

Pay for Pandemic? Treatment of COVID Effects in Pay-for-Performance Programs

Matthew Hamlin, Seattle City Light
Mrinalini Sharma, kW Engineering
Colm Otten, Seattle City Light
David Jump, kW Engineering
Michael Hamilton, Seattle City Light

ABSTRACT

Commercial pay-for-performance (P4P) programs, in which incentives are paid based on how much energy is saved, provide many potential advantages beyond the more traditional deemed savings approach. Participants who demonstrate greater energy savings (compared to a normalized baseline) will achieve greater benefits. But a perennial challenge with P4P programs is calculating the baseline from which savings will be measured. This challenge has become even more pronounced since the onset of the COVID-19 pandemic, which has caused substantial decreases in energy use among many types of commercial buildings as public health measures were implemented and building occupancy rates dropped to a fraction of pre-pandemic numbers. How should P4P baselines be adjusted to reflect this impact? For Seattle City Light's Energy Efficiency as a Service (EEaS) program, appropriate adjustments to baselines for several participants were calculated by looking at a broader population of buildings and employing a novel clustering algorithm to segment buildings into "like groups" based only on their consumption profiles. Baselines for participating buildings could then be adjusted based on index variables representing the larger scale energy trends of the similar groups to which each participating building was assigned. Inclusion of these index variables in baseline models has been shown to improve model fit for participating buildings' energy use, thereby providing a more robust and accurate baseline model that accounts for COVID-induced energy impacts. The paper also discusses methods to check and assure the relevance of the index variables over the lifetime of the projects.

Introduction

Energy Efficiency as a Service (EEaS) is a pilot program that can be thought of as a performance-based on-bill repayment program. Customers agree to be billed for the value of energy used, as well as the energy saved in their building. The value of the energy savings is paid to the party responsible for achieving the savings in the participating building through an agreement similar to a traditional power purchase agreement (PPA). To allow Seattle City Light (City Light) to facilitate these transactions, City Light must be able to quantify the savings and savings values in the participating buildings through ongoing, monthly savings analysis. The base methodology chosen for performing this analysis is an IPMVP Option C whole-building analysis which uses regression models to adjust baseline energy use to reporting period conditions. This methodology is well developed and has a successful track record of providing savings quantifications with a higher level of confidence than other methodologies employed by utilities, such as deemed savings and calculated savings.

COVID Impacts on EEaS and Whole Building Programs

One drawback of the regression analysis methodology is that it works best if the only energy changes that a building experiences, are those affected by the independent variables (e.g.: outside air

temperature) and those brought on by the energy conservation measures (ECMs). If there are other factors influencing energy use that are not included in the model, the model will have difficulty predicting what energy use would be under these conditions. In March of 2020, stay-at-home orders were issued in response to the COVID pandemic. This marked the start of a period of unprecedented changes in building use. Most commercial buildings experienced significantly lower occupancy from that point onward, as well as changes to ventilation systems and settings to reduce COVID transmission. The impact of these changes to buildings' energy use was dramatic and unforeseen. No building participating in City Light's whole building programs had relevant variables (occupancy or HVAC trend data, or appropriate stand-ins such as elevator rides or parking data) available to include in the model to track the COVID induced occupancy or operational changes. Even if those variables were available, a regression analysis model requires them to have statistical significance and their value range to have significant representation during the baseline period in order for the model to make accurate predictions of what baseline energy use would have been after the ECMs are installed.

Why This Solution Was Needed

While there sometimes are data that describe occupancy levels in buildings—for example the number of WiFi connections per day or the number of key card swipes per day—that may be used in regression models, it is time consuming to collect from program participants and often overcomplicates program requirements. Because such building-specific data were not available, we turned to non-building specific data. Using non-building specific data in regression models for Option C approaches is among the recommendations in the IPMVP Application Guide on Non-Routine Events and Adjustments (Webster/EVO 2021). Specifically, the Application Guide lists index variables developed from similar buildings as a viable option. This approach has roots in population-based evaluation approaches (Agnew and Goldberg 2017) which are used to determine exogenous effects of target populations. The population approach has traditionally been applied in the residential sector due to the availability of large groups of similar buildings.

Recent work has provided insight on defining populations in the commercial sector (Recurve 2020). This work builds on new analytic approaches for using interval data (i.e., hourly or sub-hourly consumption data) to better understand individual building trends in electricity consumption (Borgeson 2018, Scheer 2017, Scheer 2018). City Light had 15-minute interval data for roughly 550 commercial buildings. We wanted to explore the possibility of leveraging this data to determine the effect of the impacts of COVID on participating buildings. Because COVID was impacting the entire region, it was our hypothesis that it would impact buildings with similar uses in similar ways. We set out to determine if we could find similar operating buildings using electricity consumption alone, and if those groups of buildings could provide an ongoing, statistically significant, index value for how macro-scale exogenous effects (e.g., COVID) are impacting participating buildings. This paper describes the methodology we used to come up with an index value, how we applied it to the regression models used in the EEaS pilot program, and the results of applying the index value to the models and our subsequent ability to determine savings in participating buildings.

METHODS

The research team began with a dataset of hourly consumption data for roughly 1,000 commercial interval meters in roughly 550 buildings extending back to 2019. The intent was to establish groups, or *clusters*, of buildings from this data based only on similarity in their consumption profiles. In

this process we would purposefully ignore other characteristics of the buildings in order to focus only on general patterns in the consumption data. The idea was to construct clusters of sufficiently-similar buildings, and then generate a cluster-specific timeseries index that was representative of each group as a whole. As discussed later in this paper, this index should reflect the summative impact of macro-scale conditions (e.g., COVID-related work-from-home directives, etc.) that are shared in common by all buildings within a given cluster. This index could then be used as an input into the regression models to help account for COVID-related impacts that would otherwise be exceedingly difficult and costly to model.

Given the timeframe of COVID-related public safety measures in Washington State, the research team chose a data window from January 2019 and December 2020, as this covered a period of time before COVID-induced impacts on buildings in Seattle, as well as the first few months of operation during COVID. The research team believed it was important to include data from both before and during COVID because this would allow us to better assign buildings to a cluster which exhibited similar usage characteristics both before and after the pandemic. Additionally, we expected the recovery from the pandemic to last quite a long time, and it was important for the regression modeling that the baseline model training data include the entire range of the index variable.

There were four basic processing steps to prepare the data and carry out the clustering assignment process, described below. The final step was to apply this index to the building-specific regression models for participating buildings.

Meter data needed to be aggregated to a building level. To prepare the interval data for the clustering process it was important to aggregate the consumption to whole building profiles (as many buildings might have multiple meters). Buildings with large data gaps were excluded from the process at this stage, bringing the total down to roughly 500 buildings.

Data needed to be “normalized” so that the clustering process could focus on comparing profile shapes (i.e., without factoring in magnitude). The research team operated on the assumption that the *amount* of consumption was not important in determining similarity of behavior across buildings. Rather the focus should be on trends in how each building’s energy use changed over time *relative to itself*. We thus employed a normalization process which transformed each building’s hourly consumption by subtracting out a baseload value and then forcing the remaining values onto a 0-1 scale. Our data processing steps are shown in Figure 1. where first the hourly meter data for each building is summed up to a building level profile. Next we subtract the baseload (minimum KWH value) for each building and normalize the remaining shape. It is important to remove the baseload before normalizing because building shapes with large baseloads will “flatten” after normalizing if this isn’t done.

The aggregation, baseload adjustment, and normalization steps are illustrated below in Figures 1A, 1B, and 1C for an example building.

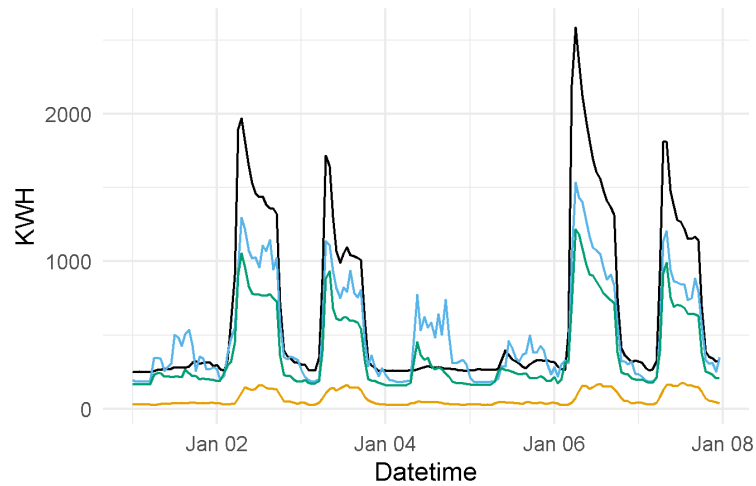


Figure 1A. Hourly electrical consumption for four individual meters in an example building. We show only one week here for illustrative purposes.

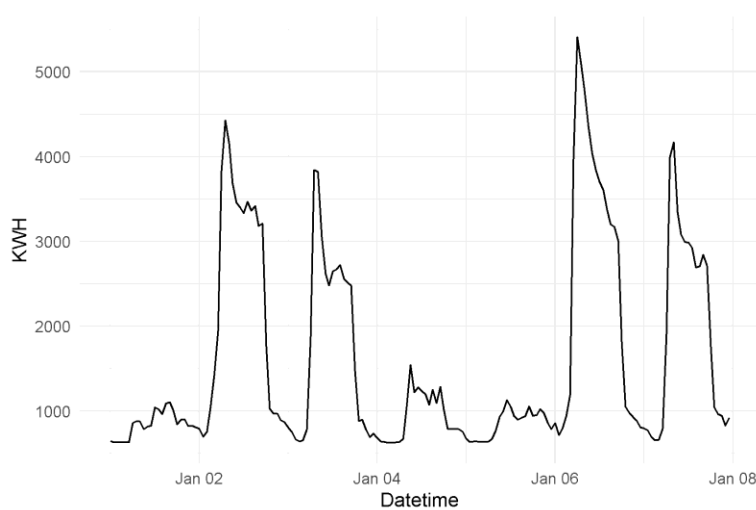


Figure 1B. Hourly electrical consumption for four individual meters in this building after they have been aggregated together and the baseload removed.

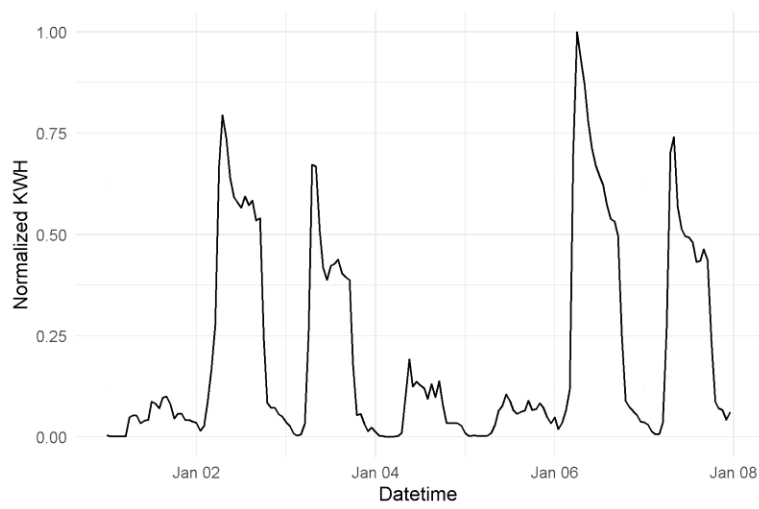


Figure 1C. This same building's hourly KWH profile for one week after normalization.

Run the actual cluster assignment process. After preparing the building-level hourly consumption shapes, there remained approximately 500 buildings to assign to clusters/groups. We employed a clustering algorithm which uses dynamic time warping (DTW) to compare the shapes of two timeseries, and a partition around medoids (PAM) method for defining the cluster's centers.¹ The benefit of using this approach as opposed to a more traditional K-means approach on the Euclidean distance is that the DTW metric allows for some flexibility *in time* as to when a certain type of curvature occurs in a profile. This concept is shown below in Figure 2 where two time series are being compared with Euclidean distance (appear dissimilar) and with the DTW distance (appear similar).

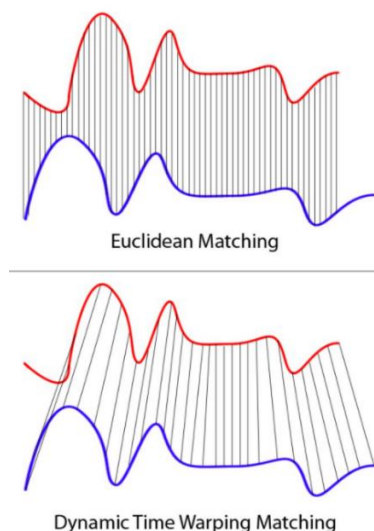


Figure 2. Illustration showing how dynamic time warping allow for increased flexibility when comparing timeseries. The timeseries are identical in both facets, but the DTW algorithm can modify the comparison lines to pick up similarities that may be slightly offset in time. *Source: Databricks.*

¹ For more detailed technical documentation of the DTW algorithm, see Databricks 2019 reference.

The DTW/PAM method can be run on a high-performance desktop machine but can take some time if the data complexity is not reduced. We found that feeding in normalized hourly profiles for several hundred buildings over 24 months was impractical. After some experimentation, we found the best performance by using average hourly weekday profiles for each month for 2019 and 2020 (576 values per building instead of the full 2x8760). Including weekends in these average monthly shapes had little effect on the outcome of these cluster assignments. Determining the optimum number of clusters was partly art and partly science. Metrics produced as part of the clustering process (elbow, silhouette, etc.) suggested that somewhere in the vicinity of 11 clusters was best. We balanced this with the desire to not have any clusters with a very small number (<20) of buildings. Figure 3 shows the general shapes for each of the clusters' normalized load shapes resulting from this process.

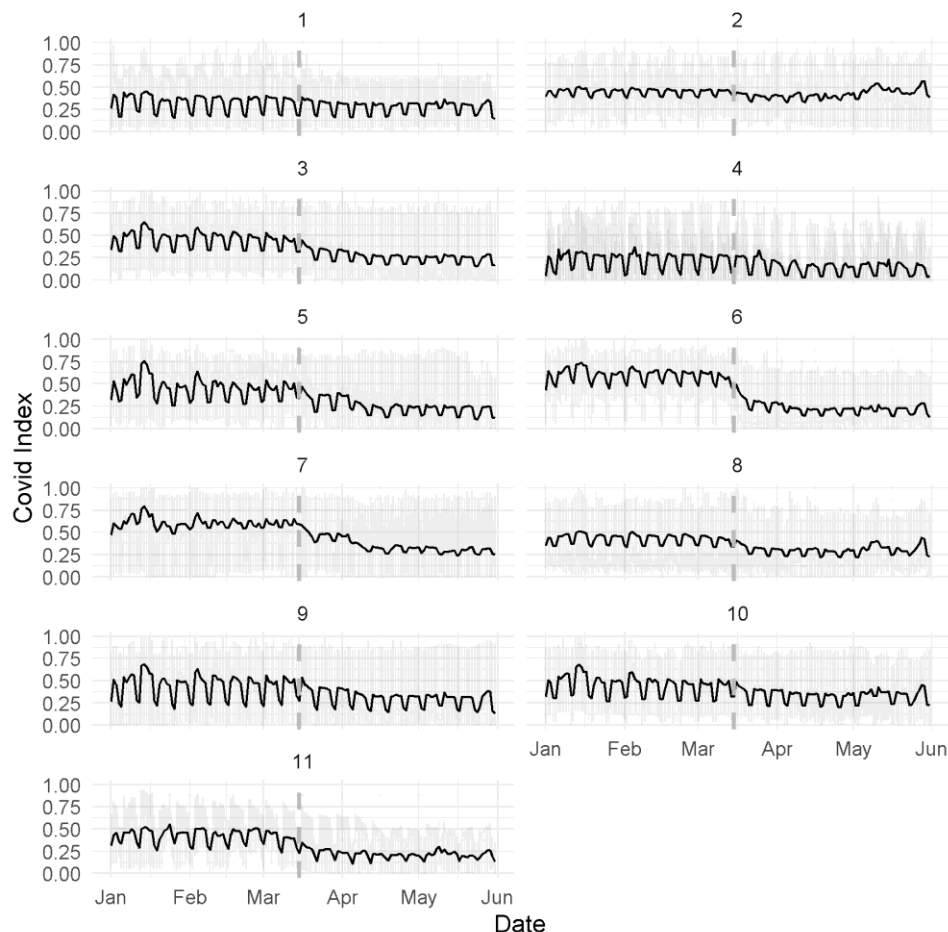


Figure 3. Cluster-level mean normalized KWH values for all buildings in the cluster (numbers above each tile represent that cluster's ID). The vertical dashed lines indicate the onset of COVID-related health precautions in Seattle in March 2020.

Generate a timeseries index value that reflects each clusters' behavior over time. Once all buildings were assigned to a cluster, we created several indices for each: a mean daily value, a 5% trimmed mean daily value, and a median daily value. Each building in the EEaS program was assigned to a cluster, and

thus the index value for that particular building was the cluster-level index. Figure 4 shows the resulting indices for the four clusters which have a participating building assigned to them.

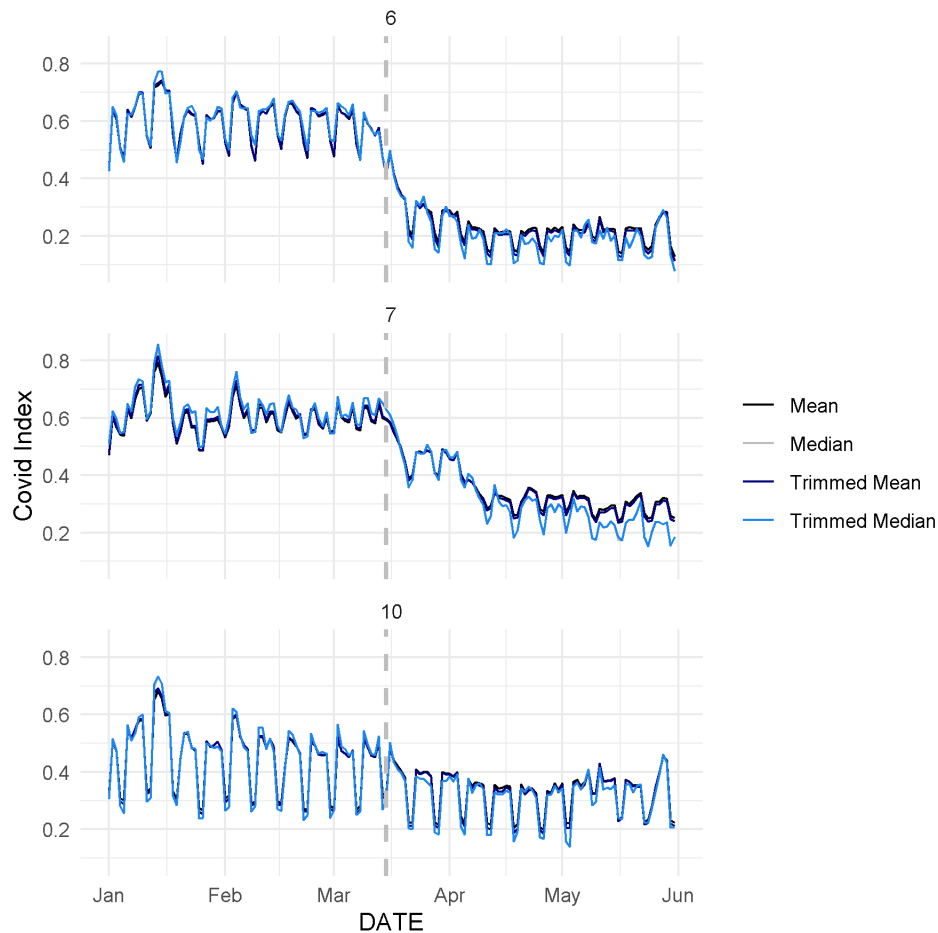


Figure 4. Cluster index values representing each of the three clusters to which a participating building was assigned. Vertical dashed lines indicate the onset of COVID-related health precautions in Seattle in March 2020.

Model building consumption with the index variable. The final step in the process was developing and testing baseline regression models for the program participants. Most common modeling algorithms use a weather variable such as ambient temperature as the single independent variable, as this data is readily available. A modeling method that accepts additional independent variables was required.

RESULTS

We found that inclusion of the cluster index variable improved our ability to model participating buildings' consumption during the program period. In the sections below we discuss the degree to which the index variable improved model results, and then discuss which particular form of the index variable (mean, median, etc.) yielded the greatest improvement.

Results With and Without the Index Variable

The tables below show the goodness-of-fit metrics over the training period and prediction accuracy (as quantified by Net Mean Bias Error (NMBE)) over a testing period for models constructed with and without the six variations of the COVID Index variable. The COVID Index variable was used as an additional independent variable in the time-of-week and temperature (TOWT) model used in this study (see Mathieu et. al., 2011, for TOWT model description, R code with additional independent variables available at <https://github.com/kW-Labs/nmecr>).

We selected the training period to run from January 2019 through September 2020 to include pre- as well as in-COVID conditions in the baseline model. The long duration of the baseline period was intentional to ensure that the baseline models can adapt to changes in building energy consumption as workers come back to the offices and the economy recovers. The test period was selected to be October 2020 through May 2021 and was a date selection was driven by data availability. Tables 1-3 provide results for three program participants. In all cases, the training period ran from January 1st, 2019 through September 30th, 2020, and the test period ran from October 1st, 2020 through June 1st, 2021.

Table 1: Model goodness-of-fit metrics and prediction accuracy using the COVID index for Site 1

Additional Variable	R2	Adjusted R2	CV(RMSE)	NMBE	Prediction NMBE
No Additional Variable	63%	62%	16.58%	0.000%	20.89%
Index Mean	76%	75%	13.25%	0.000%	4.17%
Index Median	77%	76%	13.10%	0.000%	4.23%
Index Standard Deviation	63%	62%	16.48%	0.000%	20.99%
Index Trimmed Mean	76%	75%	13.26%	0.000%	4.22%
Index Trimmed Median	63%	62%	16.47%	0.000%	20.94%

Table 2: Model goodness-of-fit metrics and prediction accuracy using the COVID index for Site 2

Additional Variable	R2	Adjusted R2	CV(RMSE)	NMBE	Prediction NMBE
No Additional Variable	80%	79%	18.54%	0.000%	15.37%
Index Mean	86%	85%	15.68%	0.000%	4.28%
Index Median	86%	85%	15.71%	0.000%	6.16%
Index Standard Deviation	81%	80%	18.19%	0.000%	9.65%
Index Trimmed Mean	86%	85%	15.63%	0.000%	3.75%
Index Trimmed Median	81%	80%	18.11	0.000%	14.08%

Table 3: Model goodness-of-fit metrics and prediction accuracy using the COVID index for Site 3

Additional Variable	R2	Adjusted R2	CV(RMSE)	NMBE	Prediction NMBE
No Additional Variable	85%	85%	10.25%	0.000%	5.63%
Index Mean	91%	91%	8.00%	0.000%	3.83%
Index Median	91%	91%	7.92%	0.000%	3.63%
Index Standard Deviation	86%	86%	9.87%	0.000%	5.37%
Index Trimmed Mean	91%	91%	7.96%	0.000%	3.54%
Index Trimmed Median	87%	87%	9.76%	0.000%	5.09%

Of the five variations of the COVID index considered, three of them (mean, median, and trimmed mean) offered greater improvement in goodness-of-fit as well as prediction accuracy metrics over the model with no additional variables. Of these three, the research team selected the median COVID index variable since this metric is more robust to the effects of outliers.

Figures 5 and 6 below show the improvement in model fit and prediction accuracy for Site 1.



Figure 5: Model fit and prediction accuracy for Site 1 with no additional variables. The dashed line represents onset of COVID-19 related shutdowns.



Figure 6: Model fit and prediction accuracy for Site 1 with the Median COVID Index as the additional variables. The dashed line represents onset of COVID-19 related shutdowns.

CONCLUSIONS/DISCUSSION

The naïve load shape clustering approach described in this paper shows some promise in allowing whole-building energy models to better reflect macro-scale changes in consumption behaviors, such as those induced by (but not limited to) the COVID-19 pandemic. This approach goes beyond using only data collected from participating buildings, thereby allowing population-level trends to be incorporated into the modeling processes. However, the methods presented here also pose methodological quandaries: how good does the clustering need to be in order for a P4P program to calculate financial incentives? And how long into the future will the results of a clustering process remain valid for a given population of buildings? Will we return to a state when the index is simply no longer needed? In theory, if there is no population-level contribution to a building's energy use, this should be reflected by the index variable becoming non-significant in the model. If we constructed a model of building energy use in the future using new data, would the index variable be significant? Such questions require some additional analysis to fully understand.

In this section we first present a few of the lessons learned to date, and then describe a few of the challenges that remain moving forward. We then suggest avenues for future research efforts.

Lesson #1: This approach is relatively simple and inexpensive to implement. We have seen that the approach of naïve clustering of similar buildings using load shapes and then generating an index value for each of these clusters can improve the fit for whole-building modeling efforts. This represents one possible solution to the problem of how to control for macro-level changes such as those induced by the COVID pandemic. Assuming that the consumption data are available to begin with, this approach is particularly attractive in its relative simplicity and cost-effectiveness. The research team did not need to collect any additional data from participants to account for COVID-related impacts.

Lesson #2: Don't make clusters too small. The research team has observed that when a cluster has more than 30-40 buildings, the performance of that cluster is generally fine. Clusters with 50-60 buildings perform even better. But when cluster membership dips below 20 buildings, we begin seeing a greater degree of variation between building profiles within the cluster, a higher silhouette value, and this leads to a correspondingly smaller degree of confidence in the resulting index value.

There are several challenges that are worth highlighting:

Challenge #1: Ensuring cluster stability over time. The initial clustering process was performed at the beginning of the program, where the population of buildings was fixed and buildings were assigned to a cluster based on similarities in their consumption profiles. But as time goes by, it is very likely that buildings within each of these clusters may start to grow “less similar” in their usage patterns. The question becomes “at what point does it make sense to re-evaluate cluster membership?” At present, the research team’s plan is to monitor the silhouette value for each cluster 1-2 times per year. Based on the starting values and any trends that we observe, we will likely need to set some type of threshold below which a decision must be made to reassign members of a certain cluster to other clusters, or else re-run the entire clustering assignment process anew. This second possibility raises important programmatic considerations, as assigning buildings to new clusters will impact the index value for that cluster moving forward. It is possible that the current clusters may remain sufficiently cohesive for a long period of time, but as COVID impacts (hopefully) wane and we establish a “new normal,” it seems likely that a reassignment or re-clustering process will be necessary at some point. Another possible way to address this is to establish a pre-specified “true-up” date when the current clustering assignments are assessed, and if necessary, redone. Doing this for all buildings at a set frequency is cleaner and more transparent, and thus more palatable to program stakeholders.

Challenge #2: Adding new participants to program at a later date can present issues in terms of cluster membership. Similar to Challenge #1 above, adding new participants midway through the program presents an interesting question: should a building’s data be removed from the cluster prior to generating the index value so that the model is not regressing a building on its own data? In practice, the difference between including a building or not when generating the index value is dependent on the size of the cluster. For clusters with greater than 50 buildings, the effect of removing the building in question may be quite small. But for smaller clusters, this is a much more important question, as each building contributes a measurable amount to the final index value.

Challenge #3: Data reporting lags. Not all meters in our sample report on the same frequency, with a small group of meters typically reporting several weeks later than most meters. This has posed problems for the clustering process, as we rely on having most of the data in place in order to generate

the index value for each cluster. Not having certain meters reporting may have particularly pronounced effects on smaller clusters where each building has an outsized influence.

Suggestions for Future Research

Future research may focus on using other data sources to corroborate the results obtained by using the cluster index value and increase confidence in its use. Sometimes data representative of a building's occupancy is available and may be used similarly to the approach used here. Comparing results from the index variable and occupancy variable approaches can provide validation. Some sources of data that may be used to reflect occupancy include heating/cooling operating data from variable air volume (VAV) equipment, as well as card swipe information (for employees required to swipe an ID card to gain building access). Another specific non-building data source that holds promise is mobility data from individuals carrying cell phones. In theory, it may be possible to use this type of mobility data as a proxy for occupancy data for a given building. It would be necessary to obtain this data for a geographic location that is sufficiently close to the building in question.

As the EEAS program proceeds, the research team will continually monitor the robustness of the cluster compositions and resulting index variables. We will look for key metrics that signal clusters must be re-defined, or that the index variable is no longer necessary. To date, the method has shown great promise and a significant step forward for energy efficiency programs that utilize meter-based approaches.

REFERENCES

- Agnew, K., M. Goldberg. 2017. "Chapter 8: Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol, The Uniform Methods Project: Methods for Determining Energy-Efficiency Savings for Specific Measures." National Renewable Energy Laboratory. NREL/SR-7A40-68564. <http://www.nrel.gov/docs/fy17osti/68564.pdf>.
- Borgeson S., A.M. Scheer, and R. Kasman. 2018. "Energy Efficiency Program Targeting: Using AMI Data Analysis to Improve At-the-Meter Savings for Small and Medium Businesses."
- Databricks. 2019. "Understanding Dynamic Time Warping." <https://databricks.com/blog/2019/04/30/understanding-dynamic-time-warping.html>
- Mathieu, J. L., P. N. Price, S. Kiliccote, and M. A. Piette. 2011. "Quantifying Changes in Building Electricity Use, with Applications to Demand Response," IEEE Transactions on Smart Grid, 1949-3053.
- Recurve. 2020, "Comparison Groups for the COVID Era and Beyond." Submitted to the United States Department of Energy. https://gridmeter.recurve.com/uploads/8/6/5/0/8650231/doe_sampling_v2.pdf
- Scheer, A.M., S. Borgeson S., and K. Rosendo. 2017. "Customer Targeting for Residential Energy Efficiency Programs: Enhancing Electricity Savings at the Meter."

- Scheer, A.M., S. Borgeson, R. Kasman et al. (2018) “Customer Targeting via Usage Data Analytics to Enhance Metered Savings.” 2018 ACEEE Summer Study.
- Webster, L. 2021, “Impacts of COVID-19 on Measurement and Verification (M&V) of Energy Savings: Options for Meter-Based Methods – IPMVP and Beyond,” 2021 Efficiency Valuation Organization (EVO). https://evo-world.org/images/corporate_documents/COVID19_in_MV_Final_20210407.pdf.