Data-Driven Examination of the Impact Energy Efficiency has on Demand Response Capabilities in Institutional Buildings

Abigail Andrews
Urban Informatics Lab,
Department of Civil and Environmental Engineering,
Stanford University,
473 Via Ortega,
Stanford, CA 94305
e-mail: andrews9@stanford.edu

Jonathan Roth
Urban Informatics Lab,
Department of Civil and Environmental Engineering,
Stanford University,
473 Via Ortega,
Stanford, CA 94305
e-mail: jonathan.roth@paconsulting.com

Rishee K. Jain[1]
Urban Informatics Lab,
Department of Civil and Environmental Engineering,
Stanford University,
473 Via Ortega,
Stanford, CA 94305
e-mail: rishee.jain@stanford.edu

Johanna L. Mathieu
Department of Electrical Engineering
and Computer Science,
University of Michigan,
Ann Arbor, MI 48109
e-mail: jlmath@umich.edu

Growing concerns over climate change and grid reliability have led to widespread adoption of energy efficiency (EE) and demand response (DR) programs at utilities. Despite such adoption, numerous questions exist regarding the interactions between EE and DR, including whether EE diminishes a building's DR potential. In this brief, we empirically examine the impact a building's EE level (quantified by traditional EE benchmarking metrics) has on its DR capabilities (quantified by a building's normalized load shed) for 194 K-12 institutional school buildings in California, USA. We found inconclusive statistical evidence that a building's EE level has an impact on its DR load shed capabilities. We provide initial evidence countering concerns that EE diminishes DR potential and thus pave the path for future work than can further support synergistic EE and DR strategies which can enhance demand-side management programs. [DOI: 10.1115/1.4054893]

Keywords: energy efficiency, benchmarking, demand response, demand-side management

1. Introduction

Utility programs offering ratepayer benefits for energy efficiency (EE) and demand response (DR) are growing in popularity across the United States and the world [1]. These programs are responding to a pivotal moment in the building energy sector's sustainable energy transition as EE programs assist in lowering absolute electricity use while DR programs help relieve grid stress during times of peak demand [2,3]. Additionally, the decarbonization benefits from EE paired with the grid reliability benefits from DR may have the potential to reshape the electricity industry, especially given the changes associated with increased building electrification and the emerging roles of distributed energy resources [4]. Studying the complex interactions between EE and DR is an emerging and growing research field. Recently, the U.S. Department of Energy established the grid-interactive efficient buildings (GEB) initiative [5] that aims to remake all of the U.S. building stock into clean and flexible energy resources by integrating EE and DR.

Few utilities coordinate EE and DR programs; a 2019 study found that only six utilities in the United States have an integrated EE and DR program for commercial customers [4]. Earlier research on the interactions and integrations between EE and DR is also limited. The lack of utility programs integrating EE and DR paired with the lack of research in the field has led to some previous work recommending against EE and DR integration [6,7]. Countering this recommendation, some emerging research [5,8] has indicated that efficient buildings will respond better to DR signals due to more sophisticated control systems and equipment that allows load flexibility. Moreover, one conceptual qualitative study [9] finds that the relationship between EE and DR could be synergetic when programs target the reduction of passive loads (i.e., the end-use load that is unmodified) and when buildings add (or retro-commission) advanced control systems.

In this brief, we aim to begin to clarify the complex relationship between EE and DR by adopting a data-driven approach and analysis of time-series energy data of 194 K-12 institutional school buildings in California, USA. We empirically examine the relationship between a building's EE level and its ability to shed load during DR events to determine whether a building's EE level has an impact on its DR load shed capabilities, and thus catalyze further research into managing, optimizing, and synergizing EE and DR in the commercial and institutional building stock.

2. Data and Methodology

In this section, we provide an overview of the datasets and methods used to investigate the impact a building's EE level (quantified by traditional EE metrics) has on the ability of a building to shed load during DR events (quantified by a building's normalized load shed).

2.1 Data. Utilizing data available from the California Energy Commission under Proposition 39 [10], we identified 197 school buildings served by PG&E, an investor-owned utility company serving Northern California, USA. The dataset has data on building characteristics, weather, and 15-minute interval meter data. All 197 buildings are enrolled in PG&E's peak day pricing (PDP) DR pricing plan. PDP is a DR plan offered by PG&E in which customers enroll to "get discounted rates throughout the year, except during 9 to 15 "events" (four-hour blocks) when the electric system is strained, and rates are higher. Events typically occur on the hottest days of summer [11]. PDP is a day-ahead notification plan. When the high temperature on the following day is expected to be above 98 °F, customers will receive an alert–text, email, or call–24 hours before the DR event. In 2016, DR events occurred between 2:00 p.m. and 6:00 p.m. Participation in DR events is optional but facility managers are incentivized to participate as the cost of electricity is surcharged per kWh during the event period. Surcharges are determined by the electric rate schedule as

[1]Corresponding author.

Manuscript received September 17, 2021; final manuscript received June 8, 2022; published online July 13, 2022. Assoc. Editor: Ming Qu.

assigned based on electric load size and building type [12,13]. Up to 15 DR events may occur per year as per the PDP plan. In the 2016 cooling season, there were a total of 12 events, all of which were on weekdays [14].

Our analysis aggregates 2016 interval meter data at the hourly level (total hourly kWh consumption), resulting in 24 readings per building per day, between May 1, 2016 and September 30, 2016 (the summer season for the PDP program). As an initial cleaning step, we remove any interval meter reading less than or equal to 0 kWh and any building without 24 complete (i.e., above 0 kWh) readings for each of the 153 data days. There is no imputation of data. Three school buildings were removed due to the cleaning process for a final dataset of 194 buildings.

2.1 Baselining and Measuring Demand Response Capability. The DR baselining method outputs the estimated amount of electricity a building would have used if no DR event had occurred. The amount of electricity a building would have consumed in the absence of a DR event (i.e., the counterfactual) is modeled as follows: $\mathrm{kWh}_{\mathrm{baseline}} = \sum_{\mathrm{HOW}} \alpha_{\mathrm{HOW}} \delta(\mathrm{HOW}) + \sum_{\mathrm{HOD}} \beta_{\mathrm{HOD}} \delta(\mathrm{HOD}) \mathrm{CDH}_{65} + \varepsilon$. Hour-of-week (HOW) is a categorical variable with levels for each of the 7×24 hours in a week. Hour-of-day (HOD) is a categorical variable with levels for each of the 24 hours in a day. The α_{HOW} terms are 7×24 hour-of-week "fixed effects" and the β_{HOD} terms are 24 hour-of-day cooling degree response (CDH_{65}). The baseline regression model was created using data from non-event days between May and September, inclusive, and then used to predict loads given event day HOW and HOD values and temperatures expressed in terms of CDH_{65}, yielding an estimate of what load would have been without a demand response event. An additive same-day adjustment was calculated and added to the baseline load estimate. To determine the change in load (i.e., DR load shed), we find the difference between the actual and expected electricity use over the event. The average impact over all 2016 DR events ($n = 12$) is computed as follows: $\overline{\mathrm{kWh}_{\mathrm{DR\ impact}}} = \frac{\sum_{i=1}^{n} \mathrm{kWh}_{\mathrm{baseline}}}{n} - \frac{\sum_{i=1}^{n} \mathrm{kWh}_{\mathrm{actual}}}{n}$. The DR load shed is presented as a percent of the total reduction of baseline load during the DR event. A positive value indicates that a building successfully reduced load during the DR event and a negative value indicates that a building used more load than expected. For our analysis, we identified outliers as DR load shed values that fall outside four times the interquartile range (IQR) of the variable.

We selected this simple regression model primarily because it is consistent with baselining methods used by program evaluators to identify DR event participants. The method is recognized and understood by efficiency practitioners as a standard practice (when paired with weather normalization) and the method is similar to the baseline that DR event participants would be evaluated against by providers and regulators. Overall, the method is similar to methods with roots in the original Princeton Scorekeeping Method (PriSM) regression models developed in the 1980s [15]. For example, those recommended by International Performance Measure & Verification Protocol (IPMVP) Option C and implemented in the CalTRACK at-the-meter evaluation standard [15–17].

2.3 Energy Efficiency. To measure the impact that EE has on DR, three EE benchmarking methods were applied to the dataset of 194 buildings. We chose to apply two benchmarking methods commonly used in the industry—energy use intensity (EUI) and Energy Star's ordinary least squares (OLS)–and a more statistically robust method recently introduced in the research literature—quantile regression (QR)-based benchmarking [18].

2.3.1 Energy Use Intensity. The calculation for each building's EUI is simple: the ratio of total summer kWh consumption–defined in this paper as the total electricity use between May 1, 2016 and September 30, 2016–to the total square footage. EUI assumes that electricity use only scales linearly with floor area but it is widely known that other factors (such as the building systems, building age, occupancy, and weather) also affect a building's energy performance [19]. We removed outliers four times the dataset's EUI IQR.

2.3.2 Energy Star's Ordinary Least Squares. We calculated Energy Star scores using a modified OLS regression model from the Energy Star K-12 schools technical reference guide [20]. The Energy Star model adjusts for six weather and business activity data features: number of workers, whether the school is open on weekends, whether energy is used for cooking, whether the building is a high school, number of heating and cooling degree days (HDD and CDD), and percent of the building that is heated or cooled. Because of data limitations, our model does not adjust for "whether the school is open on weekends" or "whether energy is used for cooking". Additionally, our model adjusts for the number of students (rather than the number of workers) and does not adjust for HDD because we are modeling EUI during the summer. The dependent variable in the OLS regression is EUI. The regression is fitted to $y_{\mathrm{predicted\ EUI}} = \beta_0 + \beta_1 x_{\mathrm{student\ per\ SF}} + \beta_2 x_{\mathrm{CDD\ floor\ area}} + \beta_3 x_{\mathrm{high\ school}}$. Next, an efficiency ratio is computed as the building's actual EUI ($y_{\mathrm{actual\ EUI}}$) to the predicted EUI ($y_{\mathrm{predicted\ EUI}}$). Finally, the efficiency ratio is compared to Energy Star's Commercial Building Energy Consumption Survey (CBECS) conditional distribution to produce each building's Energy Star score from 0 to 100 [20]. The final Energy Star score is based on the residuals of the fitted model. We removed outliers four times the dataset's Energy Star score.

2.3.3 Quantile Regression Benchmarking. For QR-based benchmarking, multiple models are constructed at each of the quantiles (τ values of 0.01–0.99) across the entire conditional distribution for a response variable–in this case, daily electricity use–as a function of covariates [18]. To maintain the same set of variables across all models at each quantile, a forward variable selection process is used in which each variable is sequentially added to every model at each quantile, and the variable that minimizes the sum of all cost functions across all quantiles is selected. For this study, there are a total of 44 variables in the dataset and we chose to include ten variables in our final QR-based benchmarking model. We found that adding more variables minimally reduces the training error while increasing the risk of overfitting. Benchmarking scores are calculated by first, constructing a suite of models at each τ value between 0.01 and 0.99 (in increments of 0.01), with the ten selected variables at each quantile. Then, identifying which of the suite of models produces a predicted value closest to each observed building energy usage value. And finally, we assign a score based on that model's quantile score = $(1 - \tau)*100$, where a higher score indicates higher levels of efficiency. The output is a benchmarking score for each building for each day from May 1, 2016 to September 30, 2016. The final benchmarking score is the average of these daily scores for each building. We removed outliers four times the dataset's QR-based benchmarking score IQR.

2.3 Assessing Impact on Demand Response Using Energy Efficiency Metrics. Our objective is to assess whether a building's EE level (quantified by traditional EE benchmarking metrics) has a statistically significant negative impact on its ability to shed load during DR events (quantified by a building's normalized load shed) as perceived by prior work [7]. After each of the three EE metrics is applied to the data, we sub-select data to form two groups: (1) buildings that performed in the top quartile (most efficient) of the EE metric and (2) buildings that performed in the bottom quartile (least efficient) the EE metric. Finally, we perform a Welch two-sample t-test to determine the nature of the relationship.

Table 1 Average percent DR load shed for the most and least efficient K-12 school buildings. Negative DR values indicate a building used more electricity than expected during the DR event window (failed to participate). All *p*-values between DR load shed values are greater than 0.05 and therefore non-significant. *p*-values were determined with a Welch two-sample *t*-test

EE metric	EE performance	Average percent DR load shed	*P*-value between DR load shed values
EUI	Most efficient	−0.45%	0.7
	Least efficient	−0.09%	
Energy Star OLS	Most efficient	2.48%	0.2
	Least efficient	−2.83%	
Quantile regression	Most efficient	0.44%	0.6
	Least efficient	0.65%	

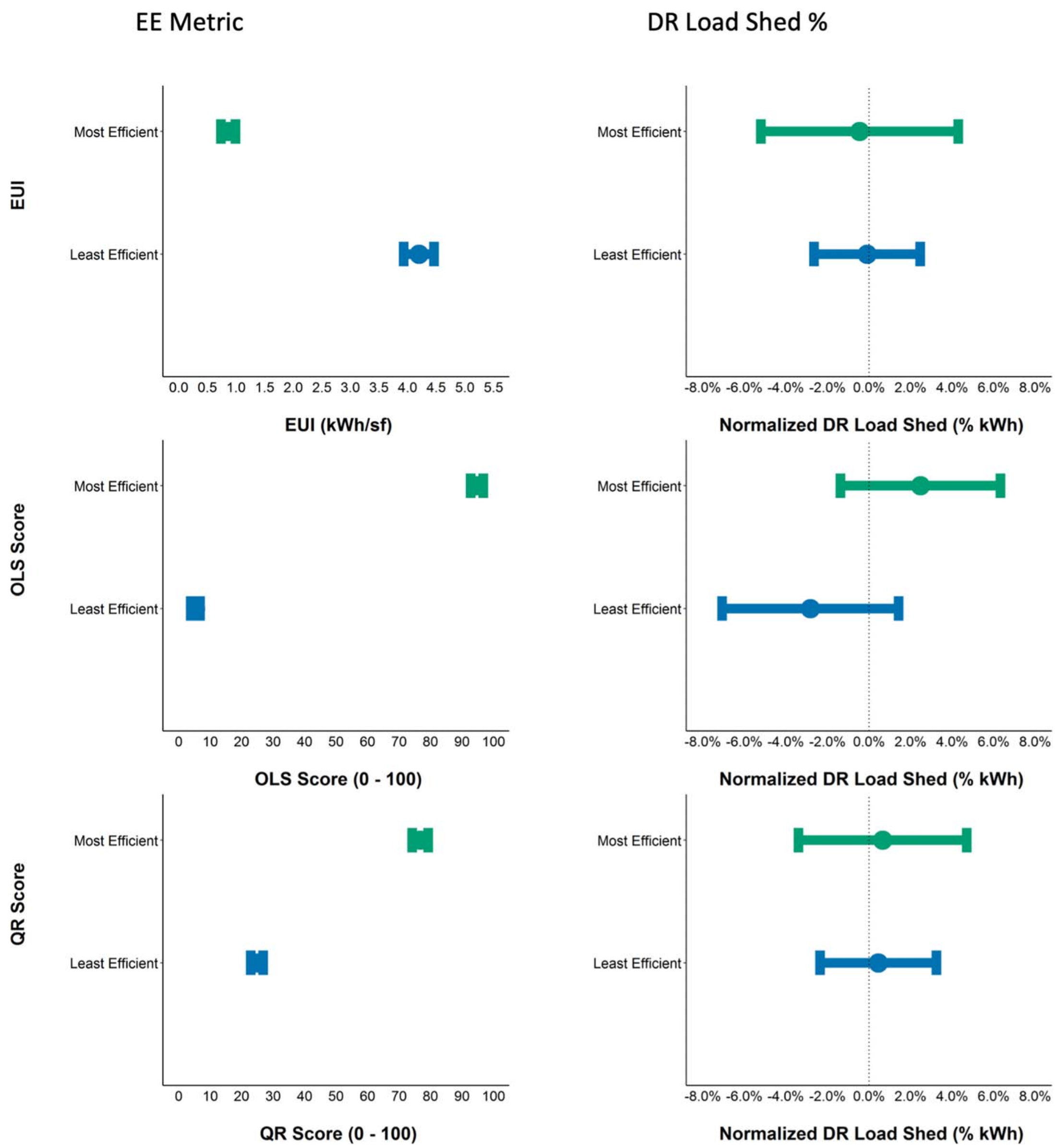

Fig. 1 Mean and confidence intervals for each EE metric and normalized building load shed for the most efficient (75th percentile) and least efficient (25th percentile) buildings. The figure is divided into three subfigures. On the left, we visualize each EE metric, and on the right, we visualize the normalized DR load shed. Results from top-to-bottom are EUI, OLS, and QR. Points represent mean values, and lines represent confidence intervals. Because the normalized DR load shed is often negative (meaning the building uses more electricity than predicted by the baseline model during the DR event) a secondary *y*-axis is placed at 0% on the DR (right side) figure. Means that are sufficiently apart from one another will have a *p*-value < 0.05 and are statistically distinct. Each EE metric mean is statistically distinct; however, the DR load shed means are not distinct.

3. Results and Discussion

Using the cleaned dataset of 194 buildings, we find no positive or negative relationship between any EE metric and DR load shed, meaning that the evidence is inconclusive as to whether or not EE has an impact on DR load shed (p-value > 0.05). This result corroborates earlier conceptual research [9] that indicated no universal relationship exists between EE and DR load shed potential. Additionally, our results fail to demonstrate that EE and DR programs are inherently cannibalistic in nature and therefore provide initial, but inconclusive, evidence contrary to the recommendation that the two not be integrated at the building level [6]. Details of the Welch two-sample t-test by EE metric are provided in Table 1.

We further examine the results and find that each EE metric mean is statistically distinct (left side of Fig. 1) but the associated DR load shed potential is not (right side of Fig. 1). As such, our method successfully isolates the buildings in terms of efficiency across all metrics but we do not observe a corresponding consistent pattern for DR load shed. We do find that the most efficient buildings consistently have a larger spread in terms of DR load shed across all our EE metrics. This indicates that traditional EE metrics are not comprehensively capturing a building's operational flexibility as highlighted in previous work [21].

The inconclusive impact that EE has on a building's DR load shed found in our empirical analysis is consistent with previous conceptual modeling that found no universal relationship exists between EE and DR [9]. EE metrics by design examine relative EE at a moment in time and therefore may be unable to capture the transient nature of DR capabilities. We postulate that this misalignment of timescales likely contributes to the inconclusive findings on the relationship between EE level and DR load shed. Traditional EE metrics and their underlying benchmarking models quantify EE by energy use, building characteristics, and occupant behavior. In all, traditional EE metrics quantify the efficiency of energy use in the building over time. Although energy-efficient buildings use less electricity for the same level-of-service, they may be able to respond equally well to DR signals as energy intensive buildings. Often, energy-efficient buildings make up for less load to shed by having greater control of that load. DR load shed captures the ability to power down when requested. Building systems, characteristics, and demographics used in traditional EE metrics are stagnant and will not be able to capture the transience of flexibility of electric load. Developing metrics that can better quantify both the efficiency and potential for grid-interactivity of a building may serve as a better alternative to existing EE metrics.

4. Limitations, Future Work, and Conclusion

Our analysis utilizes a small sample size of schools which is nonrepresentative of the commercial building stock. We recognize that the analysis would have benefited from a larger sample size and diverse set of building use types. However, de-anonymized interval meter datasets are difficult to obtain and distribute due to privacy issues. To expand on our results, future work aims to utilize larger datasets that include both interval meter data and building characteristics data. To do so, more de-anonymized meter data is necessary. However, we note that this dataset is the largest and most comprehensive dataset to date including detailed building characteristics and hourly interval data for buildings participating in demand response events. As our data only consider K-12 schools our results and metrics cannot be easily generalized to non-institutional commercial buildings. Nonetheless, our results provide initial evidence on the interactions between DR and EE and a basis for future work that considers a wider array of building types (i.e., commercial, industrial, and residential) and test our results on buildings enrolled in various DR programs (both behavioral and automated). Another extension of this work could investigate whether installing EE measures and retrofits would impact the ability of a building to shed more (or less) load during a DR event. This would require a dataset containing information on retrofits and multi-year interval meter data.

As discussed, traditional EE metrics may be unable to capture time-varying energy use patterns that are necessary for understanding DR load shed capabilities. Therefore, there is an opportunity for future work to develop new metrics that can capture efficiency levels at time-of-use. Metrics designed for measuring both operational efficiency and flexibility levels are beneficial as these metrics can also be in line with the emerging concept of GEBs [5].

Transitioning to a cleaner and more reliable grid requires developing a better understanding of the impact EE has on DR (and potentially vice versa). The increasing penetration of nondispatchable renewable energy is making the grid more difficult to operate and creating larger incentives for DR capabilities. Additionally, utilities and policymakers fear that their efforts to increase EE might hamstring their ability to operate these DR programs. In this brief, we find inconclusive evidence that a building's EE level (quantified by traditional EE benchmarking metrics) has an impact on its DR load shed capabilities (quantified by a building's normalized load shed). As a result, this work aims to catalyze future work studying the complex relationship between EE and DR and form an initial basis for an additional research examining the potential of synergistic strategies that can enhance both demand-side management programs and enable GEBs capable of providing demand flexibility to quickly changing power grid.

Acknowledgment

This work was supported by the U.S. Department of Energy Building Technologies Office under the project I-DREEM: Impact of demand response on short and long-term building energy efficiency Metrics, contract No. DE-AC02-76SF00515 and a National Science Foundation Graduate Research Fellowship (Andrews). The authors would also like to thank Dr. Sam Borgeson for his contribution to the development of the data infrastructure and initial DR analysis for this paper.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] Potter, J., Stuart, E., and Cappers, P., 2018, *Barriers and Opportunities to Broader Adoption of Integrated Demand Side Management at Electric Utilities: A Scoping Study*, Lawrence Berkeley National Lab., Berkeley, CA.

[2] Steinberg, D., Bielen, D., Eichman, J., Eurek, K., Logan, J., Mai, T., McMillan, C., Parker, A., Vimmerstedt, L., and Wilson, E., 2017, *Electrification and Decarbonization: Exploring US Energy Use and Greenhouse Gas Emissions in Scenarios With Widespread Electrification and Power Sector Decarbonization*, National Renewable Energy Lab., Golden, CO.

[3] Goulden, M., Bedwell, B., Rennick-Egglestone, S., Rodden, T., and Spence, A., 2014, "Smart Grids, Smart Users? The Role of the User in Demand Side Management," Energy Res. Soc. Sci., **2**, pp. 21–29.

[4] York, D., Relf, G., and Waters, C., 2019, Integrated Energy Efficiency and Demand Response Programs.

[5] Neukomm, M., Nubbe, V., and Fares, R., 2019, "Grid-Interactive Efficient Buildings Technical Report Series: Overview of Research Challenges and Gaps," Report No. NREL/TP-5500-75470; DOE/GO-102019-5227.

[6] Dorahaki, S., Rashidinejad, M., Abdollahi, A., and Mollahassani-pour, M., 2018, "A Novel Two-Stage Structure for Coordination of Energy Efficiency and Demand Response in the Smart Grid Environment," Int. J. Electr. Power Energy Syst., **97**(4), pp. 353–362.

[7] York, D., and Kushler, M., 2005, *Exploring the Relationship Between Demand Response and Energy Efficiency: A Review of Experience and Discussion of Key Issues*, American Council for an Energy-Efficient Economy, Washington, DC.

[8] Kiliccote, S., Piette, M. A., Mathieu, J., and Parrish, K., 2010, *Findings From Seven Years of Field Performance Data for Automated Demand Response in Commercial Buildings*, Lawrence Berkeley National Lab., Berkeley, CA.
[9] Satchwell, A., Cappers, P. A., Deason, J., Forrester, S. P., Frick, N. M., Gerke, B. F., and Piette, M. A., 2020, "A Conceptual Framework to Describe Energy Efficiency and Demand Response Interactions," Energies, **13**(17), p. 4336.
[10] Pepetone, D. S., and Services, L. M. I., 2016, "Collecting Energy Usage Data for California's Proposition 39 K-12 Program: Lessons Learned, Best Practices Defined," pp. 1–16.
[11] Pacific Gas & Electric, 2020, "Business Energy Incentive Programs" [Online], https://www.pge.com/en_US/large-business/save-energy-and-money/energy-management-programs/energy-incentives.page.
[12] Pacific Gas & Electric, 2020, "Electric Rates" [Online], https://www.pge.com/tariffs/electric.shtml.
[13] Pacific Gas & Electric, 2020, Electric Schedule A-1.
[14] Pacific Gas & Electric, 2010, Peak Day Pricing Guide.
[15] Fels, M. F., 1986, "PRISM: An Introduction," Energy Build., **9**(1–2), pp. 5–18.
[16] CalTRACK, 2019, "CalTRACK Methods".
[17] U.S. Department of Energy–Federal Energy Management Program, 2015, "M&V Guidelines: Measurement and Verification for Performance-Based Contracts -Version 4.0," U.S. Department of Energy, **3**(November), pp. 1–108.
[18] Roth, J., and Rajagopal, R., 2018, "Benchmarking Building Energy Efficiency Using Quantile Regression," Energy, **152**, pp. 866–876.
[19] Roth, J., and Jain, R. K., 2018, "Data-Driven, Multi-Metric, and Time-Varying (DMT) Building Energy Benchmarking Using Smart Meter Data," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, IFC Smith, and B Domer, eds., Springer International Publishing, Cham, pp. 568–593.
[20] Energy Star, 2018, Technical Reference: ENERGY STAR Score for K-12 Schools in the United States.
[21] Roth, J., Brown IV, H. A., and Jain, R. K., 2020, "Harnessing Smart Meter Data for a Multitiered Energy Management Performance Indicators (MEMPI) Framework: A Facility Manager Informed Approach," Appl. Energy, **276**, p. 115435.