Nathan J. Addy[1]
2557 Glen Ave.,
Berkeley, CA 94709
e-mail: nathan.addy@gmail.com

Sila Kiliccote
Environmental Energy Technologies Division,
Lawrence Berkeley National Laboratory,
1 Cyclotron Rd., MS 90R1121,
Berkeley, CA 94720
e-mail: skiliccote@lbl.gov

Duncan S. Callaway
Assistant Professor
Energy and Resources Group,
University of California, Berkeley,
310 Barrows Hall,
Berkeley, CA 94720-3050
e-mail: dcal@berkeley.edu

Johanna L. Mathieu[2]
Mem. ASME
Assistant Professor
Department of Electrical Engineering
and Computer Science,
1301 Beal Ave.,
University of Michigan,
Ann Arbor, MI 48109
e-mail: jlmath@umich.edu

How Baseline Model Implementation Choices Affect Demand Response Assessments

The performance of buildings participating in demand response (DR) programs is usually evaluated with baseline models, which predict what electric demand would have been if a DR event had not been called. Different baseline models produce different results. Moreover, modelers implementing the same baseline model often make different model implementation choices producing different results. Using real data from a DR program in CA and a regression-based baseline model, which relates building demand to time of week, outdoor air temperature, and building operational mode, we analyze the effect of model implementation choices on DR shed estimates. Results indicate strong sensitivities to the outdoor air temperature data source and bad data filtration methods, with standard deviations of differences in shed estimates of ≈20–30 kW, and weaker sensitivities to demand/temperature data resolution, data alignment, and methods for determining when buildings are occupied, with standard deviations of differences in shed estimates of ≈2–5 kW. [DOI: 10.1115/1.4028478]

1 Introduction

Through participation in DR programs and electricity markets, electric loads such as commercial buildings are becoming active resources that can help balance supply and demand on the electricity grid [1]. In traditional DR programs, system operators, utility companies, or third-party aggregators achieve system-wide demand reductions by providing financial incentives for buildings to reduce their demand during time periods when the grid is stressed [2]. One way to do this is via dynamic pricing programs, which send time-varying electricity prices to participants; these are also known as price-based programs. Electricity prices are increased when the system is operating near its peak, encouraging building operators to shed (i.e., curtail) demand or shift it to an off-peak time [3]. A second approach to DR, known as incentive-based programs, provides customers with a fixed payment in exchange for allowing DR providers to directly control their loads, for example, by thermostat setpoint adjustments. A central element of evaluating DR program impact is estimation of the size of demand sheds achieved by program participants. These estimations are typically made with "baseline models," which are used to predict what building demand would have been if a DR event had not been called. Baseline predictions are compared with actual measurements of building demand during DR events, giving us estimates of the size of demand sheds. Baseline models are usually developed with historical building demand data. Common models types include averaging models, which average demand on days similar to the day of the DR event, and regression models, which relate building demand to known parameters such as time of day, day of week, outdoor air temperature, and so on. Baseline models are used for a variety of purposes including measurement and verification (M&V), improving DR program design and operation, and, in some cases, financial settlement of DR participation rewards.

There are many examples of baseline models in the energy efficiency literature [4–9] and the DR literature [10–13]. Some of these studies compare the accuracy of predictions produced by different baseline models, e.g., Refs. [10] and [11]. However, shed estimates from the same model can differ if the model is implemented by two different building modelers. This is because specific implementation choices can affect model results. For example, different approaches to interpreting and filtering bad data, different methods for calculating model parameters, and different sources of model inputs can all affect baseline predictions. In turn, different predictions may lead to different evaluations of DR programs.

The goal of this work is to understand which types of baseline model implementation choices have the largest effect on DR performance results, specifically demand shed estimates. We use the linear regression baseline model developed in Ref. [14], which relates time-of-week, outdoor air temperature, and whether or not the building is occupied to building demand. Using 15-min interval whole-building electric demand data from a critical peak pricing (CPP) program in CA, we investigate the effect of five baseline model implementation choices: (1) source of outdoor air temperature data, (2) data resolution, (3) method for determining

[1]N. J. Addy was formally with the Environmental Energy Technologies Division at Lawrence Berkeley National Laboratory, which is where the majority of this work was conducted.

[2]Corresponding author.

Contributed by the Solar Energy Division of ASME for publication in the JOURNAL OF SOLAR ENERGY ENGINEERING: INCLUDING WIND ENERGY AND BUILDING ENERGY CONSERVATION. Manuscript received June 6, 2014; final manuscript received August 22, 2014; published online September 30, 2014. Assoc. Editor: Gregor P. Henze.

when buildings are occupied, (4) method for demand/temperature data alignment, and (5) method for filtering out bad data. This paper is an extension of our preliminary work on this topic [15].

The rest of this paper is organized as follows. Section 2 describes our methods including the baseline model, data, and modeling choices investigated. Section 3 details the results of each comparison, Sec. 4 provides a discussion, and Sec. 5 concludes.

2 Methods

2.1 Baseline Model. We briefly describe the baseline model developed in Ref. [14], which is used in this analysis. The model assumes building demand is a function of time of week and assigns a regression coefficient α_i to each 15-min interval from Monday through Friday, t_i, where $i = 1,\ldots, 480$. Additionally, it assumes that when the building is in occupied mode (e.g., during business hours), building demand is a piecewise linear and continuous function of outdoor air temperature. This choice of function is justified because we would expect that for some range of moderate outdoor air temperatures heating and cooling are minimal, and demand is not a strong function of outdoor air temperature, but as outdoor air temperature increases so do cooling needs and in turn power consumption. When outdoor air temperature is especially high, the cooling system may reach its maximum capacity, at which point demand is no longer a strong function of outdoor air temperature. These effects are more fully described in Ref. [14].

To capture the piecewise linear outdoor air temperature dependency, we divide each observed temperature $T(t_i)$ into a vector with elements $T_{c,j}(t_i)$ with $j = 1,\ldots,6$ associated with six equal sized temperature intervals that cover the full range of observed temperatures. A regression coefficient β_j is associated with each element of the vector. The temperature vector elements are computed with the following algorithm [14]:

(1) Let B_k with $k = 1,\ldots, 5$ be the bounds of the temperature intervals.
(2) If $T \geq B_1$ then $T_{c,1} = B_1$. Otherwise, $T_{c,1} = T$, $T_{c,m} = 0$ for $m = 2,\ldots, 6$, and algorithm is ended.
(3) For $n = 2,\ldots, 4$, if $T \geq B_n$ then $T_{c,n} = B_n – B_{n-1}$. Otherwise, $T_{c,n} = T – B_{n-1}$, $T_{c,m} = 0$ for $m = (n+1),\ldots, 6$, and algorithm is ended.
(4) If $T \geq B_5$ then $T_{c,5} = B_5 – B_4$ and $T_{c,6} = T – B_5$.

For example, if the minimum observed temperature is 5 °C and the maximum is 35 °C, then each bin is 5 °C wide with $B = (10, 15, 20, 25, 30)$. Then, we compute the temperature components associated with each temperature, for example, for $T = 18\,°\mathrm{C}$, we find $T_{c,1} = 10\,°\mathrm{C}$, $T_{c,2} = 5\,°\mathrm{C}$, $T_{c,3} = 3\,°\mathrm{C}$, and the remaining temperature components are 0 °C. Note that $\sum_j T_{c,j}(t_i) = T(t_i)$.

In summary, occupied mode building demand is estimated with

$$\hat{D}_o(t_i, T(t_i)) = \alpha_i + \sum_{j=1}^{6} \beta_j T_{c,j}(t_i) \tag{1}$$

Unoccupied buildings usually respond differently to outdoor air temperature than occupied buildings. Therefore, we use different temperature-related regression coefficients for unoccupied mode. As in Refs. [14] and [16], we chose to use only one temperature coefficient, β_u, in unoccupied mode because DR events when buildings are unoccupied are rare, and therefore, we did not feel the loss of degrees of freedom associated with more temperature coefficients was justified. In summary, unoccupied mode building demand is estimated with

$$\hat{D}_u(t_i, T(t_i)) = \alpha_i + \beta_u T(t_i) \tag{2}$$

Given N temperature/demand observations, Eqs. (1) and (2) can be written in matrix form

$$y = Ax + \varepsilon \tag{3}$$

where $x \in \mathbb{R}^{487}$ is the parameter vector (including all α, all β, and β_u), $y \in \mathbb{R}^N$ is the vector containing the demand observations, $\varepsilon \in \mathbb{R}^N$ is the model error, and $A \in \mathbb{R}^{N\times 487}$ contains (i) binary (i.e., 0/1) indicators corresponding to each time-of week, (ii) occupied mode temperature components, and (iii) unoccupied mode temperatures. We solve for x using an ordinary least squares estimator

$$\hat{x} = (A^T A)^{-1} A^T y \tag{4}$$

In practice, this is calculated using an algorithm from the software package that is used to implement the model.

To make a baseline prediction for a series of time-of-week/temperature pairs of length M, we generate $A \in \mathbb{R}^{M\times 487}$ and solve

$$y_{\text{predict}} = A\hat{x} \tag{5}$$

To estimate the mean demand shed over a DR event, we make a baseline prediction for each 15-min interval within the event, subtract the measured demand, and take the mean over time. We refer to these values simply as "shed estimates" throughout the remainder of this paper. Model error associated with shed estimates was explored in Ref. [16].

2.2 Data. We use 15-min interval whole-building electric demand data from 28 large commercial buildings and industrial facilities in Pacific Gas and Electric Company's (PG&E's) Automated CPP Program between 2007 and 2009. All buildings are located in Central or Northern CA. In the CPP program, on up to 12 days per year, electricity prices were raised to three times the normal price between 12 pm and 3 pm in a "moderate price period" and to five times the normal price between 3 pm and 6 pm in a "high price period." These DR events were announced day-ahead when high peak demand was expected. DR strategies included heating, ventilation, and air conditioning set point adjustments, lighting adjustments, and industrial process shifting.

We build baseline models with temperature and demand data from nonholiday weekdays when no DR event was called. We build separate models for each building in each year (referred to as a "building-year") since buildings change year to year and use only data from May 1 to Sept. 30, which is consistent with the DR season in CA. Types and numbers of buildings and building-years used in the analysis are listed in Table 1.

For each DR event and each building-year, we make separate shed estimates for the moderate and high price periods. Over the three years considered in our analysis, there were 35 DR events (12 in 2007, 11 in 2008, and 12 in 2009); however, we leave out one (Aug. 22, 2007) due to a lack of corresponding temperature data. Additionally, we do not compute shed estimates for the museum on Mondays because it was closed. In sum, we use 1108 shed estimates in our analyses.

Table 1 Types and number of buildings used in the analysis

Type	Number of unique buildings	Number of building-years
Office buildings	13	20
Industrial facilities	8	13
Retail stores	5	12
Prison and jails	1	2
Museums	1	2
Total	28	49

We acquired outdoor air temperature data from two sources: the National Climatic Data Center (NCDC) [17], a division of the National Oceanic and Atmospheric Administration, and weather underground [18], a private website that collects data from Personal Weather Stations (PWS) operated by private individuals and organizations. PWS undergo a one-time calibration but are not guaranteed to be monitored by meteorological experts.

From NCDC, we obtained approximately hourly outdoor air temperature data from the two NWS-USAF-NAVY weather stations closest to each building. From weather underground, we collected 5 - or 15-min interval outdoor air temperature data from the two PWS closest to each building. In many cases, there are multiple PWS within the same range as the closest NWS-USAF-NAVY Station. Additionally, the PWS had fewer missing data points due to a better up-time than the NWS-USAF-NAVY stations. However, we were unable to obtain suitable weather underground data from 2007 and so in our weather data source comparison (Sec. 3.1) we only use 2008 and 2009 data, resulting in 824 shed estimates.

2.3 Modeling Choices Investigated and Comparison Methods. We investigate the effect of five types of modeling implementation choices. We do not attempt to investigate all possible choices but rather focus on several that we expect could be implemented differently by different building modelers. The choices are as follows:

(1) Choice of outdoor air temperature data source: We compare the effect of using heavily curated low resolution data from NCDC versus less curated higher resolution data from weather underground.
(2) Choice of data resolution: Building demand data can be measured and/or stored at different intervals. We compare the effect of using 15 -, 30 -, and 60-min interval data.
(3) Choice of method to determine occupied/unoccupied mode transition times: Buildings usually transition between unoccupied and occupied mode in the morning and between occupied and unoccupied mode in the evening. To use the baseline model presented in Sec. 2.1, we need to estimate these transition times because a building's temperature dependency is modeled differently depending on the mode, according to either Eq. (1) or Eq. (2). We compare the effect of estimating these times manually versus with an automated heuristic algorithm.
(4) Choice of time alignment of outdoor air temperature data with building demand data: Temperature data correspond to specific time instances while electric demand data are collected over time intervals (e.g., 15-min) and then averaged. We compare the effect of aligning the data by time stamp versus offsetting the alignment by 15-min.
(5) Choice of power outage filter: Filtering out bad data and anomalous data (e.g., from power outages) is essential to building a good baseline model. We compare the effect using a simple heuristic power outage filter versus using no filter or a very sensitive filter. We provide more details about each comparison together with the respective results.

For each choice, we compare the results of a "base analysis" to those of a "variant analysis." The base analysis uses NCDC temperature data, 15-min interval building demand data, manual estimation of occupied/unoccupied mode transition times, 15-min offset alignment of temperature and demand data, and the simple heuristic power outage filter. The variant analyses always use the same choices as the base analysis except the choice under investigation.

To compare the effect of different model implementation choices, we calculate shed estimates for the base analysis and each variant analysis and then calculate three statistics on the mismatch defined as $\mathrm{shed}_{\mathrm{variant}} - \mathrm{shed}_{\mathrm{base}}$:

(a) Bias: the absolute value of the mean mismatch,
(b) Std dev.: the standard deviation of the mismatch, and
(c) Max.: the maximum absolute mismatch.

3 Results

3.1 Outdoor Air Temperature Data Source. We first compare the results of the base analysis, which uses NCDC temperature data, to the results of a variant analysis, which uses weather underground temperature data. In general, for each building, we used NCDC temperature data from only the closest weather station. We linearly interpolated the data to generate temperature estimates at each 15-min interval. Temperature estimates computed using interpolants greater than 6 h apart (meaning that more than 6 h worth of data were missing) were filtered out. In cases where exceptionally large amounts of data were missing, temperature vectors were filled in using data from the second closest weather station, as in Ref. [14]. The weather underground temperature data did not require interpolation, though in some cases it did require down sampling (from 5-min to 15-min interval data). In general, for each building, we averaged temperature data from the two closest PWS. However, when the PWS differed in distance to the building by more than 50%, we used data from only the closer PWS.

The results are shown in Fig. 1. We plot shed estimates produced by the variant analysis against those produced by the base analysis. Negative values correspond to decreases in power demand (sheds), while positive numbers correspond to increases in power demand, indicating that a building's DR strategy was not working properly. The closer a point is to the diagonal line, the more the two analyses agree. Points in the lower left quadrant reflect cases in which both analyses agree that there was a shed. Points in the upper right quadrant reflect cases in which both analyses agree that demand increased during a DR event. Points in the upper left and lower right quadrants reflect disagreement between the two analyses. Since all results are based on models, we are unable able to determine which baseline model implementation choice produces better/worse results; we can only compare the outputs of the analyses. Comparison statistics are summarized in Table 2, and a box plot of the mismatch is shown in Fig. 6.

In some cases, the base and variant analyses produce significantly different results. Different temperature data result in different model parameterizations, which can result in different demand predictions and DR shed estimates. Additionally, different measurements of DR event temperatures result in different model predictions. DR shed estimates are particularly sensitive to the model parameters that capture the demand versus temperature relationship when temperature is high. If this relationship is developed

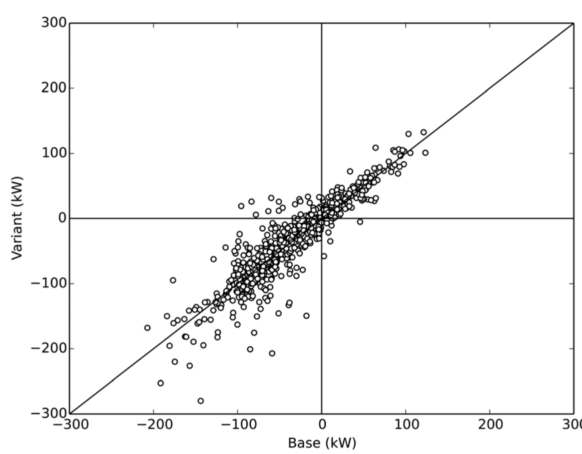

Fig. 1 Effect of outdoor air temperature data source. Base: NCDC data, variant: weather underground data. Std. dev.: 23.36 kW.

Table 2 Base versus variant mismatch statistics

Variant case	Bias (kW)	Std. dev. (kW)	Max. (kW)
Weather underground temperature data	0.39	23.36	148.3
30-min interval data	0.10	2.22	14.0
60-min interval data	0.88	4.59	22.7
Automated occupied/ unoccupied mode detection	0.56	2.39	22.4
Alignment of data	1.42	4.59	52.2
No power outage filter	4.25	30.08	394.2
Sensitive power outage filter	2.51	18.55	151.4

with insufficient data (i.e., the data set used to build the model does not contain many high temperatures), it may not be accurate. In extreme cases, demand prediction requires extrapolation, specifically, predicting demand given temperatures outside of the range of temperatures used to build the model.

3.2 Data Resolution. We next compare the results of the base analysis, which uses 15-min interval data, to the results of variant analyses, which use 30- and 60-min interval data. We computed 30- and 60-min interval data from 15-min interval data by taking the mean over each 30- or 60-min period. Care was taken to ensure that the intervals were day-aligned, meaning that the first interval of the day always represented demand during the interval from midnight to either 00:30 or 01:00. If one or more of the 15-min interval data points is required to compute a 30- or 60-min interval data point were missing, we did not compute that 30- or 60-min interval data point, but rather left that entry blank (i.e., we assume that the data point is missing). This had a minimal effect on the analysis since building-years typically had fewer than 10 h of missing data. The results of the comparison are shown in Fig. 2, Table 2, and Fig. 6.

3.3 Method to Detect Occupied/Unoccupied Mode Transitions. We now compare the results of the base analysis, which uses a manual method to detect transitions between occupied and unoccupied modes, to the results of a variant analysis, which uses an automated heuristic algorithm to detect these transitions. In the manual method, a building modeler inspects a plot of the mean daily demand profile of a building and chooses the times at which the building seems to transition to occupied and unoccupied modes. This was done for each building-year. In contrast, the automated heuristic algorithm first calculates the 2.5th and 97.5th percentiles of the demand, referred to as $D_{2.5}$ and $D_{97.5}$. These percentiles were chosen based on [19], which proposes use of these percentiles to filter outliers before conducting building demand analysis. Then, for each building-year, and each day, the transition time from unoccupied to occupied mode was determined by calculating the first time during the day that the building demand transitioned above $0.1 \times (D_{97.5} - D_{2.5}) + D_{2.5}$. The transition from occupied to unoccupied mode was determined by calculating the final time during the day that the building transitioned below this threshold. We then computed the mean transitions over the year. The results of the comparison are shown in Fig. 3, Table 2, and Fig. 6.

3.4 Data Alignment. We next compare the results of the base analysis, in which temperature and demand data are aligned with a 15-min offset, to the results of a variant analysis, which aligns data by time stamp. In the base analysis, a temperature measurement with time stamp 3:00 pm was aligned with a demand measurement with time stamp 3:15 pm since the demand measurement is actually the average demand between 3:00 pm and 3:15 pm. In contrast, in the variant analysis, a temperature measurement with time stamp 3:00 pm was aligned with a demand measurement with time stamp 3:00 pm, a simpler choice. The results are shown in Fig. 4, Table 2, and Fig. 6.

3.5 Power Outage Filter. Finally, we compare the results of the base analysis, which uses a simple heuristic power outage filter, to the results of variant analyses, which use either no filter or a more sensitive power outage filter. To build a baseline model, we do not use data from days when the minimum power consumption was less than x percent of the average minimum daily power consumption. The base analysis uses $x = 50$, while the variant analyses use $x = 0$ (no filter) and $x = 75$ (sensitive filter). The results are shown in Fig. 5, Table 2, and Fig. 6. Note that the scale of the x- and y-axes in Fig. 5 is different than that in Figs. 1–4 because the maximum differences are larger. The results of the base analysis and variant analysis with no filter look nearly identical except without filtering the results include approximately 20 outliers that all correspond to a single building-year. Inspection of the building-year's time series demand data shows an obvious power outage. In contrast, the results of the base analysis and variant analysis with a sensitive filter are significantly different because the sensitive filter removes significant amounts of data, and so many of the baseline models in the variant analysis are built with less data than those in the base analysis.

4 Discussion

We find that shed estimates are strongly sensitive to the source of the outdoor air temperature data and choice of power outage

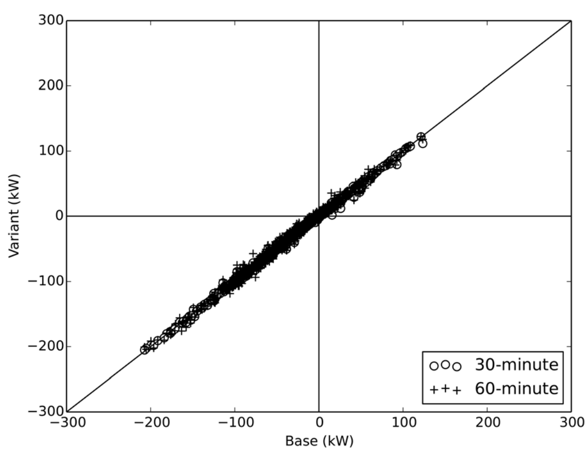

Fig. 2 Effect of data resolution. Base: 15-min interval data, variants: 30- and 60-min data. Std dev.: 2.22 and 4.59 kW.

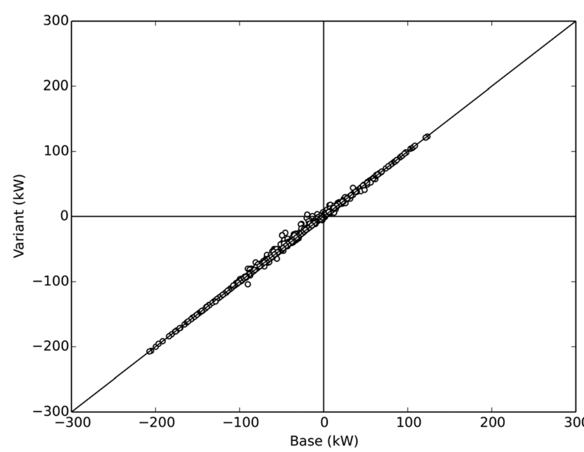

Fig. 3 Effect of method to detect occupied/unoccupied mode transitions. Base: manual selection, variant: automated algorithm. Std dev.: 2.39 kW.

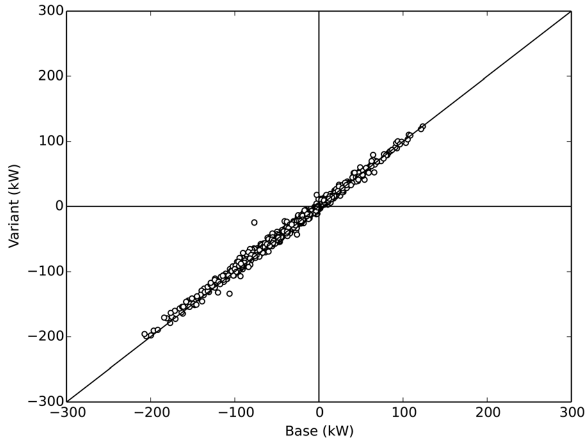

Fig. 4 Effect of data alignment. Base: 15-min offset, variant: no offset. Std dev.: 4.59 kW.

filter. They are significantly less sensitive to data alignment choices (up to a 15-min offset) and the use of 60-min interval data, and almost insensitive to the use of 30-min interval data and the choice of method to detect occupied/unoccupied mode transitions (at least for simple methods explored within this work). Based on these results, we can make a number of recommendations:

(1) Especially in areas with strong microclimates, we recommend either investing in a good source of outdoor air temperature data or acquiring multiple sources of temperature data and running multiple analyses to gain a sense for potential differences in shed estimates.
(2) We recommend investing resources into the development of good algorithms to detect bad data, e.g., power outages. It may be advantageous to measure power outages directly, rather than estimating them.
(3) It may be acceptable to use 30- or 60-min interval demand/temperature data for similar analyses. This could simplify metering needs and reduce computational burden.
(4) We recommend using automated methods to detect occupied/unoccupied mode transitions as we have shown that simple heuristic methods can produce very similar results to manual methods, as shown in Fig. 3.
(5) We recommend aligning demand/temperature data by time stamp since it is simpler than using a 15-min offset and produces similar results.

Using different sources of temperature data or different power outage filters can have significant implications for M&V. For

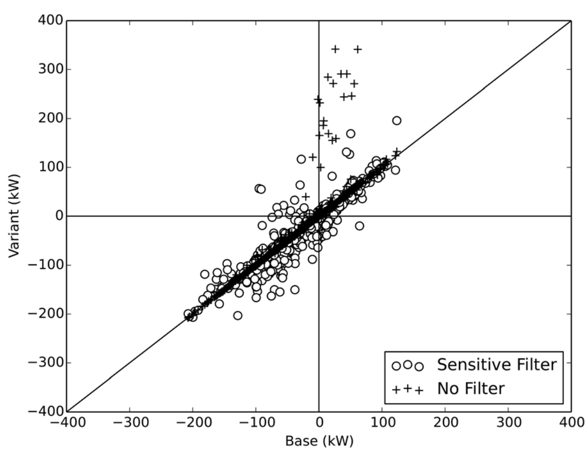

Fig. 5 Effect of power outage filter. Base: original filter, variants: no filter and sensitive filter. Std dev.: 30.08 and 18.55 kW

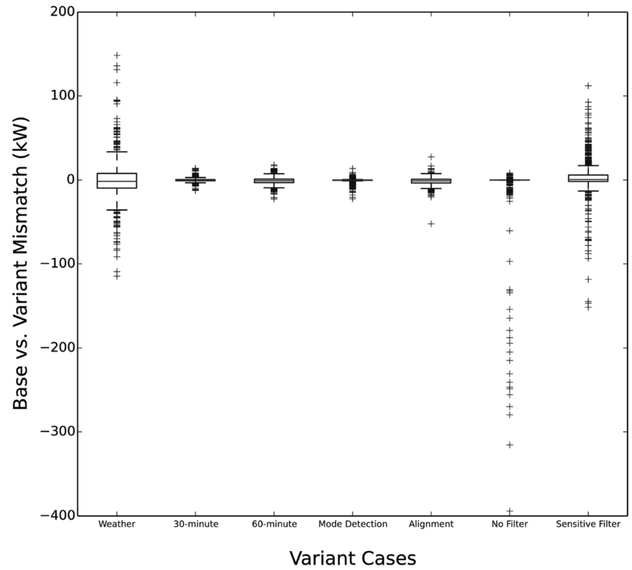

Fig. 6 Box plot of the mismatch in each variant case

example, if one was to compute the money a building saves by participating in a DR program, one could obtain results differing by hundreds of dollars depending upon which baseline model implementation choices are made. For example, assume a building pays 14 ¢/kWh for energy [20]. During PG&E's CPP high price period they would pay five times this price, or 70 ¢/kWh. From Fig. 1, we can find cases in which the two analyses produce shed estimates that are up to 150 kW different. This results in estimated savings differences of up to $150\,\text{kW} \times 3\,\text{h} \times 70$ ¢/kWh = \$315. (The savings are over 3 h because the high price period is 3 h long, as described in Sec. 2.2.) In Fig. 5, we can find shed estimate differences even greater than 150 kW when no power outage filter is used. In DR programs in which buildings are rewarded based upon the size of their demand shed (which is not the case for CPP), these baseline model implementation choices would have a large effect on a building's financial rewards.

In addition to the implementation differences detailed above, we also discovered a number of differences between implementations using different software. We implemented our analyses in both MATLAB and PYTHON and found a number of subtle differences in built-in algorithms, leading to differences in shed estimates. The differences in shed estimates were very small so we do not present detailed results here. However, these differences did make model comparison difficult. Here, we describe the sources of these differences so that other building modelers can learn from our experience.

The first difference was in rounding interpolated values. One implementation would calculate a value of 0.5 and would round up, while another would calculate a value of 0.49999… and round down. While we were not able to discern a significant difference in shed estimates, this issue made it more difficult to compare intermediate results and validate our implementations. We recommend avoiding rounding intermediate results wherever possible. A second difference was the default methods for calculating data statistics. For example, the default variance calculations in MATLAB and PYTHON have different interpretations, as either sample or population variance. It is possible to calculate either value in either software, but care must be taken to use the functions properly. A final difference was in the way each software package treated missing data (i.e., NaNs). MATLAB and PYTHON handle NaNs differently, and in some cases, it is better to filter NaNs out, rather than passing them to built-in functions.

5 Conclusion

We have investigated the effect of different baseline modeling implementation choices on DR performance evaluation. Using the linear regression baseline model in Ref. [14], we explored five different types of baseline modeling implementation choices and found that DR shed estimates are most sensitive to the choice of outdoor air temperature data source and data filtration method. They are less sensitive to data resolution (up to 60-min), data alignment (up to 15-min offset), and the method to detect transitions between occupied and occupied modes.

While it is known that different baseline models can produce different shed estimates [10,11], we have shown that the same baseline model implemented by two different building modelers can produce different shed estimates if the modelers make different baseline model implementation choices. This can result in different interpretations of DR performance, different evaluations of DR programs, and, in DR programs in which financial compensation is based upon the magnitude of shed estimates, different monetary rewards for the participating buildings. Therefore, it is important to quantify the magnitude of possible differences resulting from implementation choices and understand the range of possible interpretations that could result.

Acknowledgment

The authors would like to thank Pacific Gas and Electric Company for providing the electric demand data. This work was sponsored in part by the Demand Response Research Center, which is funded by the California Energy Commission, Public Interest Energy Research (PIER) Program, under Work for Others Contract No. 500-03-026 and by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was also partially funded by Pacific Gas and Electric Company through their Demand Response and Emerging Technologies Programs.

References

[1] Callaway, D., and Hiskens, I., 2011, "Achieving Controllability of Electric Loads," Proc. IEEE, **99**(1), pp. 184–199.
[2] DOE, 2006, "Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them," Department of Energy Report to the U.S. Congress.
[3] Borenstein, S., Jaske, M., and Rosenfeld, A., 2002, "Dynamic Pricing, Advanced Metering and Demand Response in Electricity Markets," Technical Report No. CSEMWP-105, University of California Energy Institute: Center for the Study of Energy Markets.
[4] Fels, M., 1986, "PRISM: An Introduction," Energy Build., **9**(1–2), pp. 5–18.
[5] Katipamula, S., Reddy, T., and Claridge, D., 1998, "Multivariate Regression Modeling," ASME J. Sol. Energy Eng., **120**(3), pp. 177–184.
[6] Kissock, J., Reddy, T., and Claridge, D., 1998, "Ambient-Temperature Regression Analysis for Estimating Retrofit Savings in Commercial Buildings," ASME J. Sol. Energy Eng., **120**(3), pp. 168–176.
[7] Kissock, J., and Eger, C., 2008, "Measuring Industrial Energy Savings," Appl. Energy, **85**(5), pp. 347–361.
[8] Ruch, D., Kissock, J., and Reddy, T., 1999, "Prediction Uncertainty of Linear Building Energy Use Models With Autocorrelated Residuals," ASME J. Sol. Energy Eng., **121**(1), pp. 63–68.
[9] Yang, J., Rivard, H., and Zmeureanu, R., 2005, "On-Line Building Energy Prediction Using Adaptive Artificial Neural Networks," Energy Build., **37**(12), pp. 1250–1259.
[10] Coughlin, K., Piette, M., Goldman, C., and Kiliccote, S., 2009, "Statistical Analysis of Baseline Load Models for Non-Residential Buildings," Energy Build., **41**(4), pp. 374–381.
[11] Goldberg, M., and Agnew, G., 2003, "Protocol Development for Demand Response Calculation–Findings and Recommendations," Technical Report No. CEC 400-02-017F, California Energy Commission (KEMA-XENERGY).
[12] Kozikowski, D., Breidenbaugh, A., and Potter, M., 2006, "The Demand Response Baseline, v.1.75," EnerNOC OPS Publication.
[13] Wi, Y.-M., Kim, J.-H., Joo, S.-K., Park, J.-B., and Oh, J.-C., 2009, "Customer Baseline Load (CBL) Calculation Using Exponential Smoothing Model With Weather Adjustment," Transmission Distribution Conference Exposition: Asia and Pacific, Seoul, Korea, Oct. 26–30, pp. 1–4.
[14] Mathieu, J., Price, P., Kiliccote, S., and Piette, M., 2011, "Quantifying Changes in Building Electricity Use, With Application to Demand Response," IEEE Trans. Smart Grid, **2**(3), pp. 507–518.
[15] Addy, N., Mathieu, J., Kiliccote, S., and Callaway, D., 2013, "Understanding the Effect of Baseline Modeling Implementation Choices on Analysis of Demand Response Performance," ASME Paper No. IMECE2012-86973.
[16] Mathieu, J. L., Callaway, D. S., and Kiliccote, S., 2011, "Variability in Automated Responses of Commercial Buildings and Industrial Facilities to Dynamic Electricity Prices," Energy Build., **43**(12), pp. 3322–3330.
[17] National Oceanic and Atmospheric Administration, 2009, "National Climatic Data Center," Available at: http://www7.ncdc.noaa.gov/CDO/dataproduct
[18] The Weather Channel LLC, 2012, "Weather Underground," Available at: http://weatherunderground.com
[19] Price, P., 2010, "Methods for Quantifying Electric Load Shape and Its Variability," Technical Report No. LBNL-3713E, Lawrence Berkeley National Laboratory.
[20] Pacific Gas and Electric Company, 2010, "General Schedule A-10," Available at: http://www.pge.com/tariffs/tm2/pdf/ELEC_SCHEDS_A-10.pdf