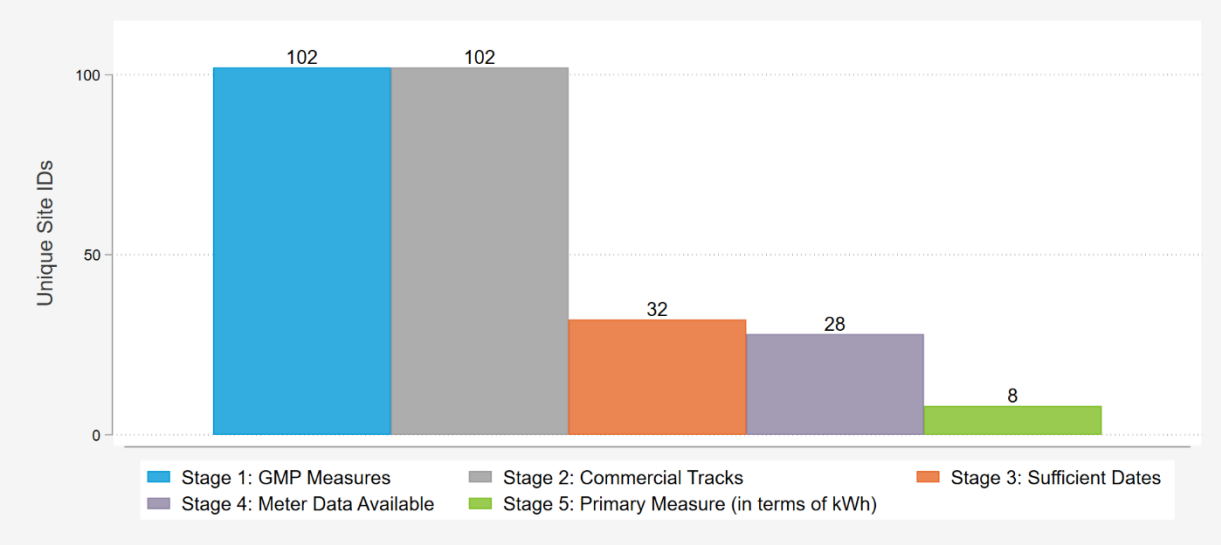
- **Stage 1:** Limited the data to only the measure of interest, as well as removed any utility providers that were not Green Mountain Power.
- **Stage 2:** Limited participation to the relevant sector (residential or commercial) using the participation "track".
- **Stage 3:** Since the AMI data begins on January 1, 2017, we removed any measures that were installed prior to January 1, 2018, and thus do not have 365 days of pre-installation data.
- **Stage 4:** Removed sites without linkage from site ID to service point ID.

- **Stage 5:** To ensure that the effect of the measure could be observed within the AMI data, DSA calculated the share of the savings that the measure of interest held in comparison to all the kWh savings associated with the site in the EVT tracking data. Sites with a share between 90% and 110% were kept in the dataset. Since some measures indicate a negative savings, this share could be above 100%.

    - ✓ For example, a site has three measures installed that produce 10, 15, and 100 kWh of annual savings. The total savings at this site is 125 kWh. If we were interested in analyzing measure three, which produced 100 kWh of savings, this site would be dropped from our analysis because it only accounts for 80% of the share of the savings at the site.

Figure 8 illustrates the process for a measure that we did not include in the final measure list. In this case, it is evident that there were not a large quantity of the outdoor, high-efficiency condensing units installed.

Figure 8: Sample Waterfall for Insufficient Measure (Outdoor High Efficiency Condesing Unit)



After applying the data-driven filters, DSA found that some of the measures would be inadequate for analysis (Table 5), so we investigated the tracking data for additional measures to analyze. The residential measures that will be used for the analysis are in Table 6, while the commercial measures are in Table 7. In total, there are 20 measures of interest listed. LED Interior Lighting is subsequently divided based upon whether the installation is commercial upstream or is a prescriptive/retrofit measure. This brings the measure list to 21 measures.

Table 5: Measures with Inadequate Data

Measure Name	Track Type	Rejection Reason
Heat Pump with Shell Improvements	Residential	Too few sites in Stage 5 (n=28)
Cold Climate HP	Commercial	Limited number of sites in Stage 5 (n=167). Much larger available sample in residential so the team elected to limit this measure to residential.
Fixture-Mount Dual Occupancy and Daylight Sensor	Commercial	No applicable sites, due to tendency of sites to install controls and fixtures together
Efficient Blower Fan	Commercial	Too few sites in Stage 5 (n=41)
Synchronous Motor Evaporator Fan	Commercial	No applicable sites, due to low share of savings
Outdoor High Efficiency Condensing Unit	Commercial	Too few sites in Stage 5 (n=8)

Table 6: Residential Measure List

Measure Name	Stage 5 Sites (n)	Vintage
Advanced Thermostat	999	Retrofit
BLPM Circulator Motor	863	Retrofit
ENERGY STAR Dehumidifier	745	Market Opportunity
ENERGY STAR Clothes Dryer	234	Market Opportunity
ENERGY STAR Clothes Washer	683	Market Opportunity
Heat Pump Water Heater – Positive Savings	909	Market Opportunity
Heat Pump Water Heater – Negative Savings	291	Market Opportunity + Fuel Switch
Variable Speed Cold Climate Heat Pump without Shell Improvements	3,462	Market Opportunity + Fuel Switch
LED Indoor Fixtures	692	Retrofit
LED Screw Base Lamp	3,107	Market Opportunity
Home Performance with Energy Star	635	Retrofit
Pool Pump	554	Retrofit
Residential Refrigerator	1,048	Market Opportunity

Table 7: Commercial Measure List

Measure Name	Stage 5 Sites (n)	Vintage
Prescriptive Compressed Air	9	Retrofit
Commercial Refrigeration	85	Retrofit
BLPM Circulator Pump	124	Retrofit
LED Exterior Fixtures	251	Retrofit
LED Agricultural Interior Fixtures	70	Retrofit
LED Interior Lighting – Building Type	1,557	Retrofit

The "vintage" column in Table 6 and Table 7 describes the underlying theory of the measure decision type and baseline assumption. Table 8 is adapted from the EVT TRM and provides an overview of the five core measure vintages used across EVT's energy efficiency programming.

Table 8: Overview of Measure Calculation Types

Vintage	Definition	Baseline Case	Efficiency Case
Market Opportunity: Time of Sale	The participant is assumed to require purchase of a new piece of equipment and is incentivized to purchase or install higher efficiency equipment than they would have done if the program had not existed.	New base level equipment, often corresponding to a federal standard or at a level representing standard industry practice.	New, high efficiency equipment meeting a program specified level.
Market Opportunity: New Construction	A program that intervenes during building design to support the use of more-efficient equipment and construction practices.	New base level equipment at the efficiency level defined in the applicable Building Energy code, federal standard level or standard practice as derived by baseline studies.	The program's prescribed level of building specification.
Early Replacement	A program that replaces existing equipment before the end of its expected life. To qualify as early replacement, there needs to have been prior contact with the customer replacing functioning equipment – e.g. during on site audit or through prior phone/email contacts, and evidence must be provided that the unit is being replaced to achieve energy savings.	Dual; for the expected remaining useful life of the existing equipment the baseline is the efficiency of the existing equipment and then shifts to represent new baseline equipment.	New, high efficiency equipment meeting a program specified level.
Retrofit	A program that upgrades or enhances existing equipment.	Existing equipment or the existing condition of the building or equipment. A single baseline applies over the measure's life.	Either new, high efficiency equipment or modifications of existing equipment to make it operate more efficiently.
Early Retirement	A program that retires duplicative equipment before its expected life is over.	The existing equipment, which is retired and not replaced.	Since the unit is retired, the efficient case consumption will be zero.

Market opportunity measures present a challenge for meter-based methods because the code minimum baseline equipment was never installed in the home or business. In the case of time-of-sale measures, there are two potential scenarios:

1) The revenue meter often observed consumption of the previous equipment, which is assumed to have reached the end of its useful life, during the baseline period. Typically we assume that a new code minimum piece of equipment is more efficient than the previous equipment, but this information is rarely collected so it is difficult to estimate relative efficiencies.

2) The home or business may not have had a certain appliance during the baseline period and the program-supported equipment is a new load. Absent program intervention, the added load would have been larger.

A pre-post regression analysis estimates the change in energy consumption at the meter. In the case of scenario #1, we expect our estimates to over-estimate savings. In the case of scenario #2, we expect our estimates to under-estimate savings and predict a load increase. New construction measures are especially problematic for meter-based methods because there is typically no baseline meter data available to analyze.

While retrofit measures are best-suited to Advanced M&V methods, we elected to pursue several market opportunity measures in the residential sector. Section 4.2 gives greater insight into how these measure-types performed.

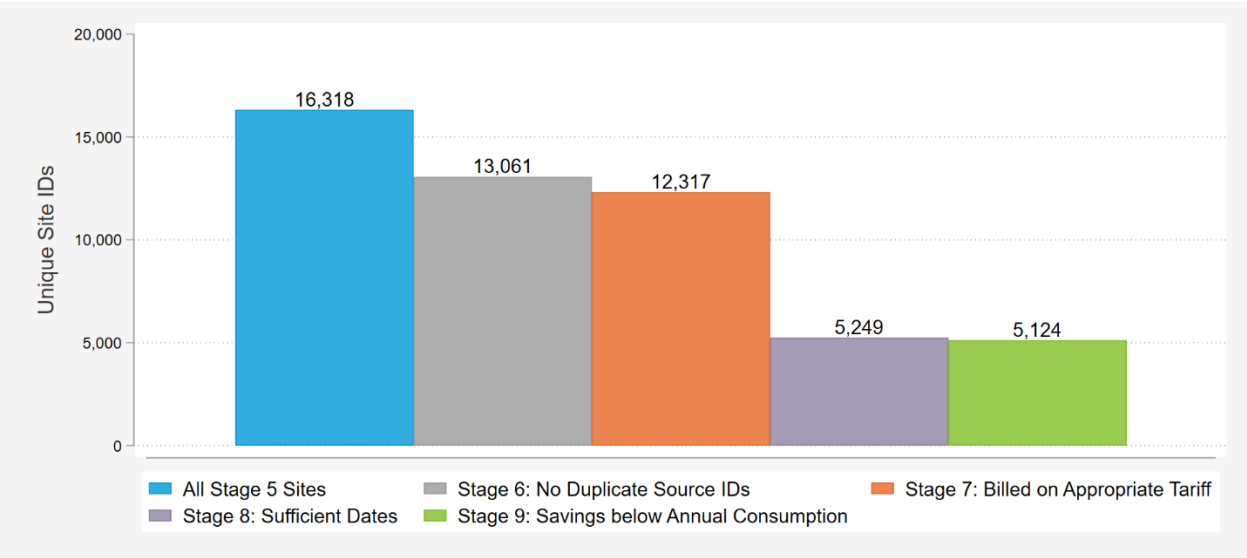
Selecting Final Prescriptive Sites

After filtering down to the stage 5 sites detailed in Section 3.2, which we were able to do exclusively using the DSM tracking data and the site ID to source ID mapping, it was necessary to refine the sites used in our analysis through a couple additional levels of AMI filtering. We begin by creating a buffer period around each installation date. For commercial sites, we remove one month of data prior to installation and one month after installation. Likewise, for residential sites, we do the same but with two weeks. This is to account for any error that may be present within logging the installation date or any measures that took more than a single day to install. This period is shorter for residential sites because installation of these measures is usually simpler and quicker. After this, we continue filtering through four additional stages.

- **Stage 6**: Drops sites where the same source ID is attached to multiple site IDs. When this occurs, we are unable to verify which site the AMI is actually metering. Using these source IDs multiple times across sites would inevitably be wrong for at least one of the sites.
- **Stage 7**: Removes sites that were tagged as receiving a measure within the appropriate track but are billed using the opposite tariff (e.g. a customer received a dehumidifer that was designated within a residential track but the customer is billed on a commercial tariff).
- **Stage 8**: Removes sites where a customer was not connected for a full year prior to their pre-installation buffer period and/or their post-installation buffer period. For these sites, we would not be able to adequately construct a predicted baseline and/or a predicted post-period.
- **Stage 9**: Drops sites that have a database savings higher than the pre-period annual consumption.

Figure 9 gives us insight into how stage 6 through stage 9 remove additional sites, across the 20 prescriptive measures that we analyze.

Figure 9: Waterfall of Prescriptive Sites



Additionally, commercial sites are subsequentlty filtered based upon their uncertainty prior to aggregation. This uncertainty metric applied is identical to the metric used for commercial projects that is defined in Section 3.7.

Selecting Final Custom Projects

Efficiency Vermont has five tracks that contain custom projects; 6012CNIR, 6013CUST, 6014CUST, 6017CUST, and 6020CUST. Green Mountain Power custom projects implemented during 2018 or 2019 accounted for 7.3 MW of summer demand reduction and 7.1 MW of winter demand reduction. After narrowing down the data to these five custom tracks, we ran each project through five distinct stages in order to identify if the projects of interest had adequate data for modeling.

- **Stage 1:** The tracks were narrowed down to only include 6012CNIR – Commercial and Industrial Retrofit, 6017CUST – Multi-Family Retrofit Custom, and 6020CUST – Multi-Family Market Rate Custom. The two additional tracks, 6013CUST – Custom Equipment Replacement and 6014CUST – Business New Construction Custom, were excluded because they contain projects that are poor candidates for this methodology.

- **Stage 2:** Removed sites without linkage from site ID to service point ID. With more comprehensive mapping of these linkages, the number of sites and quantity of demand reductions that pass through this phase could be expanded upon.

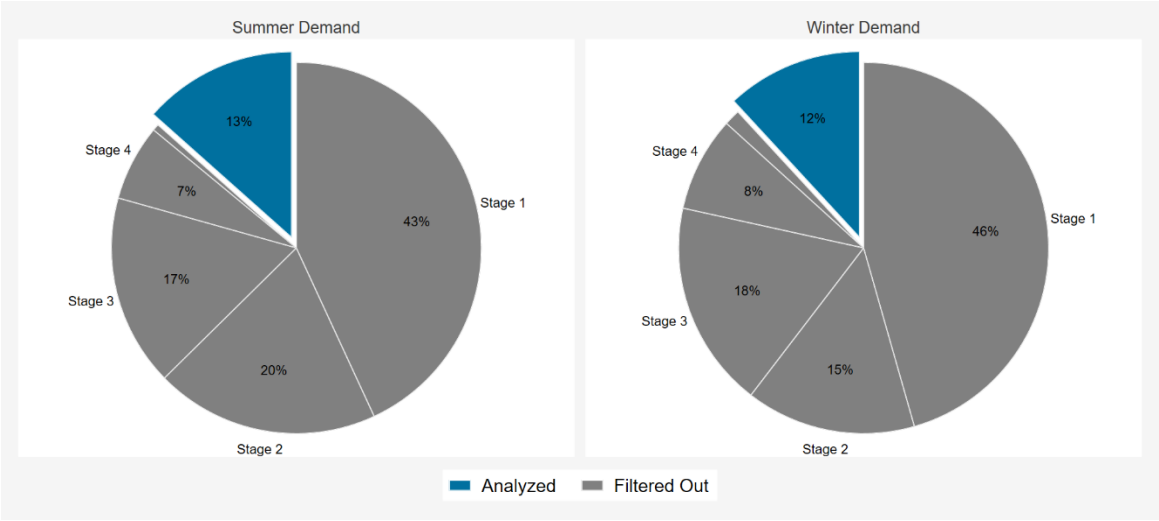
- **Stage 3:** The sites that received another intervention that accounted for more than 10% of the database claimed savings were removed from the analysis, in order to mitigate the presence of confounding factors. This is done to ensure that the impact being estimated is primarily based upon the project being analyzed.

- **Stage 4:** Sites were eliminated if they did not have an adequate amount of pre-installation and post-installation data. Due to AMI data becoming available in 2017, it would be difficult to analyze the projects implemented in 2017. In juxtaposition, projects installed in 2020 that did not have a year of post-installation data will eventually accrue enough data and be eligible for analysis.

- **Stage 5:** The final filter for custom projects is based upon the uncertainty of evaluated impacts. Using the quasi-FSU metric, which will be dicussed in Section 3.7, sites with a large uncertainty are removed from the analysis. High uncertainty indicates that the TOWT model may not adequately explain variation in energy use given the expected magnitude of savings, and thus, savings estimates are likely inaccurate.

With this full set of filters applied, Figure 10 defines the percentage of the full 7.3 summer MW and 7.1 winter MW that are analyzed using this method. While it is evident that this methodology can estimate impacts for a subset of custom projects, there is a substantial portion of projects that would not produce an accurate savings estimate.

Figure 10: GMP Custom Demand Savings Analyzed



Through this filtering of custom projects, it is evident that the stage 5 filter removes the least amount of demand. It also is the most variable, due to the discretion to change the definition of uncertainty. Without this filter, there are 124 sites that pass through the first four stages. These projects are broken down by type and the aggregate size is displayed in Table 9.

Table 9: Types of Custom Projects

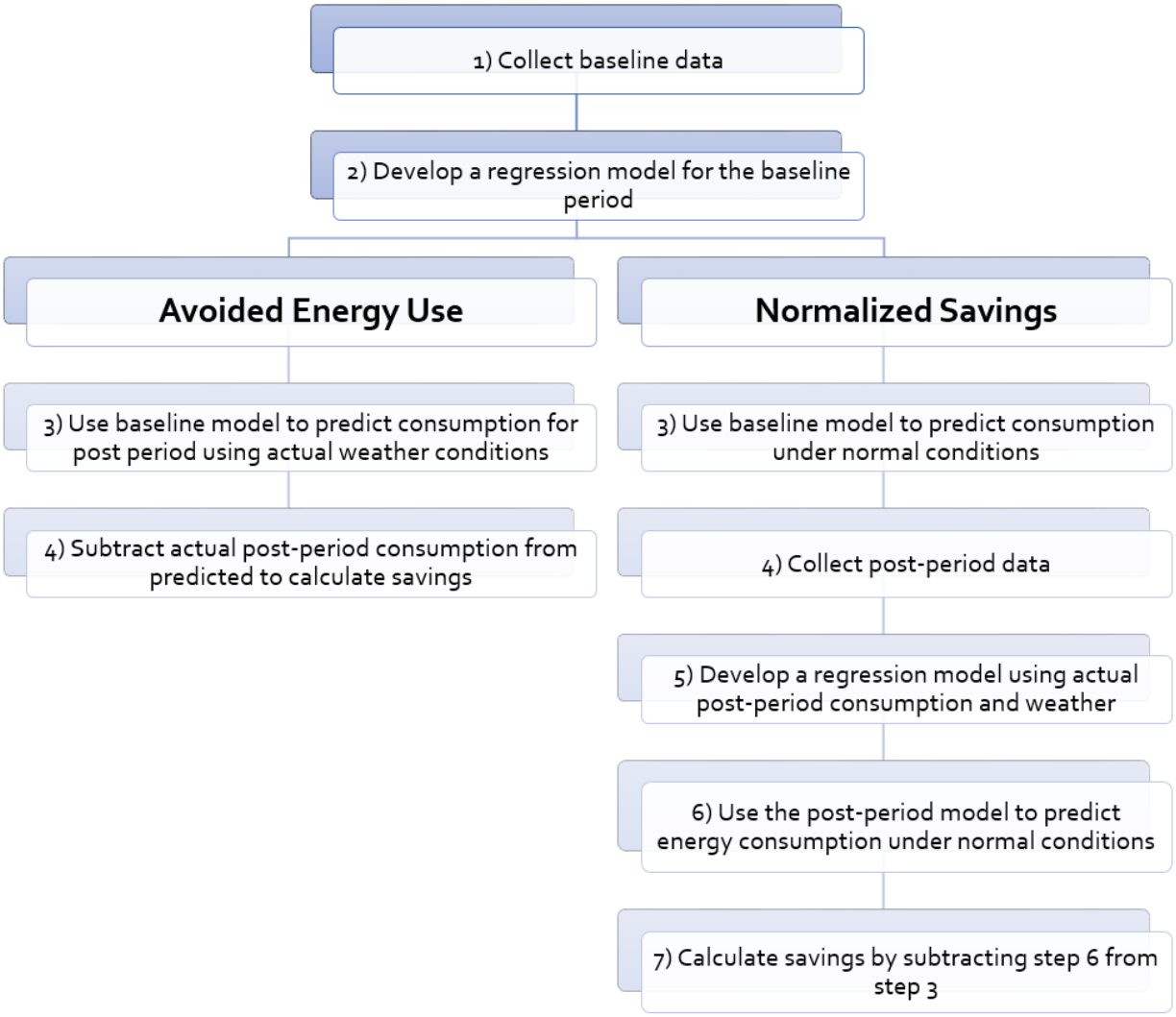
Project Type	Count	kWh/year	Summer kW	Winter kW
Lighting	57	5,192,648	633.6	511.2
Controls	9	79,798	30.7	33.1
Deep Retrofit/RCx	9	381,030	7.9	5.2
Various Measures	8	182,424	35.1	20.9
Net Metering	6	151,057	13.6	32.2
VFD	5	1,118,787	157.0	142.9
Refrigeration	4	385,425	33.6	32.3
BAS Upgrades	3	79,329	3.6	7.0
Other	23	985,588	110.1	153.5
Total	**124**	**8,556,085**	**1,025.2**	**938.2**

Various subsets of these 124 projects are analyzed, including subsets with and without COVID-19 in the post-period, as well as with and without the quasi-fsu filter. Section 4.4 presents the results of our analysis of these custom projects.

3.3 PRE POST ANALYSIS

Advanced M&V methods are a subset of the IPMVP Option C: Whole Facility approach to estimating energy savings. Figure 11 shows the two types of energy savings prescribed by IPMVP Option C methods. An avoided energy use approach produces estimates of the energy savings during the post-retrofit period using observed conditions. A normalized savings approach requires two regression models and predicts the expected impact of energy improvement under typical conditions. Our analysis procedures estimate both avoided energy use and normalized savings, but all reporting for this study focuses on normalized savings.

Figure 11: IPMVP Option C Approaches



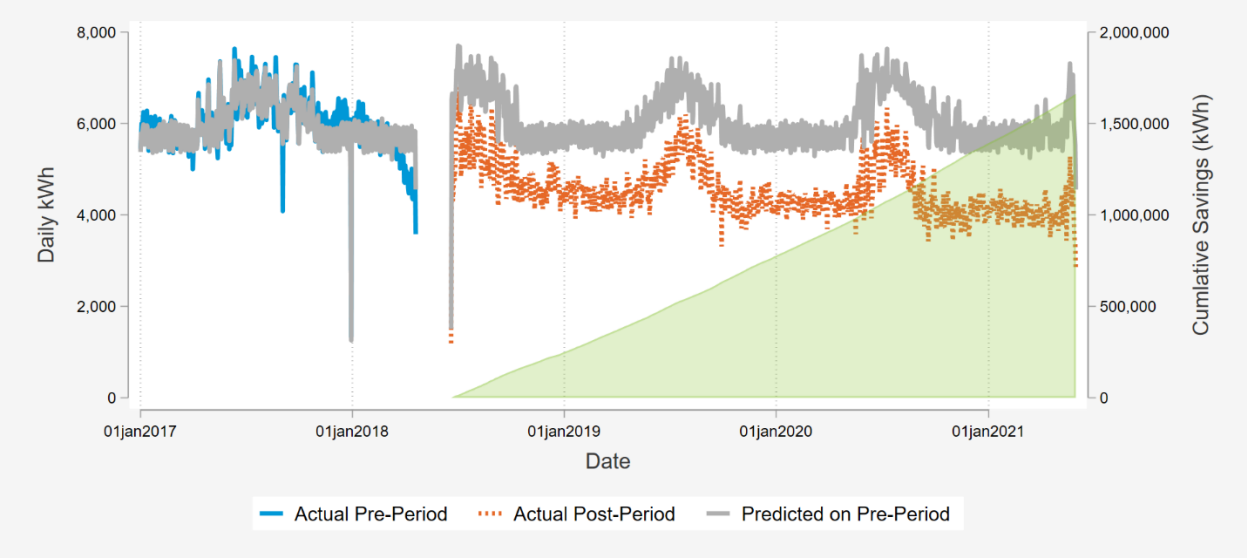
The traditional focus of Option C methods has been to estimate the total energy savings. One of the key features of using high-frequency AMI data and Advanced M&V methods is that the models produce time-differentiated energy savings estimates. Specifically, using a normalized savings approach returns an impact estimate for all 8760 hours in a year. An avoided energy use approach returns impact estimates for each hour in the post-retrofit period, which can be useful for backward looking reductions in peak loads. Armed with these granular impacts and a definition of monthly or annual peak hours, evaluator and program administrators can report on the average expected impact during specific hours of interest. These estimates are useful for a host of system-planning and valuation activities.

The general form of the regression model, as well as the set of independent variables used, follow LBNL's Time-of-Week Temperature (TOWT) Model. The dependent variable is hourly electric consumption from the GMP revenue meter. Because the relationship between outdoor air temperature and electric demand may differ depending on the temperature level, we estimated effects for six separate temperature bins, each covering a segment of approximately 19 degrees (with the total

temperature range being -19 to 96 degrees Fahrenheit). For the time-of-week variable, we used 168 categorical indicator variables – one for each hour of the week – that capture idiosyncratic patterns in occupancy and electric consumption.

Since we analyze custom projects on a site-specific basis, it is clear to see how this methodology is applied when looking at an individual site that implemented a custom project. Figure 12 displays one of the projects that performed well under this methodology. Predictions based on the pre-installation period are plotted (gray line) in addition to the actual daily consumption (blue line for pre-period and orange for post-period). In the pre-installation period, the plots sit on top of each other to display how well the model accounts for the actual variation in load. In contrast, the predicted values plotted in the post-installation period are consistently higher than the actual consumption, suggesting that the site saved energy in the post-period. The green area plot shows the accumulated difference between predicted and actual consumption in the post-period.

Figure 12: Methodological Example



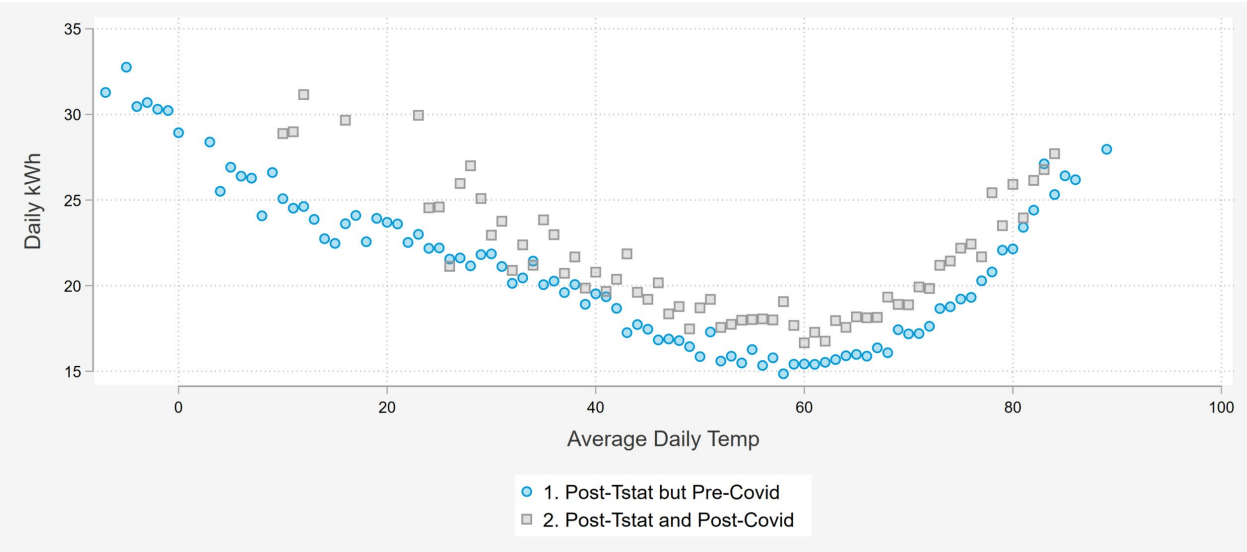
Claimed savings at this site are over 515 MWh, while the estimated annual consumption in the pre-installation period is roughly 2,000 MWh. With savings around 25% of the annual consumption, this site was flagged as a good candidate for the methodology. It does in fact perform well under this methodology, returning 512 MWh in savings, or a 99% realization rate.

3.4 MATCHING ANALYSIS

Section 3.1 laid out some of the challenges in developing a matched comparison group for the measures of interest, which represent noteworthy findings in this report. Challenges aside, the importance of testing a matched comparison group in the analysis remains. This is especially true given the effects of the COVID-19 pandemic on electric consumption. Because of the AMI data available for analysis, eligible participants are primarily limited to 2018 and 2019. For most of the 2019 participants, 12 months of post-installation data prior the pandemic is not available. Figure 13 looks at the average

daily electric consumption of Vermont homes after installing an advanced thermostat before and after March 1, 2020. The increase in consumption across all temperature conditions is clearly visible. A matched control group of residential accounts would have experienced the same shelter-in-place orders and increased work-from-home conditions as the participants that likely caused this increased consumption. To understand the effect of the measures of interest, we need to control for more than just efficiency and weather.

Figure 13: Advanced Thermostat Post-Installation Daily kWh by COVID Period



There are many ways to construct a matched comparison group, but for this study we implemented a blocked, nearest-neighbor propensity score matching process, with replacement. Matching "with replacement" means that multiple participants can have the same control that best matches with pre-period usage. Allowing for replacement improves the average quality of the matches. The more detailed matching process below is performed separately for each of the 19 measures that include a matched control group.

Matches are constructed at the meter level, due to a lack of visibility into site aggregation for our control pool. This means that if a participant has three meters, each one is matched with the best control meter, which may actually be at different locations. These meters get aggregated, much like the participants' meters, to form a matched control quasi-site.

Propensity score matching is a data pre-processing technique that identifies statistically similar non-participants for each participant. It relies on a binary outcome probit model that relates observed characteristics to whether a given meter has adopted an efficiency measure. The model used in this analysis was a propensity score model that used average daily usage (kWh) as the matching variable. Participants are then "matched" to non-participants with similar propensity scores. Effectively, propensity score matching produces a cohort of non-participants that have the same overall likelihood

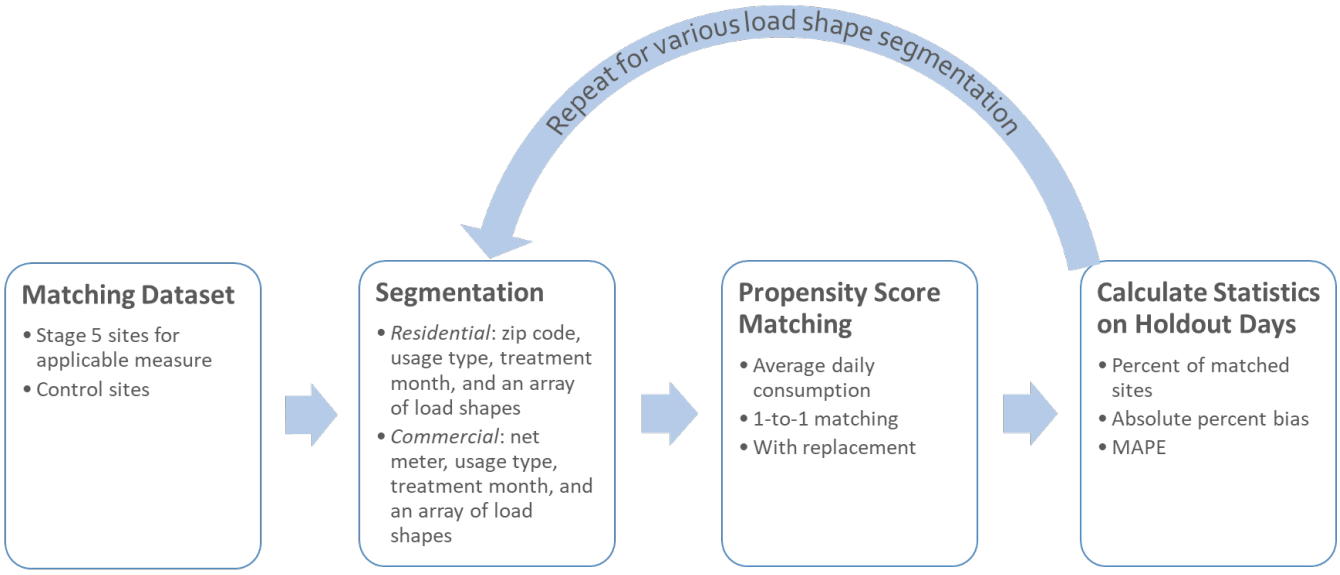
to have been treated as the participant group – the only customers that did in fact adopt the measure through an EVT program.

Since energy efficiency installations occur on a continuous basis, there is a variation in treatment timing. To ensure that the pre-period used for producing the average daily usage is clean of any potential treatment affects, a distinct pre-period is analyzed for each participant. This period is defined as the 12 months prior to installation, excluding the month of installation. From here, the 12 months of pre-period usage data for participants is compared it to an identical set of 12 months for the 196,539 control meters in the pool. This means each control candidate has average daily usage calculated 24 times, once for each of the potential pre-period intervals covered by the set of participants.

A blocked propensity score matching process performs this matching procedure for meters in each key strata separately, effectively ensuring that only participants with similar characteristics will be matched with non-participants with the same characteristics. At a minimum, we use zip code, usage type, and treatment month to segment residential customers and net metering status, usage type, and treatment month to segment commercial customers. Also, both residential and commercial meters were pre-categorized based on clustered load shape bins. To classify meters with similar behavior patterns, we create bins for monthly usage, monthly usage across the year, and average hourly usage during summer and winter seasons. In addition, we calculate the weather sensitivity of each meter and cluster it into bins. These five segmentation variables are used in various interactions, across 12 combinations.

Across these 12 tested segments, we calculate statistics based upon holdout days. These holdout days are a random sample of 3 days in the pre-installation period that are not used to calculate the average daily consumption of the meter. The full matching process is laid out in Figure 14.

Figure 14: Matching Procedures



Using these calculated statistics, the best matching model is chosen for each measure based upon three criteria:

1) Because of the segmentation combinations, not all meters were matched for each model. The model must return at least 80% matches from the original meter count. This is to ensure that an adequate amount of customers are eligible to be analyzed.

2) Models are sorted based upon the absolute percent bias, and the three models with the lowest absolute percent bias are retained.

3) The final model is selected using the lowest mean absolute percentage error.

We store the treatment meter and control meter pairs for the best matching model and, subsequently, retrieve the full set of interval data for all identified control quasi-sites.

3.5 ESTIMATION

For the commercial custom and commercial compressed air measures, the savings produced from the site-specific pre-post modeling procedure shown in Figure 11 are the goal because the underlying upgrades are so diverse. To analyze the output, we compare the Advanced M&V result to the claimed savings and assess the suitability of the project for Advanced M&V methods.

For the prescriptive measures of interest in Table 6 and Table 7, we are interested in average savings as opposed to site-specific impacts. This is why we paid careful attention to the count of stage 5 sites when developing the measure list.

To estimate the average savings for **prescriptive residential measures**, we use:

1) Panel regression models. In a panel regression, we run pooled regressions using all of the sites in the sample to estimate the average time-differentiated impacts.

2) Panel regression models with a control group. This technique relies on a "difference-in-differences" where the change in consumption observed in the comparison group is netted out of the change observed in the participant group.

The model specifications for the prescriptive residential measures are detailed in Figure 15. We use the Stata areg command[10] to fit our model, which is a regression command that handles datasets with many groupings. Using site ID as our grouping variable, we also create the temperature and time-of-week variables. The temperature spline is created using the actual weather from the Burlington International Airport from January 2017 to June 2021. Using the minimum and maximum temperature during this period, equal sized temperature bins are created, containing approximately 19 degrees (F).

[10] https://www.stata.com/manuals/rareg.pdf

The period variable details the time-of-week and ranges from 1 to 168, with each hour across the 7 days of the week being sequentially and uniquely identified.

Figure 15: Panel Regression Specifications for Pre-Installation Period

```
areg kwh temp1 temp2 temp3 temp4 temp5 i.period if post == 0, absorb(siteid)
```

Most of the residential measures listed in Table 6 are expected to be fairly homogenous across homes. Most homes will adopt one or two units of the measure. While there is some variation in capacity across the program-supported units, the size range within the residential is fairly limited. For these measures, we estimate the **mean per-household** savings. While the ENERGY STAR dehumidifier measure includes units that range from 30 pints/day to 80 pints/day, we will pool all the homes that installed the measure together and estimate average time-differentiated impacts for the measure as a whole. Much like the TRM measure characterization[11], average dehumidifier capacity amongst the homes analyzed homes will be embedded in the per-unit impacts.

As a supplement, we also provide the average number of program-supported units for our estimation sample so that readers could convert the per-household impacts to per-unit impacts. For example, the pre-post TOWT regression model estimates an average savings of 29 kWh/year amongst the 1,048 homes analyzed. These homes had an average measure quantity of 7.8 lamps so the mean per-unit savings estimate is 3.7 kWh annually.

Eight of the residential measures listed in Table 6 have a vintage of "market opportunity". For market opportunity, or replace-on-burnout, measures the baseline is a code minimum piece of equipment that was never installed in the home or business. Our modeling approach is no different for market opportunity measures, but additional context is required for reporting and interpretation of the results. Section 5.2 includes a dedicated discussion of market opportunity results.

To estimate the average savings for **prescriptive commercial measures**, we use:

1) Aggregated results from individual customer regression models.

2) Aggregated results from individual customer regression models with matched controls. To incorporate the control group in this framework, we run pre-post models on the all the controls individually and then difference out their aggregated result from the estimated impacts amongst the participants.

The model specifications for the prescriptive commercial measures are detailed in Figure 16. We use the Stata reg command[12] and specify the robust option to fit our model. The command is different than the prescriptive residential measures, due to the use of individual customer regressions as opposed to panel regressions. Besides the command, the other variables in this specification remain the same across residential and commercial measures.

[11] The Vermont TRM assumes an average water removal capability of 57.6 pints per day.

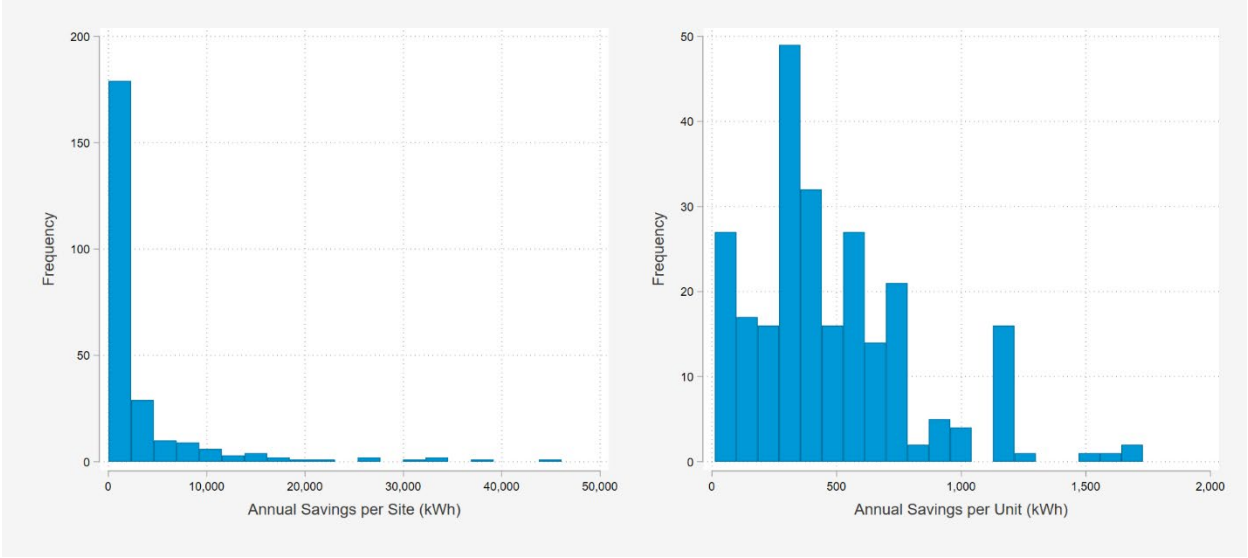
[12] https://www.stata.com/manuals/rregress.pdf

Figure 16: Individual Customer Regression Specifications in the Pre-Installation Period

```
reg kwh temp1 temp2 temp3 temp4 temp5 i.period if post == 0, vce(robust)
```

Most of our commercial measures are estimated on a **mean per-building** basis, much like our residential measures. In contrast, some of the commercial measures listed in Table 7 vary widely both in terms of measure quantity and equipment capacity. Consider the LED exterior lighting measure. The program-supported measures can range from a 15W wall pack to a 300W parking lot fixture. The left side of Figure 17 shows the large distribution of per-site impacts among the 251 sites we've identified for analysis, while the right side shows the per-unit distribution for the same sites.

Figure 17: Distribution Impacts for Among LED Exterior Fixtures Sample (n=251)



Given this type of heterogeneity, the LED exterior lighting and LED interior lighting measures mean per-site savings value would not be terribly useful. Instead, we utilize the quantity count in the EVT database to produce a **mean per-unit** savings estimate. While this does not address the variation in equipment size, it at least controls for measure quantity. Since we estimate impacts in this way, the signal size is difficult to conceptualize. To situate these savings within the context of site-level consumption, we construct an average percent savings, which examines the database savings claim in relation to the site's predicted annual consumption in the pre-installation period. This helps us to examine the signal-to-noise ratio at the site level.

There are a couple reasons we use aggregated customer regressions for commercial measures and panel regressions for residential measures, which mainly boil down to the homogeneity of residential sites and the heterogeneity of commercial sites. Aggregating the commercial measures allows us to filter out bad candidates for this methodology, while also helping to account for greater variability in commercial load. We do not implement this same methodology for residential sites, since loads are more similar across sites. When applying both of these methods to measures that perform well under Advanced M&V, similar results are returned. For example, in the pre-post TOWT panel on

dehumidifiers, we estimate a savings of -509 kWh across 223 sites. Using these same sites and applying the aggregation technique, a savings of -436 kWh is returned. Additionally, regression-based modelling is not the only way to estimate these savings. Machine learning methods could also be tested for consistency via out-of-sample testing.

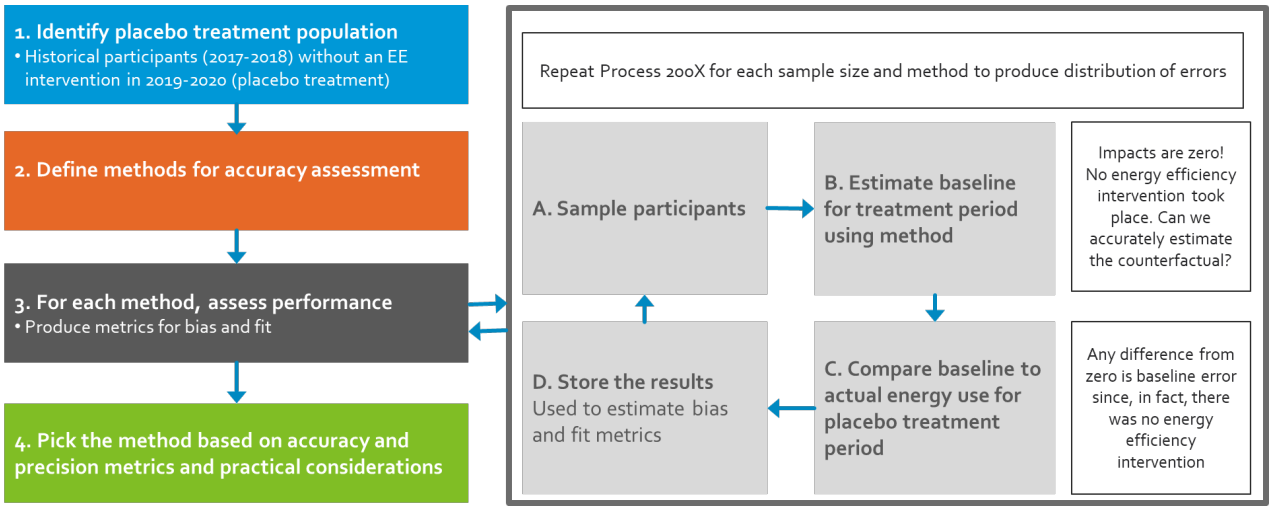
3.6 ACCURACY ASSESSMENT

The primary challenge with estimating energy savings is the need to accurately detect changes in energy consumption due to energy efficiency while systematically eliminating plausible alternative explanations for those changes, including random chance and non-routine events. Did the introduction of energy efficiency measures cause a change in energy use? Or can the differences be explained by other factors (such as the effects of the COVID-19 pandemic)? To evaluate energy savings, it is necessary to estimate what energy consumption would have been in the absence of program intervention—the counterfactual or baseline.

To assess accuracy, one needs to know the correct values. When we know the correct answers, it is possible to determine if each alternative method correctly measures energy use and, if not, by how much it deviates from the known values. Figure 18 summarizes the approach used assessing the accuracy and precision of different methods. The approach is effectively a competition, where the answer key is known, and different methods are tested repeatedly, to identify the methods that are unbiased and accurate.

We assess accuracy by applying placebo treatment periods to customers that, in fact, did not participate in EVT energy efficiency programs during the period analyzed. Because energy efficiency measures were not installed, any deviation between the counterfactual and actual loads is due to error. We repeated the process hundreds of times – a procedure known as bootstrapping – to construct the distribution of errors.

Figure 18: General Approach for Accuracy Assessment (Repeated 200x)



Before the accuracy assessment can be implemented, it is necessary to identify the proxy participant population, agree upon the baseline alternatives to include, and define the metrics for assessing performance. For this analysis, we used pseudo participants drawn from the pool of 2017 energy efficiency participants who did not participate in programs from 2018 onwards. These customers are most like actual participants in EVT programs because they were past participants in EVT programs. Critically, however, they have an extended period without participation, which allows us to conduct the accuracy assessment.

Once participants are identified, we can proceed with selecting a list of models to evaluate. In our case, we compared evaluating impacts using a pre-post TOWT methodology with the matched control group method described in the previous section. The accuracy of each method was tested both with and without COVID-19 in the post period. The "without COVID-19" case essentially maps the pre- and post-periods to calendar years 2018 and 2019. The "with COVID-19" case splits the pre- and post-periods at March 15, 2020, which falls a few days after COVID-19 was declared a pandemic by the World Health Organization and declared a national emergency in the United States.

3.7 UNCERTAINTY

With any impact estimation procedure, there is some potential for error, or uncertainty. Uncertainty can be introduced at any stage of the process. Sampling error is introduced when only a subset of the units in a population are studied. Measurement error is introduced by the instrumentation used to collect readings (even AMI meters are not perfect). Modeling, or estimation, error occurs when the mathematical model used to explain energy consumption does not explain the variation in the data perfectly. Quantifying and reporting uncertainty is an important responsibility of M&V practitioners. The simplest way to conceptualize uncertainty is a ***margin of error***. Consider a hypothetical measure which we have estimated to save 800 kWh per year on average.

- In this example 800 kWh/year is the ***point estimate***.
- If the margin of error is ± 200 kWh/year at the 90% confidence level, we infer that there is a 90% chance that the true savings value falls between 600 and 1,000 kWh/year. The range from 600 to 1,000 is referred to as the ***confidence interval***.
    - ✓ Because the confidence interval does not include zero kWh/year, we classify the estimate as ***statistically significant***.
    - ✓ If the margin of error in our example were ± 2,000 kWh/year the confidence interval would range from -1,200 to 2,800 kWh and the result would not be statistically significant. This result contains greater uncertainty and the point estimate of 800 kWh/year should be viewed with increased skepticism.
- Dividing the margin of error by the point estimate returns the ***relative precision*** or ***fractional savings uncertainty***. These metrics express the uncertainty relative to the effect size.
    - ✓ In our example, 200 divided by 800 returns relative precision of ±25%
- The confidence level at which uncertainty is displayed is a reporting decision. Once selected, the ***standard error*** of the estimate is multiplied the z-statistic or t-statistic for that confidence level to compute the margin of error.

There are a variety of industry standards regarding acceptable levels of uncertainty. The most relevant requirement for Vermont comes from ISO-NE's M-MVDR which guides measurement and verification for demand resources in the FCM and states "*The reported monthly performance shall achieve at least a 10% relative precision at an 80% confidence level.*" Technical guidance in the M-MVDR focuses primarily on sampling error, while this study deals more with estimation. We provide recommendations for reconciling these two frameworks in Section 5.

Methods for quantifying uncertainty are complex and varying by modeling procedure. This study relies on three distinct approaches – one for panel models (with and without controls), one for aggregated site-level regressions, and one for site-specific reporting. For panel models and aggregated site-level regressions, the uncertainty calculations are driven by the number of sites included in the pooled regression and follows many of statistical concepts that guide sampling. Variation is in the numerator of the uncertainty calculation and the square root of "n" is in the denominator. For pooled approaches "n" is the number of sites included in the regression model.

For site-specific reporting, only one site is included in the regression. The "n" value of interest becomes the number of observations or measurements of energy consumption and this value can be quite large when working with high-frequency interval data. It is also fundamentally incorrect in the presence of autocorrelation. Autocorrelation (or serial correlation) refers to the relationship between the past and present values of a time series. High frequency AMI data typically exhibits autocorrelation because consecutive load measurements are affected by activity within the home or business. Simply put, the load at 3:00 PM is related to what the load was at 2:00 PM and 1:00 PM. A key assumption of ordinary least squares regression is that the error terms (or residuals) are independent. In the presence of autocorrelation, this assumption is violated. Autocorrelation does not lead to biased regression estimates, but it does produce understated variance estimates. If not correctly accounted for, site-specific regression-based approaches using AMI data will significantly underrepresent the savings uncertainty and the findings will appear overly precise. Proper handling of autocorrelation is critical when using Advanced M&V methods to ensure correct inferences are drawn from the findings.

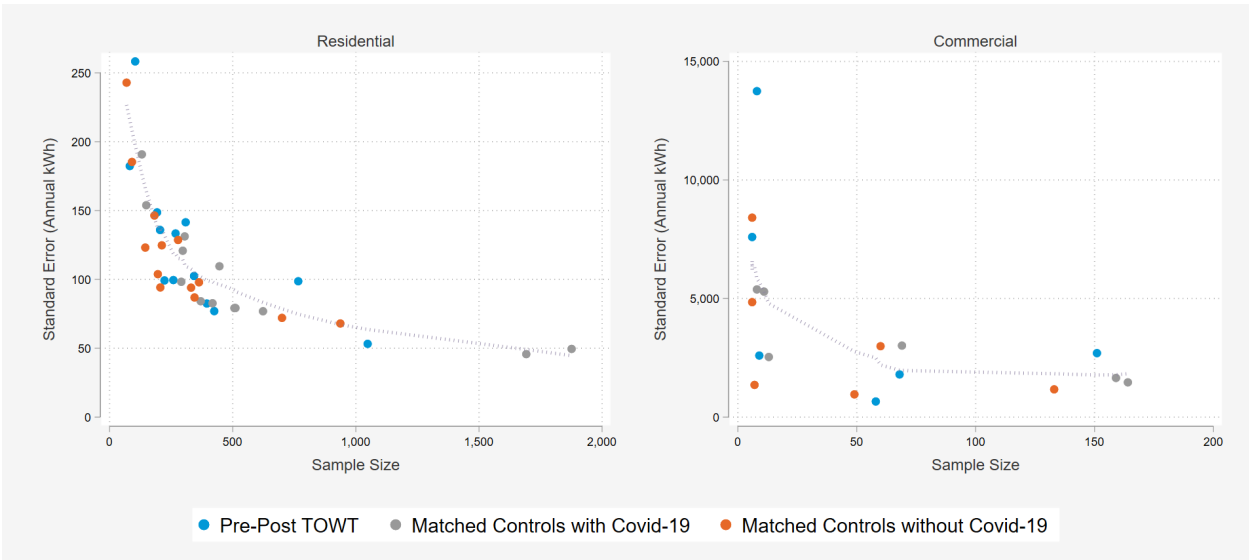
Because we use the size of the panel as the "n" value when computing uncertainty for pooled regression models, no additional adjustment for autocorrelation is needed. Using panel size is arguably a conservative assumption because the number of observations per site does not improve the uncertainty. However, the calculated uncertainty values line up well with simulation-based estimates of uncertainty in Section 4.1 so we believe the approach is appropriate.

Prescriptive Measures

For prescriptive residential measures, we report standard errors to display the uncertainty of our estimates. To calculate the standard error of these estimates, we run a supplemental TOWT fixed effects regression model that includes both pre and post data with a binary indicator variable equal to 1 for all observations in a given site's post-retrofit period. For the models with matched controls there is a second indicator variable for whether the site was treated (1) or in the matched comparison group (0).

The coefficient of the "post" term, or treatment*post interaction term when matching is used, estimates the average impact and its standard error the associated uncertainty. This procedure is run separately for annual energy, summer peak demand, and winter peak demand. The coefficients and standard errors of interest from this process represent the effect for an average hour. For energy, we multiply the standard error by 8760 to return the standard error for the full predicted year. Demand impacts are average hourly by definition so no adjustment is needed. The resulting uncertainty is largely a function of sample size, as seen on the left side of Figure 19. Measures that have a larger pool of sites are more precise in the savings estimates they produce.

Figure 19: Standard Error as a Function of Sample Size



A different method is required for the aggregated site-level regressions we use for prescriptive commercial measures. Instead, we calculate the estimated impact for each site and then the standard error of the mean of all impacts is calculated. This is performed the same way for modeling procedures those with and without matched controls. The standard error of the mean is simply the standard deviation of the individual estimates divided by the square root of "n" so these results follow the same general trend with respect to sample size. Because we aggregate the point estimate or each site level regression, no adjustment is needed for autocorrelation.

Applying Standard Errors

For the prescriptive measures we analyze, standard errors are displayed in parentheses under our point estimates, instead of a confidence interval. This statistic is presented in order to facilitate the calculation of a confidence interval, using the formula below. This allows the reader to select the confidence level that they deem appropriate (Table 10).

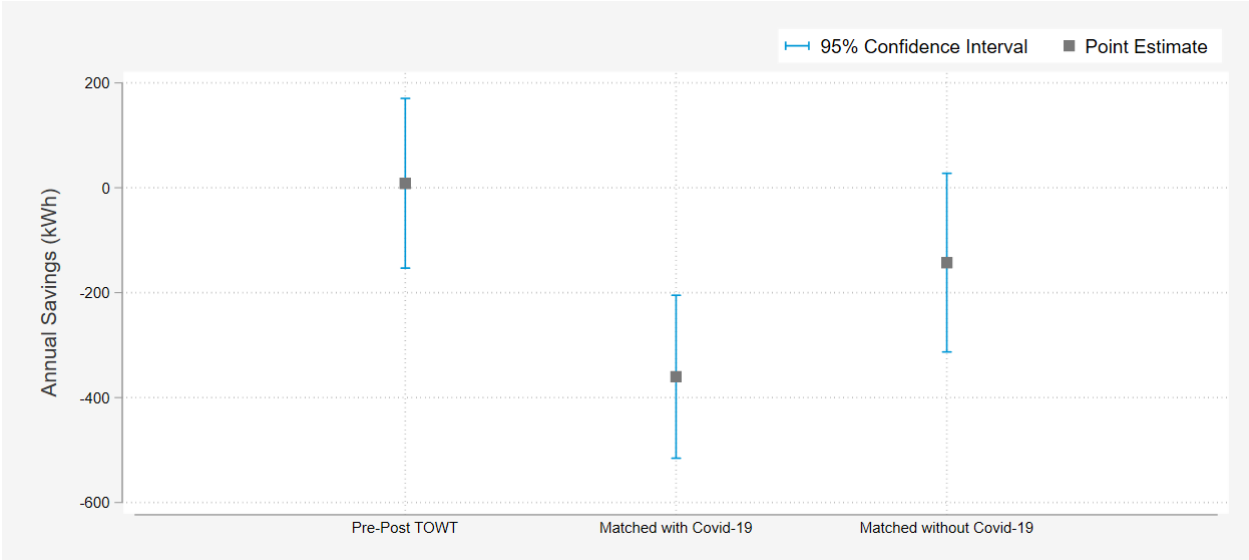
$$Confidence\ Interval = \ Point\ Estimate\ \pm\ (Z\ Statistic * Standard\ Error)$$

Table 10: Confidence Levels

Confidence Level	Z Statistic
80%	1.282
85%	1.440
90%	1.645
95%	1.960

Each prescriptive measure has savings calculated using the pre-post TOWT, the matched controls without COVID-19 period data, and the matched controls group with COVID-19. This means that we present three point estimates, as well as three standard errors for annual savings. In some cases this may be confusing, as the three point estimates may present three different stories. A good example of this is the advanced thermostat measure displayed in Figure 20.

Figure 20: Standard Error Example – Advanced Thermostat Measure



The point estimates in both of the matched control regressions suggest that customers increased electricity consumption after installing an advanced thermostat, while the Pre-Post TOWT point estimate indicates a decrease in the post-installation period. These point estimates do not tell the full story. Converting the standard errors into a 95% confidence interval, two of the three results are not statistically significant. Considering the uncertainty associated with the point estimate helps ensure that a single value is not taken out of context.

Our interpretation of this result is that the effect of advanced thermostats on energy consumption in Vermont is too small to measure with whole-building data given the available sample size. The claimed savings of this measure are small in relation to the consumption of the household. Across the 395 sites used in the pre-post TOWT regression, the average database savings claim was 85 kWh and the average pre-period weather-normalized annual consumption estimate is 6,554 kWh. This means that

the expected effect is only a 1.3% decrease in electricity consumption at the household level. The margin of error at the 95% confidence level is ± 163 kWh/year or ±2.5%. The signal is too small to detect within the noise of household consumption with this sample size. A sample size of 3,000-4,000 homes would be required to produce estimates with a 95% confidence interval narrower than the average claimed savings 85 kWh/year.

Custom Projects

Since we evaluate custom and compressed air projects individually and report site-level savings estimates, the uncertainty calculations are slightly different. We first estimate the extent of first-degree autocorrelation in the data and then use that value to adjust the n-value in the standard error calculation. Because custom project vary so widely in size, we focus on the relative precision or fractional savings uncertainty of the impact estimates. FSU is used as a filtering mechanism for determining which sites are suitable for meter-based methods. In this report, we introduce a twist on the classic definition of FSU that we call Quasi-FSU. It is constructed by dividing the uncertainty in the baseline period by the claimed database savings instead of the regression-based savings estimate.

$$FSU_{quasi} = \frac{Hourly\ Savings\ Uncertainty_{baseline}}{EVT\ Database\ Savings}$$

This metric provides insight into the applicability of Advanced M&V for the site *in advance* given the variability in building load patterns (noise) and the expected savings of the project (signal). Section 4.4 includes further discussion on the implications of this metric.

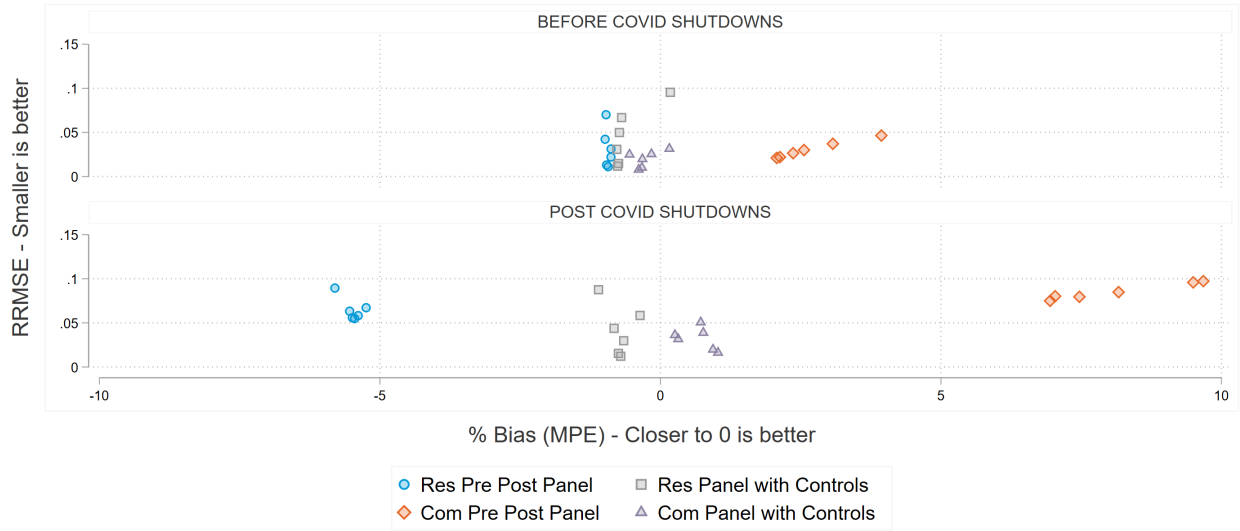
4 RESULTS

4.1 ACCURACY ASSESSMENT

This section summarizes the results of the accuracy simulation. Recall that for each estimation method, we measured accuracy across multiple sample sizes and we looked at two distinct periods: one period that is entirely unaffected by COVID-19 and one period where the post-installation aligns with the state of emergency declaration in Vermont. All modeling uses 12 months of "pre" data and 12 months of "post" data. Figure 21 summarizes the accuracy assessment results for all sample sizes and methods for both residential and commercial customers. Within each customer segment and method, there are six points, each representing the average error for a different sample size. The top half of the figure shows results based on the period that is unaffected by COVID-19, and the bottom half shows results based on the analysis with COVID-19 in the post-installation period. The x-axis depicts the bias of the results (mean percent error) and the y-axis depicts the accuracy of the results (relative root mean squared error). The best methods are at the bottom-center of the graphs, where both the bias and average errors are close to zero.

All of the methods perform relatively well in the period that is unaffected by COVID shutdowns. For the analysis that aligns the post-installation period with COVID-19 shutdowns, the pre-post models without a matched control group consistently under-predict usage for residential customers and consistently over-predict usage for commercial customers. The methods that include a matched control group are able to significantly reduce the bias for both residential and commercial customers. This results makes sense intuitively because the pandemic affected non-participants and participants alike.

Figure 21: Method Comparison



Residential

Figure 22 through Figure 24 depict the results of the residential accuracy simulation for each sample size with and without COVID-19 in the post period. We measured errors separately for the three key study metrics (annual kWh, summer kW, and winter kW). For each metric, the x-axis lists the sample sizes and the boxplots represent the distribution of errors for the 200 simulation runs for each sample. As the pseudo-participants did not receive any intervention during the period analyzed, a "good" method should have impact estimate results that are centered around zero (no consistent upward or downward bias) and the distribution of the errors should be relatively narrow (the results should be consistent for different subsets of customers). Boxplots with the center line above zero indicate the model is over-predicting usage and savings on average, while boxplots with the center line below zero indicate the model is under-predicting usage and savings on average. One key pattern that emerges is that the boxplots become narrower as the sample size increases (moving from left to right within each pane), indicating that the precision of the methods increase when larger samples are used. The second key takeaway is that the pre-post method without controls consistently under-predicts energy and demand after the COVID-19 shutdowns.

Figure 22: Residential Annual kWh Simulation Results

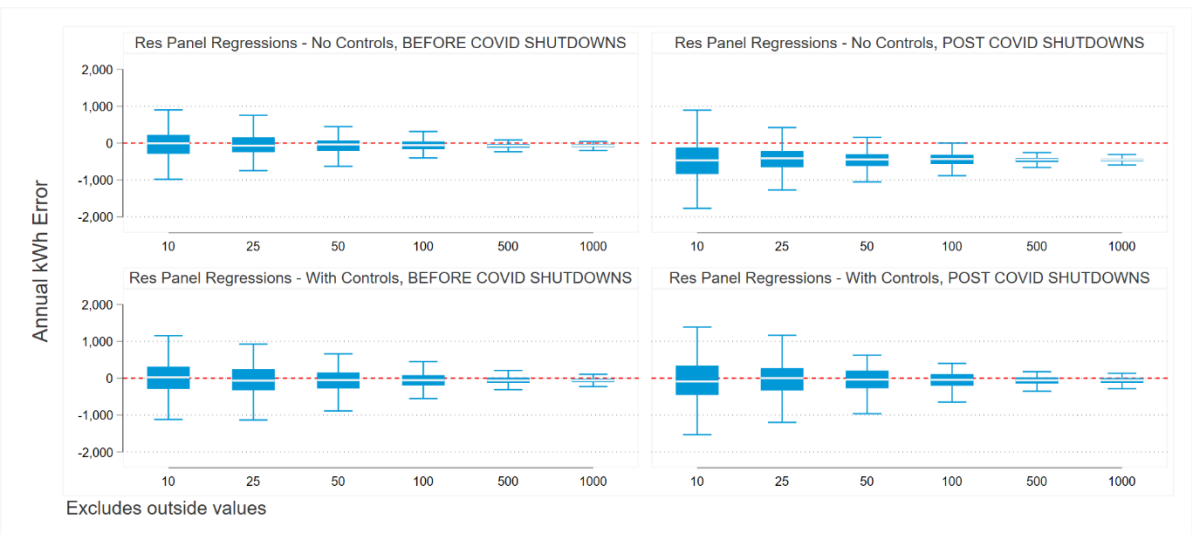


Figure 23: Residential Summer kW Simulation Results

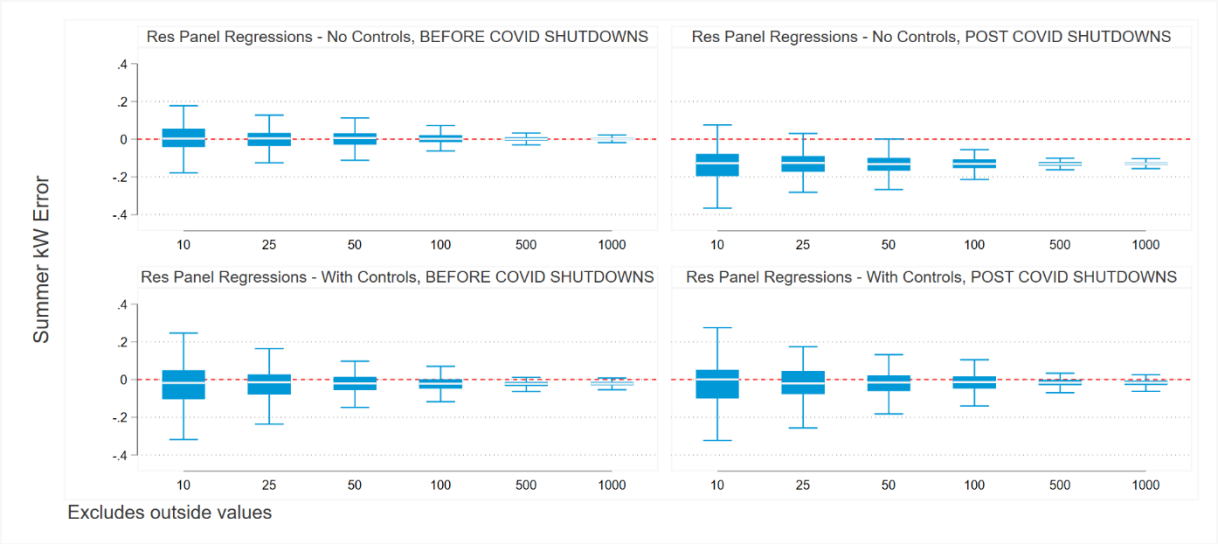
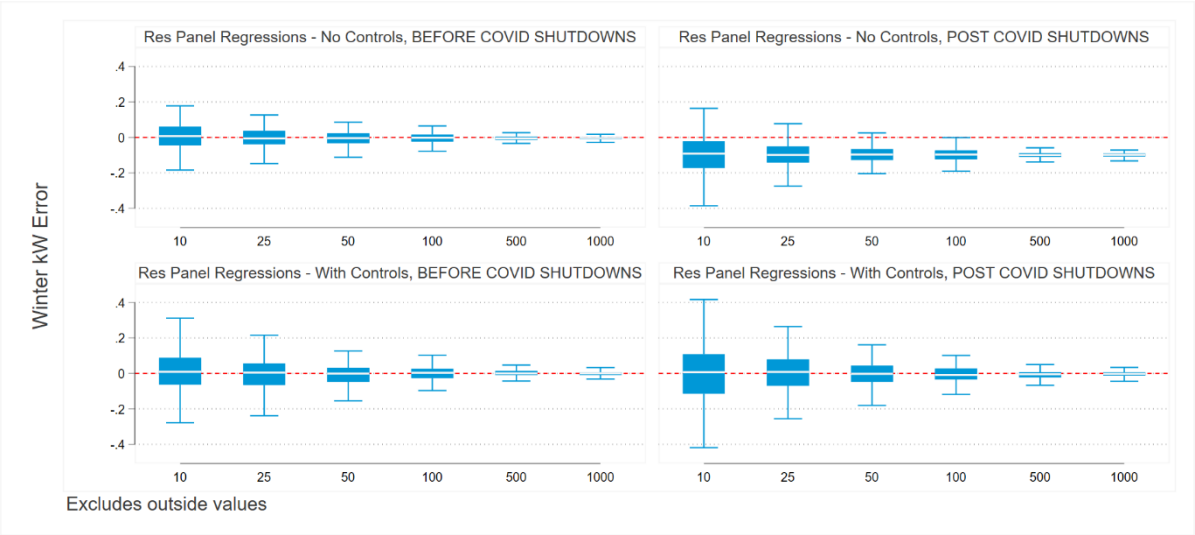


Figure 24: Residential Winter kW Simulation Results



Commercial

Figure 25 through Figure 27 depict the results of the commercial accuracy simulation for each sample size before and after the COVID-19 shutdowns took place. As with residential, we measured errors for annual kWh, summer kW, and winter kW to determine the accuracy for all three key metrics in the primary analysis. Similar to the residential simulations, precision increases across all methods when larger sample sizes are used. Using a matched control group reduces bias both before and after COVID-19 shutdowns, although the matched controls did not improve the overall accuracy of the estimates.

Figure 25: Commercial Annual kWh Simulation Results

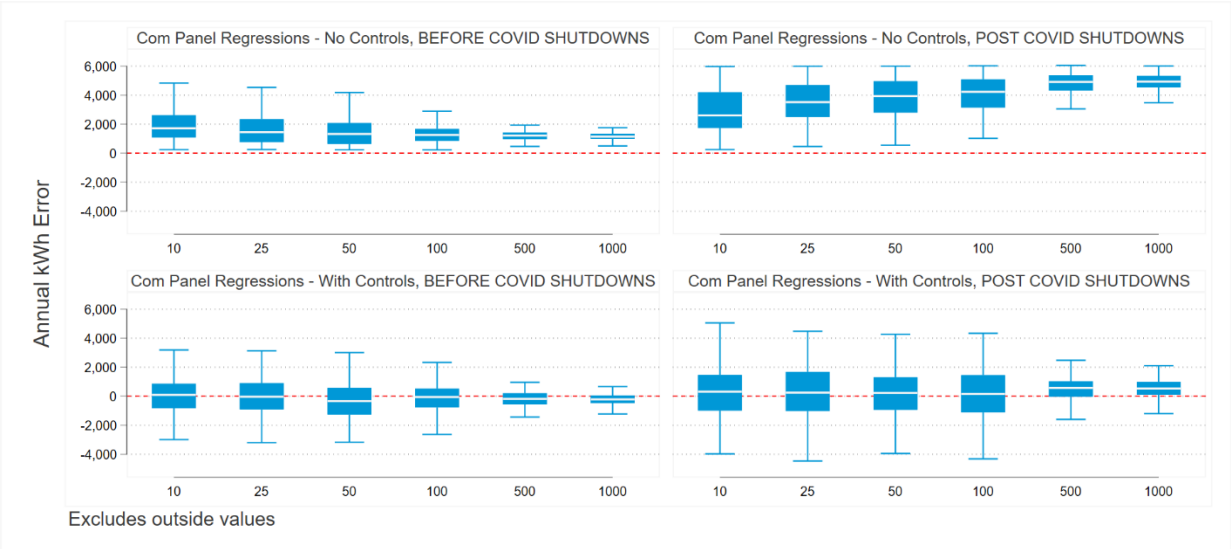


Figure 26: Commercial Summer kW Simulation Results

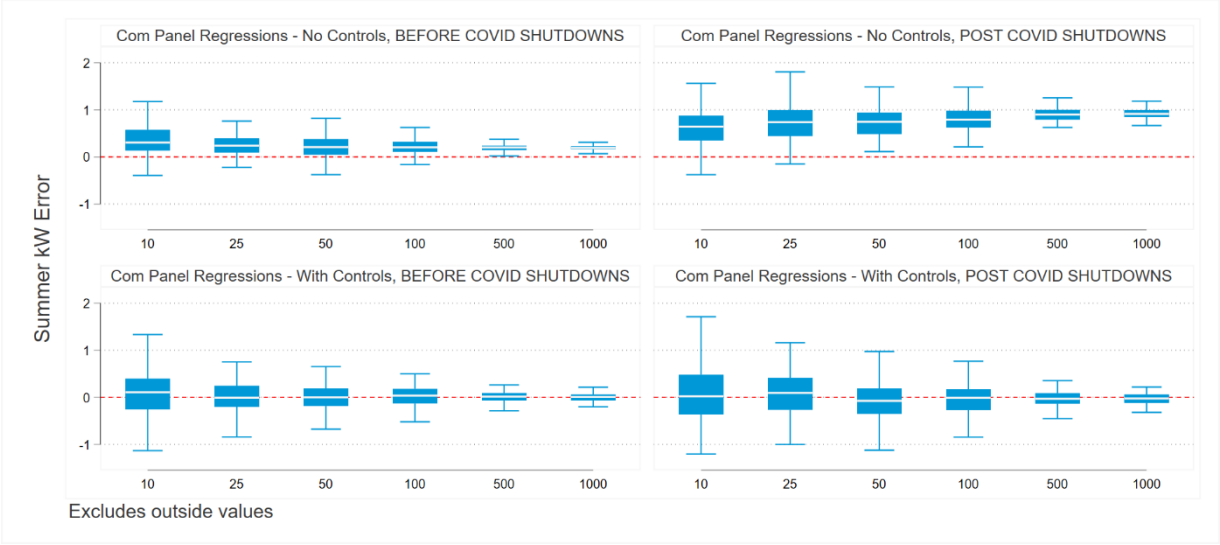
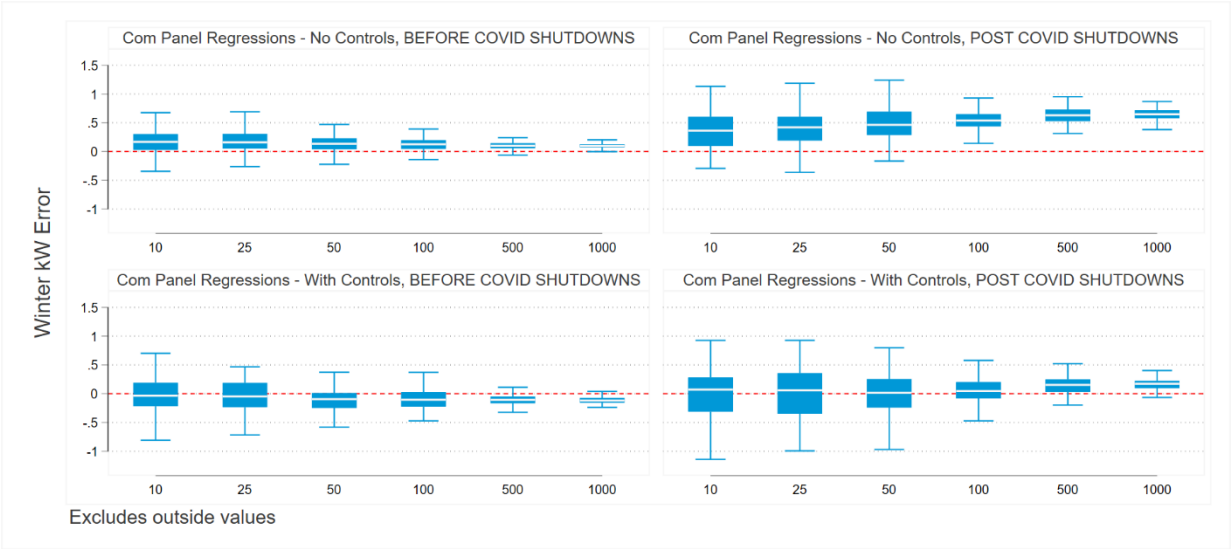


Figure 27: Commercial Winter kW Simulation Results

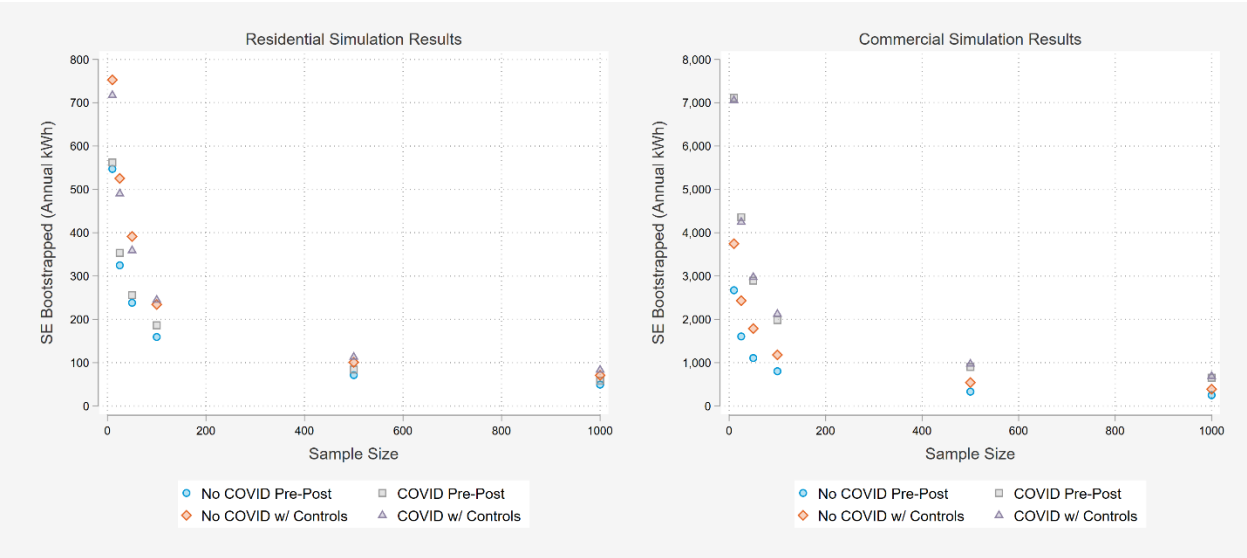


Standard Errors

This simulation exercise provides a useful check on the uncertainty calculations described in Section 3.7. Figure 28 shows our estimated standard errors from the simulation and aligns closely with the uncertainty estimates from our core approach shown in Figure 19.[13] For example, both approaches return an estimated standard error of 100 kWh (or a margin of error of ± 165 kWh/year at the 90% confidence interval) for a residential panel model with sample size of 500 homes. This consistency across methods validates our primary approach and indicates that the uncertainty estimates presented in the following measure-specific results tables are reasonable.

[13] For a given sample size and methodology, the simulation results in 200 impact estimates. The standard deviation of these impact estimates is an estimate of the standard error of our regression-based estimates.

Figure 28: Bootstrapped Standard Errors



4.2 PRESCRIPTIVE RESIDENTIAL

ADVANCED THERMOSTAT

The advanced thermostat measure promotes installation of new Wi-Fi connected smart thermostats. Advanced thermostats reduce heating and cooling consumption through a configurable schedule of temperature set-points and automatic variations to that schedule to better match HVAC system runtimes to meet occupant comfort needs. These schedules may be defaults, established through user interaction, and be changed manually at the device or remotely through a web or mobile app. Automatic variations to that schedule could be driven by local sensors and software algorithms, and/or through connectivity to an internet software service. The majority of the claimed savings for this measure are fossil fuel based on the saturation of heating types and climate in Vermont. EVT claims electric savings for summer cooling in homes with air conditioning and during the heating season for homes with electric heat. Homes with fossil fuel heat and no air conditioning claim a limited amount of electric savings based on reduced operation of the circulating fans and pumps.

TRM Characterization

The advanced thermostat measure is a fast-moving area in the industry and the TRM characterization evolved over our period of investigation. The measure vintage is **retrofit** and the TRM assumes that 67% of the replaced thermostats are manual and the remaining 33% are programmable. The EVT TRM includes a new construction version, but no instances are included in our analysis due to the required duration of pre-retrofit AMI data. The measure characterization was updated for 2019 to reduce the electric savings claim from pumps and fans during the heat season. The most common energy savings claim was 41 kWh per-unit in 2017-2018 and 25 kWh per-unit in 2019-2020. It is not uncommon for

Vermont homes to have two program-supported thermostats at one premise and the homes in our analysis averaged 1.4 thermostats.

Modeling Results

The estimates of annual savings for this measure vary across modeling approaches, ranging from positive savings to negative savings. Table 11 compares the savings estimates produced by the pre-post TOWT and the matched control groups, as well as the claimed database savings. The standard errors are similar across modeling types are similar but the savings estimates are quite different. Despite the reasonable sample size, this measure is inherently difficult to estimate due to the small signal-to-noise ratio. The average annual per-household savings claim for the advanced thermostat measure is only about 1% of household electricity consumption. The margin of error from our regressions are approximately ±1.5% at the 90% confidence level.

Table 11: Advanced Thermostat Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	85	0.006	0.029
Pre-Post TOWT (N = 395)	9 (83)	-0.017 (0.013)	-0.002 (0.022)
Matched Controls without COVID-19 (N = 345)	-143 (87)	-0.012 (0.015)	-0.027 (0.025)
Matched Controls with COVID-19 (N = 511)	-360 (79)	-0.032 (0.014)	-0.071 (0.020)

*Reported for the sites used in the Pre-Post TOWT.

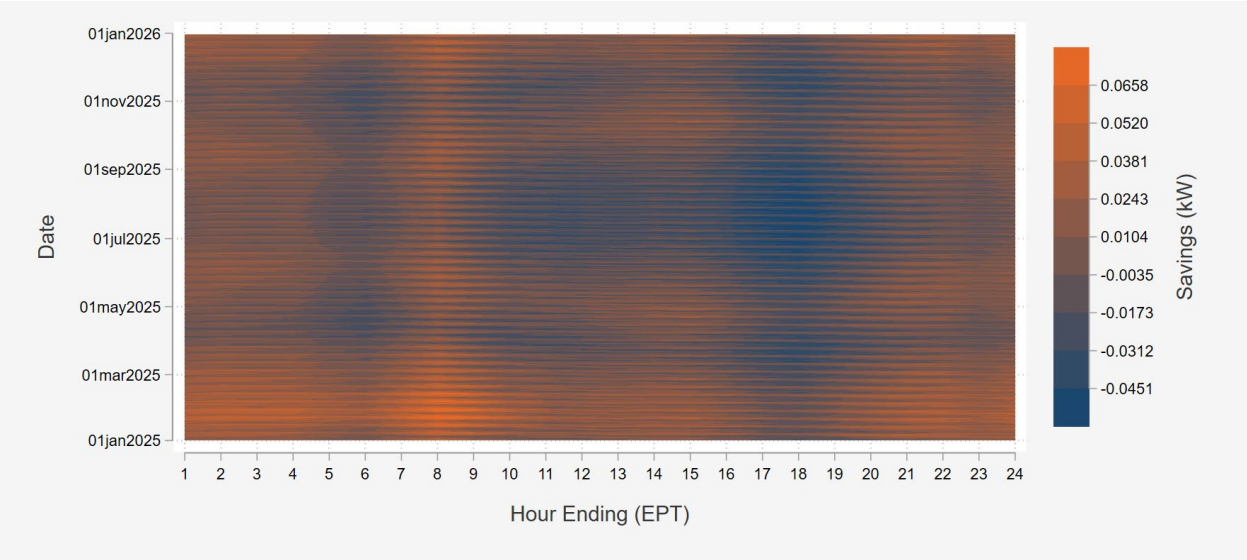
The noisy estimates for advanced thermostats may also be a function of the nature of the measure. For example, a good time to install an advanced thermostat is with the installation of a new air conditioner. This could represent savings (relative to the old AC unit) or a load increase (if the home did not have Central AC previously). There is likely also considerable variation in how customers use their advanced thermostat and how they used their previous thermostat. It's also possible that some matched control group homes have advanced thermostats that purchased without an EVT rebate.

To get a more granular look at how the annual savings are distributed across the 8760 hours of the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used.

The heat map for the advanced thermostat is displayed in Figure 29. Using the pre-post TOWT model, the limited savings are concentrated most heavily during winter mornings. Summer afternoons show modest load increases. It is certainly possible that some homes installed their advanced thermostat along with adding central air conditioning (without an EVT rebate) and that added air conditioning is confounding our thermostat analysis. Unfortunately, the shape of this measure does not convey much,

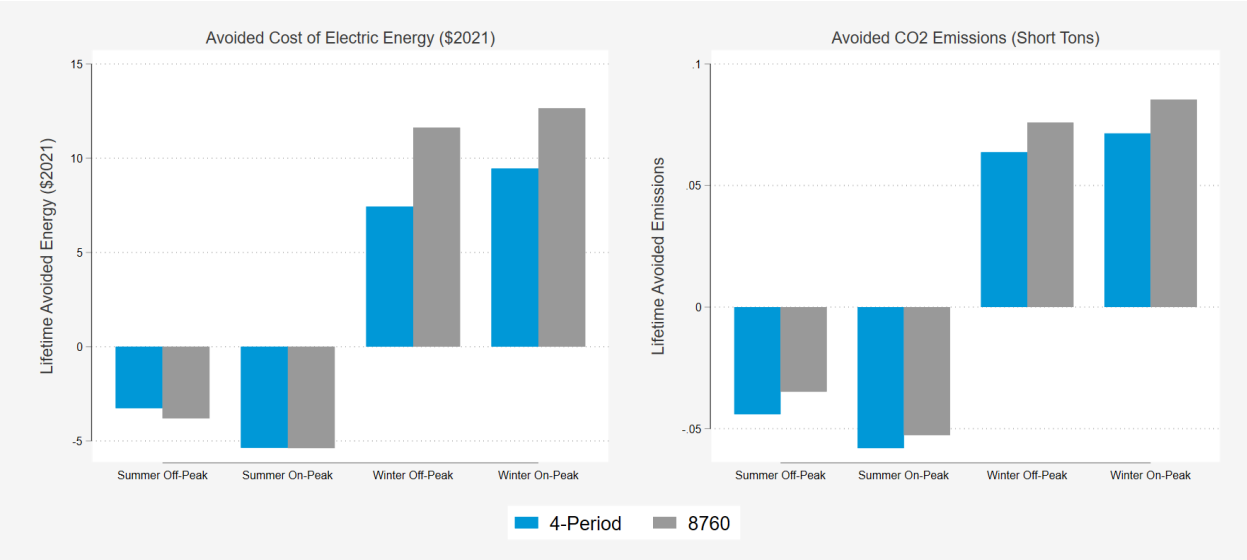
as both the savings estimates and shape of the savings changes as different modeling techniques are implemented.

Figure 29: Advanced Thermostat Heat Map



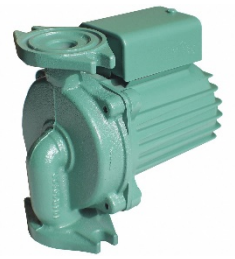
Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a ten year measure life. These were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $15.08, which is 45.3% greater than the four period avoided costs. Similarly, the avoided emissions sum to 0.07 short tons of CO_2 for the 8760 profiles, which is 55.2% higher than if the four-period assumptions were used. A more detailed summation of these avoided costs/emissions are presented in Figure 30.

Figure 30: Advanced Thermostat Valuation Comparison



BLPM Boiler Motor – Residential

Brushless permanent magnet (BLPM) circulator pump motors are found in homes with baseboard and radiant floor hydronic heating systems. This is the motor that moves hot water from the boiler to various areas in the home. BLPM pumps are more efficient because they lack brushes that add friction to the motor and their ability to modulate their speed to match the load. As the system flow demand changes (zones open or close), the drive senses the torque difference at the impeller via the change in the magnetic field difference and adjusts its speed by altering the frequency to the motor. BLPMs are especially efficient in no-load/low-load applications.



TRM Characterization

The residential BPLM measure characterization was largely unchanged over our period of investigation. It is a fully deemed **retrofit** measure with claiming 87.5 kWh of annual savings per-unit. The homes in our analysis typically have 2.1 program-supported units.

Modeling Results

The annual kWh savings estimates for this measure fluctuate based on the modeling procedure applied, ranging from positive savings to negative savings. Table 12 provides more detail into the savings estimates produced by the pre-post TOWT and the matched control groups, as well as the claimed database savings. It is important to be mindful of the standard errors when comparing the point estimates across models. Like the advanced thermostat measure, this measure is challenging to model with whole-building data due to the small signal-to-noise ratio. The average annual savings claim per household that installed a BLPM boiler motor is only about 2% of household electricity consumption. The margin of error of these pooled regression is also approximately ±2% of annual electric consumption.

Table 12: BLPM Motor Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	185	0.000	0.057
Pre-Post TOWT (N = 343)	190 (102)	-0.010 (0.017)	0.031 (0.024)
Matched Controls without COVID-19 (N = 331)	19 (94)	-0.015 (0.013)	0.006 (0.024)
Matched Controls with COVID-19 (N = 507)	-253 (79)	-0.035 (0.012)	-0.052 (0.019)

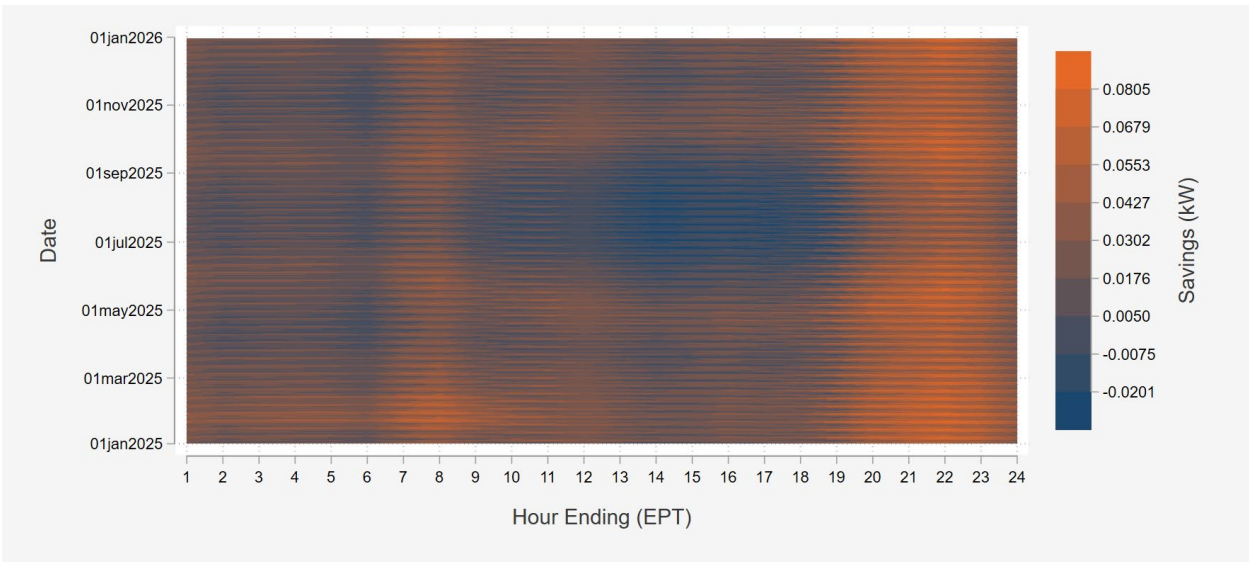
*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed across the 8760 hours of the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the

Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used.

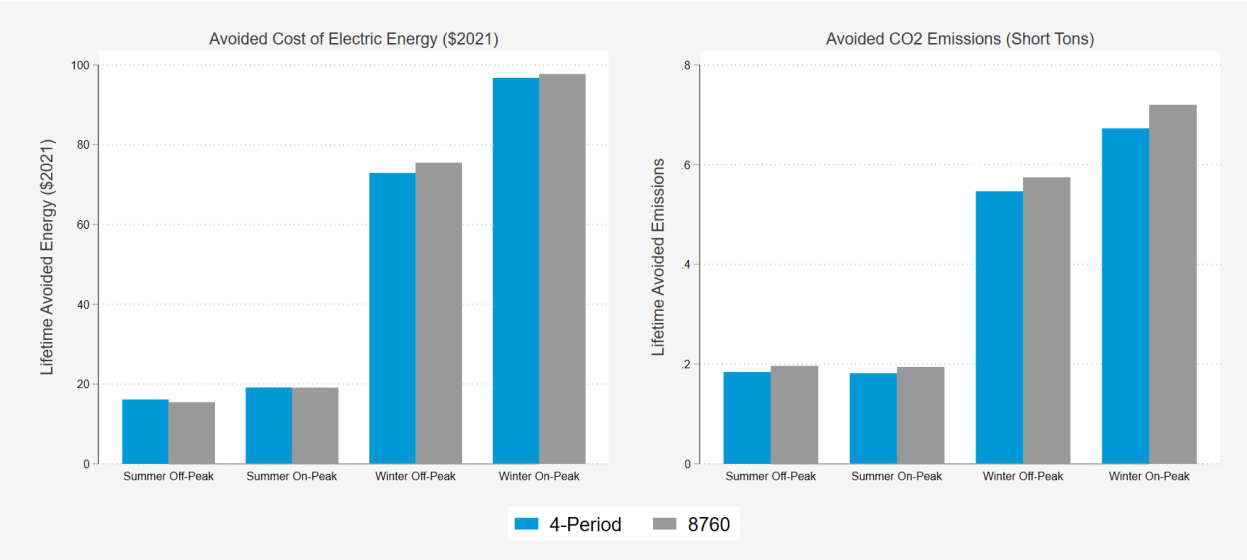
Figure 31 displays the heat map for this measure. Using the pre-post TOWT, the savings produced are concentrated most heavily in the evenings of winter and shoulder months. As expected given the heating focus of the measure, the impacts are close to zero on summer days when there is no heating load. Both the magnitude and shape of this measure's impacts vary across modeling approaches.

Figure 31: BLPM Motor Heat Map



Using the predictions produced for 30-year climate normals, we estimate avoided costs and avoided emissions over the 20-year life of the measure. These were calculated using both the 8760 valuation profiles, as well as the four-period definitions of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $207.71, which is 1.3% greater than the four-period avoided costs. Similarly, the avoided emissions sum to 1.68 short tons of CO_2 for the 8760 period, which is 6.0% higher than if the four-period assumptions were used. A more detailed summation of these avoided costs/emissions are presented in Figure 32.

Figure 32: BLPM Motor Valuation Comparison



ENERGY STAR DEHUMIDIFIERS

Dehumidifiers remove moisture from the air by drawing air across a coil with a fan. The water collected can be stored in a receptacle inside the unit that needs to be emptied manually or have a hose that connects to a permanent draining location. They are used most often during summer months when relative humidity is high and in damps areas like a basement.

TRM Characterization

The ENERGY STAR dehumidifier measure primarily follows a **market opportunity** characterization. The baseline is a new dehumidifier of the same size that meets the federal standard. In 2019, an early replacement path was added to the TRM, but most volume still follows the market opportunity path. EVT updated this measure characterization each year from 2017 to 2019, but the core electric outputs stayed the same with virtually all units assigned 229 kWh of annual electric savings. In late 2019, both the federal standard and ENERGY STAR specification for dehumidifiers changed, so the 2020 measure characterization necessarily evolved to reflect these standards. The key assumption about annual hours of use has remained constant at 1,632 hours. Within our analysis, households have an average of 1 unit.

Modeling Results

Regardless of the modeling type used for this measure, there is always an evident increase in the electric load of the sites analyzed, compared to the positive savings claim detailed by the database. Table 13 provides more detail into the savings estimates produced by the pre-post TOWT and the matched control groups, as well as the claimed database savings. Each modeling procedure estimates a statistically significant increase in electric consumption. The consistent nature of these findings suggests that participating homes likely did not have a dehumidifier previously and the ENERGY STAR dehumidifier represents an added load at the meter. Of course, if the home had added a code minimum

dehumidifier instead, the load increase would have been larger. The TRM characterization is appropriate, it just doesn't reflect the same change in consumption observed by the revenue meter.

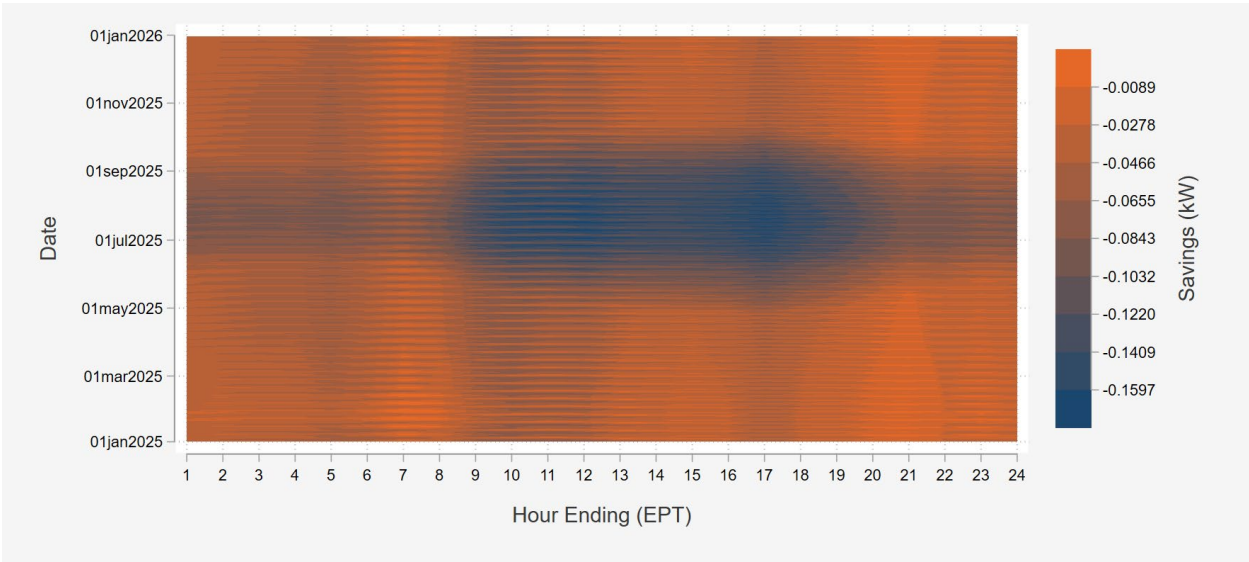
Table 13: ENERGY STAR Dehumidifier Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	230	0.050	0.000
Pre-Post TOWT (N = 223)	-509 (99)	-0.135 (0.019)	-0.027 (0.025)
Matched Controls without COVID-19 (N = 206)	-355 (94)	-0.111 (0.015)	0.023 (0.025)
Matched Controls with COVID-19 (N = 418)	-267 (83)	-0.102 (0.013)	0.013 (0.020)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed across the 8760 hours of the year, we generate a heat map using the pre-post TOWT model. Figure 33 plots the impacts for all 8760 hours in a hypothetical year using 30-year climate normal weather for the Burlington International Airport. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The results table and the heat map indicate a consistent negative savings, while the database is claiming positive savings. The heat map helps to tell a more cohesive story regarding why this measure behaves the way it does and how it actually behaves as expected. Load increases are concentrated during the daytime hours of summer months when the weather is hot and humid.

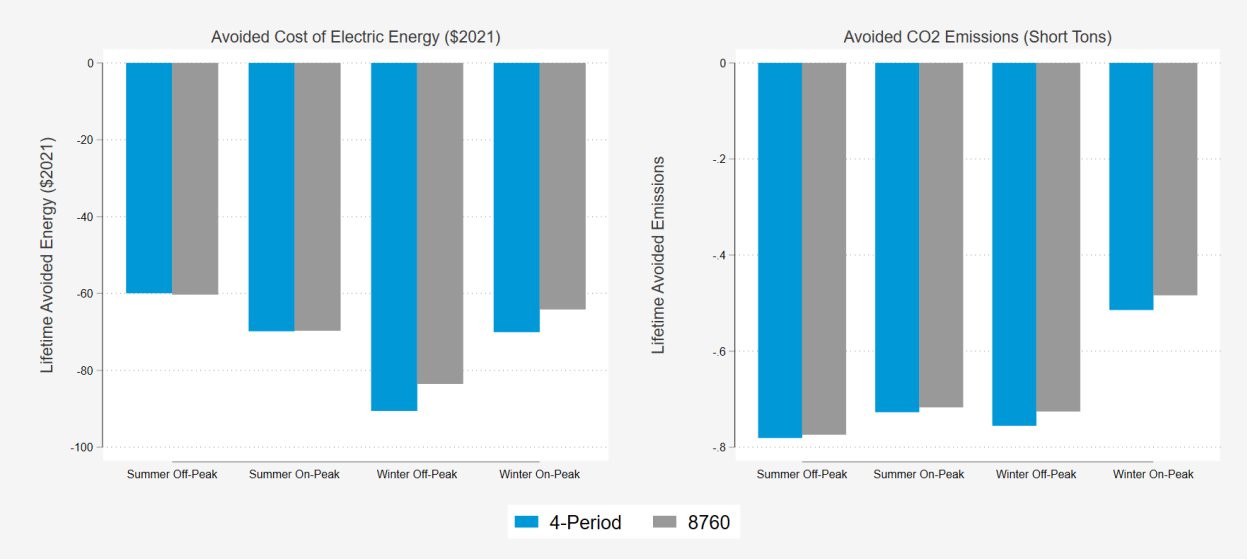
Figure 33: ENERGY STAR Dehumidifier Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a twelve year measure life. These were calculated using both the 8760 valuation profiles and the four

costing periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is -$277.58, which is 4.6% lower than the four-period avoided costs. Similarly, the avoided emissions sum to -2.70 short tons of CO_2 using 8760 profiles, which is 2.9% lower than if the four-period assumptions were used. In this case, "lower" means a smaller negative number. Figure 34 presents a more detailed summation of these lifetime avoided costs/emissions.

Figure 34: ENERGY STAR Dehumidifier Valuation Comparison



ENERGY STAR Clothes Dryer

A high efficiency clothes dryer uses less electricity per pound of clothing dried compared to a standard efficiency dryer. The heat source in the dryer can be electric resistance, heat pump technology, or a hybrid of the two. Qualifying models can be vented or ventless, compact or standard size, and use a 120V or 240V outlet.

TRM Characterization

The ENERGY STAR clothes dryer measure primarily follows a **market opportunity** characterization. The baseline is a new clothes dryer of the same size that meets the federal standard. The TRM includes an early replacement path, but almost all instances in the EVT tracking data follow the market opportunity characterization. EVT updated the characterization for 2019, but the annual kWh and peak kW savings only changed slightly. The sites within our analysis indicate an average of 1 unit per household.

Modeling Results

For all modeling approaches, there is an evident decrease in the electric load of the sites analyzed, which mirrors the database predictions. Table 14 compares the detailed savings estimates from the pre-post TOWT and the matched control group analyses to the claimed database savings. It is important to note that, while each estimate indicates a positive annual savings, the standard errors for

these estimates are fairly large (none of the savings estimates are statistically significant). The size of these standard errors is directly linked to the small sample sizes.

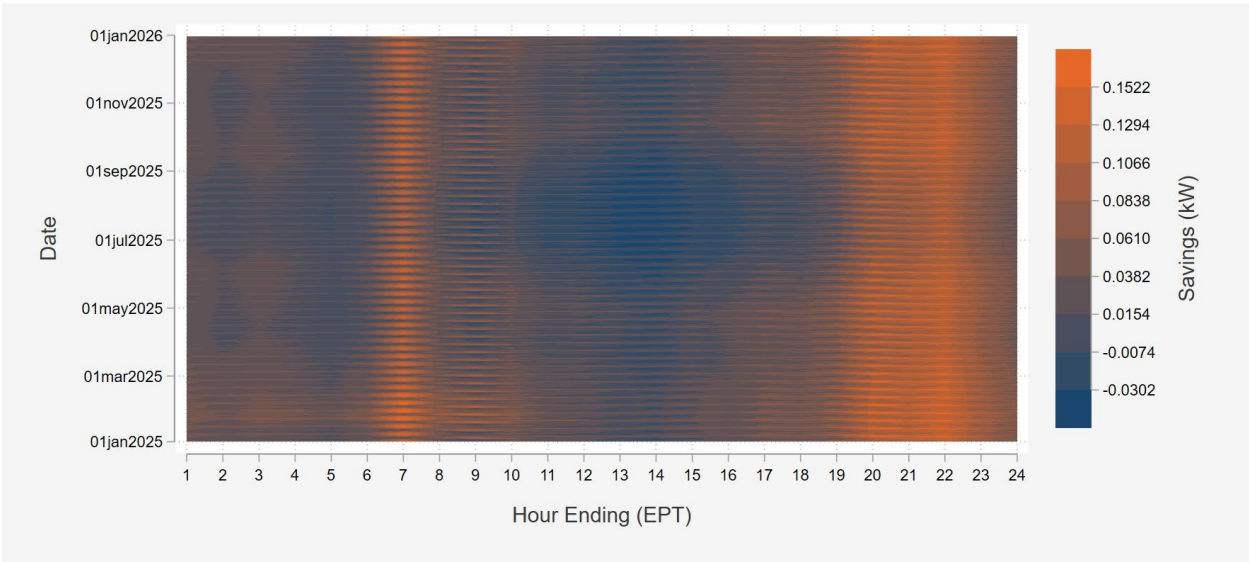
Table 14: ENERGY STAR Clothes Dryer Modeling Summary

Source	kWh/year	Summer kW	Winter kW
EVT Database*	221	0.023	0.030
Pre-Post TOWT (N = 82)	337 (182)	-0.006 (0.031)	0.056 (0.045)
Matched Controls without COVID-19 (N = 69)	104 (243)	-0.014 (0.043)	0.025 (0.053)
Matched Controls with COVID-19 (N = 131)	221 (191)	0.010 (0.026)	0.049 (0.043)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed over the 8760 hours in the year, we use the pre-post TOWT model to generate a heat map. Figure 35 plots the 8760 points in the prediction period, where there are 365 days and 24 hours in each day. This prediction is based on the 30-year climate normal weather for the Burlington International Airport. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The heat map suggests that clothes dryer savings are concentrated in the morning and in the evening. In addition, there appears to be variation between weekends and weekdays, which is visually evident by the stripes of blue between the orange at 7:00 AM (EPT). The shape of the savings for this measure are defined and consistent across modeling approaches, much like the savings estimates.

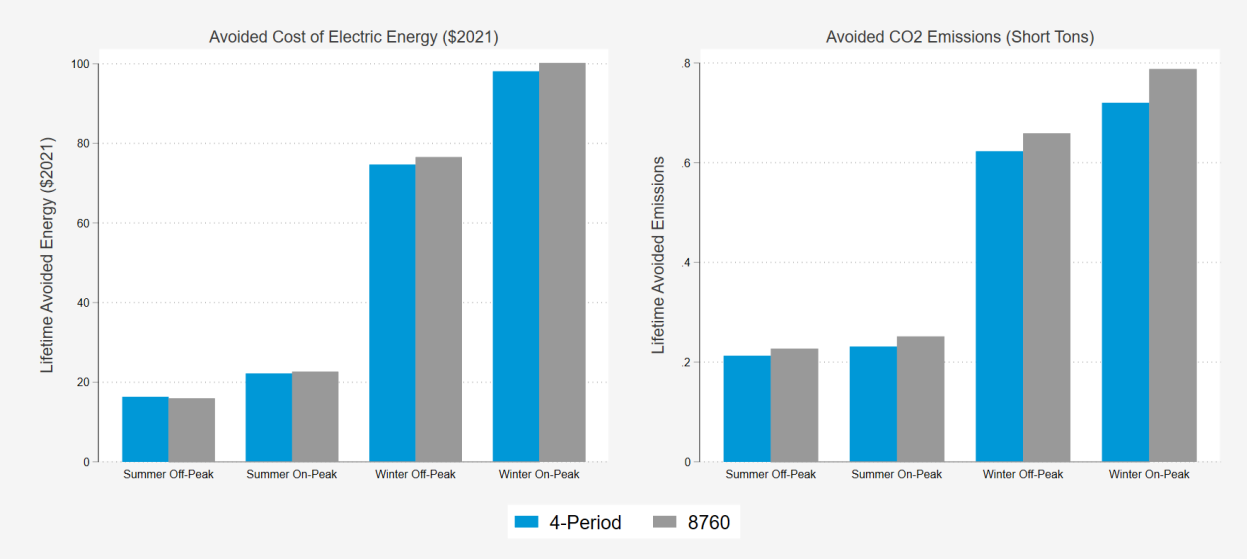
Figure 35: ENERGY STAR Clothes Dryer Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a twelve year measure life. These were calculated using both the 8760 valuation profiles, as well as the

four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $215.32, which is 1.9% greater than the four-period avoided costs. Similarly, the avoided emissions sum to 1.93 short tons of CO_2 for the 8760 period, which is 7.2% higher than if the four-period assumptions were used. A more detailed summation of these avoided costs/emissions are presented in Figure 36.

Figure 36: ENERGY STAR Clothes Dryer Valuation Comparison



ENERGY STAR Clothes Washer

High efficiency clothes washers use less water to clean clothes per pound. This saves water heating energy for cycles that use warm or hot water. ENERGY STAR clothes washers also leave clothes drier at the completion of the cleaning cycle. This means less energy is required by the clothes dryer to remove the remaining moisture. In addition to the ancillary reductions from the water heater and clothes dryer, high efficiency clothes washers are also more efficient at agitating and spinning the clothes.

TRM Characterization

The Efficient Clothes Washer measure is an incredibly complex TRM characterization with various efficiency levels, washer types, and assumptions about the domestic hot water and clothes dryer. There are multiple permutations but it is overwhelming utilized via a **market opportunity** path. Nearly all program-supported clothes washers 2017-2020 claimed between 120 and 159 kWh per unit in the EVT tracking database. This represents approximately 1.5% of annual electric consumption for a typical Vermont household. Households in our analysis installed an average of 1 unit.

Modeling Results

This measure demonstrates small annual savings estimates across the various modeling types. Table 15 provides more detail into the savings estimates produced by the pre-post TOWT and the matched

control groups, as well as the claimed database savings. None of the savings estimates are statistically significant. Since it is a market opportunity measure and most homes would have had a clothes washer previously, we expected Advanced M&V methods to return a larger savings estimate than the TRM characterization, but that was not the case.

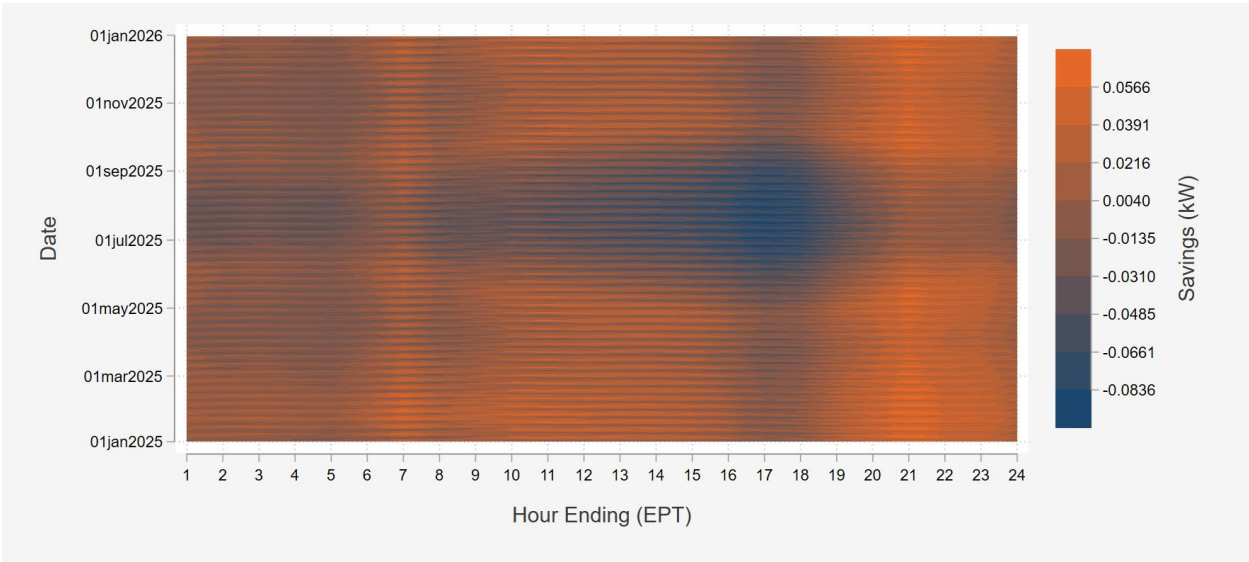
Table 15: ENERGY STAR Clothes Washer Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	130	0.013	0.018
Pre-Post TOWT (N = 259)	0 (100)	-0.060 (0.024)	0.007 (0.024)
Matched Controls without COVID-19 (N = 196)	76 (104)	-0.055 (0.021)	0.012 (0.026)
Matched Controls with COVID-19 (N = 370)	37 (84)	-0.039 (0.015)	0.009 (0.019)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed over the 8760 hours of the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The heat map for the ENERGY STAR clothes washer is displayed in Figure 37. The aggregate impact totals zero kWh so the heat map is not terribly useful.

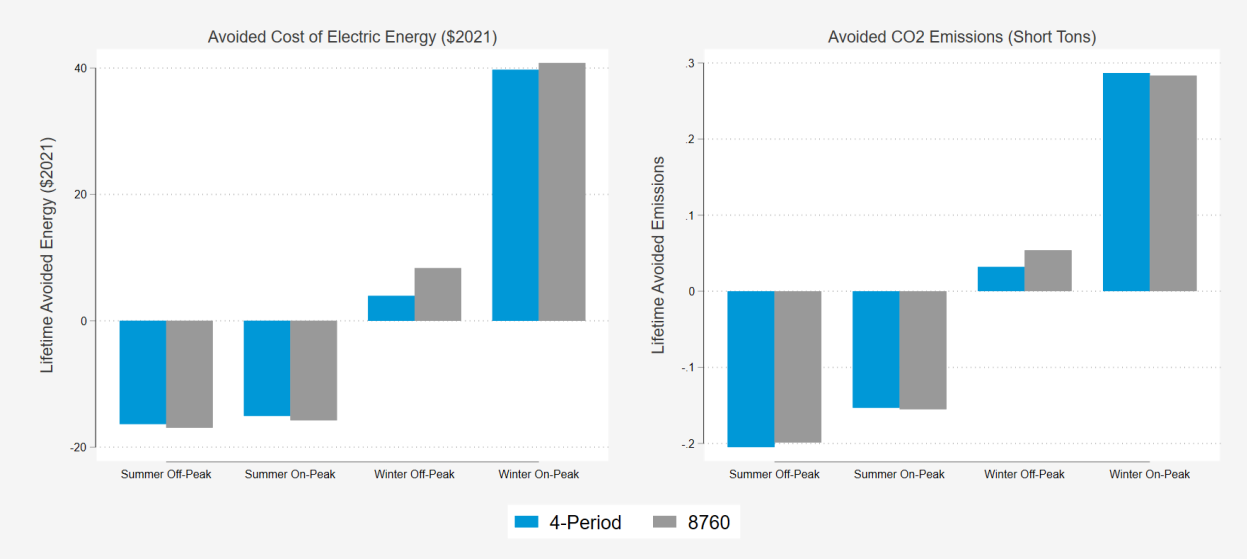
Figure 37: ENERGY STAR Clothes Washer Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 14-year measure life. These were calculated using both the 8760 valuation profiles, as well as the four

periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. Figure 38 shows the results. Interestingly, the avoided costs are positive and the emissions are negative for both valuation profiles. While we encourage readers to not read too much into these clothes washer results, it is a useful illustration of how interventions that are purely load-shifting can create real impacts with time-differentiated valuation profiles.

Figure 38: ENERGY STAR Clothes Washer Valuation Comparison



HOME PERFORMANCE WITH ENERGY STAR

Home Performance with ENERGY STAR (HPES) is not a single measure but an entire program track. HPES is a contractor-driven semi-custom offering to improve the thermal efficiency of homes. Eligible projects achieve minimum air leakage reduction of 10% (as measured by a pre- and post-blower door test) and installation of all recommended health and safety improvements. The main measures are air sealing and insulation. While the electric savings are modest (200-250 kWh per home), the fuel savings are large (10+ MMBtu). This makes sense given the prevalence of fossil fuel heat and the climate in Vermont. HPES does include some water heating and space heating fuel-switching measures, but the EVT tracking database never shows negative savings. It's unclear if any of the fuel switching is from fossil fuel to electricity (and the database just doesn't record negative values), or if fuel switching is exclusively between fossil fuels or from electricity to fossil fuel.

TRM Characterization

HPES is a classic **retrofit** offering. The pre-retrofit condition of the home is the baseline. This type of offering is well-suited to meter-based methods because the savings should be highly weather-sensitive. Unfortunately, expected electric signal size in Vermont is small given the low prevalence of electric heat and limited need for air conditioning.

Modeling Results

The impact estimates for HPES vary widely across modeling methods, ranging from positive savings to negative savings. Table 16 provides more detail into the savings estimates produced by the pre-post TOWT and the matched control groups, as well as the claimed database savings. Estimates for the pre-post TOWT and the matched controls without COVID-19 are not statistically significant. As discussed elsewhere in the report, a matched control group is needed to have any chance of modeling meter-based results during the pandemic, but it is not a perfect solution and matched control group results with COVID in the post-period likely still have an upward bias.

Table 16: Home Performance with ENERGY STAR Modeling Results

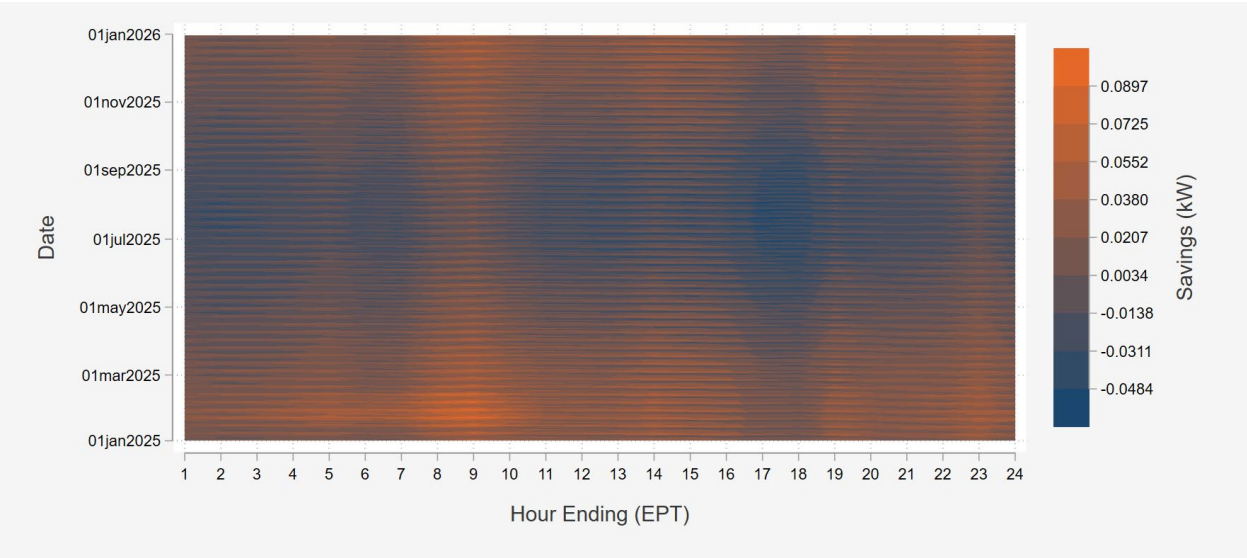
Source	kWh/year	Summer kW	Winter kW
EVT Database*	199	0.000	0.044
Pre-Post TOWT (N = 193)	37 (149)	-0.011 (0.022)	0.023 (0.028)
Matched Controls without COVID-19 (N = 145)	-187 (123)	-0.022 (0.020)	0.000 (0.030)
Matched Controls with COVID-19 (N = 291)	-332 (98)	-0.021 (0.017)	-0.064 (0.025)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed over the 8760 hours of the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used.

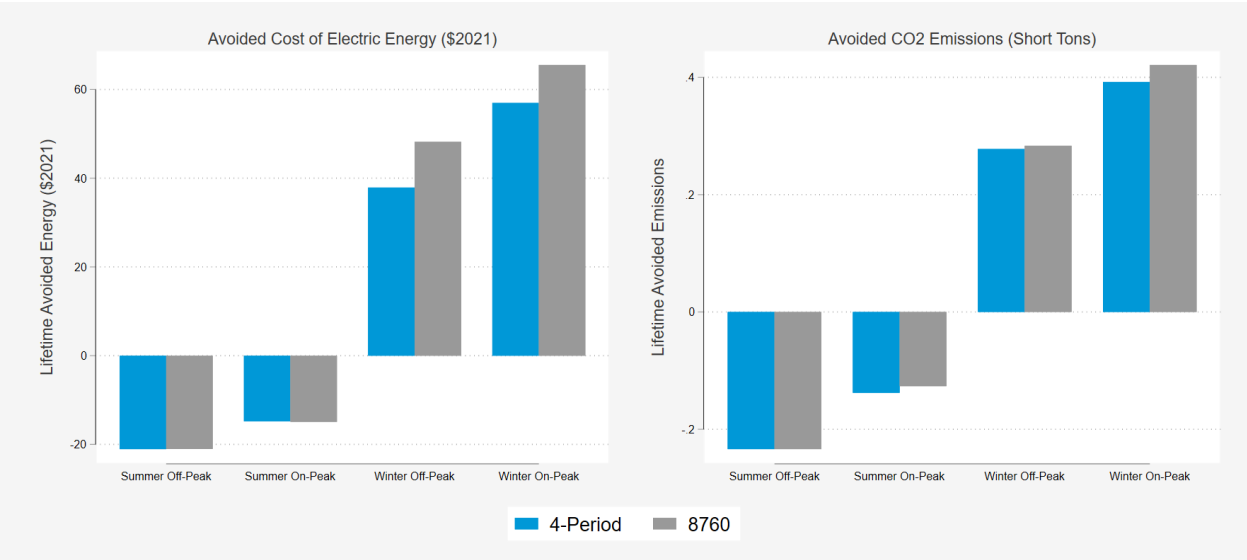
Figure 39 shows the heat map for HPES. Using the pre-post TOWT, the negligible savings produced are concentrated most heavily in the winter morning. The shape of the savings for this measure does not convey much, as the savings estimates and shape of the savings invariably changes across the different modeling techniques. Additionally there is not an anticipated saving shape for this measure since it is an amalgamation of various home performance with energy start measures.

Figure 39: Home Performance with ENERGY STAR Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over an average measure lift of 22 years. These were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $77.76, which is 24.2% greater than the four period avoided costs. Similarly, the avoided emissions sum to 0.34 short tons of CO_2 for the 8760 period, which is 13.4% higher than if the four-period assumptions were used. A more detailed summation of these avoided costs/emissions are presented in Figure 40.

Figure 40: Home Performance with ENERGY STAR Valuation Comparison



Heat Pump Water Heater

Heat pump water heaters (HPWH) use electricity to move heat from one place to another instead of generating heat directly from resistance coils. They operate like an air conditioner or refrigerator, but in reverse. HPWHs are two to three times more energy efficient than conventional electric resistance water heaters, but the efficiency level varies depending on the ambient temperature of the space they operate in. HPWHs are more efficient in warm spaces than cold spaces because it is easier to remove heat from warm air than cool air.

TRM Characterization

Aside from an infrequently used low-income direct install path, heat pump water heaters follow an end-of-life measure characterization in the EVT TRM. The TRM algorithms take interactive effects into account and the measure is bifurcated into units less than 55 gallons and greater than 55 gallons of tank capacity. Homes with natural gas water heaters are not eligible. Savings assumptions do depend on the existing water heater fuel though.

- **Fossil Fuel Baseline**: For cases where this measure is installed in a home with an existing electric resistance water heater or in a new construction project, electric savings account for the improvement in performance of a HPWH over a baseline electric resistance water heater. From an electric standpoint this measure permutation follows a **market opportunity** path because the baseline is a code-minimum resistance water heater rather than the replaced electric water heater.
- **Electric Baseline**: For homes with existing fossil fuel water heaters, the installation of a HPWH results in an electric penalty equal to the annual electricity use of the water heater to represent the added electric load. From an electric standpoint this measure permutation behaves like a **retrofit** measure because the baseline is no electric water heating and the efficient case is the program-supported HPWH.

DSA divided the HPWH analysis sample into sites with a fossil fuel baseline (negative electric savings claim) and sites with an electric baseline (positive electric savings claim), where both have an average installation of 1 per household. The following sections present the results.

Modeling Results – Fossil Fuel Baseline (Negative Savings)

Regardless of the modeling approach used for this measure, we see a substantial increase in the average electric load of the sites analyzed. This is consistent with the measure characterization and represents the newly electric water heating load of the program-supported HPWH. Table 17 compares the savings estimates produced by the pre-post TOWT and the matched control groups with the claimed database savings. While the sample size of between 91 and 149 homes is one our smaller residential pooled regressions, the signal size is large. This leads to statistically significant energy and winter demand estimates for all three models. Our regression-based estimates are generally smaller (less negative) than the claimed savings stored in the EVT database. This could suggest that installed HPWHs are more efficient than the assumptions in the TRM measure characterization or that Vermont

water heating needs are smaller than assumed. Our results align well with the assumed shape in the EVT TRM and show larger peak demand impacts in the winter.

Table 17: HPWH with Fossil Fuel Baseline Modeling Results

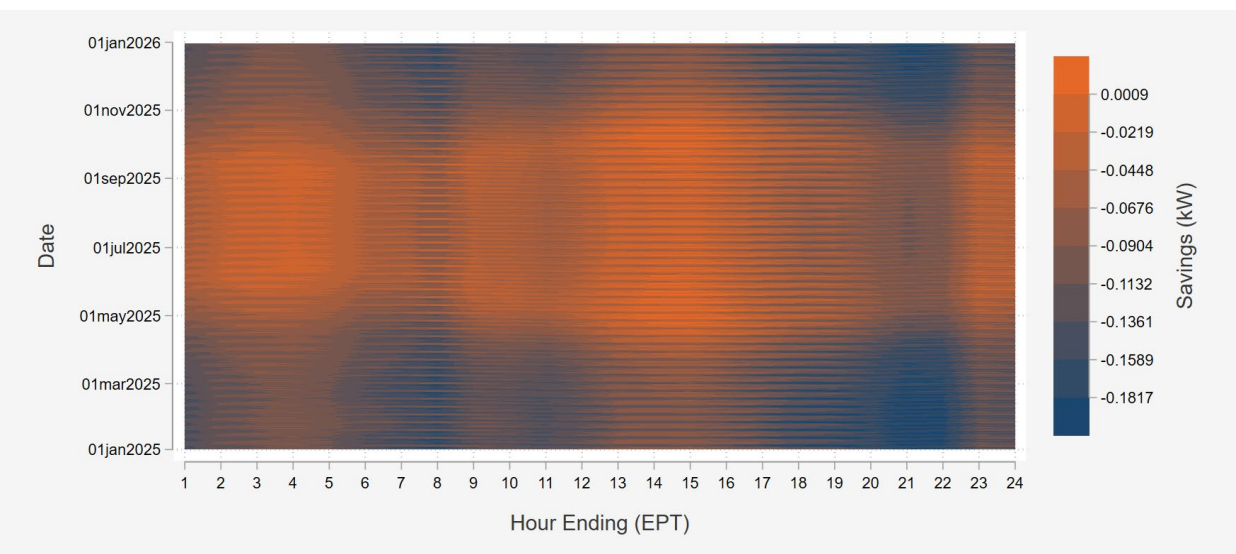
Source	kWh/year	Summer kW	Winter kW
EVT Database*	-922	-0.074	-0.146
Pre-Post TOWT (N = 104)	-723 (258)	-0.017 (0.031)	-0.126 (0.069)
Matched Controls without COVID-19 (N = 91)	-675 (185)	0.004 (0.028)	-0.107 (0.037)
Matched Controls with COVID-19 (N = 149)	-619 (154)	-0.024 (0.026)	-0.115 (0.033)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed over the 8760 hours of the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used.

Figure 41 is made up of almost all negative values, which is expected for a fuel switching measure. Load increases (blue regions) are concentrated during mornings and evenings. They are also larger in the winter months – possibly because HPWHs operate less efficiently when ambient temperatures are low.

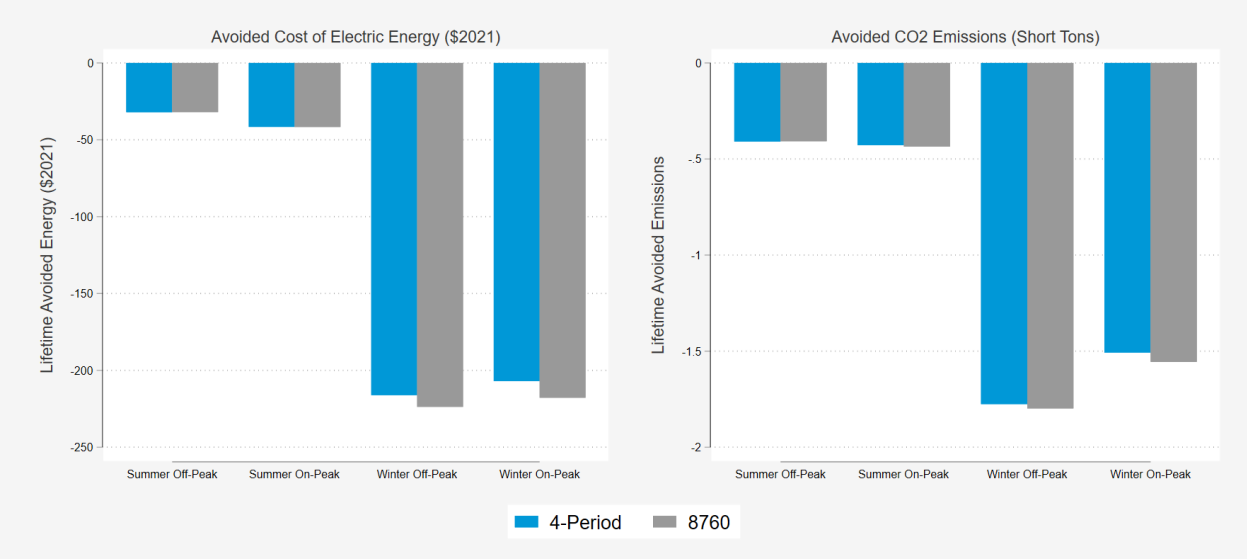
Figure 41: HPWH with Fossil Fuel Baseline Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 13-year measure life. These were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results

for avoided costs of electric energy is -$515.27, which is 3.6% greater than the four period avoided costs. Similarly, the avoided emissions sum to 4.20 short tons of CO_2 for the 8760 period, which is 1.8% higher than if the four-period assumptions were used. Figure 42 provides a detailed summary of these avoided costs/emissions by costing period.

Figure 42: HPWH with Fossil Fuel Baseline Valuation Comparison



Modeling Results – Electric Baseline (Positive Savings Claim)

In contrast with the households that participate in the fuel switching version of the heat pump water heater measure, households that replace an electric resistance water heater with a HPWH claim positive annual savings. Table 18 compares the savings estimates produced by three regression methods to the claimed database savings. We would expect the regression-based estimates to be larger than the EVT database claim if the replaced electric water heater was less efficient than current code-minimum, but that is not what the results show. While these annual savings are positive, the meter-based estimates are a fraction of the impacts stored in the EVT tracking database. Interestingly, the summer demand reductions are as large, or larger, than the winter kW savings for this group of homes. This may be driven by the fact that HPWHs operate more efficiently in warm ambient

temperatures. If HPWHs are installed in an unconditioned or semi-conditioned space, we would expect more efficiency gains over electric resistance in the summer than in the winter.

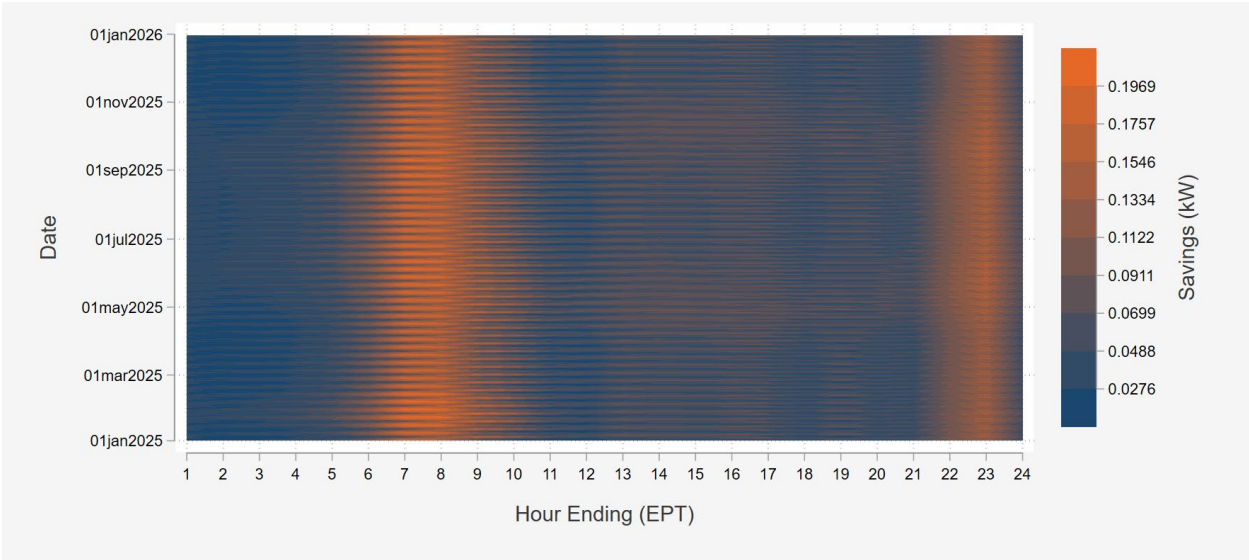
Table 18: HPWH with Electric Baseline Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	1,888	0.151	0.299
Pre-Post TOWT (N = 309)	623 (141)	0.062 (0.017)	0.058 (0.029)
Matched Controls without COVID-19 (N = 278)	350 (129)	0.081 (0.019)	0.020 (0.028)
Matched Controls with COVID-19 (N = 446)	242 (110)	0.076 (0.017)	-0.023 (0.024)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed over the 8760 hours of the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. HPWHs with an electric resistance baseline see a concentration of savings in the morning and the evening, which is when residential water heating usage is highest.

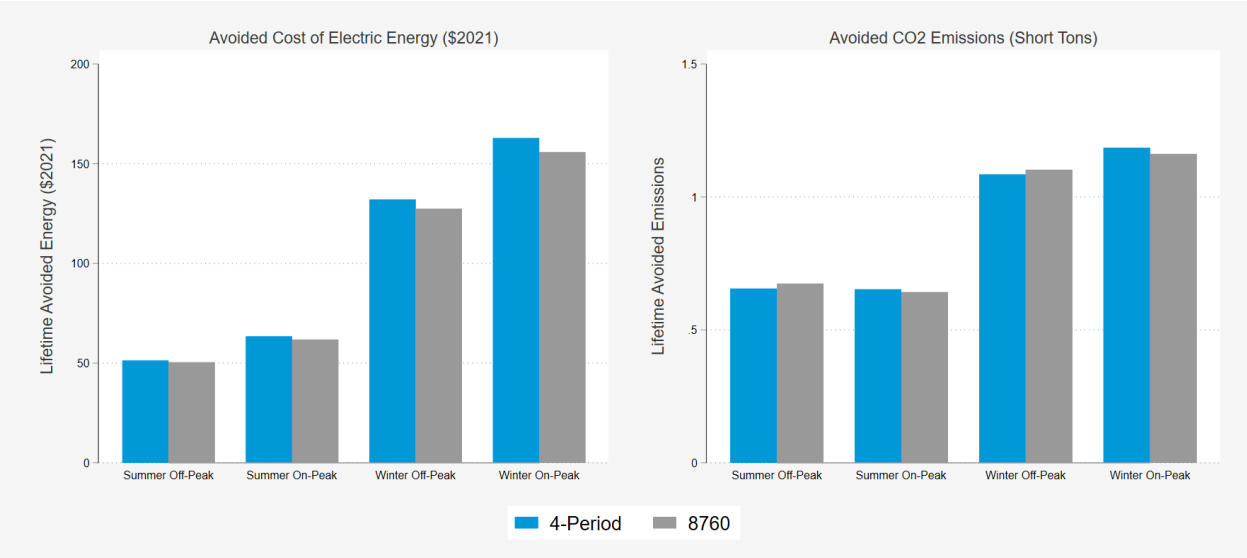
Figure 43: HPWH with Electric Baseline Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 13-year measure life. These were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $395.44, which is 3.6% lower than the four period avoided costs.

Similarly, the avoided emissions sum to 3.58 short tons of CO_2 for the 8760 period, which is 0.0% higher than if the four-period assumptions were used. Figure 44 presents a more detailed summary of the lifetime avoided costs and emissions.

Figure 44: HPWH with Electric Baseline Valuation Comparison



LED Indoor Fixture

This residential measures includes residential lighting that is wired rather than screw-based. Common applications include track lighting, wall-wash, sconces, ceiling downlight kits, and fan lights. The equipment is generally sold through participating lighting distributors. Since the equipment is installed inside homes, the interactive effects on the homes heating and cooling system are considered.

TRM Characterization

This is a **market opportunity** measure which relies on EISA-equivalent wattages as a baseline. Functionally this means a halogen or incandescent baseline is assumed for each fixture. The details of the TRM characterization evolved over our period of investigation, but between 58 and 72 kWh/year were claimed for nearly all fixtures. Households in our analysis installed an average of 4.4 units.

Modeling Results

Table 19 compares savings estimates produced by the pre-post TOWT and the matched control groups to the claimed database savings. Our meter-based estimates for this measure are consistently lower than the annual savings estimates stored in the EVT tracking data, and none of the estimates are statistically significant. Our regression-based inherently model the existing condition of the home as the baseline as opposed to the new code minimum fixture assumed in the TRM characterization. EVT does not collect the details of the replaced equipment, which could be more or less efficient than the

TRM baseline assumption. It is also possible that some program-supported fixtures are entirely new to the home (e.g., an added load the meter).

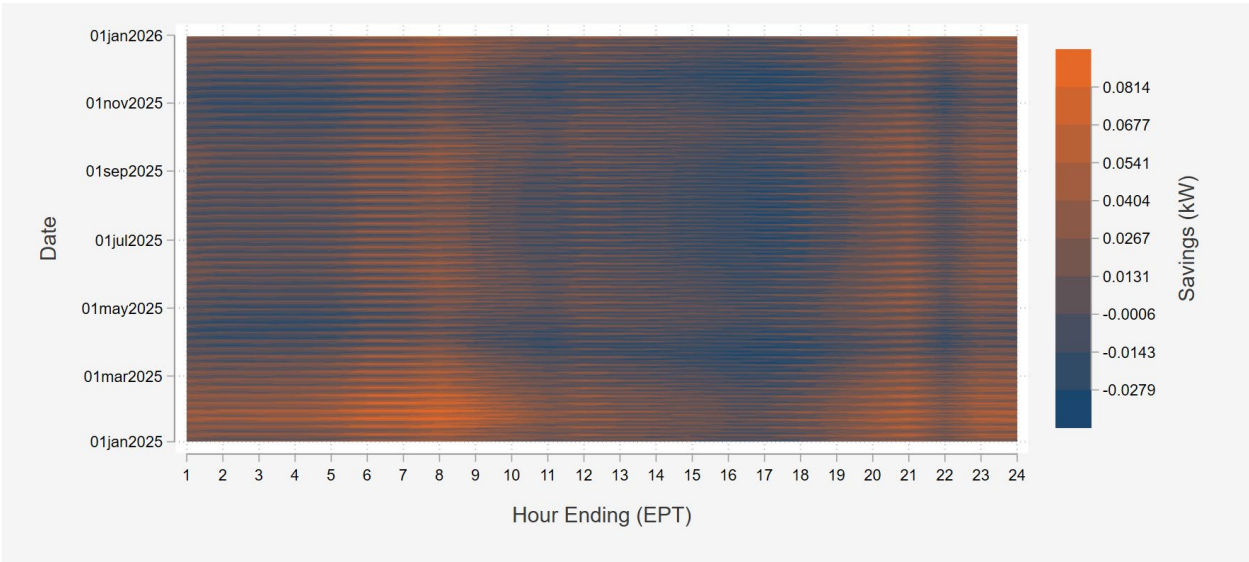
Table 19: LED Indoor Fixture Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	263	0.022	0.079
Pre-Post TOWT (N = 205)	82 (136)	-0.008 (0.025)	0.014 (0.031)
Matched Controls without COVID-19 (N = 182)	40 (146)	0.039 (0.023)	-0.006 (0.032)
Matched Controls with COVID-19 (N = 305)	-116 (131)	0.009 (0.025)	-0.032 (0.025)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual impacts are distributed over the 8760 hours in the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The heat map for LED indoor fixtures is displayed in Figure 45. Using the pre-post TOWT, LED Indoor Fixture savings are concentrated in the morning and evening. Unfortunately, the magnitude and direction of savings estimates for this measure change with different modeling techniques, as does the overall 8760 load shape.

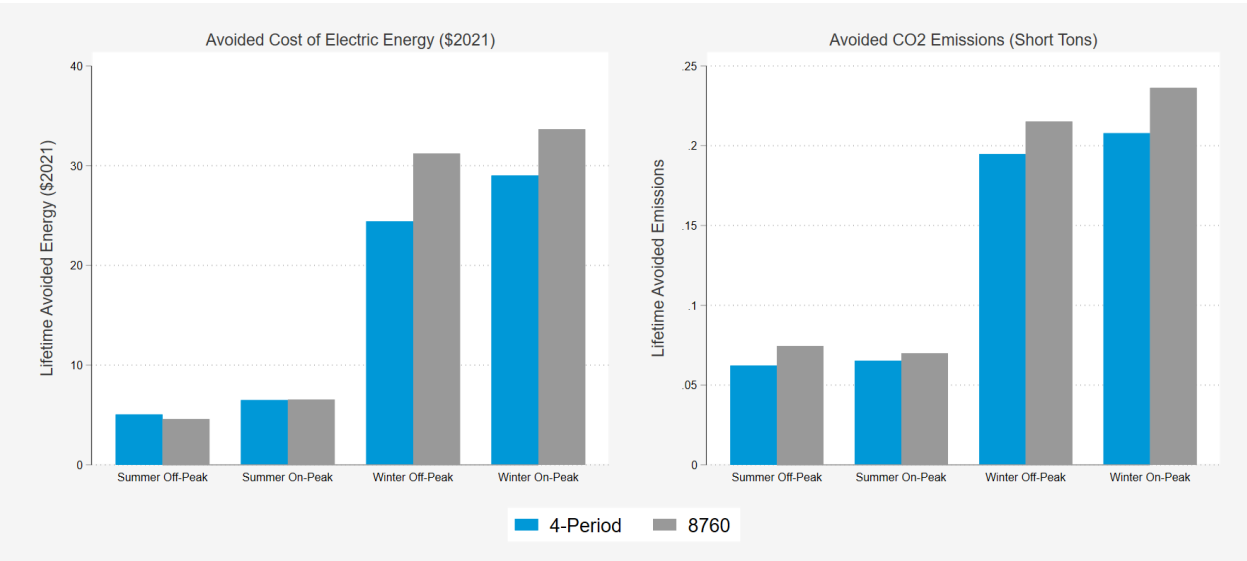
Figure 45: LED Indoor Fixture Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 15-year measure life. These resource impacts were calculated using both the 8760 valuation profiles,

as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $76.00, which is 14.5% greater than the four period avoided costs. Similarly, the avoided emissions sum to 0.60 short tons of CO_2 for the 8760 period, which is 11.0% higher than if the four-period assumptions were used. A more detailed summation of these avoided costs/emissions are presented in Figure 46.

Figure 46: LED Indoor Fixture Valuation Comparison



LED Screw-Base Lamp

This measure includes most screw-based LED light bulbs, including standard A-lamps, reflectors, globes, and candelabras. EVT offers low or no cost LED light bulbs through a variety of delivery mechanisms including retail, direct install, and dropship. LEDs produce light much more efficiently than traditional lighting technologies, last far longer, and generate less waste heat. During our period of investigation, this measure was by far the largest residential measure in terms of electric resource savings.

TRM Characterization

The LED screw-base lamp measure includes multiple permutations, but most volume follows a **market opportunity** path, which assumes a code-minimum halogen or incandescent lamp as the baseline. Operating assumptions such as hours-of-use, in-service rate, and interactive effects combine to create a limited roster of per-unit savings that range from 30 to 50 kWh annually per lamp, with households in our analysis installing an average of 7.8 units. This measure uses a "dual baseline" to account for an assumed change in codes and standard over the course of the 15-year mechanical life of the measure. The mechanics of this dual baseline changed over our period of analysis so, for simplicity, our valuation calculations just consider the full life of the lamp.

Modeling Results

Table 20 compares savings estimates produced by the pre-post TOWT and the matched control group regressions to the claimed database savings. The standard errors of this measure are much smaller than the standard errors of the LED indoor fixtures, due to the much larger sample size. The low per-home energy savings estimates are noteworthy for this measure given the precision of the estimates, and the contribution of residential LED lamps measure to the portfolio. Our hypothesis is that some of the program-supported lamps actually replaced CFLs rather than the code-minimum halogen or incandescent lamps assumed in the TRM baseline. Of course, the homeowner could have replaced the lamps with a code minimum lamp – hence the market opportunity characterization. Whether that is likely or not is an attribution question and outside the scope of this study.

Table 20: LED Screw-Base Lamp Modeling Results

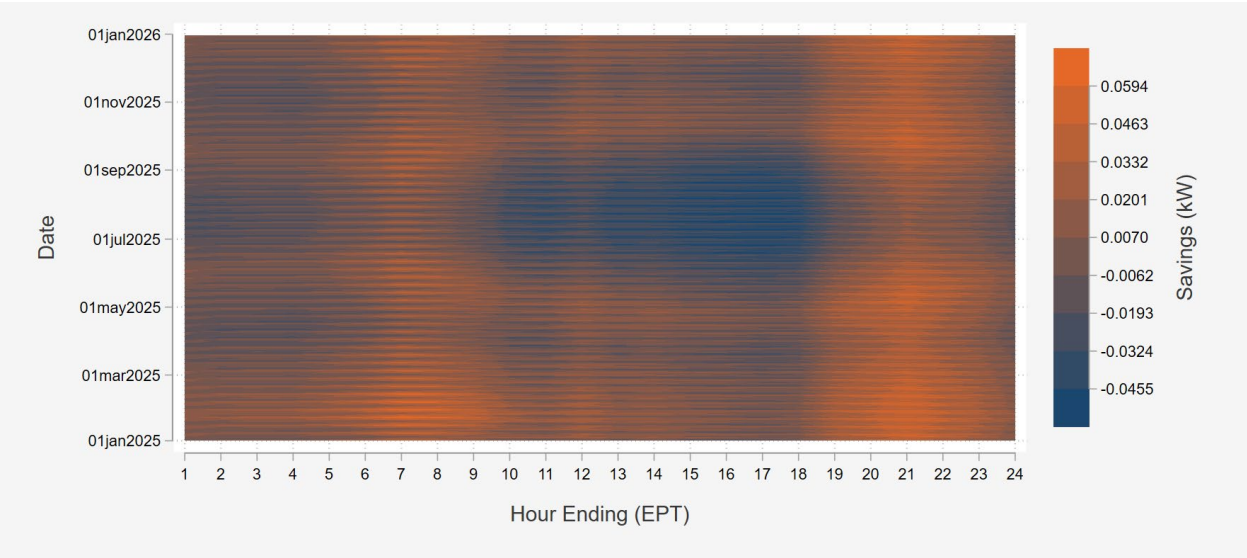
Source	kWh/year	Summer kW	Winter kW
EVT Database*	336	0.028	0.101
Pre-Post TOWT (N = 1,048)	29 (53)	-0.035 (0.009)	0.015 (0.013)
Matched Controls without COVID-19 (N = 937)	134 (68)	0.001 (0.012)	0.045 (0.015)
Matched Controls with COVID-19 (N = 1,692)	-2 (46)	-0.001 (0.008)	0.012 (0.011)

*Reported for the sites used in the Pre-Post TOWT.

To get a more granular look at how the annual savings are distributed over the 8760 hours in the year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The heat map for LED screw base lamps is displayed in Figure 47. Savings are concentrated in the evening – which is typical for residential lighting. Unfortunately, the savings load shape of this measure is inconsistent across the different modeling techniques. We would recommend Vermont rely on the 8760 load shape from an end-use light logging study[14] rather than an estimated shape from whole-building data.

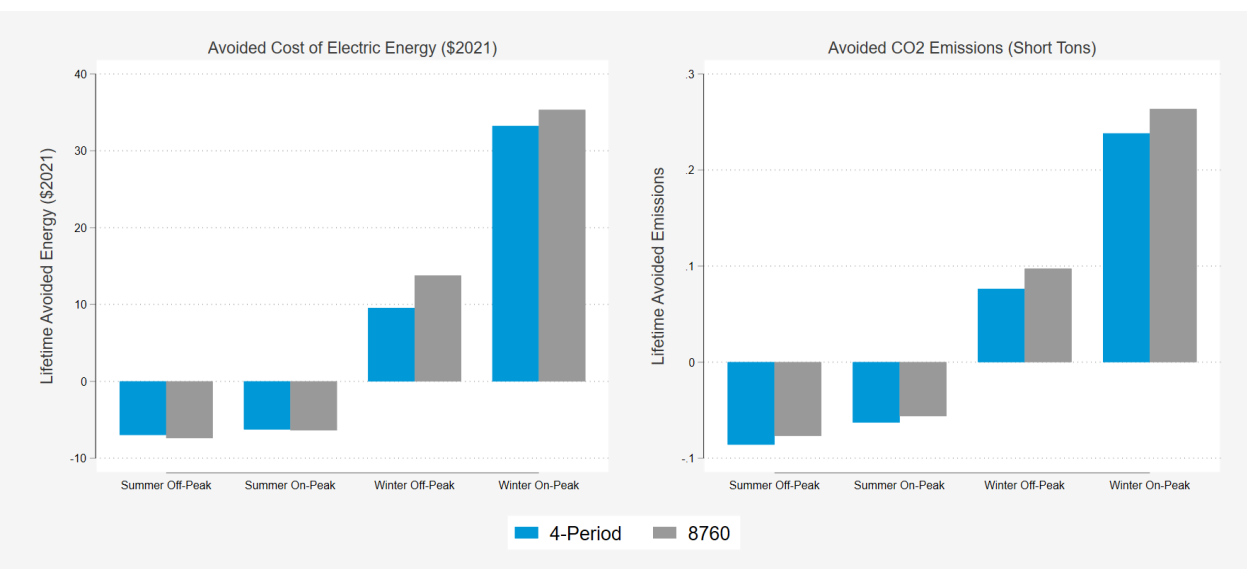
[14] Link to 8760 load shapes from a 2015 Pennsylvania study https://www.puc.pa.gov/Electric/xls/Act129/SWE-Light_Metering_Study-AppendixA-Residential_Load_Shapes.xlsx

Figure 47: LED Screw Base Lamp Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 15-year measure life. These were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $35.30, which is 16.3% greater than the four period avoided costs. Similarly, the avoided emissions sum to 0.23 short tons of CO_2 for the 8760 period, which is 27.4% higher than if the four-period assumptions were used. Figure 48 presents a more detailed summary of these avoided costs and emissions by costing period and method.

Figure 48: LED Screw Base Lamp Valuation Comparison



High Efficiency Pool Pump

High efficiency pool pumps circulate pool water more efficiently through use of a two-speed or variable speed motor. Variable speed pumps use a pressure sensor to modulate speeds appropriate for the water pressure rather than always running at full-speed. Pool pumps need to run frequently, but not at full speed so variable speed motors save both energy and wear and tear on the pool pump.

TRM Characterization

The efficient pool pump measure follows a **retrofit** characterization in the EVT TRM and assumes that program supported units replace a single-speed motor. In 2017 and 2018, virtually all units claimed 2,111 kWh per year of energy savings. In 2019, EVT expanded the measure characterization to include above ground pool pumps and reflect upcoming changes in the federal standard and ENERGY STAR specification. Per-unit savings for the most common in-ground pumps also climbed to 2,362 kWh annually. Across our analysis sites, all households only installed 1 unit.

Modeling Results

Across all three modeling approaches, this measure produces modest annual savings compared to the database claims. Table 21 provides more detail into the savings estimates produced by the pre-post TOWT and the matched control groups. The variation between the database claim and meter-based estimates is most pronounced for summer peak demand savings with the database claim being 10-20 times higher than the modeling results. With large claimed savings and a decent sample size, this measure should be an excellent candidate for this methodology, so these results are unexpected. One possibility is that a subset of homes did not have a pool or pool pump previously so the high-efficiency pool pump was a new load at the meter.

Table 21: Pool Pump Modeling Results

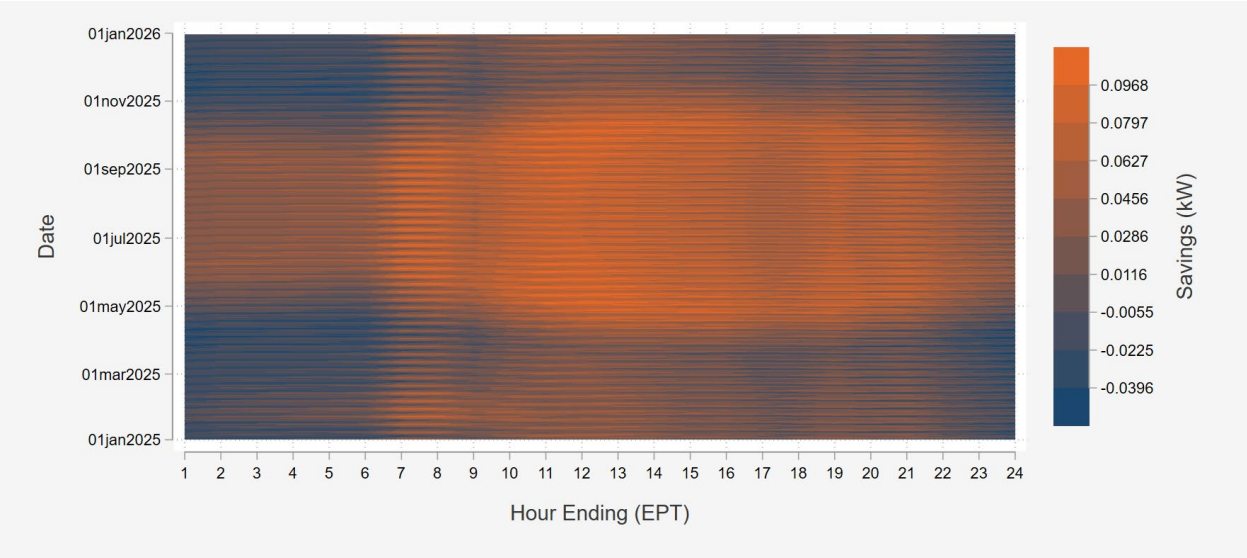
Source	kWh/year	Summer kW	Winter kW
EVT Database*	2,122	1.363	0.000
Pre-Post TOWT (N = 268)	276 (133)	0.065 (0.032)	0.016 (0.024)
Matched Controls without COVID-19 (N = 212)	293 (125)	0.096 (0.019)	0.026 (0.032)
Matched Controls with COVID-19 (N = 297)	150 (121)	0.090 (0.022)	-0.019 (0.028)

*Reported for the sites used in the Pre-Post TOWT.

To visualize impacts across the 8760 hours of a year, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. The y-axis shows future dates to make it clear that actual weather is not used. Figure 49 indicates that the savings for the pool pump motor is concentrated over the summer months, even more so during

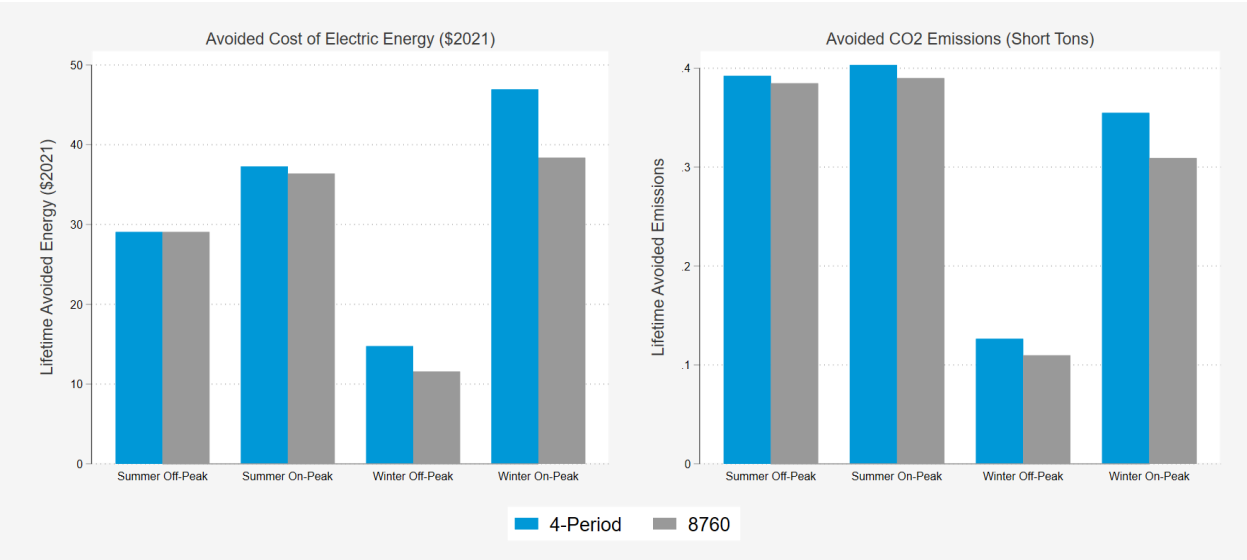
daylight hours, when pools would be in use. This shape is consistent with the TRM measure characterization even though regression-based savings estimates are approximately one-tenth of the TRM-based savings.

Figure 49: Pool Pump Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 10-year measure life. These resource impacts were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is $115.40, which is 10.9% lower than the four period avoided costs. Similarly, the avoided emissions sum to 1.19 short tons of CO_2 for the 8760 period, which is 7.0% lower than if the four-period assumptions were used. A more detailed summation of these avoided costs/emissions are presented in Figure 50.

Figure 50: Pool Pump Valuation Comparison



REFRIGERATOR

Refrigerators are a ubiquitous household appliance used to store food and beverages and prevent spoilage. Most refrigerators have two compartments; one for fresh food and one for frozen food. Household refrigerators use a vapor compression cycle where refrigerant is compressed and passed through coils or tubes that that are passively cooled by exposure to air in the room. A fan blows air over the coils to deliver cool air to the insulated compartments where food is stored.

TRM Characterization

The EVT TRM includes both an early replacement and market opportunity measure characterization but savings were claimed almost exclusively using the **market opportunity** during our period of investigation. Savings are specified separately for three different efficiency tiers, with most units accounting for less 100 kWh of claimed annual electric savings. The measure characterization was the same for 2017 and 2018 and then updated for 2019. The 2019 characterization follows the same general architecture and has slightly more conservative per-unit savings. Across our analysis sites, households installed an average of 1 unit.

Modeling Results

Table 22 compares normalized savings estimates from the pre-post TOWT and the matched control groups to the claimed database savings. Each modeling procedure produces an annual savings estimate that is positive and statistically significant at the 95% confidence interval and generally larger than the

database claims. This is likely due to the replaced refrigerator being less efficient than the assumed code minimum baseline.

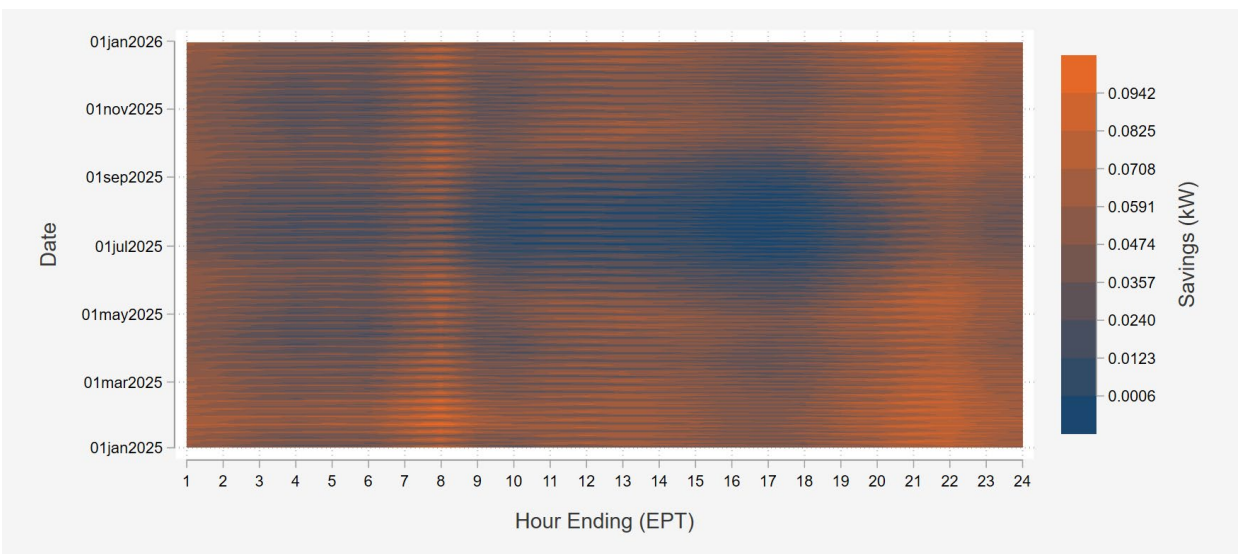
Table 22: ENERGY STAR Refrigerator Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	70	0.008	0.007
Pre-Post TOWT (N = 425)	379 (77)	0.011 (0.013)	0.054 (0.018)
Matched Controls without COVID-19 (N = 363)	232 (98)	0.003 (0.015)	0.063 (0.019)
Matched Controls with COVID-19 (N = 623)	211 (77)	0.007 (0.013)	0.051 (0.016)

*Reported for the sites used in the Pre-Post TOWT.

To visualize the load shape of the annual impact estimate, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. Figure 51 shows a more seasonal trend than expected. Refrigerators operate year-round, so we would expect the savings to be more evenly spread out across the year. Or it may be the case that the shape is a result of efficiency gains being lost when temperatures are higher, such as in the summer afternoon.

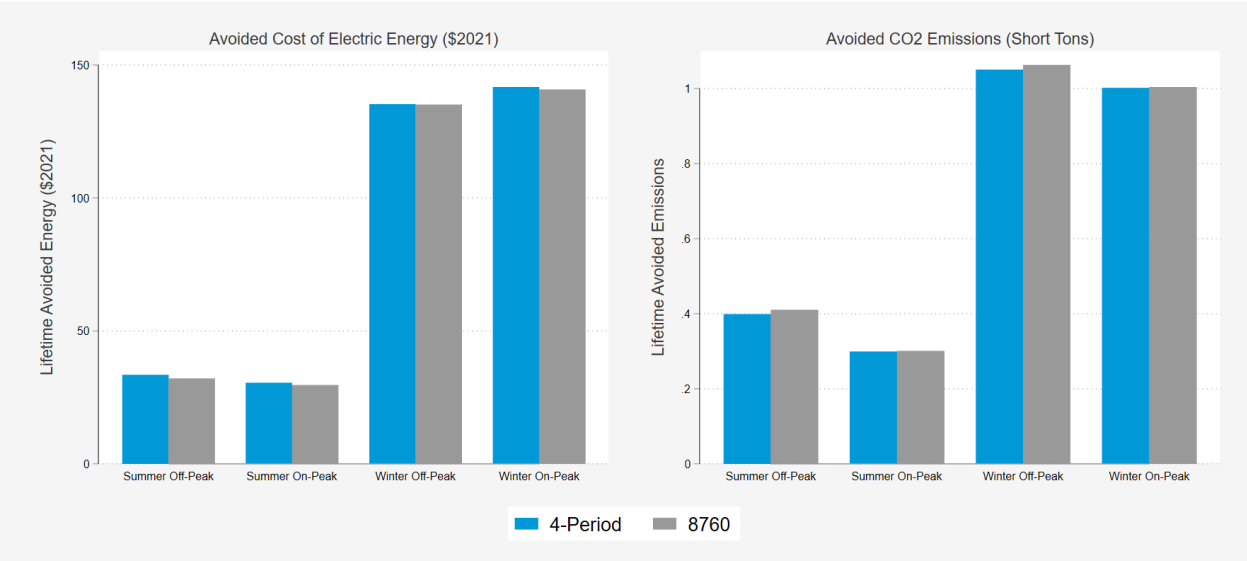
Figure 51: ENERGY STAR Refrigerator Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 17-year measure life. These benefit streams were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak.

The 8760 results for avoided costs of electric energy is $337.47, which is 1.0% lower than the four-period avoided costs. Similarly, the avoided emissions sum to 2.78 short tons of CO_2 for the 8760 period, which is 1.0% higher than if the four-period assumptions were used. Figure 52 presents detailed summation of benefits by costing period and valuation profile type. .

Figure 52: ENERGY STAR Refrigerator Valuation Comparison



Variable Speed Cold Climate Heat Pump

The variable speed heat pump is an energy efficiency measure that provides both heating and cooling. The means of efficiency comes from the heat pump's ability to transfer heat using a compressor, instead of generating the heat itself. During cooler months, the system draws heat from the environment and distributes it indoors. Conversely, in the hotter months, the system's function reverses and moves heat that is concentrated indoors and distributes it outdoors. Additional benefits include: these units do not require venting or ductwork, making installation costs lower, and they allow for zone-specific temperature control, reducing operating costs. The installation of this efficiency measure tends to coincide with an increase in electric consumption, due to the process of customers fuel switching from fossil fuels to electricity. While this increases demand on the grid, it also replaces these inefficient fuels with a cleaner alternative.

TRM Characterization

The TRM categorizes the variable speed mini-split heat pump in terms of both a market opportunity and retrofit. The retrofit characterization can be better understood as fuel switching. In these cases, the home's existing heating system is fossil fuel or electric resistance. Retrofit measures are classified as

having negative savings, or increasing electric consumption. This is to account for the heating offset, where the heat pump is being used to provide supplemental heat. The added electric load is counted as a penalty. On the other hand, the market opportunity characterization claims a positive savings, or decreased electric consumption, as it accounts for the incremental savings of an efficient heat pump, in juxtaposition to the installation of a baseline heat pump.

In the 2017 and 2018 tracking data, market opportunity and retrofit records come in pairs for each participant. The net consumption of this pair of records is negative, indicating increased consumption with the installation of the heat pump. There is a significant change in this characterization between the 2018 and 2019, which affects how the variable speed mini-split heat pumps are claimed. Beginning in 2019, the claims no longer come as a pair of market opportunity and retrofit. Instead, the measure is only claiming the positive savings of the market opportunity. This differentiation is highlighted in Table 23. While the claimed savings approach changed in 2019, the effect at the meter does not. Adding a heat pump increases electric heat consumption whether those negative savings are claimed or not.

Table 23: EVT Database Claim Summary

Year	Average Database Savings (kWh)	Count of Negative Savings (N)	Count of Positive Savings (N)
2017	-2,549	2,546	2,549
2018	-811	1,463	2,673
2019	1,001	0	4,174
2020	965	0	6,762

Modeling Results

Our modeling efforts consistently show an increase in the electric load at the sites analyzed – particularly during the winter peak demand period. Most of the homes in our TOWT analysis sample followed the two-part savings claim. As the TRM began to only claim the market opportunity savings of installing a heat pump, the average claimed savings is skewed more positively. When we allow post-retrofit data from the COVID period into the modeling the average database claim becomes positive. Table 24 details the savings produced by the pre-post TOWT and the matched control group in comparison to these claimed database savings. All savings estimates are statistically significant.

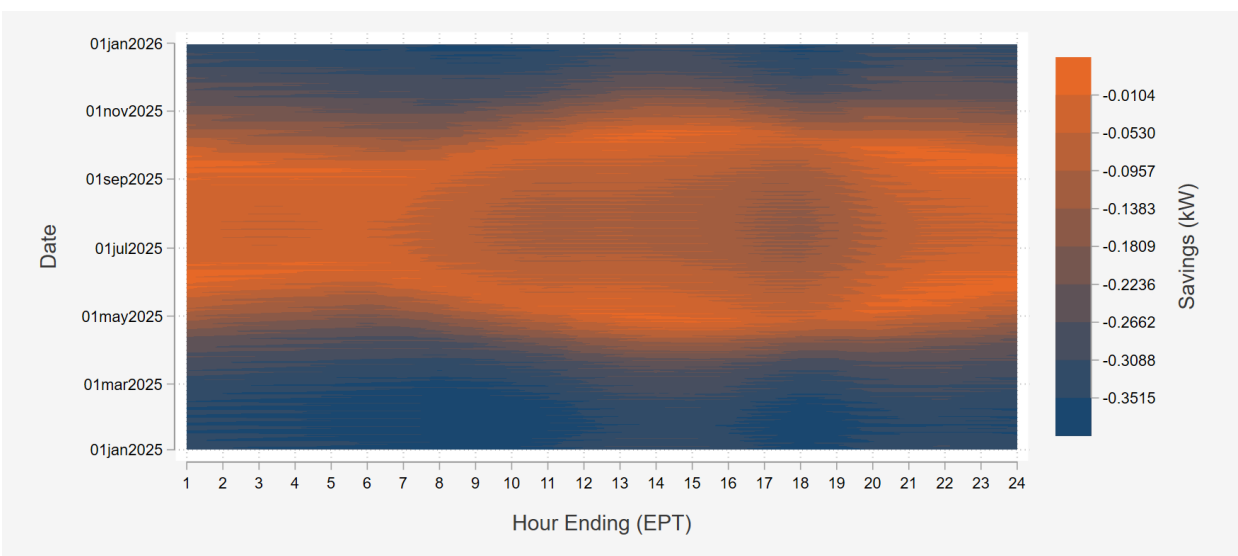
Table 24: Variable Speed Heat Pump Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	-415	-0.040	-0.203
Pre-Post TOWT (N = 766)	-1,539 (99)	-0.111 (0.012)	-0.350 (0.024)
Matched Controls without COVID-19 (N = 700)	-1,252 (72)	-0.016 (0.010)	-0.315 (0.017)
Matched Controls with COVID-19 (N = 1,876)	-1,272 (50)	-0.036 (0.008)	-0.296 (0.012)

*Reported for the sites used in the Pre-Post TOWT.

We use the pre-post TOWT model to generate a heat map of hourly impacts. This visual plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The variable speed heat pump model estimates a large load increase during the winter months. Negative savings is a good thing in this case because it means Vermont homes are using the heat pump for heating and displacing delivered fuel consumption. There is also a pocket of denser load increase in the summer months during the middle of the day, presumably due to homes that may not have had air conditioning previously using the heat pump as an air conditioner.

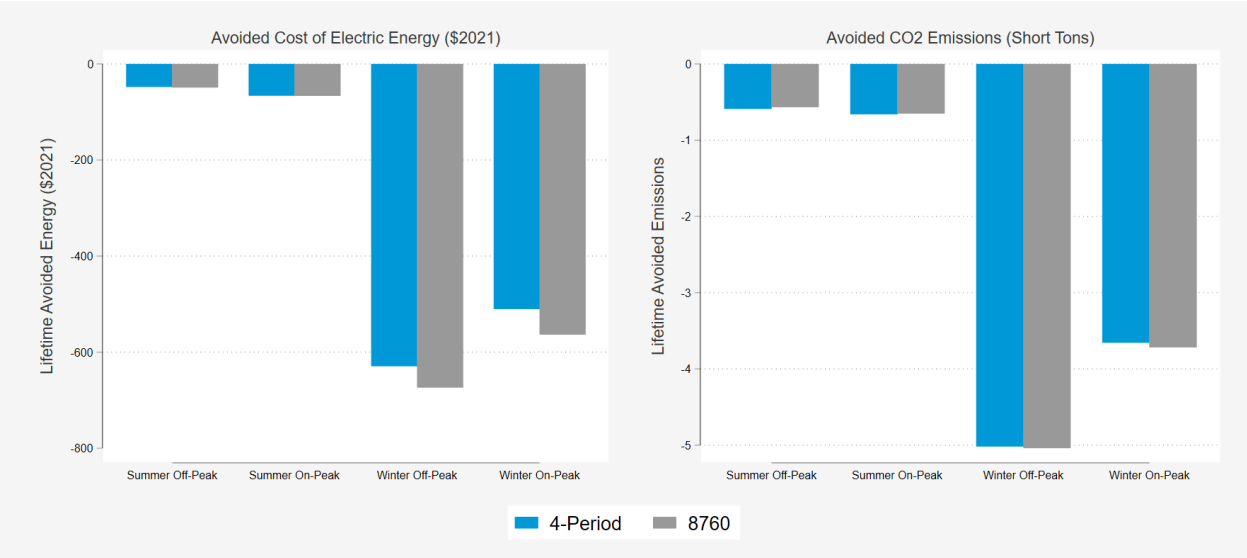
Figure 53: Variable Speed Heat Pump Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 15- year measure life. These benefit streams were calculated using both the 8760 valuation profiles, as well as the four periods of summer on-peak, summer off-peak, winter on-peak, and winter off-peak. The 8760 results for avoided costs of electric energy is -$1,352.44, which is 7.4% greater than the four period avoided costs. Similarly, the avoided emissions sum to 9.98 short tons of CO_2 for the 8760

period, which is 0.5% higher than if the four-period assumptions were used. Importantly, that estimate does not account for the reduced emissions from on-site fossil fuel combustion. Figure 54 provides a more detailed summary of these avoided costs and emissions by costing period and method.

Figure 54: Variable Speed Heat Pump Valuation Comparison



Overview of Prescriptive Residential Results

Table 25: Overview of Prescriptive Residential Results

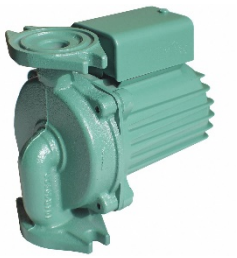
Measure	Vintage	kWh/year		Key Takeaway
		Database	Pre-Post TOWT	
Advanced Thermostat (N = 395)	Retrofit	85	9 (83)	Effect size is likely too small for this approach.
BLPM Boiler Motor – Residential (N = 343)	Retrofit	185	190** (102)	Effect size is likely too small for this approach.
ENERGY STAR Dehumidifier (N = 223)	Market Opportunity	230	-509*** (99)	This approach produces consistent findings across modeling methods.
ENERGY STAR Clothes Dryer (N = 82)	Market Opportunity	221	337** (182)	Potential to use this approach with a larger sample size.
ENERGY STAR Clothes Washer (N = 259)	Market Opportunity	130	0 (100)	Potential to use this approach with a larger sample size.
Home Performance with ENERGY STAR (N = 193)	Retrofit	199	37 (149)	Potential to use this approach with a larger sample size.
Heat Pump Water Heater – Fossil Fuel Baseline (N = 104)	Market Opportunity + Fuel Switch	-922	-723 (258)***	This approach produces consistent findings across modeling methods.
Heat Pump Water Heater – Electric Baseline (N = 309)	Market Opportunity	1,888	623*** (141)	This approach produces consistent findings across modeling methods.
LED Indoor Fixture (N = 205)	Retrofit	263	82 (136)	TRM characterization likely overstates the in-situ baseline.
LED Screw Base Lamp (N = 1,048)	Market Opportunity	336	29 (53)	TRM characterization likely overstates the in-situ baseline.
High Efficiency Pool Pump (N = 268)	Retrofit	2,122	276*** (133)	This approach produces consistent findings across modeling methods.
Refrigerator (N = 425)	Market Opportunity	70	379*** (77)	This approach produces consistent findings across modeling methods.
Variable Speed Cold Climate Heat Pump (N = 766)	Market Opportunity + Fuel Switch	-415	-1,539*** (99)	This approach produces consistent findings across modeling methods.

* Significant at an 80% confidence interval.
** Significant at a 90% confidence interval.
*** Significant at a 95% confidence interval.

4.3 PRESCRIPTIVE COMMERCIAL

BLPM Circulator Pump – Commercial

Brushless permanent magnet (BLPM) circulator pump motors are found in businesses with baseboard and radiant floor hydronic heating systems. This is the motor that moves hot water from the boiler to various areas in the building. BLPM pumps are more efficient because they lack brushes that add friction to the motor and their ability to modulate their speed to match the load. As the system flow demand changes (zones open or close), the drive senses the torque difference at the impeller via the change in the magnetic field difference and adjusts its speed by altering the frequency to the motor. BLPMs are especially efficient in no-load/low-load applications.



TRM Characterization

The commercial BLPM measure follows a **retrofit** characterization in the Vermont TRM and has been unchanged since January 2017. The TRM characterization is limited to BLPM motors less than or equal to three horsepower and is restricted to systems that use high mass boilers (>300,000 Btu/h) where the primary loop circulator runs constantly during the heating season. The measure has three size bins with deemed per-unit annual electric savings of 401, 780, and 1,924 kWh.

Modeling Results

Table 26 compares our regression-based estimates to the claimed savings in EVT's tracking database. Small sample sizes lead to large standard errors and imprecise impact estimates for the commercial BLPM measure. While the point estimate for each procedure indicates an increase in electricity consumption in the post-installation period, the impacts are not statistically significant. As a retrofit measure with sizable per-unit impacts, this measure would be a good candidate for meter-based methods if not for the limited sample size.

Table 26: Commercial BLPM Motor Modeling Results

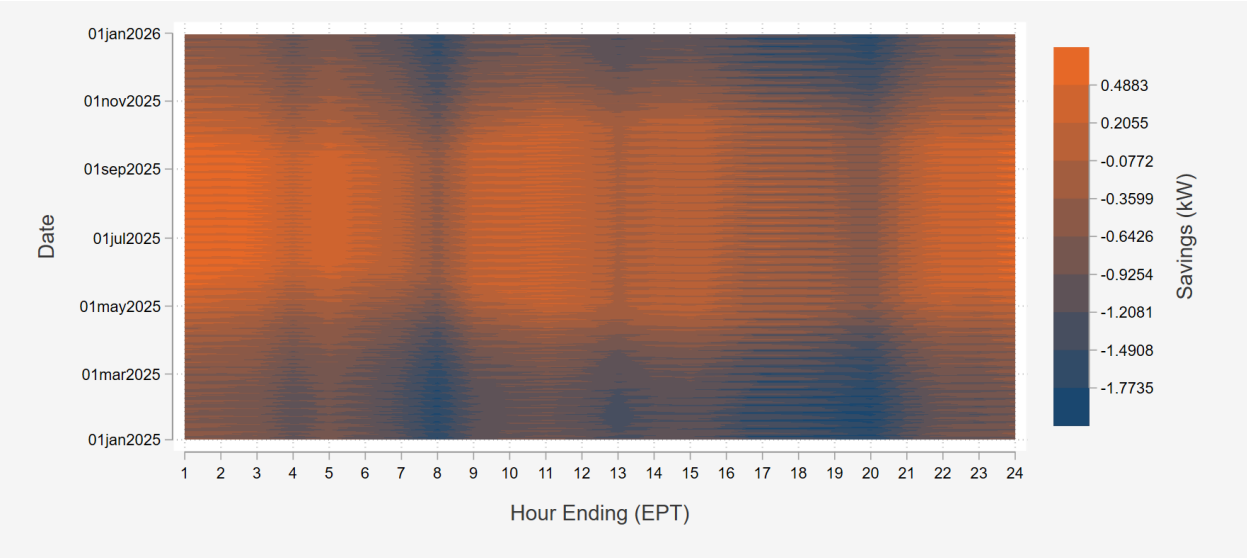
Source	kWh/year	Summer kW	Winter kW
EVT Database*	2,214	0.001	0.133
Pre-Post TOWT (N = 9)	-3,803 (2,598)	-0.045 (0.422)	-1.353 (0.841)
Matched Controls without COVID-19 (N = 7)	-2,462 (1,359)	0.848 (0.597)	-1.581 (1.064)
Matched Controls with COVID-19 (N = 13)	-3,810 (2,536)	-0.083 (0.644)	-0.935 (0.658)

*Reported for the sites used in the Pre-Post TOWT.

Figure 55 shows the 8760 load shape for the nine sites included in the pre-post TOWT model. This prediction is based on the 30-year climate normals for the Burlington International Airport weather

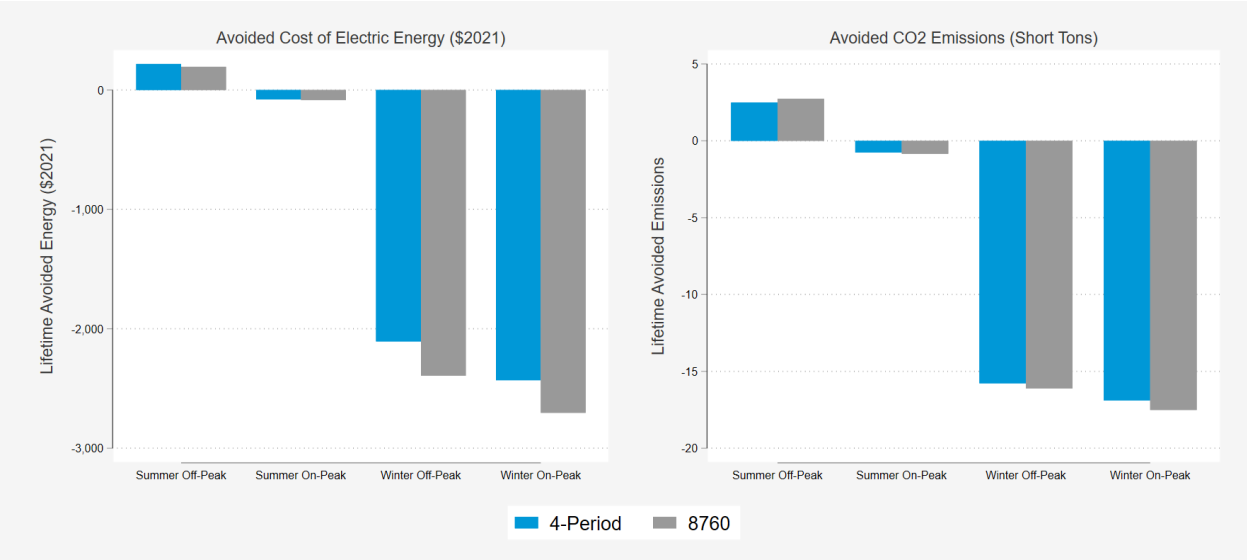
station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The largest estimated load increases for this model happen in the winter months during the morning and evening (blue regions).

Figure 55: Commercial BLPM Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 20-year measure life. These were calculated using both the 8760 and four-period valuation profiles. The 8760 results for avoided costs of electric energy is -$4,987.46, which is 11.8% greater than the four period avoided costs. Similarly, the avoided emissions sum to -31.74 short tons of CO_2 for the 8760 period, which is 2.5% higher than if the four-period assumptions were used. Figure 56 provides a full comparison.

Figure 56: Commercial BLPM Valuation Comparison



Commercial Refrigerator

Commercial reach-in refrigerators and freezers are used to store perishable items in food service and grocery establishments. The doors can be solid or glass and are available in a wide range of volumes and configurations. Program-supported units must meet or exceed the ENERGY STAR 4.0 specification.

TRM Characterization

The commercial reach-in refrigerators and freezers measure follows a **market opportunity** characterization in the EVT TRM. It is unique amongst market opportunity in that the baseline is not the current federal standard, but the prior standard that was effective as of January 2010. Per-unit savings vary by size and door type.

Modeling Results

Table 27 compares the meter-based estimates to the average values stored in the EVT database. The results for this measure are underwhelming across all modeling procedures due to the sample sizes. The point estimates fluctuate greatly between the three modeling procedures, and the standard errors suggest a margin of error 5-10 times the size of the claimed savings. Like the BLPM circulator motor, small sample size thwarts the commercial refrigerator analysis.

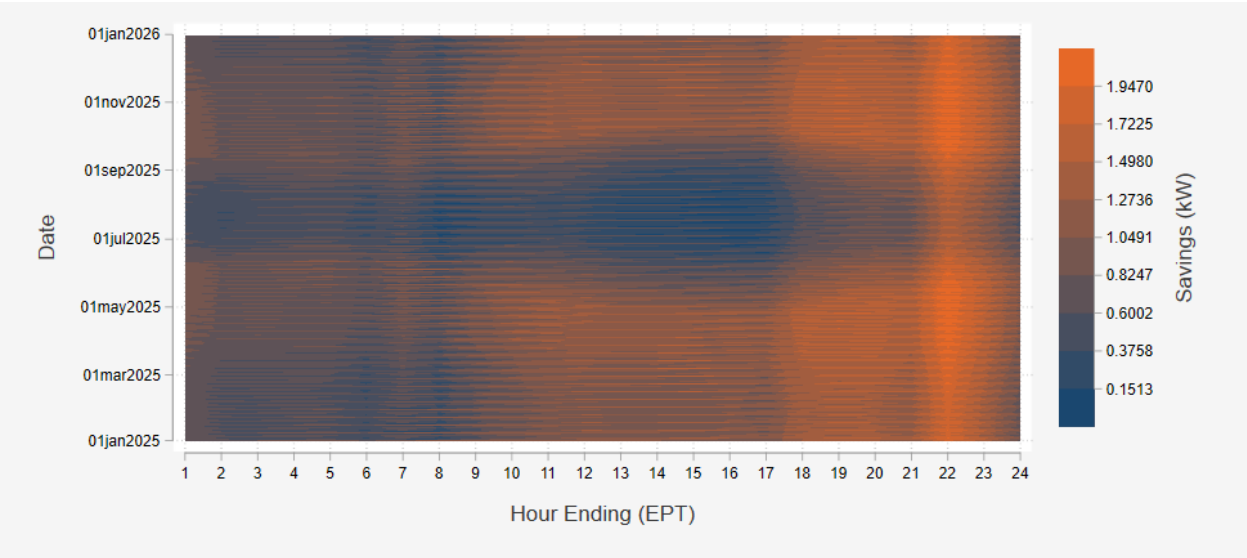
Table 27: Commercial Refrigerator Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	1,688	0.229	0.189
Pre-Post TOWT (N = 6)	8,369 (7,595)	0.354 (1.455)	1.389 (0.979)
Matched Controls without COVID-19 (N = 6)	15,886 (8,413)	1.232 (2.185)	2.063 (0.692)
Matched Controls with COVID-19 (N = 8)	4,949 (5,384)	-0.604 (0.980)	0.568 (0.725)

*Reported for the sites used in the Pre-Post TOWT.

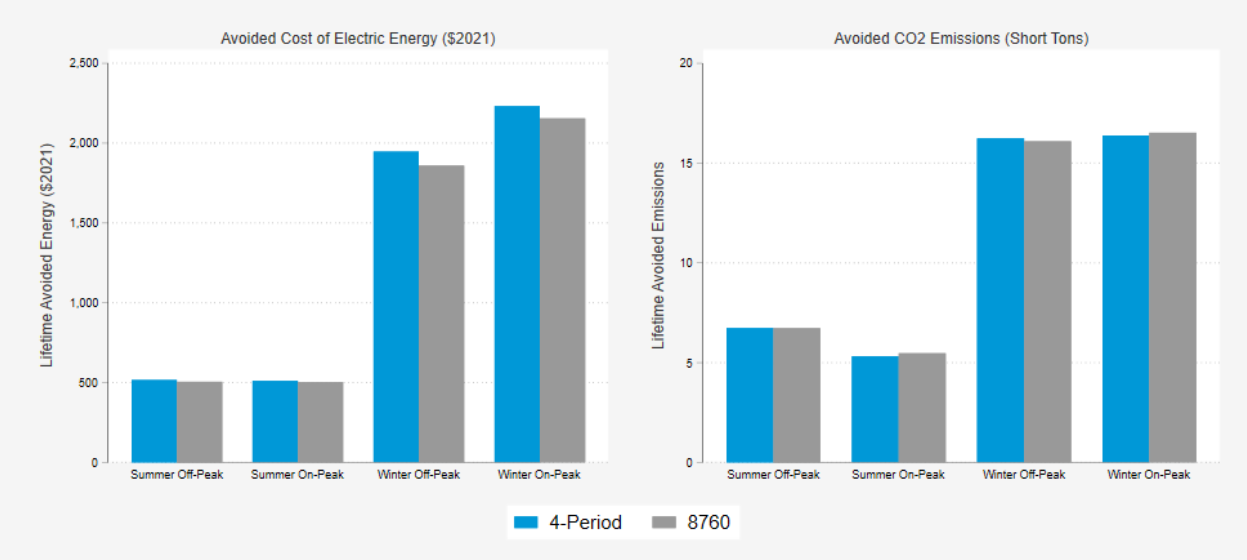
For completeness, we use the pre-post TOWT model to generate a heat map. This visual plots the impacts for each of the 8760 hours in a hypothetical year using 30-year climate normals for the Burlington International Airport weather station. Since the predictions are based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. Figure 57 presents the savings shape using the six sites that construct the savings estimates for the pre-post TOWT. Estimated load increases are most pronounced during summer afternoons (blue regions).

Figure 57: Commercial Refrigerator Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 12-year measure life. These were calculated using both the 8760 and four-period valuation profiles. The 8760 results for avoided costs of electric energy is $5,022.48, which is 3.7% lower than the four-period avoided costs. Similarly, the avoided emissions sum to 44.81 short tons of CO_2 for the 8760 period, which is 0.3% higher than if the four-period assumptions were used. A more detailed summation of these avoided costs/emissions are presented in Figure 58.

Figure 58: Commercial Refrigerator Valuation Comparison



Compressed Air

The compressed air measure is treated more like the custom projects and less like the other prescriptive commercial measures. This is due to the large savings claims for the installation of these measures. Because the compressed air systems can account for a significant portion of an industrial facility's electricity consumption, there is room to optimize this technology to yield large savings. Table 28 summarizes the savings size of the four sites that we could analyze using the TOWT methodology. The savings claims for the four sites in the table vary considerably, but this may be related to the variability in the implementation of this measure (can be comprised of various efficiency practices). The demand claims either represent a weather-resistant claim, where the summer and winter demand or identical, or a summer only claim, where the claimed demand savings in the winter is zero.

Table 28: Compressed Air Description

SIC Description	kWh/year	Summer kW	Winter kW
Nonresidential Building Construction	6,627	1.57	0.00
Groceries and Related Products	28,277	6.72	6.72
Sawmills and Planing Mills	13,705	1.57	1.57
Cut Stone and Stone Products	32,516	8.60	0.00
Total	**81,125**	**18.46**	**8.28**

Modeling Results

Much like the custom projects discussed in Section 4.4, the compressed air sites are simply analyzed using a pre-post TOWT model on a site-specific basis. The compressed air sites were large and had unique load patterns that would make it difficult to find an adequate matched control site or group of sites.

Table 29: Compressed Air Results

	Without COVID-19			With COVID-19		
SIC Description	kWh/year	Summer kW	Winter kW	kWh/year	Summer kW	Winter kW
Nonresidential Building Construction	-1,580	0.13	0.27	472	0.87	0.29
Groceries and Related Products	-4,868	-1.17	-1.21	-3,886	-1.71	-0.39
Sawmills and Planing Mills	277,679	18.56	46.46	433,613	40.04	58.15
Cut Stone and Stone Products	-6,933	-24.33	0.76	-37,190	-25.37	-1.58

In the 'Without COVID-19' results, three of the four sites return negative savings estimates. The fourth site returns large positive savings claim, but the estimate is not statistically significant. In the 'With COVID-19' results, most of the sites see a slight increase in their savings estimates but the estimates still fall short of the savings that are claimed in the database.

Generally, this measure does not appear to be a good candidate for using the TOWT methodology. This is due to two factors: there are few applicable sites that installed this measure, making it difficult to create an aggregate model for the measure as a whole, and the few applicable sites do not appear to be impacted by the independent variables used in the modeling procedure.

LED Agricultural Interior Fixtures

Indoor agriculture can occur in a greenhouse or inside a building and gives the farmer increased control over lighting levels, temperature, and humidity compared to outdoor agriculture. LEDs are a more efficient lighting technology than fluorescent or high intensity discharge technologies like metal halide or high pressure sodium. LEDs also emit less waste heat which can reduce the cooling and ventilation requirements of an indoor grow operation.

TRM Characterization

LED Agricultural Interior Fixtures are a subset of the broad LED Lighting Systems measure in the EVT TRM which generally follows a **retrofit** characterization. The baseline equipment is assumed based on the lumen output of the installed LEDs. Smaller lumen bins assume a blended incandescent and fluorescent baseline and the larger lumen bins rely on a metal halide baseline.

Modeling Results

Table 30 compares our meter-based estimates to the average claimed savings in the EVT database. It is difficult to draw conclusions for this measure because of the limited sample size and large standard errors. The point estimate for each modeling procedure indicates a decrease in electricity consumption in the post-installation period, but the magnitude is smaller than the claimed savings.

Table 30: LED Agricultural Interior Fixtures Modeling Results

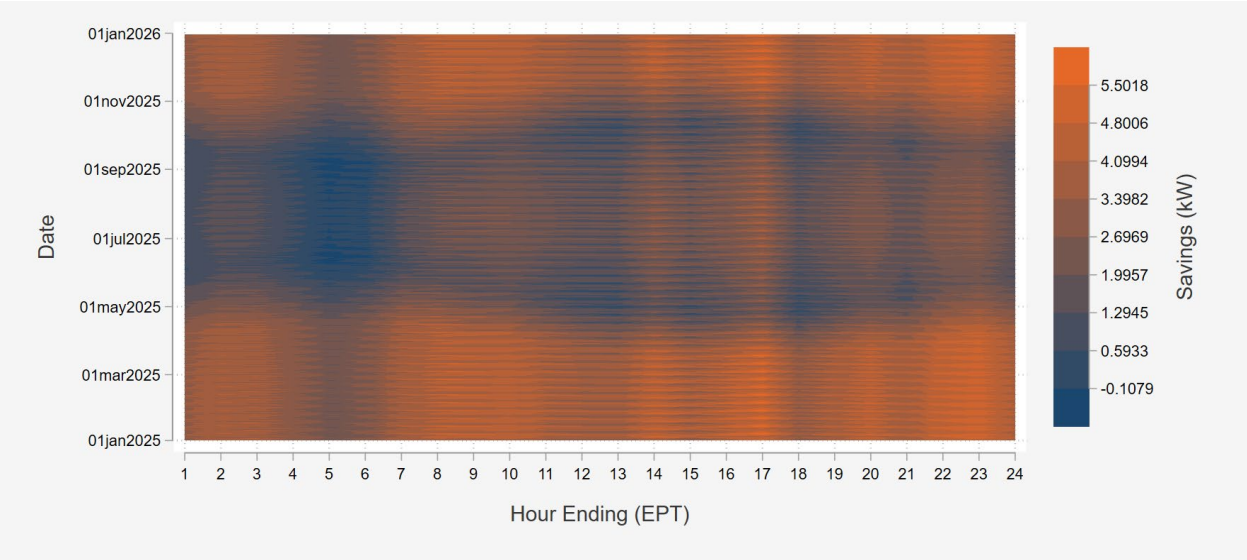
Source	kWh/year	Summer kW	Winter kW
EVT Database*	47,429	4.542	12.188
Pre-Post TOWT (N = 8)	23,581 (13,746)	2.527 (1.060)	3.731 (1.887)
Matched Controls without COVID-19 (N = 6)	15,733 (4,849)	1.893 (0.884)	3.556 (0.982)
Matched Controls with COVID-19 (N = 11)	18,256 (5,293)	2.423 (0.899)	3.081 (1.411)

*Reported for the sites used in the Pre-Post TOWT.

We use the pre-post TOWT model to generate a heat map and visualize the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. The 8760 profile in Figure 59 shows savings concentrated in the winter months. This is consistent with the current TRM

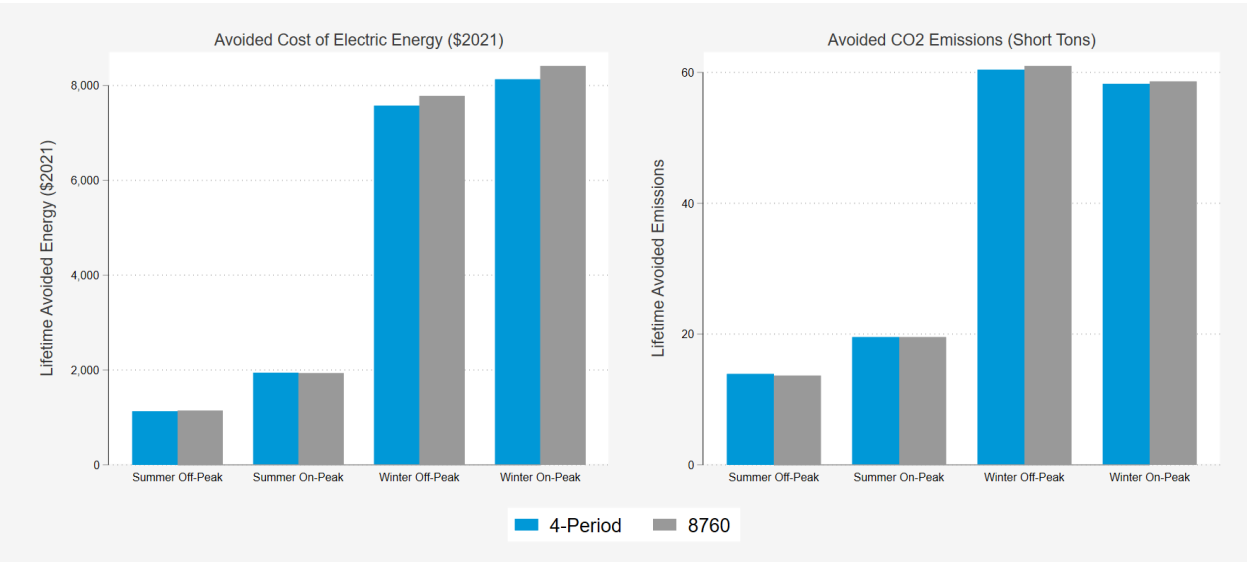
characterization which has winter peak demand savings three times higher than the summer peak demand savings.

Figure 59: LED Agricultural Interior Fixture Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 15-year measure life. We compute benefits using both the 8760 and four-period valuation profiles. The 8760 results for avoided costs of electric energy is $19,272, which is 2.6% greater than the four period avoided costs. Similarly, the avoided emissions sum to 152.8 short tons of CO_2 for the 8760 period, which is 0.4% higher than if the four-period assumptions were used. Figure 60 provides a complete comparison by costing period and valuation profile.

Figure 60: LED Agricultural Interior Fixture Valuation Comparison



LED Exterior Fixtures

Exterior fixtures include wall packs mounted on outside walls or canopies of buildings and freestanding parking lot and area lights. Often exterior lighting fixtures are controlled by a photocell or timer that turns the light on at dusk and off at dawn. Because these fixtures are located outdoors, there are no interactive effects with the heating and cooling system of the participating facility.

TRM Characterization

LED Exterior Fixtures are a subset of the broad LED Lighting Systems measure in the EVT TRM which generally follows a **retrofit** characterization. The baseline equipment is assumed based on the lumen output of the installed LEDs. Smaller lumen bins assume a CFL baseline and the larger lumen bins rely on a metal halide baseline.

Modeling Results

Looking at the pre-post TOWT and the matched controls without COVID-19, this measure analysis returns statistically significant savings estimates which align well with database claims. Table 31 compares estimates from the pre-post TOWT and the matched control group regressions to the claimed database savings. Because measure quantities are so variable for this measure, we estimate impacts on a per-unit basis. Since we estimate impacts in this way, the signal size is difficult to conceptualize. To situate these savings within the context of site-level consumption, we construct an average percent savings. This measure's annual claimed energy savings is 7% of site's annual consumption, on average, one of the larger commercial measures in terms of signal size. Additionally, exterior fixtures tend to operate from dusk to dawn, when there is less electric load in the building, making the change easier to detect. We view LED exterior fixtures as one the most successful measures, in terms of viability for estimating meter-based savings with Advanced M&V, despite a limited sample size.

Table 31: LED Exterior Fixture Modeling Results

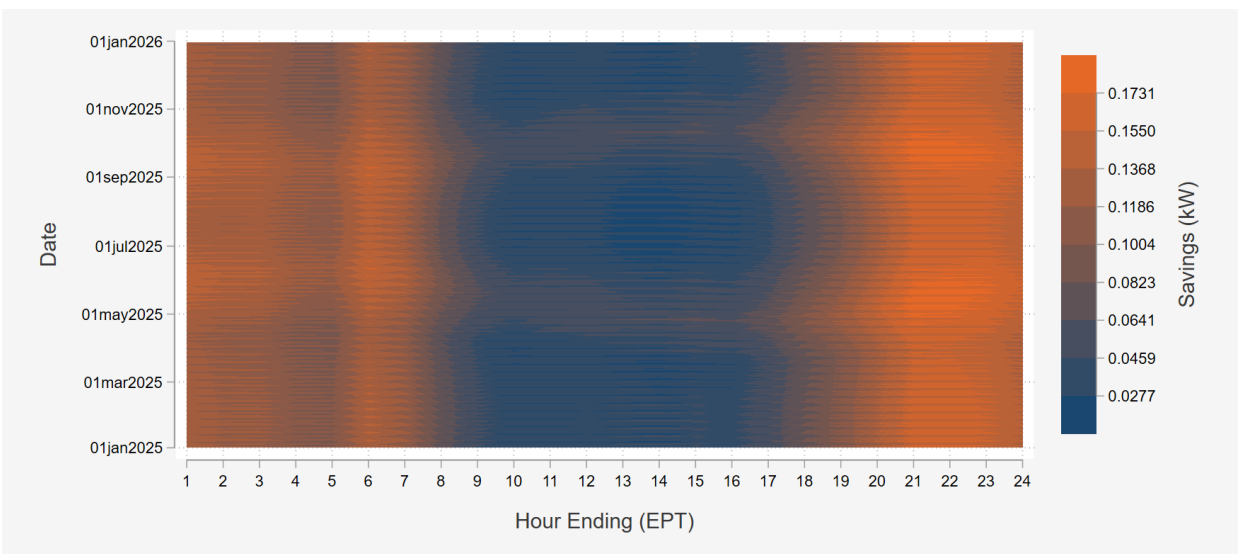
Source	kWh/year	Summer kW	Winter kW
EVT Database*	730	0.006	0.111
Average Percent Savings**	7%	0%	9%
Pre-Post TOWT (N = 68)	823 (278)	0.032 (0.051)	0.093 (0.033)
Matched Controls without COVID-19 (N = 60)	921 (353)	-0.005 (0.057)	0.111 (0.044)
Matched Controls with COVID-19 (N = 69)	477 (362)	-0.076 (0.057)	0.048 (0.050)

*Reported for the sites used in the Pre-Post TOWT.

** The annual database claim divided by the predicted annual consumption in the pre-installation period for all sites used in the Pre-Post TOWT.

A heat map plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. Figure 61 shows that LED exterior fixture savings tend to be concentrated at night. Intuitively, this makes sense as exterior fixtures are typically switched on as night approaches and turned off in the morning. The shape is distinctive and clear; in fact, it even shifts during the summer months when the daylight lasts longer. Stable savings estimates and load shapes across modeling methods suggest that commercial exterior lighting as a viable candidate for meter-based evaluation.

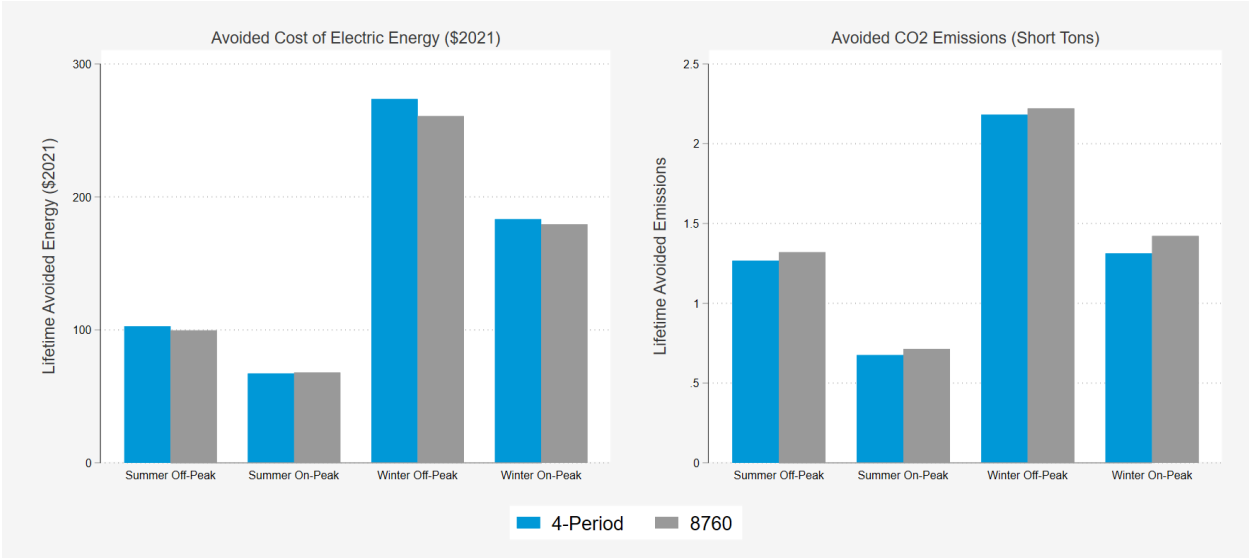
Figure 61: LED Exterior Fixtures Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 15-year measure life. We calculate benefits using both the 8760 and four-period valuation profiles and

compare the results in Figure 62. The 8760 results for avoided costs of electric energy is $607.75, which is 3.1% lower than the four-period avoided costs. Similarly, the avoided emissions sum to 5.68 short tons of CO_2 for the 8760 period, which is 4.2% higher than if the four-period assumptions were used.

Figure 62: LED Exterior Fixture Valuation Comparison



LED Interior Lighting

Interior lighting fixtures were the largest electric offering in the EVT portfolio during our period of analysis. Interior fixtures can be high-bay or low-bay and vary widely in terms of wattage and lumen output depending on application. This measure category also includes refrigerated case lighting. We split our analysis by track, with projects from the 6013 – Prescriptive Equipment Replacement assigned to the "Prescriptive" measure and projects from the 6013 – Upstream Commercial assigned to the "Upstream" measure. The distinction deals with program delivery model rather than the equipment itself. EVT collects more information about the participating facility in the prescriptive path than its upstream counterpart where transactions are between EVT and participating lighting distributors.

TRM Characterization

LED Interior Fixtures are a subset of the broad LED Lighting Systems measure in the EVT TRM which generally follows a **retrofit** characterization. The baseline equipment is assumed based on the lumen output of the installed LED fixture. Baseline wattages are generally mapped to T8 and T5 fluorescents fixtures with comparable lumen output. The TRM includes HVAC interactive effects to account for the reduced waste heat emitted by LEDs.

Modeling Results – Prescriptive LED Interior Fixtures

The savings estimates for this measure fluctuate across modeling procedures and none of the estimates are statistically significant. Table 32 compares the per-unit savings estimates from the pre-post TOWT and the matched control group regressions to the claimed database savings. Since we estimate impacts on a per-unit basis, the signal size is difficult to conceptualize. To situate these savings within the context of site-level consumption, we construct an average percent savings. This measure's annual claimed energy savings is 3% of site's annual consumption. This is a rather small signal, which in part explains why the regression models fail to produce statistically significant evidence of savings.

Table 32: Prescriptive LED Interior Lighting Modeling Results

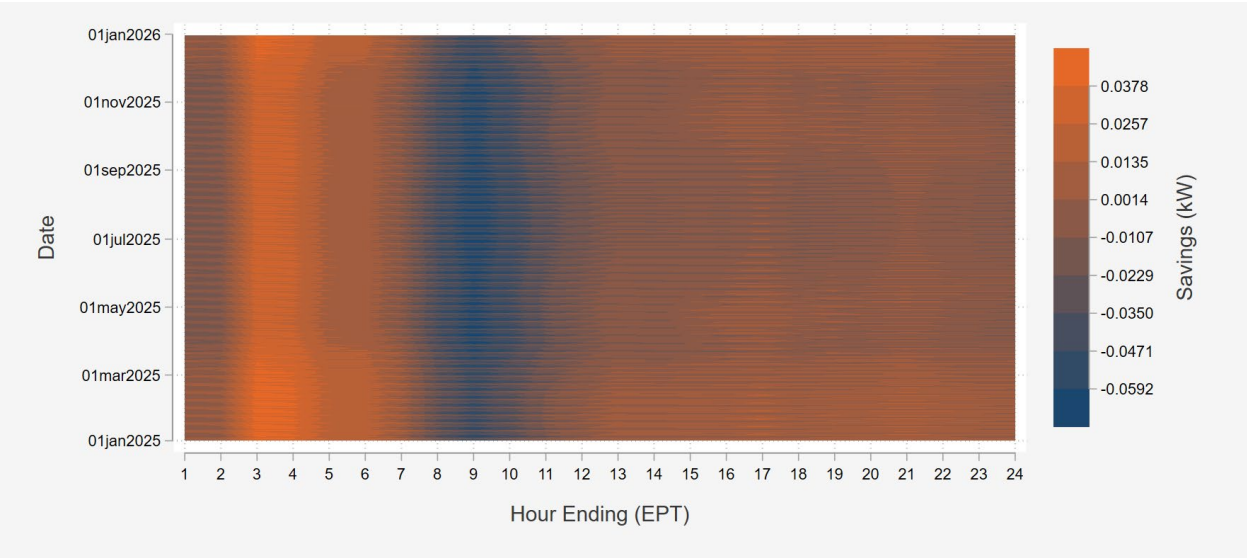
Source	kWh/year	Summer kW	Winter kW
EVT Database*	211	0.043	0.027
Average Percent Savings**	3%	4%	3%
Pre-Post TOWT (N = 151)	-45 (47)	-0.005 (0.008)	0.006 (0.007)
Matched Controls without COVID-19 (N = 133)	45 (63)	0.014 (0.010)	0.010 (0.008)
Matched Controls with COVID-19 (N = 164)	18 (55)	0.011 (0.011)	0.002 (0.009)

* Reported for the sites used in the Pre-Post TOWT.
** The annual database claim divided by the predicted annual consumption in the pre-installation period for all sites used in the Pre-Post TOWT.

A heat map plots the impacts for each of the 8760 hours in a hypothetical year. This prediction is based on the 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. Figure 63 displays the heat map for the prescriptive LED interior measure using the pre-post TOWT regression. Unfortunately, the load shape of the savings from this measure is inconsistent across the different modeling techniques. We would recommend Vermont rely on the 8760 load shape from an end-use light logging study[15] rather than an estimated shape from whole-building data.

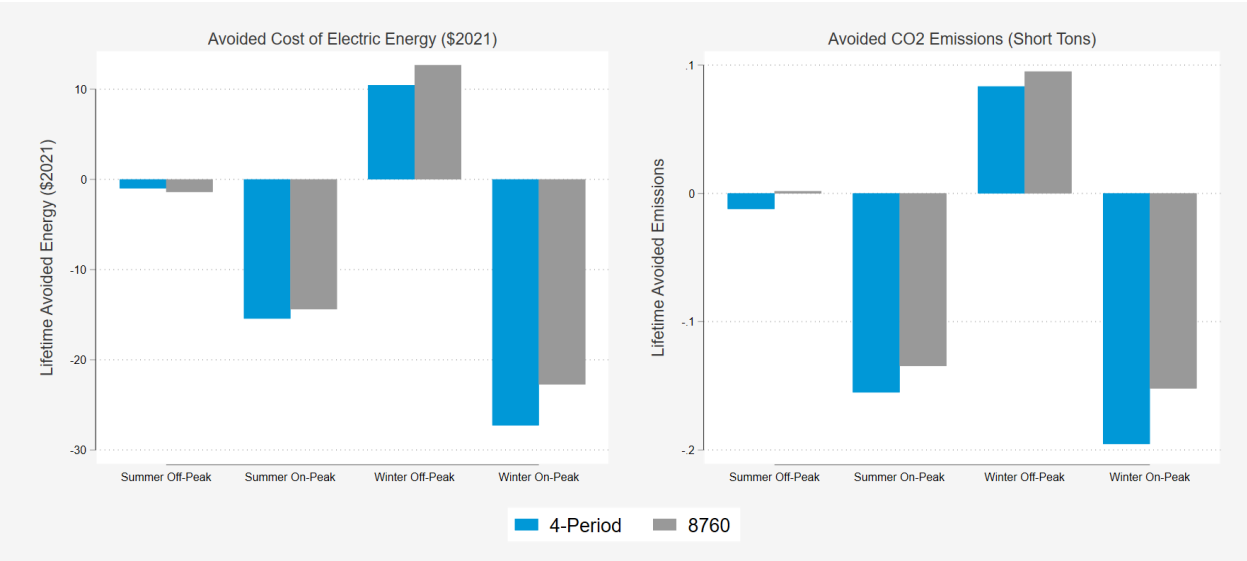
[15] Link to 8760 load shapes from a 2015 Pennsylvania study. https://www.puc.pa.gov/Electric/xls/Act129/SWE-Light_Metering_Study-AppendixB-Commercial_Load_Shapes.xlsx

Figure 63: Prescriptive LED Interior Lighting Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 15-year measure life. We calculate benefits using both the 8760 and four-period valuation profiles and compare the results in Figure 64. The 8760 results for avoided costs of electric energy is -$25.87, which is 28.5% lower than the four period avoided costs. Similarly, the avoided emissions sum to 0.19 short tons of CO_2 for the 8760 period, which is 47.0% lower than if the four-period assumptions were used.

Figure 64: Prescriptive LED Interior Lighting Valuation Comparison



Modeling Results – Upstream LED Interior Fixtures

The savings estimates for the upstream version of this measure also fluctuate between modeling procedures. Table 33 compares the per-unit outputs of the pre-post TOWT and the matched control group regressions to the claimed database savings. This measure's annual claimed energy savings is

13% of site's annual consumption, which sits above the 10% threshold that is recommended for whole-building analysis. The point estimates for the upstream LED interior lighting measure are more consistent than its prescriptive counterpart despite a smaller sample size, but the regression-based savings estimates are still not statistically significant.

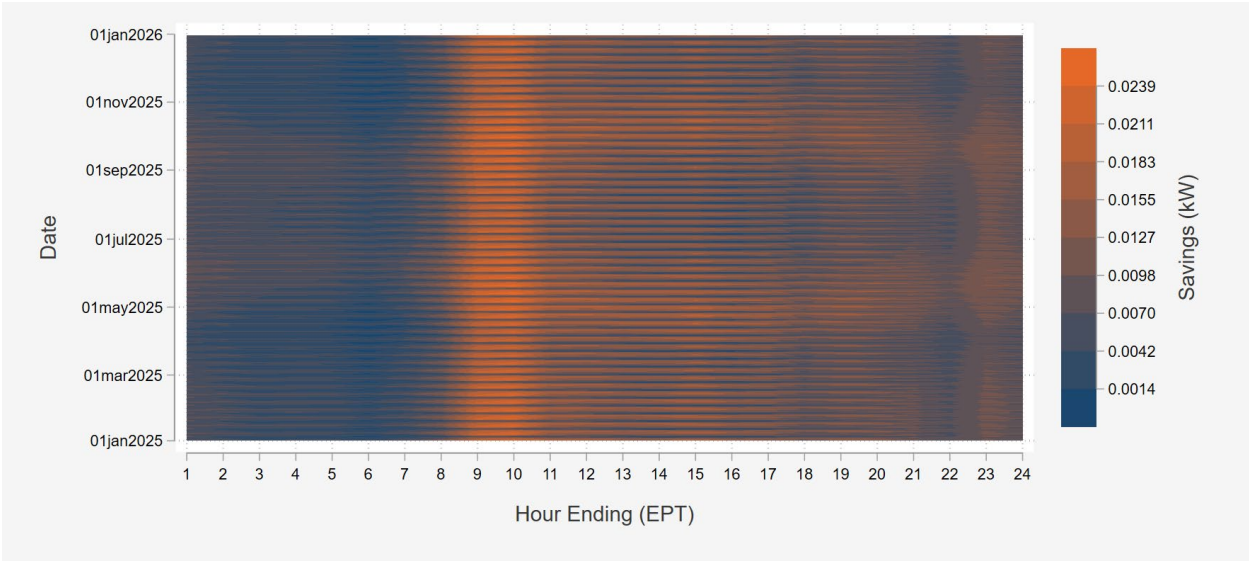
Table 33: Upstream LED Interior Lighting Modeling Results

Source	kWh/year	Summer kW	Winter kW
EVT Database*	184	0.037	0.019
Average Percent Savings**	13%	15%	10%
Pre-Post TOWT (N = 58)	80 (72)	0.012 (0.015)	0.010 (0.011)
Matched Controls without COVID-19 (N = 49)	83 (103)	0.014 (0.022)	0.016 (0.019)
Matched Controls with COVID-19 (N = 159)	-16 (307)	0.005 (0.111)	0.002 (0.062)

*Reported for the sites used in the Pre-Post TOWT.
** The annual database claim divided by the predicted annual consumption in the pre-installation period for all sites used in the Pre-Post TOWT.

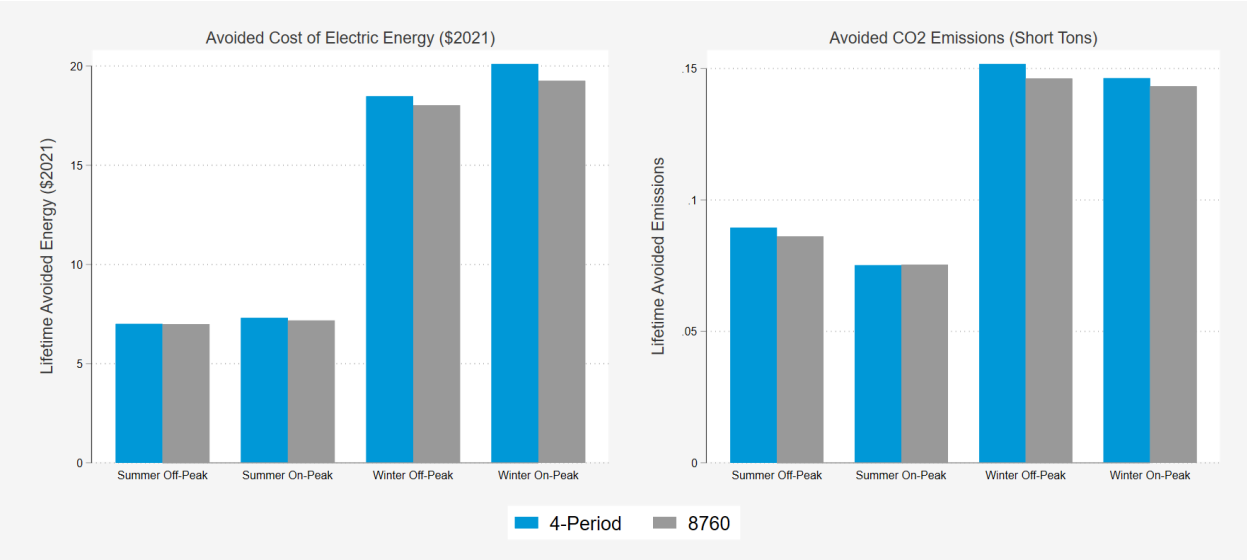
A heat map plots the impacts for each of the 8760 hours in a hypothetical year. These predictions use 30-year climate normals for the Burlington International Airport weather station. Due to these predictions being based on normalized weather, the y-axis shows future dates to make it clear that actual weather is not used. Figure 65 shows the heat map for the upstream LED interior lighting measure using the pre-post TOWT regression models. This profile matches the expected shape for interior lighting better than prescriptive LED interior fixture measure.

Figure 65: Upstream LED Interior Lighting Heat Map



Using the normalized impact estimates, we compute avoided energy costs and avoided emissions over a 15-year measure life. We calculate benefits using both the 8760 and four-period valuation profiles and compare the results in Figure 66. The 8760 results for lifetime avoided cost of electric energy is $51.46, which is 2.8% lower than the four-period avoided costs. Similarly, the avoided emissions sum to 0.45 short tons of CO2 for the 8760 period, which is 2.6% lower than if the four-period assumptions were used.

Figure 66: Upstream LED Interior Lighting Valuation Comparison



Overview of Prescriptive Commercial Results

Table 34: Overview of Prescriptive Commercial Results

Measure	Vintage	kWh/year		Key Takeaway
		Database	Pre-Post TOWT	
BLPM Circulator Pump – Commercial (N = 9)	Retrofit	2,214	-3,803* (2,598)	Potential to use this approach with a larger sample size.
Commercial Refrigerator (N = 6)	Retrofit	1,688	8,369 (7,595)	Potential to use this approach with a larger sample size.
LED Agricultural Interior Fixtures (N = 8)	Retrofit	47,429	23,581** (13,746)	Potential to use this approach with a larger sample size.
LED Exterior Fixtures (N = 68)	Retrofit	730	823*** (278)	This approach produces consistent findings across modeling methods.
LED Interior Lighting – Prescriptive (N = 151)	Retrofit	211	-45 (47)	The average effect size is likely too small for this approach.
LED Interior Lighting – Upstream (N = 58)	Retrofit	184	80 (72)	Potential to use this approach with a larger sample size.

* Significant at an 80% confidence interval.
** Significant at a 90% confidence interval.
*** Significant at a 95% confidence interval.

4.4 CUSTOM

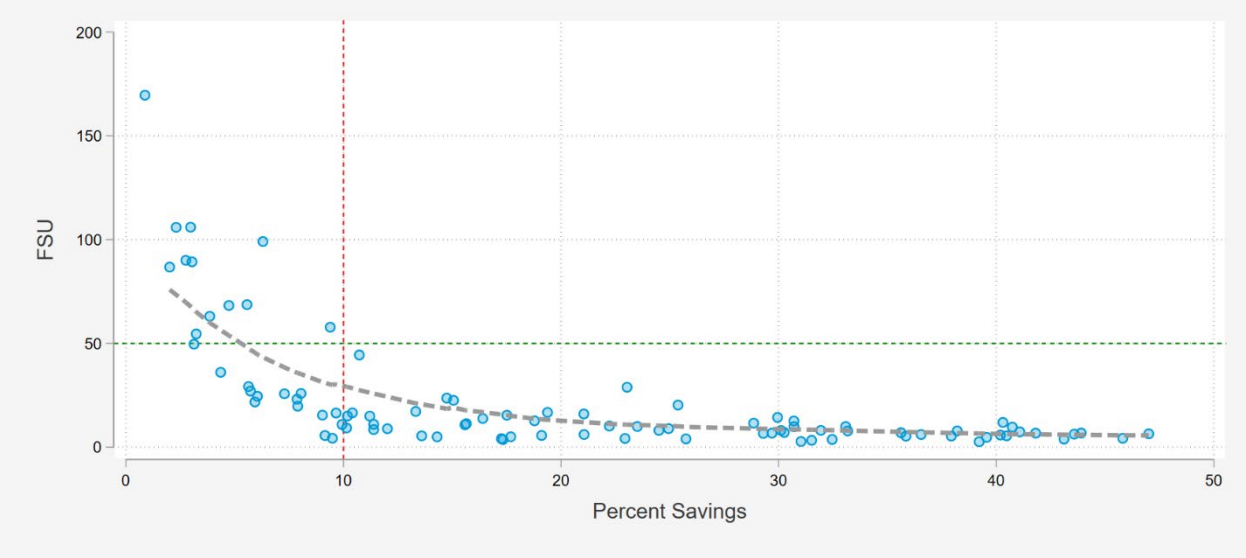
Prescriptive projects' energy and peak demand savings estimates are based upon generalized operating and efficiency assumptions. Because prescriptive measures tend to be adopted by numerous customers and have a limited per-unit savings, there is an "evening out" effect. We expect that any given measure characterization in the EVT may overstate savings in some homes and business and understate savings in others. However, the intent is to provide an unbiased projection of average performance. In contrast, custom projects' savings estimates are based on unique and varying sets of measures that are installed as one comprehensive project and often generate large project savings. To develop these claimed savings estimates, site-specific equipment and operational considerations are taken into account. Often custom projects are made up of multiple measures that interact with one another, so it is integral to accurately account for the interactions of the measures.

Due to the heterogeneity of these projects, to garner impacts at the meter, we run TOWT site-level regression for each custom project. Importantly, we did not develop a matched control group for the custom analysis. The custom sites were large and many had unique load patterns that would make it difficult to find an adequate matched control site or group of sites.

Goodness of Fit

Goodness of fit is the final filter that is applied to the custom projects. This is done using the fractional savings uncertainty (FSU), which expresses uncertainty as a percentage of the estimated savings. FSU is useful metric because it takes into account both signal size (expected change in consumption) and noise in the model. ASHRAE Guideline 14 suggests that the maximum allowable standard for this type of analysis be 50% FSU at the 68% confidence level. This threshold is indicated by the green line in Figure 67. As the savings increases in relation to consumption, the FSU tends to get better, or closer to zero. Due to these limitations, this metric is not ideal for measures where the savings is a small portion of the whole-building consumption, typically those with under 10% savings, indicated by the red line in Figure 67. Therefore, this metric is only used for our custom projects, which tend to have higher percent savings than do prescriptive measures.

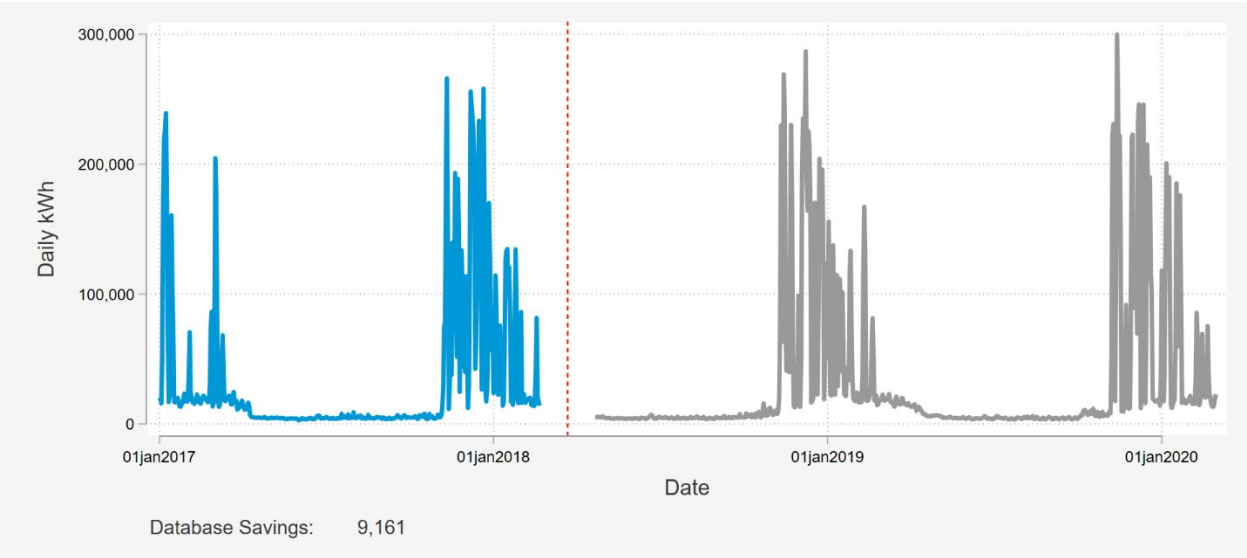
Figure 67: FSU Curve



For this analysis, we implement a function of fractional savings uncertainty that is based solely on pre-period estimation, called quasi-FSU. It is calculated as the function of hourly savings uncertainty and claimed database savings. This approach helps to determine applicable sites that should perform well with the methodology, prior to implementing any changes or enrolling customers into a pay-for-performance program. More generally, it gives insight into how well this model would perform given the types of load volatility these commercial buildings experience.

For example, Figure 68 displays a site that successfully passed through every filter, but is subsequently filtered out using the uncertainty metric. In this case, it is evident that weather does not explain the variation in this site's energy consumption, making this methodology inadequate for estimating annual savings. Even if it were to pass through this final layer of filtering, the claimed savings are extraordinarily small as compared to the daily consumption. This signal-to-noise ratio would make it near impossible to perceive a change.

Figure 68: Site without Adequate Uncertainty



Aggregate Impact

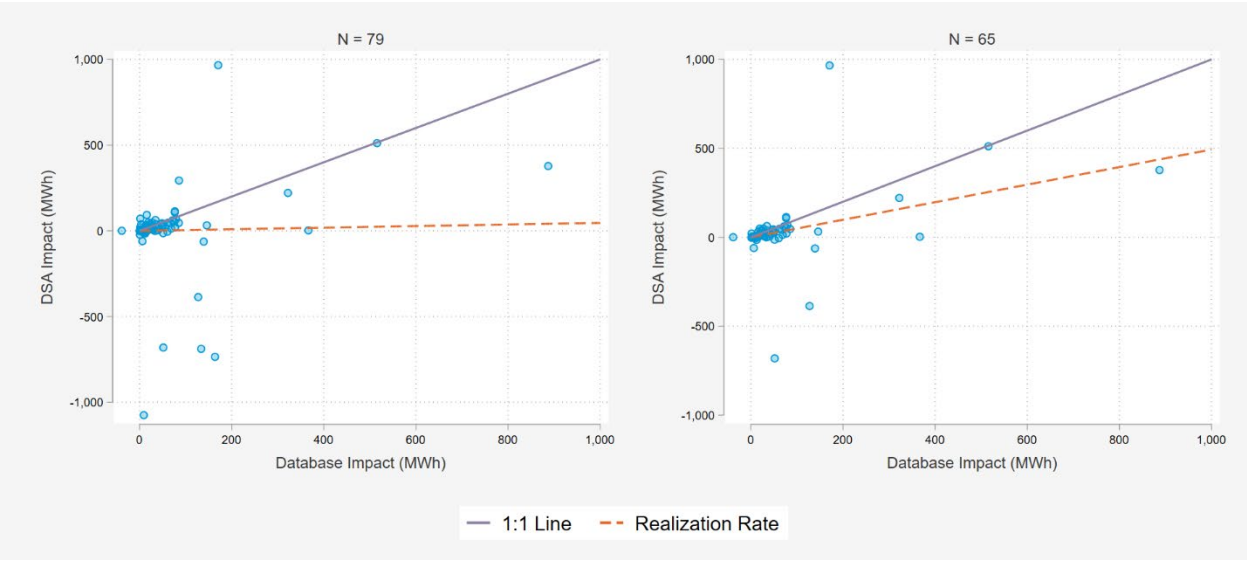
Of the 124 identified custom projects, only 79 contain twelve months of post-period data prior to the start of COVID-19 shutdowns. Using the 79 sites that had adequate post data, 65 of them pass through the quasi-FSU filter. We get a 5% realization rate for energy when we use all 79 sites versus a 49% realization rate when we use just the 65 sites that pass the goodness of fit test. This highlights the importance of applying this goodness of fit filter because without it, we would include sites with high uncertainty that bias the custom results on aggregate. Table 35 shows the aggregate kWh yearly impact, summer demand reduction, and winter demand reduction for the 65 sites.

Table 35: Aggregate Custom Impacts

Source	kWh/year	Summer kW	Winter kW
EVT Database	4,503,880	682	536
DSA	2,222,375	264	289
Realization Rate	49%	39%	54%
Margin of Error (90% Confidence)	815,719	256	213
Relative Precision (90% Confidence)	±37%	±97%	±74%

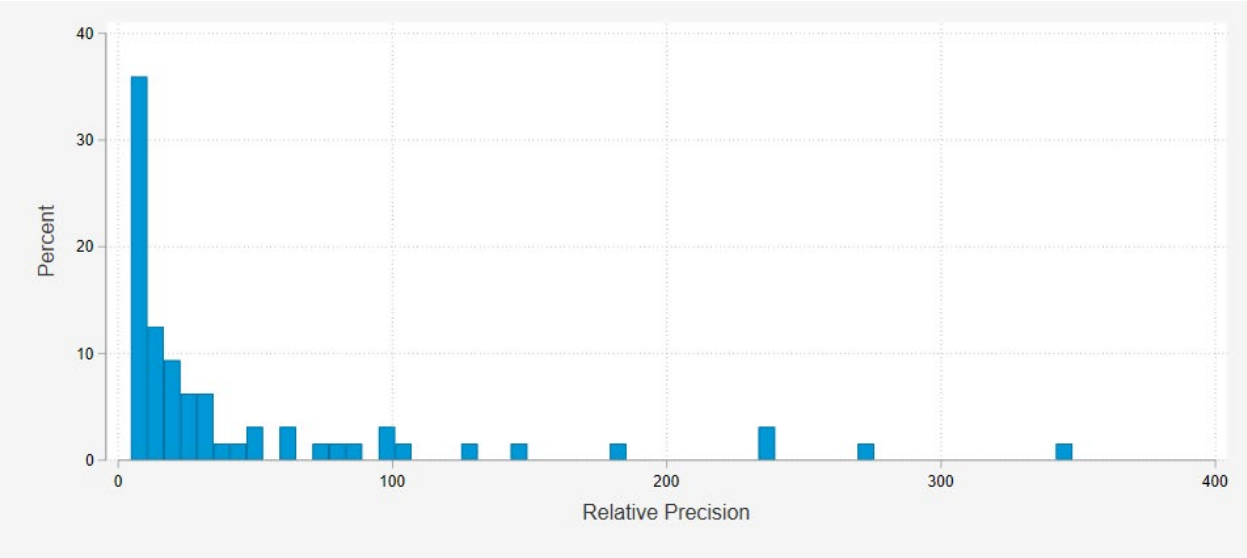
The left side of Figure 69 highlights that running this method on each site can be very noisy, even though a trend is prevalent within the data. As we zero in on well-behaved buildings, or building that pass through our quasi-FSU filter, on the right side of Figure 69, we see the database and the calculated savings are better aligned. This brings to the surface the issue of how buildings that are not well-behaved should be estimated. It also helps to emphasize the importance of not over-automating, as there are outliers that cannot be lumped in together.

Figure 69: Scatter of Database and DSA Impacts



The relative precision reported in Table 35 is based on the normalized FSU. This is not the same as sampling uncertainty, but it is instead the aggregate estimation uncertainty from the sites analyzed. We believe that this differentiation in uncertainty could be ignored in an ISO-NE certification. Figure 70 displays the distribution of relative precision across the 65 sites analyzed. The majority of these sites have a relative precision below 20% (at a 90% confidence level), but the distribution is greatly skewed by the outliers that occur over 100%. These outliers are not balanced out in aggregate, due to relative precision being strictly positive. Even with this variability, each of our aggregate estimates in Table 35 are statistically significant.

Figure 70: Distribution of Relative Precison (N = 65)



Aggregate Impact with COVID

It is important to understand how the COVID-19 pandemic may have influenced savings estimates. Results from our simulation exercise (Section 4.1) indicated that commercial loads were down during COVID-19 shutdowns. This conclusion would lead to the belief that savings estimates would be inflated when COVID-19 shutdowns overlap with the post-period. In our analysis, we estimated a realization rate of 69% for energy when we use all 124 sites versus a realization rate of 76%, seen in Table 36, when we use just the 102 sites that pass the goodness of fit test.

While it is evident that the realization rate has increased with the inclusion of COVID-19 in the post-period, we are unable to make a direct comparison because there has also been the inclusion of more sites. The right side of Table 36 includes only the 65 sites that were analyzed in the without COVID-19 aggregate impact, but this time includes COVID-19 in the post-installation regression. When we strictly use these 65 sites that are included in both analyses, the with COVID-19 analysis produces an energy realization rate of 72%, as opposed to the 49% energy realization rate in Table 35.

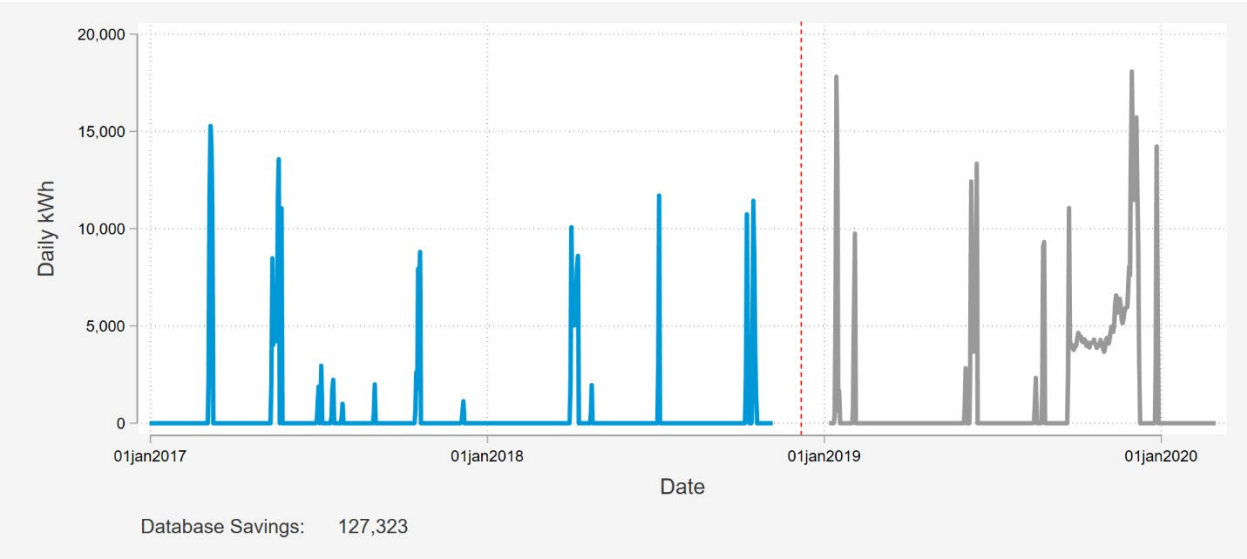
Table 36: Aggregate Custom Impacts with COVID-19

	All (N = 102)			Overlap (N= 65)		
Source	**kWh/year**	**Summer kW**	**Winter kW**	**kWh/year**	**Summer kW**	**Winter kW**
EVT Database	8,044,303	982.30	839.36	4,503,880	682.18	535.59
DSA	6,104,733	779.19	895.20	3,237,281	310.95	422.56
Realization Rate	76%	79%	107%	72%	46%	79%

Understanding Outliers

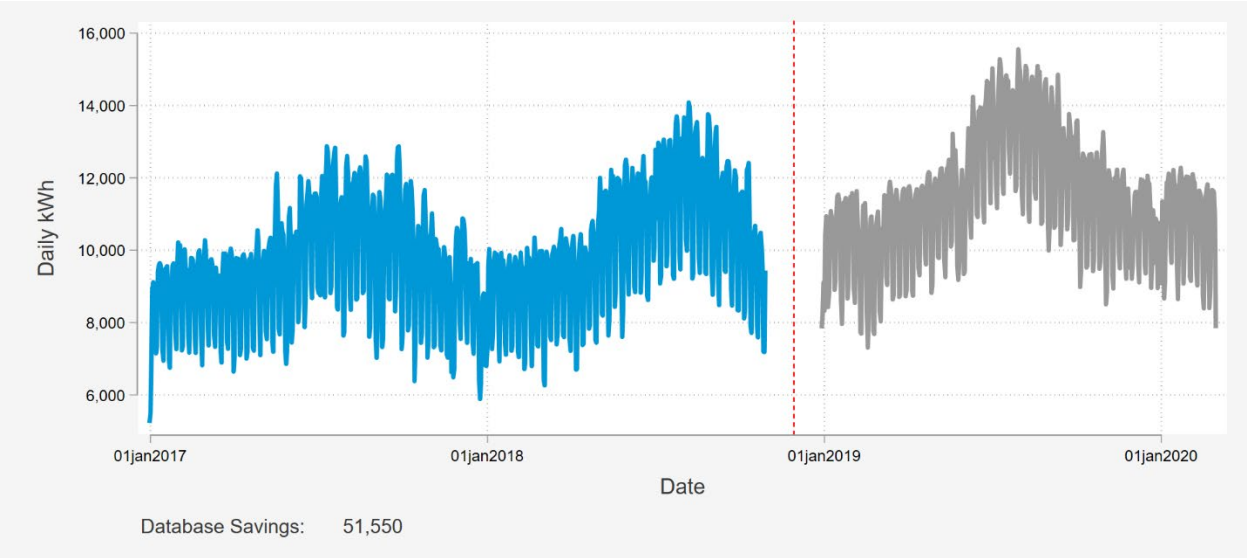
Even with filters in place, sites that are poor candidates for regression-based savings estimates can make it through to the analysis sample. For example, Figure 71 and Figure 72 depict sites that are aggregated in both the 65 sites in Table 35, as well as the 102 sites in Table 36. While volatile, the loads in the baseline period of Figure 71 were routine enough to pass through the quasi-FSU uncertainty metric. The volatility of the loads change in the post-period with a long duration of increased usage, hence the large increase in the negative savings.

Figure 71: Site with Volatile Loads



Additionally, due to evolving occupancy and usage patterns of commercial buildings, it is also possible that load did indeed increase, as in Figure 72, even with the installation of the energy efficient measure, due to circumstances that changed the production or usage time of the facility. In fact, the ideal time to add energy efficiency measures would be as a site has started to grow in size or production.

Figure 72: Site with Increased Loads



These outliers highlight the fact that the aggregation of custom measures does not tell the full story. There are many caveats when trying to methodologically explain consumption, especially of large

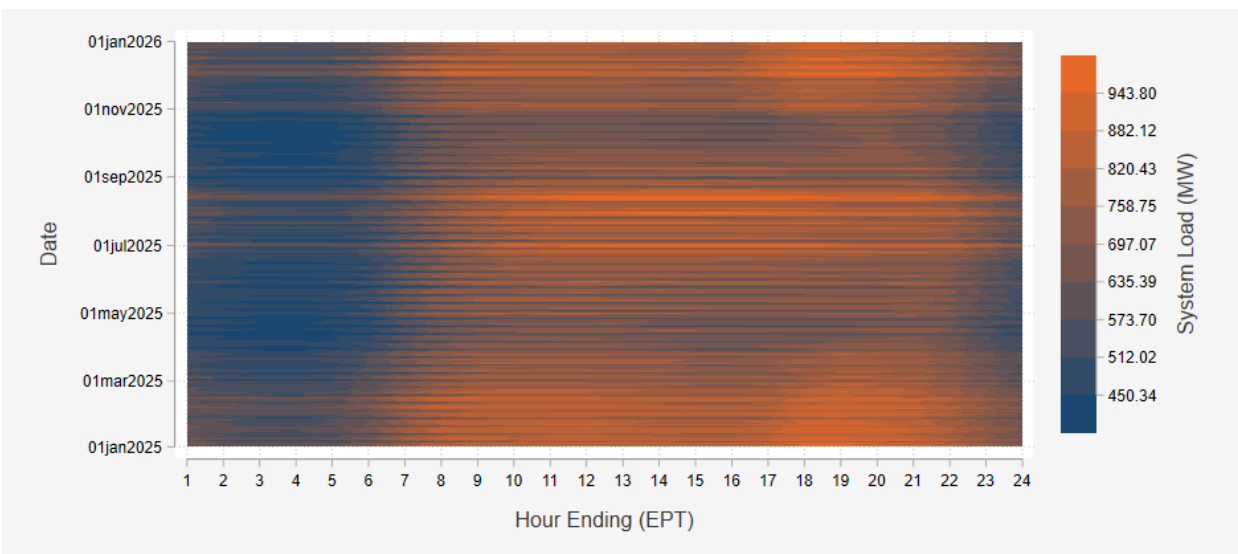
commercial buildings. Automating this process could lead to discrepancies that can be attributed to site-specific considerations.

4.5 TIME-VALUE OF EFFICIENCY

For each prescriptive measure, we produce two avoided cost of energy estimates and two avoided emissions estimates based on the measure life of the technology. One of these is based on the granular 8760 valuation profiles from AESC 2021, while the other relies on the status quo four-period valuation framework. The four-period valuation profiles come from the same underlying AESC 2021 data, just collapsed into a weighted average within the summer on-peak, summer off-peak, winter on-peak, and winter off-peak costing periods. For some measures, the resulting lifetime benefits are hardly any different, while others are noticeably different. By comparing the results, we are able to get insight into how different measure economics can be under the two approaches. The underlying premise of this analysis is that the 8760 valuation profiles are more accurate. If the two results are similar, then the four-period approach is a harmless simplifying assumption. If the results are very different, we conclude that the four-period profiles tend to over or under-value the measure economics.

System load is the weighting variable use to calculate four-period valuation profiles from the underlying 8760 data in the AESC study. The heat map in Figure 73 shows the forecasted system load for Vermont from the AESC 2021 study. System load is highest during the day and peaks on summer and winter evenings.

Figure 73: System Load Heat Map (2025 Forecast)



The status quo valuation approach in most of New England is to differentiate the avoided cost of electric energy into four costing periods. Figure 74 displays the distribution and the mean for each of these periods. From this figure, it becomes clear that the winter season exhibits more variability in prices. The cost of electric energy in the summer ranges from $26 to $48 per MWh, while the winter ranges from $16 to $82 per MWh. The red reference lines denote the simple average for each of the

four periods and indicates that winter costs are higher. While seasonality plays a large role in the cost of electric energy, the on peak and off peak periods within seasons tend to display less variability, but the on peak periods have a higher average cost in each season. As discussed in Section 3.1, this "Retail" cost of energy is not the actual retail cost to ratepayers, but a wholesale price adjusted for risk, line losses, and the cost of renewable energy credits to load.

Figure 74: Retail Price across the Four Costing Periods

Visualizing the distribution of avoided costs within a costing period helps us to understand the risk associated with a four-period valuation framework. Risk stems from the fact that the savings load shape of a measure may not follow the load shape of the system. A good example of this is the LED exterior lighting measure. While the system peaks during the day, this measure realizes savings during the night. Figure 75 shows the share of annual energy consumption/savings in each hour of the day. In cases like this, where system load does not match the savings profile load of the measure, the system-weighted average may not reflect the weighted average savings produced by the specific measure.

Figure 75: Annual Share of System Load and Measure Savings for LED Exterior Lighting

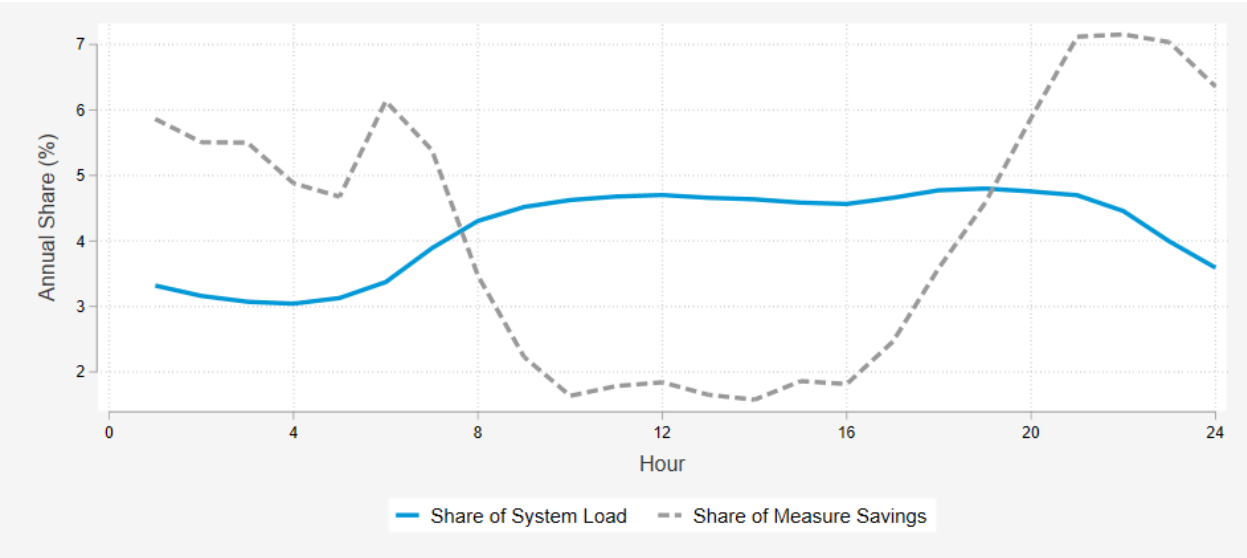


Table 37 gives additional detail into how the shares are distributed during the winter on-peak period, since this period has the largest variability and highest average cost. While the share of the system load has a slight decrease between hour 20 and 21, the measure savings have a large increase.

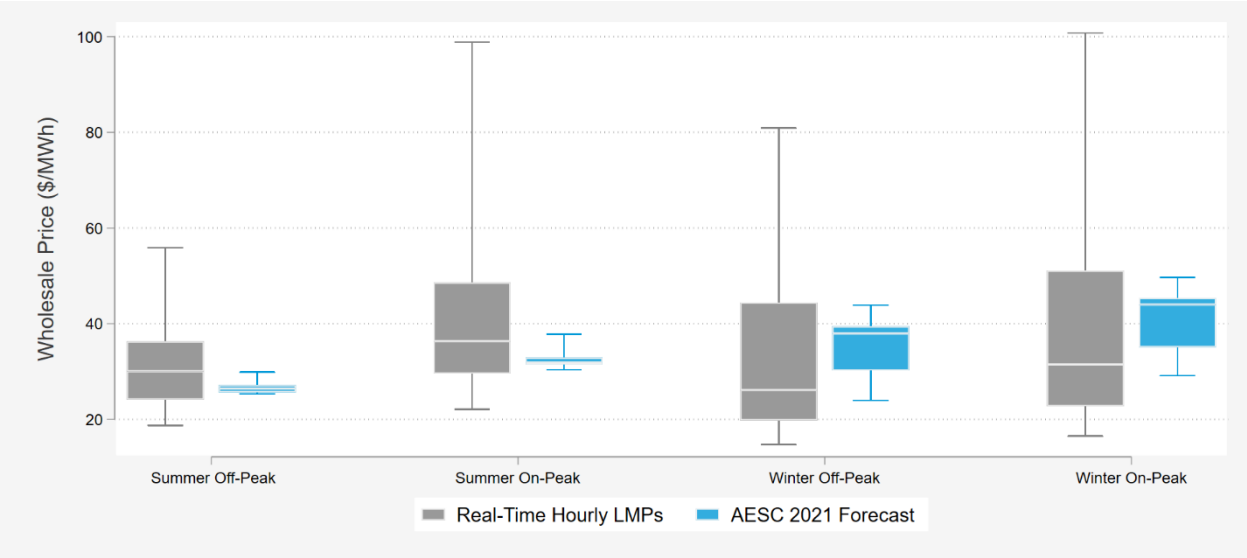
Table 37: Winter On-Peak Share of System Load and Measure Savings for LED Exterior Lighting

Hour Ending	Average Retail Price	Share of Measure Savings	Share of System Load
8	$50.57	6%	6%
9	$50.20	4%	6%
10	$49.15	3%	6%
11	$48.64	3%	6%
12	$48.70	3%	6%
13	$48.65	3%	6%
14	$48.86	3%	6%
15	$49.03	4%	6%
16	$50.12	3%	6%
17	$52.04	5%	6%
18	$52.74	7%	7%
19	$52.79	8%	7%
20	$52.52	10%	7%
21	$52.27	13%	6%
22	$51.88	12%	6%
23	$51.32	12%	5%

Avoided emissions follow the same logic as avoided costs, but as we can see in Figure 6, avoided emissions are more variable over the forecast horizon. Even with this variability, there is potential for the four-period approach to either overstate or understate the granular estimates.

It is important to understand that while the granular estimates may provide more specificity, they remain forecasts. Figure 76 compares the real-time, hourly wholesale prices from January 1, 2021 to August 24, 2021 across the four costing periods with the AESC 2021 forecast. Across each period, it is evident that the real-time wholesale prices tend to be more variable than the forecasted prices (taller box and longer whiskers). This increased volatility would likely cause greater departures between the four-period and 8760 avoided costs.

Figure 76: Actual vs. Predicted Wholesale Prcies (January 1, 2021 to August 24, 2021)



Factors such as season, time of day, demand for electricity, and day of week all contribute to the variation in avoided costs and emissions. With granular measure load shapes, it is possible to assign economic preference to measures that conserve energy when the value stack is greatest. To better reflect the time value of efficiency, Vermont could move to more granular load shapes or develop separate four-period avoided costs by measure using more granular load shapes rather than system load.

5 COMPARISON WITH TRADITIONAL M&V METHODS

The results of this study demonstrate the power of Advanced M&V methods as a tool for estimating granular energy impacts at scale. For less than $100,000 DSA developed 8760 savings estimates for twenty prescriptive measures and over 100 custom projects. DSA staff never visited the state of Vermont or even spoke with a participating Vermont customer by phone about their efficiency upgrade. The results presented in Section 4 also suggest that caution is needed regarding when Advanced M&V works well and when it does not. In summary, Advanced M&V is a powerful tool – but it is not the right tool for every job.

Additionally, when considering the time required to produce results, it is important to understand that the results of this Advanced M&V study focus on **normalized savings**, rather than **avoided energy use**. For normalized savings, a second regression model is developed for the post-retrofit period and generally requires a full year of post-retrofit meter data. Traditional M&V is a broad category with a range of methods, but if normalized savings are the desired output, Advanced M&V methods are likely no faster or take longer to produce than Traditional M&V.

The meter-based estimation techniques applied to garner the normalized results produced in this study are a departure from the techniques traditionally used in Vermont for prescriptive measures, which rely on complex measure characterizations in the TRM that blend equipment characteristics and assumptions about operating characteristics into per-unit values. Evaluation of prescriptive measures and programs has historically relied on the verification of proper application of TRM inputs and calculations, supplemented by studies of key parameters where appropriate. Custom project evaluations rely on a mix of engineering calculations, end-use metering, and whole-building analysis. We understand that use of whole-building AMI analysis has increased in recent years so the methods in this study are less of a departure for custom programs.

Most summer and winter peak demand savings from EVT's programs are nominated into the ISO-NE Forward Capacity Market (FCM) auctions. Due to FCM proceeds being an important funding stream, ISO-NE's M&V requirements are an important driver of the evaluation procedures in Vermont. One of the key requirements in ISO-NE's Manual for Measurement and Verification of Demand Reduction Value from Demand Resources[16] (M-MVDR) is that the aggregate demand reduction estimates derived from sampling have at least ±10% relative precision at the 80% confidence level. Chapter 5 of the M-MVDR lists Option C: Whole Facility Regression as an acceptable measurement and verification methodology but cautions that "*Option C should not be used if the performance value is expected to be small relative to the total facility load, due to the small signal-to-noise ratio.*"

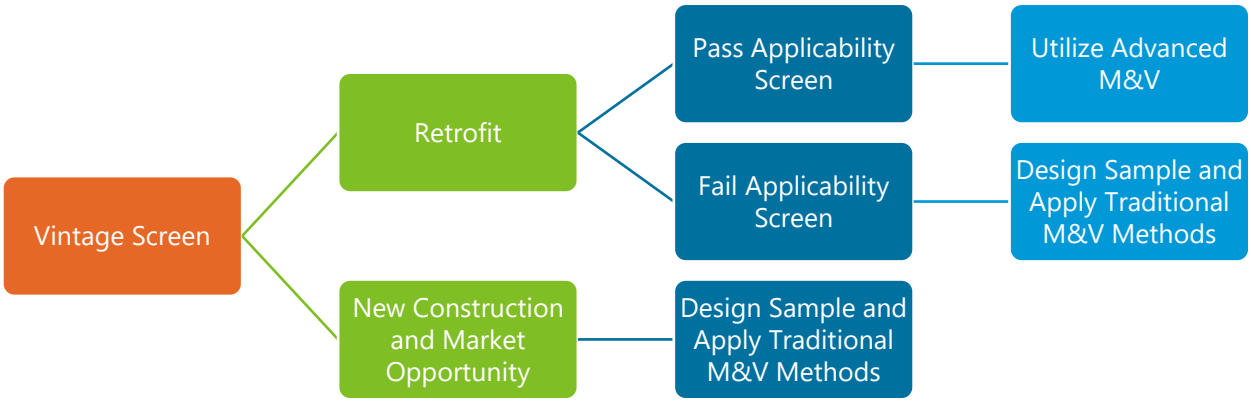
5.1 CUSTOM

The M-MVDR does not address the distinction between sampling error and estimation error. Our interpretation of the manual is that estimation error, or the uncertainty associated with a site-specific

[16] https://www.iso-ne.com/static-assets/documents/2017/02/mmvdr_measurement-and-verification-demand-reduction_rev6_20140601.pdf

regression analysis, can be ignored in the calculation of aggregate sampling error, but that evaluators should only apply Option C methods to projects that pass a "signal-to-noise" threshold such as FSU. The challenge with applying Advanced M&V methods more broadly to custom projects in the ISO-NE context is, if projects are screened for a "signal-to-noise" threshold, the random sampling procedure has been violated and the precision formulas in Section 7.2 of M-MVDR are not appropriate. There is an opportunity to increase the use of Advanced M&V in the ISO-NE evaluations, as outlined in Figure 77. We envision a process where projects that are suitable for Advanced M&V are evaluated that way, while random sampling and traditional M&V are reserved for the share of the custom population deemed unsuitable for Advanced M&V. The key difference between Figure 77 and our understanding of the current process is order of operations. Currently, sample design happens first and then sampled projects are screened for suitability of meter-based methods.

Figure 77: Custom C&I Evaluation Process



The cost savings potential from increased use of Advanced M&V in the custom C&I component of Vermont's FCM evaluation hinges on what share of the peak demand savings are suitable. Figure 11 showed that less than 15% of the relevant summer and winter demand from GMP's C&I projects made it through the applicability screening. This share could be bolstered with improved meter correspondence or by looking at the impact of multiple projects at the same site as a "bundle". However, some amount of traditional M&V will always be needed to address custom new construction and custom market opportunity or to deal with retrofit sites that fail data availability and technical screening.

The TOWT regression model used in this study relies on weather and time factors as the primary explanatory variables. For large custom projects in an industrial setting, weather is often not nearly as important as production. It is likely the EVT or a third-party evaluation contractor could gather production data from a manufacturing site and significantly improve the regression model. This approach falls somewhere in-between Advanced M&V and Traditional M&V because of the

correspondence needed with the site to identify key input/output variables, gather the data, and normalize it.

While custom C&I evaluation is a logical starting point for increased incorporation of Advanced M&V in Vermont, we urge PSD and EVT to think carefully about using individual customer regression estimates for settlement. In an Energy Savings Guarantee program, where participant incentives are based on the results of an individual customer regression model, rebate amounts will be a mixture of performance and random noise. The example site in Figure 72 showed that even sites that pass a full regiment of screening checks can return erratic savings estimates. In Figure 69, 11 of the 65 sites analyzed return an estimated increase in consumption (negative savings) and would receive no incentive under an Energy Savings Guarantee program design. At the program or portfolio level, site-level over and under estimates will cancel out. However, the benefits of aggregation will be a small consolation to facilities that get little or no rebate from a good project due to random modeling noise. If Vermont chooses to implement an Energy Savings Guarantee produce, we offer the following recommendations.

- Clearly explain to participants how the mathematical model of their facility energy was developed. Provide them the raw data, regression coefficients, and weather data.
- Limit participation to sites with a large expected reduction in facility consumption and predictable load patterns. The quasi-FSU metric used in the analysis is useful because it considers both elements.
- Communicate the uncertainty associated with their regression model in clear energy and financial terms. Participants should understand that their incentive amount is driven by both performance and an element of chance, but the chance is symmetric. The likelihood of over-payment and under-payment are the same if the model is unbiased.
- Establish a process by which participants can provide supplemental modeling variables such as production or operating schedules to improve the model's explanatory power. Make it clear that these data must also be furnished for the post-retrofit reporting period.
- Create a mechanism for documenting and accounting for non-routine events. Bonneville Power Administration's Verification by Energy Modeling protocol[17] includes a useful discussion on this topic.
- Report on project performance at pre-defined intervals. We recommend every three months during the performance period. This approach could help program administrators and participants identify issues early on and avoid surprises and contentious settlement discussions.

The last bullet point in the list above touches on a key consideration for comparing Advanced M&V and Traditional M&V. Advanced M&V methods can produce estimates of avoided energy use quite quickly. Consider the example site shown in Figure 12. The mathematical model of facility energy use is established using pre-retrofit data. With near real time access to AMI data, it is possible to estimate

[17] https://www.bpa.gov/EE/Utility/measurementandverification/Documents/7_BPA_MV_Energy_Modeling_Protocol.pdf

impact one day into the post-retrofit period. While this quick feedback is useful, caution should be applied before projecting performance in the first week or month to an annual basis.

Between EVT and a third-party evaluation team, annual M&V expenditures for custom C&I programs are approximately $500,000. DSA has minimal visibility into the allocation of these costs but estimates that approximately $100,000 of those costs are for labor and instrumentation to sub-meter 20-25 projects annually. Table 27 illustrates how annual M&V costs might be reduced if the number of site requiring end-use metering could be cut in half through increased prioritization of Advanced M&V methods.

Table 38: Cost Savings from Increased Use of Advanced M&V

M&V Framework	Current FCM Process	Increased Use of Advanced M&V
Total Sample Size (annual)	75	100
Sites with End-Use Metering	20	10
Labor, Equipment, and Travel Expense for Option A/B (per site)	$5,000	$5,000
Annual Metering Expenses	$100,000	$50,000

While the current Vermont M&V process seeks to leverage AMI data where possible and only falls back on end-use metering where necessary, DSA believes the revised evaluation design procedure shown Figure 77 could further reduce site visits and metering by screening for suitability earlier in the process and reducing the share of peak demand savings entering the "sample design and traditional M&V" pathway. We estimate this change could save approximately $50,000 per year. As long as the Advanced M&V infrastructure costs are less than avoided labor and equipment fees, the transition would be a net cost savings.

5.2 PRESCRIPTIVE

With Advanced M&V methods, the need for sampling is largely eliminated, but there are established statistical procedures to quantify the estimation uncertainty. Prescriptive measures that follow a TRM characterizations typically involve a mixture of engineering formulas and operating assumptions from a variety of sources. Figure 78 shows an extract from the residential low flow showerhead measure in the Efficiency Vermont TRM to illustrate.

Figure 78: Excerpt from Low Flow Showerhead Measure in the EVT TRM

GPM_{base}	=	Flow rate (gpm) of baseline showerhead = 2.5 gpm for direct install or free giveaways[6] = 2.4 gpm for RNC[7]
GPM_{low}	=	Flow rate (gpm) of low flow showerhead = 1.5 gpm for direct install or free giveaways[8] = 1.9 gpm for RNC[9]
HOURS	=	Annual full load hours = 3,427.1 hours[1]
ISR	=	In service rate, or the percentage of units rebated that are actually installed = 100% for direct install or RNC = 56%[10] for free products
SH/home	=	Average number of showerheads per household = 1.3 for existing homes or multifamily buildings[11] = 2.1 for RNC[12]
$TEMP_{in}$	=	Assumed temperature of water entering house = 51.9 F[13]
$TEMP_{sh}$	=	Assumed temperature of water coming from showerhead = 101 F[14]
Tshower	=	Average shower length in minutes = 7.8[15]
usedays/year	=	Days showerhead is used per year = 365

Like other TRMs, the EVT TRM provides the point estimate of the input parameters but not the associated uncertainty from the source analysis. The ISO-NE M-MVDR refers to these factors as stipulated parameters and allows them to be used provided certain statistical sampling criteria are met when reviewed by an independent evaluator. This is typical for traditional M&V frameworks which focus on the uncertainty from *sampling* as opposed to estimation. Reconciling these two frameworks for prescriptive measures is much more challenging than with custom C&I. We offer the following suggestions for consideration.

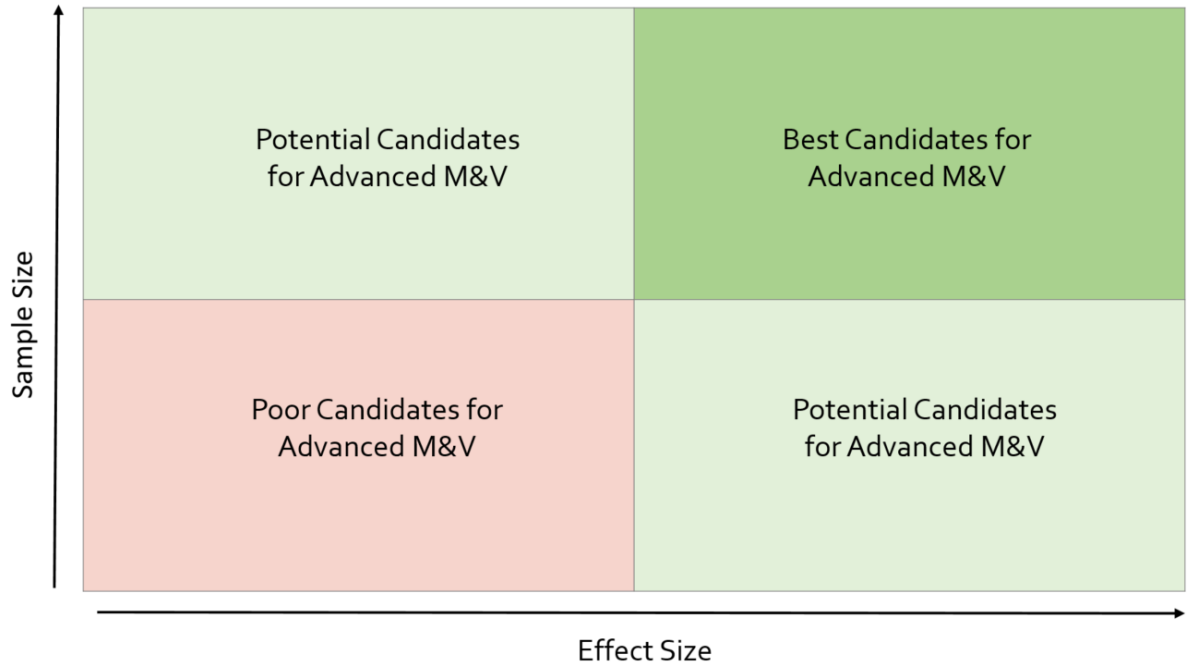
- For measures like LED exterior fixtures where Advanced M&V methods return stable and statistically significant impact estimates, it may be advisable to replace the TRM characterization with the outputs of a pooled regression-based analysis. This could simplify the TRM entry and make verification of the savings straightforward, but there are also drawbacks.
  - ✓ PSD might anticipate some amount of initial resistance from stakeholders with respect to transparency. The current TRM characterizations have each assumption and input sourced and readers can duplicate the results with a bit of arithmetic.
  - ✓ This transition would likely result in some loss of specificity. For example, the LED exterior fixture contains three lumen bins with different savings assumptions. DSA did

not attempt to model impacts separately by lumen bin because sites often install fixtures from multiple bins in a single project. Segmentation presents a variety of challenges so measure characterizations based on pooled regressions would be simplified both with respect to number of parameters and number of permutations.

- From a precision standpoint, we would recommend treating the outputs of this type of measure characterization no differently from a traditional amalgamation of equations and assumptions. In other words, classify the per-unit savings as stipulated parameters in the ISO-NE measurement and verification plan.
- DSA has limited visibility into the cost of TRM maintenance and prescriptive measure characterizations. We assume it is only a small component of the approximately $3.6 million of annual EVT Development and Support Services. Increased use of Advanced M&V outputs certainly has the potential to streamline these activities.
    - ✓ We understand the third-party cost to verify the EVT savings claim is approximately $150,000 per year. Absent a significant shift to regression-based measure characterizations, we see limited opportunity for cost reduction in this area.
- Using Advanced M&V results to verify EVT's annual savings claim or ISO-NE certification for prescriptive measures would create significant risk. There is some amount of risk that ISO-NE could reject the results, but the wide range of potential outcomes is our primary concern. Consider the residential LED Screw-Based bulb measure. From 2017-2020, this measure accounted for approximately 21% of EVT's kWh savings, 17% of the summer kW savings, and 33% of the winter kW savings. Our regression-based estimates return significantly lower energy and demand savings per-unit for this measure. As discussed in Section 4.2, there are both signal size and baseline challenges associated with modeling this measure with whole house data. Using Advanced M&V to evaluate this measure could create a FCM shortfall penalty and a variety of other potential disruptions.

Our results for prescriptive measures in Sections 4.2 and 4.3 showed mixed success. The successful measures all had one common feature: energy impacts were large as a percentage of whole-building load. It is easier to detect large changes in energy consumption than it is to detect small ones. The measures with a large number of participating homes or businesses available for analysis tend to have a smaller standard errors. Figure 79 summarizes the overarching theme across the pooled measure analysis.

Figure 79: Advanced M&V Applicability Quadrants



Market opportunity measures also tended to present challenges in Sections 4.2 and 4.3. Advanced M&V methods produce estimates of the change in energy consumption at the meter. For market opportunity, or replace-on-burnout measures, the baseline is typically a code-minimum piece of equipment that was never installed in the home or business. We observed an interesting set of outcomes for market opportunity measures in Section 4.

1) For measures like the residential ENERGY STAR refrigerator and clothes dryer, our meter-based estimates are ***larger*** than the savings claim in the EVT tracking database. Presumably, this is because most homes had a refrigerator and clothes dryer prior to the ENERGY STAR purchase. The old inefficient unit that was replaced likely used more energy than the code-minimum new appliance assumed in the EVT TRM measure characterization. Figure 80 below illustrates the distinction for refrigerators. While the dominant measure permutation in EVT's programs claims the difference between kWh_{base} and kWh_{eff}, a meter-based approach estimates the difference between $kWh_{baseOLD}$ and kWh_{eff}.

Figure 80: Energy Efficient Refrigerators TRM Characterization

kWh_{base} = Annual kWh consumption of baseline models meeting Federal Standard. Note different value for each category since the average capacity of available product in each category varies[4]

Efficiency Level of Unit Installed	kWh_{base}
ENERGY STAR	693
ENERGY STAR Most Efficient	460
CEE Tier 3	455

For Shift Model Approach only kWh_{base} = Actual Rated kWh [5]

$kWh_{baseOLD}$ = Assumed consumption of existing unit being replaced. Note different value for each category since the average capacity of available product in each category varies.

Efficiency Level of Unit Installed	Pre-1993 or unit in poor condition	Age Unknown
ENERGY STAR	1454[6]	817[7]
ENERGY STAR Most Efficient[8]	965	542
CEE Tier 3	955	537

kWh_{eff} = Annual kWh consumption of efficient models. Note different value for each category since the average capacity of available product in each category varies[9].

Efficiency Level	kWh
ENERGY STAR	632
ENERGY STAR Most Efficient	409
CEE Tier 3	349

2) For the ENERGY STAR dehumidifier measure, our meter-based estimates are negative. This suggests homes typically did not have a dehumidifier previously so the addition of a dehumidifier increased consumption at the meter. Of course, the load growth would have been larger if the home had purchased a code minimum dehumidifier instead of an ENERGY STAR unit.

3) For the LED Screw-Base Lamp measure, our modeling results consistently return lower savings values than the TRM-based values stored in the program tracking data. In this case, we believe the lamps being replaced are actually more efficient than the code-minimum halogen or incandescent assumed in the TRM characterization. Just based on the duration and intensity of Vermont's CFL campaigns, it is likely that many program-supported LEDs are replacing CFLs in residential sockets.

The annual energy and peak demand savings estimated returned by our Advanced M&V analysis likely should not be considered "evaluated" or "verified" savings for these measures and we would not recommend EVT adopt them as per-unit savings assumptions in the TRM. However, elements of an Advanced M&V analysis can still be quite valuable. Specifically:

- **The load shapes of the measures can be used for valuation**. While the annual energy savings may be too high or low, the distribution of impacts is unbiased. Converting the load shape to shares (e.g. what percentage of the annual impact occurs in each hour of the year) would allow for use across a variety of planning and valuation applications.
- **Updating TRM Operating Assumptions**. The TRM characterization for ENERGY STAR dehumidifiers assumes that units run an average of 1632 hours per year. This assumption comes from an ENERGY STAR calculator. The estimated consumption of a new ENERGY STAR unit is given by the following equation.

$$kWh_{eff} = \frac{Avg\ Capacity * 0.473}{24} * Hours * \frac{1}{Integrated\ Energy\ Factor}$$

  If we insert our Advanced M&V result of 509 kWh/year as kWh_{eff} and the TRM default values of 38.4 pints/day for average capacity and 1.76 as the Integrated Energy Factor and solve for the Hours term, the results is 1184 hours per year.

  - ✓ While this exercise provides a mechanism and indication of the directional change called for in the TRM, it relies on several strong assumptions. This calculation assumes all participants did not have a dehumidifier previously and that the dehumidifier capacity and efficiency of the homes analyzed align with TRM defaults.
- **Calibration of TRM inputs for early replacement versions of measures**. For example, our modeling for the ENERGY STAR refrigerator measure returns estimates ranging from 211 kWh/year (matched controls with COVID data) to 379 kWh/year (pre-post without COVID data). These results could be added to the existing kWh_{eff} assumption to update the $kWh_{baseOLD}$ assumption. Doing so would require the Advanced M&V results be divided by the average number of rebated units per household because occasionally homes purchase multiple refrigerators. Program administrators would need to weigh the tradeoffs between more recent data and the efficiency-level specificity in the current TRM characterization.
- **Account for the impacts of future demand side management in a more informed way**. The three prior bullets certainly apply to Advanced M&V results from variable speed cold climate heat pumps, but there are several important nuances with fuel switching measures that require additional caution when interpreting results.
  - ✓ In the EVT TRM, ductless heat pump installations are claimed in two parts. The retrofit measure characterizes the switch from fossil fuel heating to a code minimum heat pump and the impacts are claimed by Vermont Distribution Utilities' Tier III programs. The market opportunity measure is claimed by EVT and represents the electric efficiency savings between a code minimum and high efficiency heat pump. Impacts must be summed across the two measures to arrive at the TRM's estimate of the impact of the ductless heat pump on the grid. In Table 39, the annual energy impact is equal to the sum of the kWh Penalty column for the retrofit measure and the ΔkWh Total column for the market opportunity measure. The same logic applies to peak demand savings, but requires an additional step of applying the summer and winter coincidence factors to the kW Penalty and ΔkW values.
  - ✓ The meter-based estimates in Table 24 are consistently lower than the kWh penalties and winter kW increases implied in Table 39. We were unable to extract heat pump

capacity from the EVT tracking database for our sample, but the average is likely in the 18,000-24,000 BTU/hour range.

Table 39: Variable Speed Mini-split Heat Pumps – Retrofit Entry

Retrofit

Type	Capacity	kWh Penalty	kW Penalty
Single Zone	6,000	(1,493.77)	(1.08)
Single Zone	9,000	(1,805.86)	(1.31)
Single Zone	12,000	(2,051.28)	(1.48)
Single Zone	15,000	(2,891.69)	(2.09)
Single Zone	18,000	(2,990.76)	(2.16)
Single Zone	24,000	(3,753.36)	(2.71)
Multi Zone	18,000	(2,873.75)	(2.08)
Multi Zone	24,000	(3,914.81)	(2.83)
Multi Zone	30,000	(4,864.97)	(3.52)
Multi Zone	36,000	(6,760.13)	(4.89)
Multi Zone	42,000	(8,785.04)	(6.35)
Multi Zone	48,000	(6,509.66)	(4.71)

Market Opportunity

Type	Capacity	Capacity Ranges	ΔkWh Total	ΔkW
Single Zone	6,000	5,000 - 8,000 BTU/h	612.27	0.41
Single Zone	9,000	8,001 - 11,000 BTU/h	619.09	0.41
Single Zone	12,000	11,001 - 14,000 BTU/h	607.89	0.40
Single Zone	15,000	14,001 - 17,000 BTU/h	872.45	0.58
Single Zone	18,000	17,001 - 22,000 BTU/h	680.27	0.44
Single Zone	24,000	22,001 - 32,000 BTU/h	792.64	0.51
Multi Zone	18,000	16,000 - 20,000 BTU/h	680.74	0.44
Multi Zone	24,000	20,001 - 24,000 BTU/h	1,160.58	0.77
Multi Zone	30,000	24,001 - 32,000 BTU/h	1,323.92	0.89
Multi Zone	36,000	32,001 - 38,000 BTU/h	1,759.29	1.17
Multi Zone	42,000	38,001 - 44,000 BTU/h	2,268.67	1.54
Multi Zone	48,000	44,001 - 65,000 BTU/h	1,791.01	1.16

Any pooled analysis that includes multiple permutations must be interpreted with caution, but this finding suggests that the Equivalent Full Load Hours for heating term in the EVT TRM warrants further review. If Vermont residential customers are using ductless heat pumps to serve less heating load than assumed, that affects both electric and fossil fuel resource savings. Perhaps more importantly, the meter-based results provide insight into the planning implications for a highly electrified future as Vermont takes aggressive measures to combat climate change. The meter-based estimates may also reveal a teaching opportunity with Vermont contractors and customers about using heat pumps as the primary heating strategy in homes to displace as much delivered fuel consumption as possible.

6 SOFTWARE OPTIONS

Table 40 lists potential open-source, Advanced M&V tools. Each tool has strengths and weakness in terms of applicability to the broader state of Vermont.

Table 40: Advanced M&V Tools

Tool	Creator	Software Required	Open Source	User Interface	Customizability	Best Applications
ECAM	SBW Consulting	Microsoft Excel	Yes	Yes	Medium	Site-Specific Commercial and Industrial
RMV2.0	LBNL	R-Studio & Web browser	Yes	Yes	Low	Commercial & Industrial
OpenEE Meter	Recurve	Jupyter Notebook	Yes	No	High (in code)	Residential & Commercial
NMECR	kW Engineering	R-Studio (developer environment)	Yes	No	High (in code)	Commercial

ECAM is a Microsoft Excel add-in and requires no statistical programming experience. It includes functionality for importing meter data and gives the user a wide variety of useful visualization tools. ECAM is best used for a limited number of individual customer analyses. Large pooled analyses or programmatic studies that require hundreds or thousands of site-level analyses call for a more scalable solution.

Within the RMV2.0 tool, a screening is applied to each building to identify whether it is suitable for analysis, which is an important step of the process in developing accurate estimates. Additionally, this tool identifies potential non-routine events within the time-series and will display a visual representation. But across the implementation of this tool, the only modifiable parameter is the time-scale weighting function, which heavily limits the customizability. Pre-processing is also required to fit the specific format for input data.

The OpenEE Meter tool has features to better fit the nuances of residential buildings through an adjusted approach where 12 weighted monthly TOWT models are created to account for seasonal variation. This allows this tool to accurately estimate savings for both commercial and residential buildings. Comparison group functionality was added in response to the COVID-19 pandemic. Unfortunately, with the lack of user interface, this tool requires a steeper learning curve. While open source, the Recurve platform to implement OpenEE is a paid product.

The benefits of the NMECR tool include the calculation of both avoided energy consumption and normalized savings, as well as a high level of customizability within the code. In addition, there are

various day types that are allowed to be input into the dataset. On the other hand, the level of automation is below the other two tools, and the ease of usability is not very high.

Efficiency Vermont has also developed significant in-house Advanced M&V capabilities. On one hand, EVT is the ideal party to develop an Advanced M&V tool for Vermont because of its relationships with electric utilities and program participants. However, their role as the program administrator means that the outputs of their procedures would not satisfy the ISO-NE requirement for an independent third party reviewer without some sort of external review or replication.

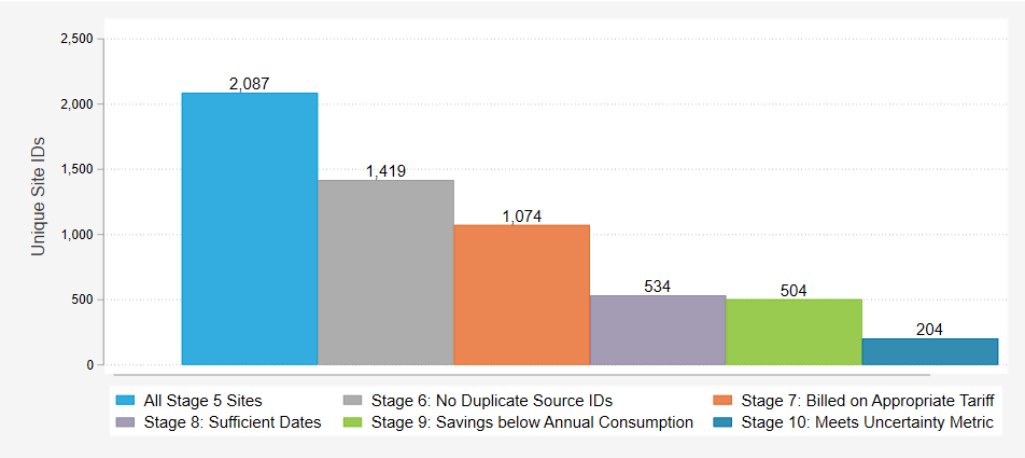
While each of these options provide a viable path toward to increased use of Advanced M&V in Vermont, the right fit will ultimately come down to the use case. The use of automated software is most valuable when the number of sites/projects to analyze is large. It is especially important for a small state like Vermont to carefully consider the likely volume of applicable projects, sites, and measures as it considers software options in this space.

7 CONCLUSIONS

The usefulness of the measure-specific analyses in this study was varied. Advanced M&V techniques are a powerful tool, but it is important to understand the applications where they work well and the applications where they do not. We identified several areas where increased use of Advanced M&V can improve accuracy and lower cost, but we do not see these methods replacing the TRM or widely replacing the need for Traditional M&V techniques. The key themes of the study include:

- **Data preparation is key.** While the correspondence EVT has between energy efficiency participation and GMP meter and account infrastructure is impressive for an external organization, it is imperfect. Advanced M&V procedures require precise linkages to achieve scale and the 'efficiency utility' construct in Vermont likely makes this more challenging than a state where IOUs are the program administrator.
  - ✓ Imperfect linkages can lead to sites that do not have meters appropriately aggregated and, therefore, do not display accurate consumption records. Or it can lead to inadequate matching, since we are unable to guarantee that a matched control did not receive treatment.
  - ✓ If the role of Advanced M&V in Vermont grows, efforts to improve correspondence may be warranted. Address matching and GIS mapping of meter coordinates could help but will require resources from both EVT and the electric utilities.
- **Filters identify those eligible for analysis.** We apply filters, most notably uncertainty filters, so that the sites we analyze are limited to those that qualify as good candidates for Advanced M&V methods. Without these filters, the results would likely be inaccurately quantified.
  - ✓ These filters remove a significant portion of sites. This is problematic for custom and commercial projects as those eliminated are unable to be estimated using the techniques described in this report. Figure 81 depicts the AMI-level filters that are applied to all the commercial measures. Of the sites that make it to stage 9, we see that 60% are lost after we apply the quasi-FSU filter. This mirrors the story told in Figure 10 for custom projects.

Figure 81: Waterfall of AMI Filters Applied to All Commercial Measures



- **Signal-to-noise ratios are important.** Measures that are best suited for Advanced M&V methodology have large sample sizes and produce a large change in electricity usage. Large sample sizes ensure that we are adequately able to approximate the savings of the population of participants who installed that measure. Additionally, since we are using whole-building AMI data, the measures must produce enough savings so that they can be detected within the noisy consumption of the building.
- **Precision needs to be a consideration.** We use an accuracy assessment to assess bias across our various methods. This is a useful step as error in a model can never be less than the bias. For example, a model that over predicts by 2% cannot have a margin of error of less than 2% in the real world.
    - ✓ Standard error is a function of sample size, so the more customers that adopt a measure, the more precise the estimate will be.
- **Matching helps a lot.** In the face of a disruption like COVID-19, matching is the only chance we have of producing unbiased estimates that separate the effect of efficiency from exogenous changes in energy consumption related to the pandemic. However, it is not clear that our matching procedures remove the COVID-19 bias entirely.
    - ✓ We cut off our post-period data at March 2020 to analyze projects and measures with exclusively pre-COVID data. Going forward, it will be important to decide when the pandemic has subsided enough to consider meter data "normal" enough to act as a baseline without adjustment or a matched comparison group.

These findings give us insight into the necessary components for applying Advanced M&V methods in Vermont. To have a chance at accurately and precisely estimating savings from efficiency measures, the guidance above must be taken into consideration. Measures must be appropriately evaluated prior to implementation of this method. Without evaluation, estimates may grossly overstate or understate energy savings.

8 TECHNICAL APPENDIX

The hyperlink below will download a Microsoft Excel workbook containing 8760 and 864 load shapes for each of the prescriptive measures investigated in this report. The workbook includes separate tabs for each modeling approach: pre-post TOWT, matched controls without COVID period data, and matched controls with COVID period data. The 8760 load shapes include an estimate for each hour of the year while the 864 load shapes contain average hourly load impacts for a peak day, average weekday, and average weekend/holiday by month. The 8760 load shapes are mapped to the 2025 calendar. Readers should be mindful of the 'days included' field when applying the 864 load shapes as this indicates the number of days of that day type in the month.

- https://demand.files.com/f/3b91d830cf01b64f

The hyperlink below will download a zip file containing the code that was developed to complete the analysis in this report. It is divided into five folders: pulling data, pre-post analysis, matching analysis, accuracy assessment, and valuation. The pulling data folder contains code regarding the EVT tracking database, as well as the AMI data. The pre-post analysis, matching analysis, and valuation folders contain the code that generates the results in Section 4.2 through Section 4.5, while the accuracy assessment code generates the results in Section 4.1. Throughout this code, we implement user written commands, which were written by Demand Side Analytics. These commands will begin with `dsa_', and the underlying ado files (or the text files that contain the Stata program) can be provided upon request.

- https://demand.files.com/f/d5ad16cf9fc8d787