Analysis Techniques for Evaluating Energy Conservation Programs Using Utility AMI Data

Richard Raustad, Danny Parker, Robin Vieira, Karen Fenaughty
FSEC Energy Research Center, University of Central Florida

ABSTRACT

With the evolution of Advanced Metering Infrastructure (AMI) electric meters, utility companies now have direct access to whole building electricity use at a granular time scale. AMI data can be used for a variety of purposes beyond billing, for example, to evaluate the efficacy of energy conservation (EC) programs. Historical methods for calculating EC program savings include building simulation models and laboratory and/or field testing. With big data now available, which analysis methods are more likely to yield quality results?

In a recent project, the Orlando Utility Commission provided monitored AMI data from Oct 1, 2015 – Sep 27, 2018 for 2,832 Orange Country, Florida rebate participants. These participants had either already enrolled in a rebate program or had signed up to participate. The project objective was to analyze this large AMI data set to estimate the savings from a heat pump retrofit program in energy (kWh) and coincident peak demand (kW) relative to baseline efficiency levels.

This paper illustrates five methods of predicting EC program savings for a utility company's rebate program using: 1) side-by-side groups of current and future participants, 2) before and after evaluation of a stable group of participants, 3) evaluation of a before and after group using pooled regression, 4) regression on individual accounts with results then averaged, and 5) building simulation model results are used as a comparative baseline. This effort was pursued to see if the bias between the various participant segments could be reduced to focus on energy differences within the retrofit equipment itself.

Introduction

The objective of this paper is to describe the energy use evaluation and analysis methods used for utility company rebate programs. The primary intention is to estimate the energy savings (kWh) and coincident peak demand (kW) so that future segments of the programs can be most efficiently applied.

The challenge was to analyze rebate program savings using only AMI meter data with whole building energy use rather than the end-use energy for that program. The ability to use whole building energy data to evaluate rebate program savings depends on the specific rebate program and whether or not the resulting energy savings are statistically significant. When analyzing large samples of whole building electricity, which method of analysis provides the most reasonable results? The analysis methods used here followed the recommended protocols suggested by U.S. DOE for evaluation of utility programs as described in an NREL report by Agnew and Goldberg (2013). A brief summary of each of the five energy analysis methods investigated are:

1. Current vs. Future Participants: Evaluation of the average of each retrofit group segment in a year compared with another control group in that same year that would then install the specific rebate in a future year. Annual energy use is then compared.
2. Before/After Protocol: This is a conventional approach where the average of a single group of rebate participants is tracked for a year before and then after the rebate installation. Annual energy use is then compared.
3. Evaluation of Before/After Sample with Pooled Fixed Effects Regression: This is similar to the second approach, but the pooled averages of the segments before and after retrofit are then regressed against weather.[1] Each group serves as its own control.[2] This involves summing the 15-minute demand data and then collapsing the series into daily average kWh per day and then regressing average energy use for the group using nearby representative weather data. Weather normalized annual energy use is then compared.
4. Individual Account Regression with Summation: This is a regression of each account against weather after sites are screened for data completeness and confounding influences (e.g. incomplete data, multiple efficiency measures installed simultaneously). The results for individual cooling, heating and annual energy consumption are then summarized to provide the minimum, average and maximum statistical results.
5. Simulation Evaluation: This is a conventional approach and uses recognized building simulation models to estimate the retrofit savings based on housing characteristics and hourly weather with detailed energy models.

AMI Data Availability

AMI data analysis begins with a review of available data. Although thousands of customer accounts were provided, a review of the data and consideration of the analysis technique eliminated a significant portion of available data. An account was eliminated if the rebate installation occurred too early or too late in the available time series or the quality of the available data appeared suspect (i.e., missing data or unexpected changes in energy consumption such as a refurbishment or changes in building usage). The amount of data available for analysis is also impacted by the uncertainty in the rebate installation date and can result in large attrition in the accounts that had installed the measures by the target evaluation date. The point here is that utility companies can only analyze a rebate program where a sufficient data set exists and that data is of sufficient quality to provide reasonable results. For this analysis, the utility company provided 15-minute energy (kWh) data over a four-year period. Since the energy analysis relied on daily usage the electric meter collection frequency was less important. However, evaluation of demand savings required a collection frequency of 1 hour or less. For these reasons, a robust AMI meter data collection protocol is required. If the data retrieval process misses data for various reasons, that historical data needs to be collected at the next opportunity. Of course some data will be lost, but efforts to minimize missing data are valuable.

---

[1] Another powerful before/after analysis method is a two-stage regression with heating and cooling slopes to a variable temperature and a separate intercept term estimated for each site. However, this analysis is time consuming for large datasets.

[2] See Section 5.0 of Agnew and Goldberg (2013).

Data quality is important. When a rebate program installs or replaces a building component, the analyst expects that the change in energy use is only attributed to the rebate program. However, with a change in building ownership, operational control of electric equipment is typically altered or renovations may be done which can alter the magnitude of the electric signals. To identify changes in signal magnitude, the easiest assessment technique is to review a graphical representation of the AMI data before and after rebate measures are installed as described later in the paper.

Building Characteristics

Building characteristics are also important when analyzing AMI data sets. Comparing a retrofit and control group that have significant demographic and building specific differences can reduce confidence in the results. To understand the building stock in a certain geographical area a property database can be included into the analysis.

For the evaluation we conducted, a property appraisal database with 95,926 electric utility customer records was incorporated. This customer demographic data was provided by the University of Florida and was used to help interpret and evaluate the processed AMI data as well as determine simulation parameters. Key demographic data were used to estimate influences such as square footage, swimming pools and market value. There were some cases where a key variable was not in the property database or there was no match for a specific rebate customer.

The building characteristics of single family utility customers are summarized in Table 1. Commercial buildings also have varying characteristics that should be reviewed or understood prior to analysis. Of interest is that older buildings are constructed to less stringent building codes and larger buildings or buildings with pools have higher energy use. Multi-story buildings have ceilings that are protected from solar gains on all but the top floor which affects air-conditioner energy use compared to single-story buildings of the same total conditioned floor area. These differences can be eliminated using a before and after analysis method for these particular segments.

Table 1. Territorial building characteristics

Single Family			Conditioned Floor Area			Bedrooms			Floors	Pools
Vintage	Group Designation	total n 95,926	25% quartile	Median	75% quartile	25% quartile	Median	75% quartile	Median	%
<1982	Older	47,063	1,203	1,481	1,904	2	3	3	1	15
1982-2000	Mid-aged	17,702	1,388	1,748	2,198	3	3	3	1	24
>2000	Newer	31,161	1,894	2,337	2,990	3	4	4	2	19

Analysis Method Comparison

The following analysis uses time series average AMI meter data for a group of utility customers. The entire group of AMI meter data is collapsed into a single average time series for the first two analysis methods. For example, the AMI data for day 1, time 00:15 were averaged for all customers within a group. Then time 00:30 was averaged for all customers, and so on throughout the analysis time period. This provided a single time series of data for each group – either the retrofit and control groups or the before and after analysis of the rebate groups. The latter two analysis methods analyze AMI data per customer.

These time series data were analyzed in a statistical analysis program (Stata). Additional post-processing was necessary to organize information in report quality format. This effort proved to be very time consuming. However, our analysis experience provided distinct improvements to the analysis methodology as the evaluation progressed, leading to more robust results with greater confidence.

Utility Company Residential Heat Pump Rebate Program

Orlando Utility Commission's heat pump rebate program replaces older less-efficient central air-conditioners or heat pumps with above code equipment.[3] Significant energy reductions can be achieved with newer more efficient systems. Replacement equipment of 15 SEER or greater qualify for rebates based on system size (tons AC) and efficiency (SEER). Rebates currently range from $90 to $1,630 and require an Air-Conditioning, Heating, and Refrigeration Institute (AHRI) certificate or reference number for the replaced system.

Current versus Future Participants

The specific residential heat pump rebate program which replaced existing heat pumps with above code efficiency had segments with varying degrees of efficiency levels. In the first evaluated method, the results of the rebate and control group method is compared to the before and after protocol results. The entire AMI data set was subdivided into SEER groups of varying efficiency levels (15, 16, 17, 18-19 and 20+) to identify savings relative to the previously installed equipment efficiency. The efficiency of the replaced system was unknown.

AMI data from 2016 for rebate participants that had installed the heat pump prior to 2016 were compared with future program participants that had not yet installed the equipment. The demographics and rebate customer information review verified a general expectation of increasing energy use by group in the baseline condition. The advantage of the current and future participants evaluation is that demographic differences are controlled within each group. However, splitting up the available AMI data set into separate efficiency groups further reduced the sample size with the highest efficiency groups having the lowest sample size -- presumably due to higher equipment costs. In the tables below the annual energy use and savings are relative to the specific group.

Table 2 shows the results for the unpaired rebate participants that had not yet installed the equipment (Control) and those that had installed the equipment (Rebate) when customer accounts with duplicate rebate measures and swimming pools are removed from the analysis. Evaluating the overall program suggested that the heat pump retrofit program drops annual electricity consumption by roughly 10% per year in some groups while showing an energy increase in others. Additionally, these are the "gross savings" levels and do not account for the efficiency of the original equipment (which was unknown) or that consumer energy use behavior may have changed over time. Utility companies would be well served to document existing equipment models and specifications, and other pertinent information, during the rebate installation.

[3] Orlando Utilities Commission Residential Rebate Program. Retrieved June, 2020.

Given observed variances, sample sizes smaller than about 40 for an unpaired evaluation as in Table 2 cannot be considered valid at a reasonable confidence level unless energy differences are very large. For instance, the seeming lack of savings for SEER 17 and SEER 18 are likely due to the retrofit sample building characteristics significantly differing from the control sample which consists of only 25-35 sites. Little can be concluded beyond the findings for the higher efficiency groups as the sample sizes are too limited.

Table 2. Rebate and control group results

	Annual Energy (kWh, participants)				Savings	
SEER	Control	n	Rebate	N	%	Comments
15	16,156	191	14,629	537	9.45	Lowest efficiency group
16	16,871	104	15,800	224	6.35	
17	15,911	25	15,932	57	-0.13	Small sample size
18	17,850	35	17,966	70	-0.65	Small sample size
20+	18,109	13	16,101	17	11.09	Inadequate sample size

The advantages of this method is that the bias is potentially eliminated as the non-participants in the target year theoretically have the same biases in demographic factors as the participants and that there is no need to correct for weather (as there is with pre and post schemes) because each group, participants and non-participants, is evaluated in the same year.

The disadvantages are that groups are being compared side-by-side and the associated statistical variances are higher since the samples cannot be paired, leading to loss in precision in the estimates of smaller groups, with only larger samples able to attain acceptable levels of precision. The contradictory apparent negative savings for the higher efficiency segments is likely caused by this shortcoming.

Before/After Protocol

Before/after sample evaluations are powerful statistically in that the buildings and occupants are paired and largely unchanging, leading to reduced variance at a given sample size. A comparison of the control group of 368 buildings was evaluated both in 2016 before retrofits were installed and in 2018 after the retrofits were installed. This was accomplished by estimating consumption for the available weather related periods in the pre period (January – August 2016) with the matching weather period in the post period (January – August, 2018). This methodology uses the same homes in the pre- and post-period (a paired sample), thus the variances between the periods are naturally lower than the current vs. future analysis method and the savings estimate more statistically robust. However, there are weather differences to be accounted for.

The overall annual savings of the heat pump program evaluation applying the before/after protocol is about 4-9% per year as shown in Table 3. However, unlike the comparison of current and future participants with undersized sample sizes, this evaluation shows that each of the groups produce savings, but not equally. This method may be generating more realistic results for this program. The higher efficiency SEER 15 and SEER 18 groups appear to produce the largest savings, although this will be better evaluated next by comparing the response of the various groups to prevailing weather in the pooled regression analysis.[4]

Table 3. Before and after protocol results

SEER	n	Annual Energy, kWh		Savings	Comments
		Before	After	%	
15	191	15,297	14,033	8.26	Lowest efficiency group
16	104	15,926	15,281	4.04	
17	25	15,364	14,185	7.67	
18	35	16,654	15,135	9.12	
20+	13	17,192	16,075	6.50	Marginal sample size

The advantages of this method are that the same group of sites are tracked for at least a two-year period and the associated variance in the evaluation estimates is potentially lower as much of the statistical uncertainty in comparing groups arises from the unequal variance in comparing unlike groups. A before/after group approach is much more statistically efficient. Typically, the buildings are exactly the same in the pre- and post-group evaluations. Generally, variances are reduced by more than three times with a pre- and post-evaluation scheme.

Disadvantages to this method is that weather differs in the years before and after the retrofit. Thus, weather normalization is likely necessary, but problematic. Where seasonal data is sparse, such as the very short heating seasons in Central Florida, models may not be correlated to the same temperature range from one year to the next. Similar to the side-by-side group, because of the uncertainty in the rebate install dates, there was a larger amount of attrition in the accounts that we could assume had installed the measures by the target evaluation date. Even with a graphical review of AMI data the actual install dates cannot be identified with certainty unless a marked difference in energy use is apparent (see Figure 4). Another disadvantage to this method is the length of evaluation period, which allows more time for changes in occupancy, behavior, building improvements, and system degradations that may be unobserved by evaluators.

Evaluation of Before/After Sample with Pooled Fixed Effects Regression

To address limitations in the before-after protocol, a pooled regression analysis was performed to evaluate a group's pre- and post-data to see how they differed before and after the retrofits were installed when considering weather factors.

The pooled regression evaluated daily energy use of the pooled building groups in 2016 and 2018 against weather using linear regression for heating and cooling. Because much of the energy savings variation within groups comes from weather, these results tell us about the relative efficiency of the groups in baseline condition, the differences or similarities in baseload electricity use, and the savings from pre- to post-condition. The pooled regression intrinsically normalizes for weather as well. Also, as this method uses regression, with resulting statistics that may be used to further eliminate buildings with energy use poorly associated with weather (these can be vacant buildings, buildings with seasonal occupants, or

[4] One possible explanation for the large savings of the well sampled SEER 15 group is that free-riders may have been more often replacing very old and poorly operating heat pumps than in the higher efficiency groups.

rentals). Most buildings will show a discernable correlation to weather (see Figure 1 and 4). However, some buildings may be highly efficient which reduces the ratio of air conditioning energy use to total building energy use (lower signal magnitude) or have non-weather related energy consumption variation which reduces the correlation to weather. In this analysis, if the R-squared value of the regressed cooling and heating season data was less than 0.3 and 0.15 (lowered to allow more heating data in a cooling dominated climate), respectively, the account was removed from the analysis. Note that this correlation statistic is very low and most sites will be retained for analysis unless other artifacts of energy consumption dominate the result. These limits may differ between residential and commercial buildings and eliminated sites should be reviewed to support the correlation limits.

Of the nearly 2,900 rebate participants, there were 257 participants with complete AMI meter data where retrofits were installed by December 31, 2017. These were screened to assure good data quality and limited data variance between groups. Table 4 shows the same sites that were evaluated in 2016 before retrofit and then again after the retrofit in 2018 to address concern with bias in selecting the groups for the current vs. future participants. Although the sample sizes of the higher efficiency groups are small, they are paired, pre and post, which dramatically reduces the variances experienced in the small samples in the current versus future evaluation. Based on a previous evaluation of utility data by Parker[5], sample sizes larger than 12 could yield good significance with this approach. Obviously, estimation errors will be reduced with larger samples, even in a paired scheme.

Table 4. Before and after sample with pooled fixed effects regression results

SEER	n	Annual Energy, kWh		Savings %	Comments
		Before	After		
15	148	16,400	16,436	-0.22	Lowest efficiency group
16	70	16,299	14,997	7.99	
17	18	15,551	13,621	12.41	
18	18	18,510	15,412	16.74	
20+	3	--	--	--	Inadequate sample size

For this evaluation, the daily building energy use is calculated over time and then regressed against the average daily temperature at the local weather station for the period involved. Existing protocols show that regressing daily average energy consumption against weather reduces variances that result from the natural rhythms of temperature and solar radiation on building loads that follow a diurnal cycle (Haberl et al., 2005). Including weather in the analysis is effective because 80% or more of the daily variation in electricity use in large aggregations of buildings comes from weather. This approach has been demonstrated by previous FSEC research as able to discern differences in weather-related consumption as low as 3% at a 90% confidence level (Fenaughty and Parker, 2018) when using a two-stage regression with a variable temperature determination for best fit.

Regressions explained about 85% of the variation of daily electricity use beyond a 66°F average daily outdoor temperature and regressions for average daily outdoor temperature less

[5] Appendix A: Comparative Statistical Efficiency of Paired and Unpaired Sampling Designs in Utility Program Evaluations was included in the final report to the utility company, Parker, et. al.

than 65°F explain about 54% of the variation. This shows the strong relationship of heating, ventilation, and air conditioning (HVAC) energy use to outdoor temperatures and the sporadic heating season in Florida. Figure 1 shows an example of the pooled regression analysis for the SEER 16 group for the 70 accounts pooled together. Note the obvious cooling energy savings, but the possible increase in space heating.

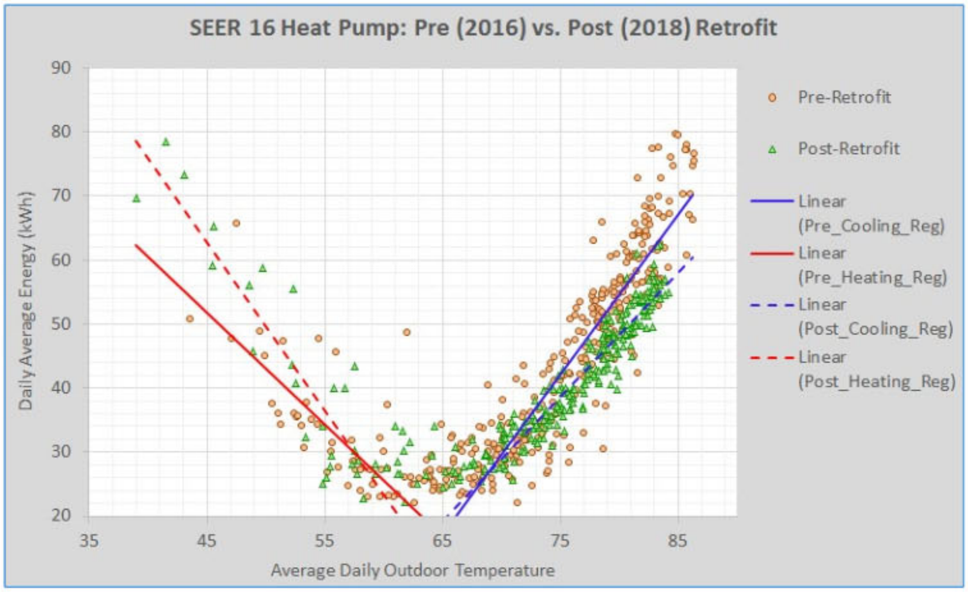

Figure 1. SEER 16 group with daily electricity use plotted pre and post with evaluated regression lines, n = 70. *Source*: Parker, 2019.

Individual Account Regression with Summation

One problem with the pooled regression approach is that it can hide large deficiencies with individual accounts that muddle the overall statistical rigor. This can include sites with long periods of missing data or those where there appears to be no predictable space conditioning use at all (e.g. vacant rental, vacation homes etc.). The pooled regression approach is recognized as a powerful method as described by Agnew and Goldberg (2013). However, we found the individual account regression approach to be the most reliable in that it evaluates a regression as shown in Figure 1 for the overall group, but does this for the individual accounts in an automated fashion. Sites with large portions of missing data or poor weather responsiveness can be excluded from the evaluation (see Figure 4).

To evaluate the performance of each efficiency group, we estimate a four-part regression for each rebate participant. This method first regresses daily building energy use against the average daily outdoor temperature in the baseline condition before retrofits were installed in 2016. A linear regression was performed on this data, however, a non-linear regression may also be appropriate. The cooling and heating season data is regressed against weather using a balance point approach where data is eliminated from the regression below or above an average daily outdoor temperature depending on season. For the cooling season, the best balance point is chosen by first starting at an appropriate temperature to capture extremes but avoid unrealistic results, and includes only days with average daily outdoor temperature greater than the balance point temperature, then marching this balance point forward one degree Fahrenheit at a time until the temperature is found which has the greatest correlation to the energy use data set, known as the goodness-of-fit or R-squared value.

An upper and lower limit on the balance point is used to avoid unrealistic results. The balance point is a surrogate for the thermostat temperature where an approximate 10°F offset is caused by internal heating from lighting and appliances per Agnew and Goldberg. In residential buildings this was often 66°F for the cooling season, meaning that an average daily outdoor temperature of 66°F suggests an interior average set point of 76°F during that season and is the point at which the building began to need cooling. The intercept term is an indicator for the base appliance load that does not respond to weather. The regression coefficients are then captured for each customer account. The same procedure is then performed for the heating season. The evaluation is then repeated for each account within each efficiency group.

This analysis is also performed for the post-retrofit data set and regression coefficients and balance points are retained for both cooling and heating. Annual energy use is estimated for each building by applying the regression coefficients to a typical regional weather dataset. Daily cooling energy is estimated for average daily outdoor temperatures greater than the balance point temperature and heating energy for average daily outdoor temperature less than the balance point. Baseload energy use (i.e., no HVAC operation required) between the cooling and heating balance points is estimated as the cooling energy use at the cooling balance point outdoor temperature. The cooling energy use model is used for baseload estimation because it is typically much stronger than the heating energy use model in our cooling-dominated climate. The before and after retrofit averaged results are presented in Table 5. In most cases the cooling and heating energy savings are rather large and simply represent the regression analysis of the pre- and post-data with regard to the previously installed system efficiency (unknown), as well as any changes in occupied dwelling energy use profiles over the analysis period. Careful review of the example shown in Figure 4 also show significant savings and support these results which may include savings unrelated to the specific rebate program.

Table 5. Individual account regression with heating, cooling and annual summations

System Efficiency	Seasonal Segment	Before kWh	After kWh	Savings kWh	Savings %
SEER 15	Heating	1,148	814	334	29.1
	Cooling	5,898	3,864	2,034	34.5
	Annual	14,902	13,367	1,535	10.3
SEER 16	Heating	752	577	175	23.3
	Cooling	6,520	4,097	2,423	37.2
	Annual	14,442	13,803	640	4.4
SEER 17	Heating	1,141	646	495	43.4
	Cooling	6,290	5,189	1,101	17.5
	Annual	13,942	12,287	1,655	11.9
SEER 18	Heating	535	629	-94	-17.5
	Cooling	6,169	4,224	1,945	31.5
	Annual	16,706	14,332	2,374	14.2

Advantages are that this method produces explicit estimates of the cooling and heating related changes for each segment before and after retrofit and differences in HVAC consumption before and after are identified, as well as changes coming from the non-weather responsive baseload. Also, broader savings impacts can be estimated by applying different weather locations to the model.

Disadvantages of this approach include some potential error arising from non-HVAC consumption that is correlated with HVAC use itself and the temperature terms that govern it.[6] This can be reduced somewhat by limiting the analysis to non-pool buildings. The length of the evaluation period may also introduce changes to building energy use where non-program related changes to base load consumption can come from changing ownership and patterns of use of building electronics, home office equipment and other amenities. In spite of these shortcomings, one can expect that using this regression approach will better approximate changes to heating and cooling. It will also estimate changes to baseload that likely point to specific differences between the various rebate control segments or appliance saturation and load growth.

Evaluated Impact on Summer Peak Demand by SEER segment

We also attempted to evaluate summer peak demand impacts of the various SEER segments using two analysis techniques. One was to compare the Control group of 267 homes before and after retrofit on succeeding summer peak days. These were evaluated on the OUC peak summer day of July 28, 2016 before they possessed the new heat pumps and then compared against the same homes evaluated on the summer peak day in 2018 (September 18th). The two-year span between the days compared subjects this evaluation to the same disadvantages mentioned above, e.g. occupancy/behavior change, building modifications and systems degradation. The utility generation plant peaks in both days came at Hour 17. Figure 2 shows the comparison of the various groups pre and post.

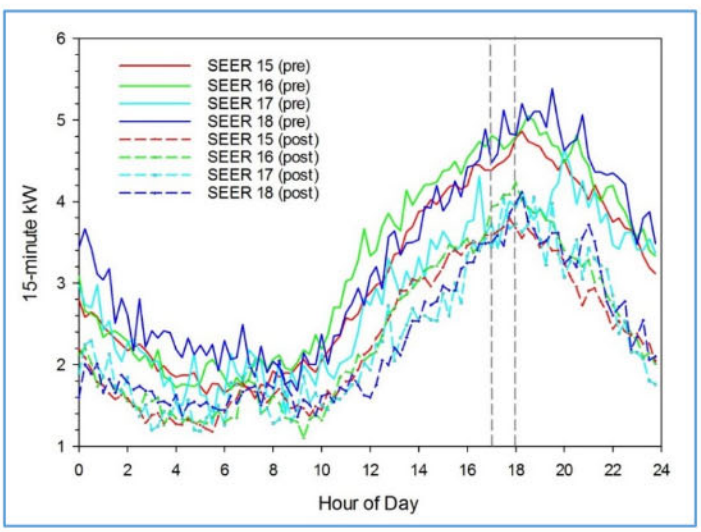

Figure 2. Comparison of 267 homes demand profiles pre retrofit (July 28, 2016) against post retrofit (September 18th, 2018). *Source*: Parker, 2019.

---

[6] For instance, in colder weather more hot water is used and lighting electrical needs are greater. Similarly in summer, pool pump energy can be greater as well as energy use for refrigeration.

Here, we compared the average kW demand of July 28, 2016 with September 18th, 2018 (summer peaks in 2016 [95.0 °F] and 2018 [91.9 °F], see Figure 3) with the averaged AMI data for the groups. Both group segments are evaluated at hour 17 (daylight savings time). Results are shown pre and post for each SEER group in Table 6.

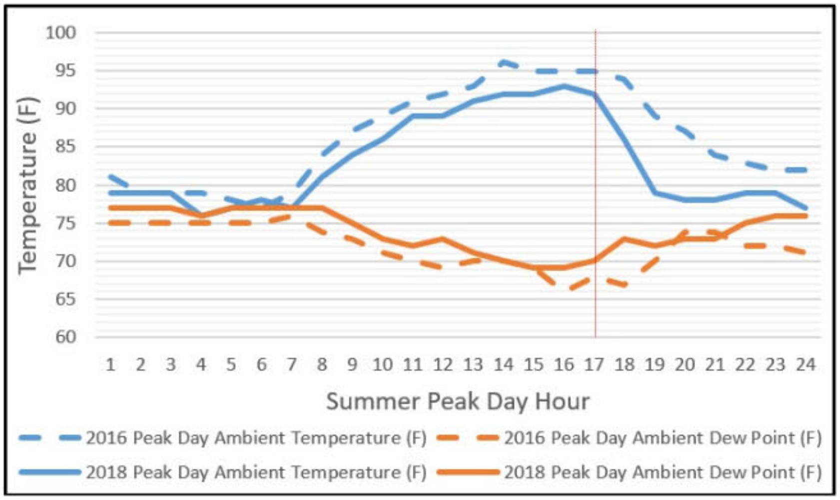

Figure 3. Summer Peak Day Conditions.

Table 6. Comparison of Peak kW between SEER Groups on Summer Day in 2016 vs. 2018

Description	SEER 15	SEER 16	SEER 17	SEER 18
Sample Size	148	70	17	18
Pre kW	4.58	4.73	3.78	4.84
Post kW	3.69	4.09	3.75	3.72
Saved kW	0.88	0.63	0.02	1.12
% Saved kW	19%	13%	0%	23%
kW Saved relative to SEER 15	NA	-0.25	-0.86	0.24

Except for the SEER 17 group, sizable reductions in utility coincident peak are shown for each segment. However, we know that this kind of comparison is subject to the same issues with small samples of less than 40 – and results can be off on a given day with a small number of sites. This means that the results above are likely more valid for the SEER 15 and 16 groups, but not as reliable for the under-sampled SEER 17 and 18 segments.

Given that concern, we used a different method to estimate the peak impacts based on evaluation of a time-temperature matrix which utilizes energy use to estimate demand for each weather sensitive hour (Moezzi et al., 1992). This method uses linear regression of the measured AMI average kWh for the peak hour during the cooling or heating season against the established weather. In this case, the regression is run for each cooling hour at hour 17. The temperature at hour 17 on July 28th, 2016 was 95 °F at Orlando International Airport (MCO). The temperature at hour 17 on September 18th, 2018 was 91.9 °F. Table 7 shows the pre and post kW reductions for the various groups when evaluated at 95 °F. The SEER 20+ group was too small to include.

Similar to the pooled energy evaluation, we discovered the individual account regression method works much more reliably than simply comparing the profiles which is subject to problems with outliers for a few sites, particularly with small samples. Using this analysis method, all the groups show demand reductions of more than 0.65 kW coincident with the OUC system peak. Additionally, the SEER 18 group achieves the largest kW reduction. Also, as with the first analysis method, the demand reductions for the SEER 15 group appears slightly higher than the SEER 16 and 17 groups. However, it is important to keep in mind that the baseline cooling equipment (or duct systems) could be easily skewing this result if the SEER 16 and 17 equipment is going into newer homes with better baseline systems.

Table 7. Comparison of Temperature Adjusted kW between SEER Groups on Summer Peak Day in 2016 vs. 2018

Description	SEER 15	SEER 16	SEER 17	SEER 18
Sample Size	148	70	17	18
Pre kW	4.41	4.63	4.27	4.64
Post kW	3.68	3.95	3.62	3.58
Saved kW	0.73	0.68	0.65	1.06
% Saved kW	17%	15%	15%	28%
kW Saved relative to SEER 15	NA	-0.05	-0.08	0.33

The results of Table 7 show reasonable peak demand savings for all groups. The SEER 18 group consistently showed both large energy savings and demand reductions across the various analysis methods described. Only the peak day data was analyzed for this project, however, it is anticipated that these peak savings occur in other seasons albeit somewhat proportional to building shell temperature difference and system runtime. Similar savings across the SEER groups is not unexpected since higher SEER groups may have selected higher baseline efficiency. The simulation results described in the following section estimates a 0.38 kW reduction in summer coincident peak demand for SEER 18 relative to SEER 15 in Table 8. This AMI meter data analysis shows 0.33 kW, a reasonable agreement.

Simulation model energy analysis

The historical method for predicting rebated program energy savings is the use of building simulation models. This method is preferred due to the reliability and repeatability of results. A simulation model allows for all other building parameters to be held constant while the rebate program performance aspect can be altered as needed to determine savings potential. The limitations can be severe, of course, if assumptions are biased or incorrect.

For this OUC project a simulation analysis was performed to analyze the heat pump rebate program with varying levels of efficiency. As noted in Table 8, the annual energy use for a SEER 14 heat pump, the baseline comparison (Before), remains constant. Greater energy savings are estimated as higher efficiency levels are simulated.

Simulations have the advantages of repeatability and that small differences in any parameter can be estimated as the simulation can maintain all other parameters at a constant level. Sample size or quality of metered data is of no concern and typical meteorological year (TMY) data can be applied for normalization. Typical site characteristics for program participants can be simulated if those characteristics are known.

Table 8. Simulation analysis results for annual energy use and coincident peak demand for 3.5 ton heat pump

SEER	Annual Energy, kWh		Energy Savings %	Demand Savings, kW	
	Before	After		Summer	Winter
15	15,342	15,042	1.95	0.16	0.15
16	15,342	14,787	3.64	0.30	0.26
17	15,342	14,563	5.08	0.43	0.31
18	15,342	14,360	6.40	0.54	0.41
19	15,342	14,227	7.27	0.62	0.43
20+	15,342	13,836	9.82	0.85	0.58

Disadvantages are that this method does not use empirical data to evaluate a program and it will not provide any insight to real demographic differences in retrofit program uptake. For instance, the participants for higher efficiency heat pumps may already be higher energy users, and may already have higher efficiency HVAC equipment prior to replacement. This method also ignores "takeback" where those installing better equipment may choose to maintain lower interior summer temperatures. Also, simulation results may provide little insight beyond what the utility has already done in the past to create most programs.

Discussion of method results

The four AMI data energy analysis methods described in this paper suggest that choosing a weather-based regression approach for individual accounts aggregated into larger segments provided a higher level of confidence and improved accuracy as the chosen method progresses from the *rebate and control group method* with filtering of appropriate accounts based on data quality. [7]

For the first method (current vs future participants) the sample size needs to be very large (>100 for each segment) to account for the variances in the rebate and control groups. Similarly, the pre- and post-analysis of the rebate group also requires a large sample size. Although the before and after pooled regression method may further reduce the sample size if regression statistics show poor correlation with customer energy use, smaller sample sizes are possible since pooled regression against weather eliminates some of the variability in the samples. Finally, the individual account regression method allows characterization of cooling, heating, and baseload energy use which allows disaggregation of space conditioning segments as well as non-weather responsive baseloads.

The simulation results presented here show a trend of increasing savings at increased efficiency levels. The before and after pooled regression method also showed this same trend. However, there is an important distinction between AMI data analysis and simulation models. The AMI data shows actual energy use, including changes in occupancy or occupant behavior, replacement of equipment for more efficient products, elimination or addition of equipment

[7] This paper was written to document the analysis methods used to identify utility rebate program efficacy -- a full year after the analysis was completed. When comparing these analysis methods during the assembly of this paper, the statistical correlation metrics (e.g., variance, confidence interval, etc.) were no longer available and would require a significant effort to recreate. This note is a suggestive reminder to save important information used during an analysis, documenting methods and influences since that information may be needed in the future.

over time, etc. The results of the AMI data analysis shows a specific rebate program's true impact on the generation plant while simulation models focus solely on the specific rebate. The available data and goal of the analysis will help indicate which methods may be most appropriate.

While the objective is to provide the most accurate statistical estimate of utility program influence, examining different ways of data evaluation can provide additional insight as a project is laid out. For instance, the more efficient heat pump rebate segments tend to be concentrated within more affluent and larger homes. This may influence realistic program results and expectations, but may also hide a real influence of the evaluated measure within programs – such as more efficient baseline systems being replaced.

Computerized version of integrated regression method

Evaluation of AMI data for hundreds, if not thousands, of customers would be impossible without automation, as evaluating individual accounts can be very time consuming. The methods described in this paper took one analyst approximately six months to complete, for a single rebate program. The amount of time needed for rebate program analysis can be significantly reduced if the methodology is programmatically applied. A computer program was developed, which replicated the before and after individual account regression method, where each step of the manual evaluation method was included in the computer program. Open source regression routines were added, as were graphical plotting routines. Analysis results were automatically reported to comma separated variable (csv) files for post processing, error checking and reporting.

An example of an automated graphical presentation is provided in Figure 4. As the computer program processed AMI data and calculated regression results, each customer's energy use could be plotted for easy review.

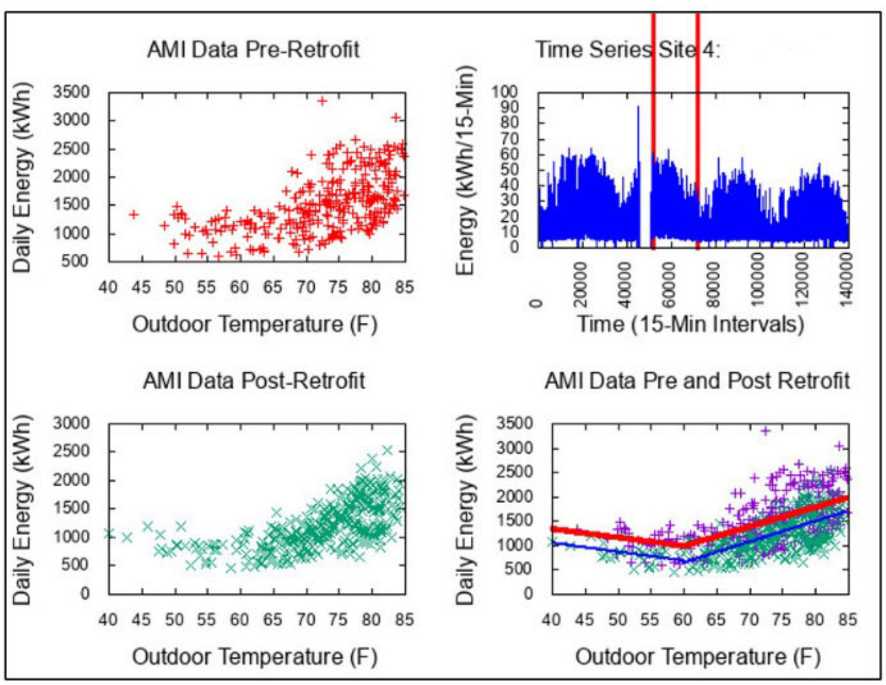

Figure 4. Automated presentation of customer energy use. *Source*: Raustad, 2019.

The left two plots show before and after energy use versus average daily outdoor temperature. The upper right plot shows the time series data with the red vertical lines indicating the times between which the retrofit was installed. Ideally there would be one vertical line representing the install date, otherwise, a window of installation is required. This time series identifies missing data and if energy use has significantly changed over time. The bottom right plot shows pre- and post-data on the same graph and includes the regression results for the cooling and heating season. This data presentation is used to determine if the regressions match the AMI data and whether energy savings are evident. A freely available graphical program (Gnuplot) was used for this project.

The generalized steps in creating a pooled regression analysis program is as follows. The same methodology applies to energy and coincident peak demand analysis.

- Procure property appraiser database for desired territory
- Identify rebate program participants, installation date, and other pertinent information
- Pull AMI meter data, consider data format needed for processing, time stamp required
- Acquire local weather data for period of analysis
- Separate data into pre- and post-records or include an installation date (or range) as input
- Create a customer account input list file to be read by the program that includes customer premise/meter identification and other necessary processing information (e.g., installation data, rebate identifier, size, efficiency, etc.)
- Process either a single customer at a time or read in entire customer data set
- If used, link property appraiser database to customer premise and meter identification
- Process time series data to provide daily average energy use and average daily temperature, energy data is available for 95% of day or discard (100% for coincident peak demand analysis), adjust daily energy use based on availability fraction for that day
- Either at the same time as energy processing or later, save the daily coincident peak hour energy use and local average hourly outdoor temperature, peak data is available for entire hour or discard
- Loop on a balance point temperature to
  - Remove data below/above balance point depending on season and regress (linear and/or non-linear) against corresponding outdoor weather, save coefficients and R-squared values
  - If R-squared is greater than previous, save coefficients and R-squared values
  - Increment balance point temperature and repeat to balance point limit
- Using regression coefficients and balance point temperature, calculate estimated daily energy use using local annual weather file (daily average temperatures), use cooling season coefficients for daily weather temperatures greater than cooling balance point and heating season coefficients for daily weather temperatures less than heating balance point, use cooling season coefficients and cooling balance point temperature when daily weather temperatures fall between or equal to balance points (in cooling dominated climate)
- If used, apply graphical 4-plot reporting for pre-retrofit and post-retrofit daily energy use versus outdoor temperatures, time series raw data versus time, and combine daily pre- and post-data versus outdoor temperatures and regression results (Figure 4)
- Apply regression R-squared value limitations to invalid customer accounts to remove poorly formed regressions either automatically or manually, repeat analysis if performed manually

- Remove customer accounts from input files based on graphical review and repeat analysis
- Sum cooling season, heating season, baseload and total energy use
- Report calculated information to comma separated variable (csv) or other format output file
- Post process results in report quality format

Acknowledgements

This effort would not have been possible without the great support and guidance from our partner staff, Orlando Utility Commission, namely Tom Gross, Peter Westlake, Justin Kramer, Marcus Rozbitsky, and Sheila Rivera. We are grateful to Nick Taylor with University of Florida's Program for Resource Efficient Communities for providing demographic data to OUC.

References

Agnew, K. and M. Goldberg, Whole Building Retrofit with Consumption Data Analysis Evaluation Protocol: The Uniform Methods for Determining Energy Efficiency Savings for Specific Measures, DNV KEMA, NREL/SR-7A30-53827, National Renewable Energy Laboratory, April 2013. https://www1.eere.energy.gov/wip/pdfs/53827-8.pdf

Fenaughty, K. and Parker, D., "Evaluation of Air Conditioning Performance Degradation: Opportunities from Diagnostic Methods," Proceedings of the 2018 Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, American Council for an Energy Efficient Economy, August 2018.

Gnuplot portable graphic utility program, http://www.gnuplot.info/

Haberl, J. S., C. Culp, and D .E. Claridge. 2005. "ASHRAE's Guidelines 14-2002 for Measurement of Energy and Demand Savings: How to Determine What was Really Saved by the Retrofit," Fifth International Conference for Enhanced Building Operations, Pittsburgh, PA: American Society of Heating, Refrigerating, and Air-Conditioning Engineers.

Moezzi, M., Eto, J, Brook, M., Mattson, W. "Improving Utility Peak Demand Forecasts with Measured Data: An Application of PG&E Residential End-Use Data," Proceedings of the 1992 Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, American Council for an Energy Efficient Economy, August 1992.

Parker, D., R. Vieira, K. Fenaughty, R. Raustad and P. Fairey, Evaluation of the Heat Pump Retrofit Program for the Orlando Utilities Commission, FSEC-CR-2089-19, Florida Solar Energy Center, Cocoa, FL, 23 January 2019.

Stata, StataCorp LLC, https://www.stata.com