Curtailment Estimation Methods for Demand Response

Lessons learned by comparing apples to oranges

Charalampos Chelmis, Muhammad Rizwan Saeed, Marc Frincu, Viktor K. Prasanna
Ming Hsieh Department of Electrical Engineering Department
University of Southern California
{chelmis, saeedm, frincu, prasanna}@usc.edu

ABSTRACT
Accurate estimation and evaluation of consumption reduction achieved by participants during Demand Response is critical to Smart Grids. We perform an in-depth study of popular estimation methods used to determine the extent of consumption shedding during DR, using a real-world Smart Grid dataset from the University of Southern California campus microgrid. We provide insights to the process of selecting a reasonable baseline with respect to potential misinterpretation of the estimation of electricity consumption reduction during DR.

Categories and Subject Descriptors
H.4 [**Information Systems Applications**]: Miscellaneous

General Terms
Measurement, Performance, Reliability, Verification

Keywords
Baseline Models; Data Analysis; Demand-Response; Load Forecasting

1. INTRODUCTION
Demand Response (DR) is a load management technique which provides a cost-effective alternative to traditional supply-side solutions meant to address demand increase during times of peak electrical load. With the rapid integration of advanced metering infrastructure, Smart Grids enable real-time implementation of dynamic demand programs. In fact, consumption curtailment may be used to: 1) reduce chances of black-outs during peak electricity usage periods, when electrical generation systems may not always meet peak demand requirements, 2) reduce the need for utilities to build and maintain capital-intensive power plants which provide idle capacity to be used in response to high peak demand, which happen just a few times a year, 3) accommodate the increasingly penetration of renewable energy sources, which

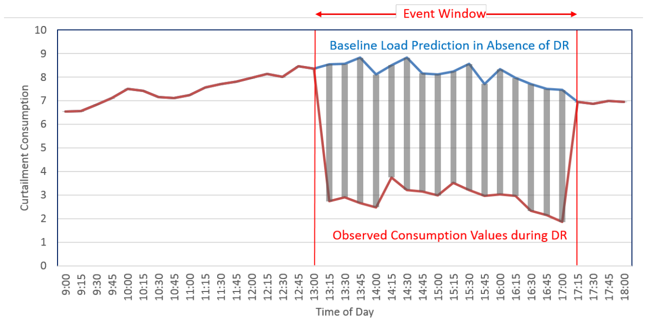


Figure 1: Conceptual diagram of consumption reduction estimation

incur uncertainty at the generation side due to their intermittent and unpredictable characteristics, and 4) allow utilities to proactively maximize their profits, or else minimize their loses, from buying energy to cover a generation-demand mismatch.

Despite the savings possible through DR, the success of such programs essentially hinges upon user participation and their timely response to DR signals [25]. To compensate for inconveniences in customers' activities due to load curtailment, utilities make monetary payments to affected customers over the curtailment period (hereafter referred to as DR period) [17]. Accurate estimation and evaluation of consumption reduction achieved by participants during curtailment is therefore critical to DR programs. One of the main barriers in involving households to participate in DR is the difficulty in automating response to DR signals as direct home appliances management from the utility might not be possible. In cases where DR participation is voluntary [2] estimating the extent of curtailment is even more important for DR programs. To summarize, accurate forecasting and analysis of electricity demand reduction is crucial for two reasons: (i) help consumers understand their curtailment footprints during DR and to receive "appropriate" compensation, and (ii) enhance utilities' ability to perform informed selection of "appropriate" customers for participation in future DR programs.

Typically, baseline models estimate what the consumption would have been in the absence of DR, i.e., Baseline Load Profile (BLP), and compare such counterfactual predictions with observed electric consumption during DR to determine the extent of curtailment. Figure 1 provides a conceptual diagram for estimating consumption shedding as a reult of participating in DR. The amount of computed curtailment (i.e., shaded area) depends on the the accuracy of the baseline model used. As many baseline models exist, such models can produce different curtailment estimates. The problem with calculating BLP model accuracy, lies mainly in the fact that there is no actual reference value to compare against. We argue that without careful consideration, utility providers can end up with erroneous data on the actual curtailment which can in turn lead to billing or rewarding issues.

In this work, we statistically analyze the effect of various BLP models on consumption shed estimation, with the objective of improving the accuracy of estimating electricity demand reduction due to participation in DR programs. Using real-world data from the University of Southern California (USC) microgrid, we show that choosing a good baseline depends on both intrinsic (e.g., DR strategy, day of week) and extrinsic (e.g., temperature, human behavior) factors. To the best of our knowledge, our work is the first to provide an in-depth comparative analysis of the effect of BLP models for post DR analysis in a real-world, large-scale setting.

The rest of the paper is structured as follows. We present related prior work in Section 2. We provide the details of our dataset in Section 3. In Section 4, we study the effectiveness of five popular methods currently in use for post-DR analysis by utilities to determine the extent of curtailment undertaken by consumers participating in DR. We measure estimated curtailment for the various baselines under a real-world DR scenario with realistic energy consumption reduction goals in Section 5. Particularly, we shed light on the effect of baseline selection on the interpretation of consumption reduction as a result of DR programs. We discuss the implications of our findings in Section 6. We summarize our conclusions and future research directions in Section 7.

2. RELATED WORK

Curtailment analysis has recently started to receive attention, as it is an integral component of DR programs [13, 14]. Most related work in this area focuses on energy consumption prediction [13, 14, 28, 15]. Utilities generally use simple averaging techniques for consumption forecasting [3, 21, 27], as such techniques seem to provide good estimates despite of their simplicity. Such models make predictions based on linear combinations of consumption values from recent or "similar" days [3, 21, 27]. To adjust for weather and other conditions on DR days, predictions can be multiplied by a morning adjustment factor [14]. Regression methods [11, 18, 7, 9, 12] and time series [4] approaches have also been proposed. Several methods for residential load prediction were described in [10]. Recently, large scale analysis of automated meter reading (AMR) data with the objective of improving the forecasting accuracy for household electricity demand has been conducted [15, 28]. The main challenge with choosing a method for electricity demand forecasting is there is no clear advantage in selecting one over the other, and their efficiency depends on many factors including: prediction interval, and number of features used for prediction (e.g., temperature, building surface, day of the week/year, type of the day, etc.).

The problem of planning short-term load curtailment in a dense urban area was discussed in [17, 26]. The main assumption of such works was that customers sign up to curtailment programs and always comply when asked to curtail their loads. In our work, we find that estimated consumption reduction varies during DR, as a direct result of the methods used to estimate reduction. For this reason, we assume curtailment to be variable, and venture to study reduction behavior as a function of alternative methods reported in the literature that are used for baseline load estimation during DR. Our analysis could be used as a compliment to the work presented in [17, 26], to drive more informed solutions to the optimization problem of short-term load curtailment planning.

Several studies on alternative methods for calculating the impact of DR programs have been performed [14], emphasizing mostly on methods used by utilities in the US. In fact, [14] argued that averaging methods are well-suited for low load varying buildings, whereas seasonal regression models are better fit for buildings with weather-sensitive loads. Prior work however relied on a relatively small number of calendar days and buildings for their experiments. We instead report our observations based on (i) a much larger selection of proxy event days, (ii) a great variety of building types (including commercial, residential, and non residential buildings) serving a variety of functions (e.g. office buildings, dormitories, and classrooms), and (iii) a large collection of ADR strategies.

3. DATASET

For our experiments, we consider a real-world Smart Grid dataset from the University of Southern California campus microgrid[1]. The dataset comprises of a collection of observed electricity consumption values (measured in kWh at every 15 minutes) from 35 buildings, collected over a one year period (November 2012 - December 2013). The dataset contains a diverse set of building types: academic buildings with teaching and office space, residential dormitories, and administrative buildings. Building names have been obfuscated for privacy issues. The USC Facility Management Services (FMS) used the Open Automated Demand Response (OpenADR) [23] communication specification to send curtailment requests to buildings on the USC campus microgrid. Particularly, the controlled microgrid experiments detailed in this report apply a variety of Fully-Automated Demand Response strategies [24]: Global Zone Temperature Reset (GTR) [20], Variable Frequency Drive Speed Reset (VFD) [20], Equipment Duty Cycling (Duty), and their combinations. Such strategies directly reduce the heating, ventilation, and cooling (HVAC) loads, which make up a significant portion of the overall energy consumption of buildings.

Out of the 35 buildings, 33 participated in at least one DR event during this period. The number of DR events across buildings is not homogeneous. Some buildings participated in more than 40 events in total, while others were rarely selected (less than 10 events). The choice of strategy is also heterogeneous. This results in a total of 380 events for different building-strategy pairs. The distribution of event

[1]The dataset is available upon request for academic use from the USC Facility Management Services (FMS).

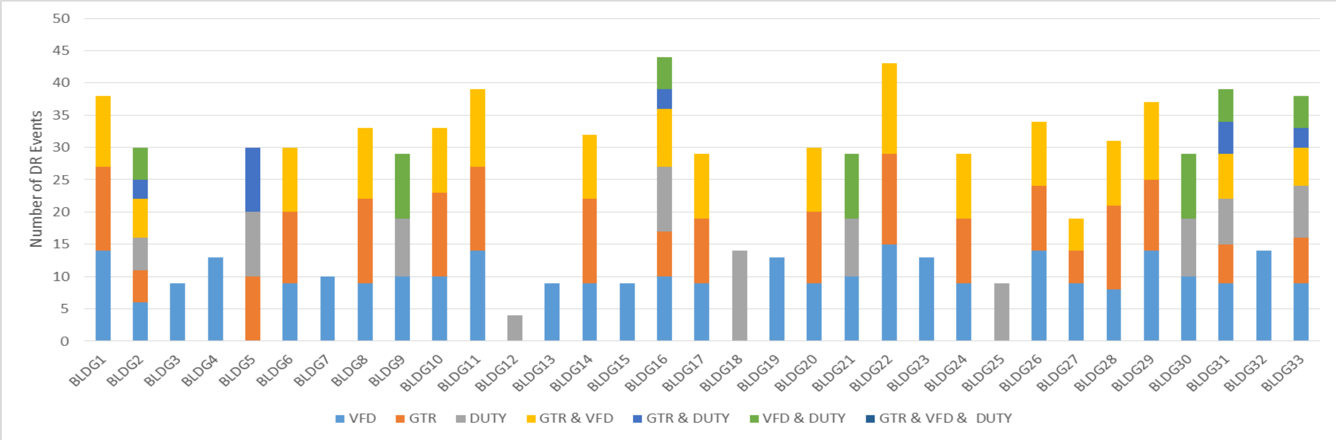


Figure 2: Distribution of DR events across buildings.

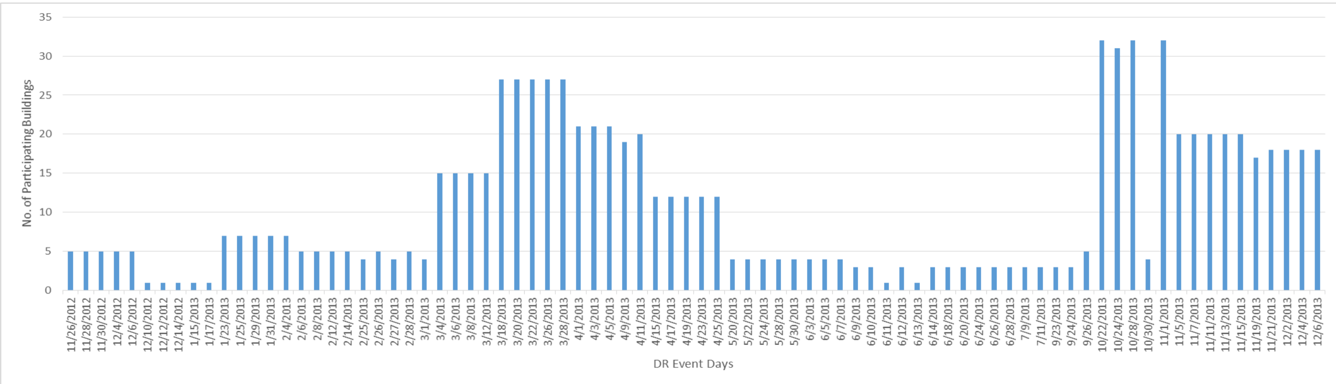


Figure 3: Number of participating buildings in DR events between November 2012 - April 2013.

participations per building is shown in Fig. 2. The total number of buildings participating in a DR event fluctuates over time, as shown in Fig. 3. Experiments were conducted while school was in session, allowing building responses to each strategy to be characterized during standard operation. Due to climate particularities DR events in the microgrid were conducted during the 1:00-5:00PM time frame when demand peaks and temperature is high. Elsewhere [1], we have analyzed the intrinsic properties of this dataset and performed an extensive empirical quantitative evaluation of various very short-term energy consumption prediction methods, showing the correlation between electricity demand forecasting accuracy and factors including weather and customer characteristics.

4. BASELINE SELECTION FOR CONSUMPTION SHEDDING ESTIMATION

Baseline methods are currently used in post-DR analysis to determine the extent of curtailment undertaken by consumers [14]. For an accurate estimation of consumption shedding as a result of participating in DR, baseline accuracy is important. In this section, we evaluate five popular methods used for baseline load profile estimation. We present our analysis in the following subsections.

4.1 Baseline Prediction Methods

Auto Regressive Integrated Moving-Average [8]. The Auto Regressive Integrated Moving-Average model (ARIMA) predicts future electricity consumption values based on a linear combination of previous, equally spaced univariate time series data. Its advantage lies in the fact that it is simple to use, and that it does not require knowledge of the underlying domain. However, parameter estimation for ARIMA requires human expertise to examine the partial correlogram of the time series. ARIMA has been used to forecast real world time series data such as stock [22] or fuel prices [19], as well as electricity load [6]. In our experiments, ARIMA is trained on a two months window (60 days) of preceding data sampled at 15 minute intervals, to make predictions for the following 16 values during DR, i.e., 1:15-5:00PM, at 15 minute granularity.

New York ISO [21]. The New York ISO (NYISO) is calculated from previous five days with the highest average kWh value. These days are chosen from a pool of ten previous days, which are selected starting two days prior to the event day, and excluding weekends, holidays, past DR event days or days on which there was a sharp drop in the energy consumption. In addition, a day is included in the pool only if the average consumption on that day is more than 25% of the last selected day. The process repeats until all ten days have been placed in the pool of days for baseline calculation.

Days are then ranked based on average hourly consumption and five days with the highest value are selected. Finally, the baseline is calculated by taking hourly averages across these days. For baseline calculation on a DR event day, a morning adjustment factor can also be calculated from the two hour values prior to DR event by comparing calculated baseline consumption and actual measured data.

Southern California Edison ISO [27]. The Southern California Edison ISO model (CASCE) estimates baseline consumption by averaging past ten days. These days cannot include weekends, holidays or past DR event days. Once ten days have been selected, the baseline is calculated as their hourly average. similar to NYISO, a morning adjustment factor is applied to the calculated baseline.

California ISO [3]. According to California ISO model (CAISO), the baseline is the hourly average of three days with the highest average consumption value among a pool of ten selected previous days. Selected days cannot be weekends, holidays, past DR event days. CAISO's performance can be considerably improved by introducing a morning adjustment factor [14]. We denote this modified version of CAISO as **CAISOm**. In our experiments we consider both versions.

Fixed Value: The simplest way to determine baseline consumption during DR is to measure the consumption value just prior to the beginning of the DR event and use this value throughout the DR event window. We expect this method to provide good results when consumption exhibits low variability. When this assumption is invalid, using a fixed value may result in a poor approximation of the baseline consumption.

4.2 Evaluation Metrics

To evaluate the ability of baseline models to estimate what the actual consumption would have been in the absence of DR, we measure average deviance between predicted consumption, fc_t^{15}, and actual consumption, ac_t^{15}, sampled at a 15 minute granularity, as follows:

$$MPE = \frac{100}{n} \sum_{t=1}^{n} \frac{fc_t^{15} - ac_t^{15}}{ac_t^{15}}, \tag{1}$$

where n represents the interval (here $n = 16$) over which the average is computed. When prediction is perfect, this metric is zero, however its upper limit is unbounded.

To measure model bias, we measure the median of the distribution of errors. Intuitively, the closest to zero the median of the error is, the more unbiased the model. Conversely, a positive (negative) median indicates that the model has a tendency to over (under) predict, i.e., predict values larger (smaller) than the actual values. The median is selected instead of average because of its insensitivity to outliers. In our evaluation, we forecast consumption on non-DR days (between 1-5pm, for consistency with DR days), for which the actual consumption is known. We then calculate the percent error median and the mean absolute percent error (MAPE). When DR days are considered, consumption in the absence of DR is unknown. Instead, we calculate baseline consumption (b_t^{15}) as a proxy of ac_t^{15}, and use Equation 1 to measure deviance between the baseline prediction and observed curtailed consumption, sampled at a 15 minute granularity.

4.3 Baseline Analysis

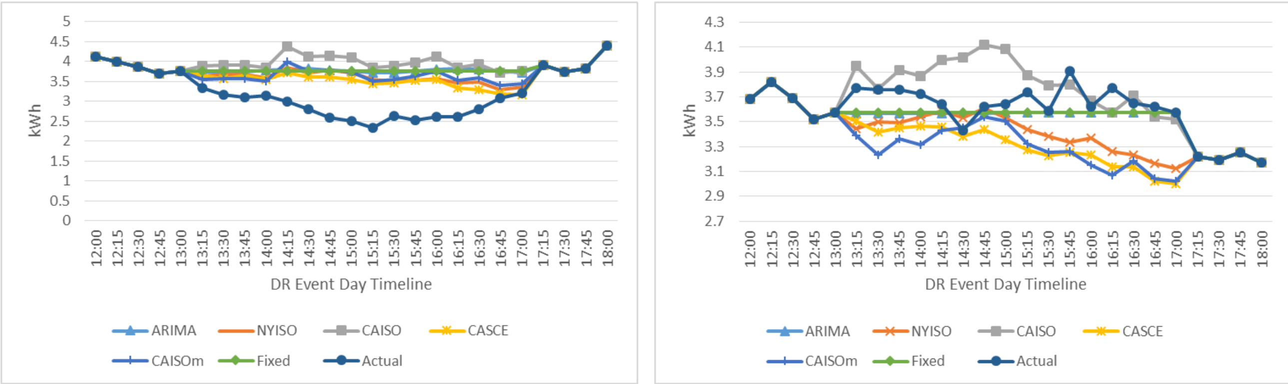
Fig. 4a shows a sample profile of observed electric consumption during DR for a building in our dataset, overlaid with estimated baseline consumption in the absence of DR (see Sect. 4.1). The difference between baseline predictions and actual consumption is positive in all cases, indicating that consumption shedding was achieved during the DR event. Fig. 4b demonstrates a more complicated scenario. In this case, observed consumption during the DR period does not exhibit a sharp drop as in Fig. 4a. Instead, it fluctuates during the DR event window, and exhibits a decrease afterwards. This leads to negative kWh difference between most baseline predictions and actual consumption.

Our dataset contains on-campus buildings which participate in Automated DR, i.e., building equipment is automatically controlled by a central control system based on predefined strategies. In this context, negative curtailment, i.e., higher consumption than the baseline prediction during DR, is not to be expected. We assume such an effect to be the result of equipment failure or control measures taken to ensure that temperature will not exceed comfort limits for humans (e.g., HVAC units turn on to cool down a building). More importantly, bad baseline consumption assessment can result in erroneous measurements (cf. Fig. 4b).

Contrary to consumption prediction for non-DR days where we can directly compare predicted values with observed values, estimates of what consumption would have been in the absence of DR are difficult to verify in the presence of DR. Hence, a crucial question arises: "*which baseline gives the closest approximation of consumption in the absence of DR?*", as improper baseline selection might lead to misleading results interpretation. For example, CAISO estimates electricity consumption shedding (cf. Fig. 4b) in contrast to the rest of the methods which result in negative kWh difference between predicted consumption in the absence of DR and observed consumption. We approach this research question through an empirical study. Our statistical analysis is based on consumption shedding calculations from profiles such as these shown in Fig. 4) for all buildings, all strategies, and all baselines. We examine the performance of a baseline in terms of bias, i.e., dominance of positive or negative predictions, and accuracy, i.e., average absolute percent error.

We begin by selecting a number of non-DR days and estimate baseline consumption between 1-5 PM window. We compare the results to the actual consumption values to calculate the error (kWh difference and percent kWh difference). For this, we have selected a total number of approximately 350 days per building from instructional days in Spring 2012, Fall 2012, Spring 2013, Summer 2013, and Fall 2013.

We argued in Sect. 4.2 that a reasonable model should have zero median and exhibit minimum variance. Fig. 5 shows that among all baseline prediction methods that we consider in this work, CASCE performs best, consistently achieving



(a) Positive curtailment (b) Negative curtailment

Figure 4: Dependence of estimated electricity consumption shedding on accurate baseline consumption forecasting.

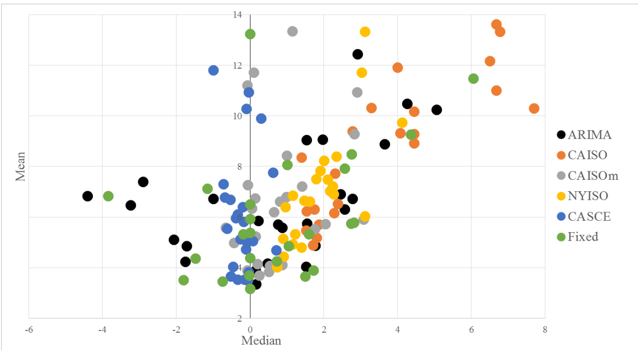


Figure 5: Baselines comparison.

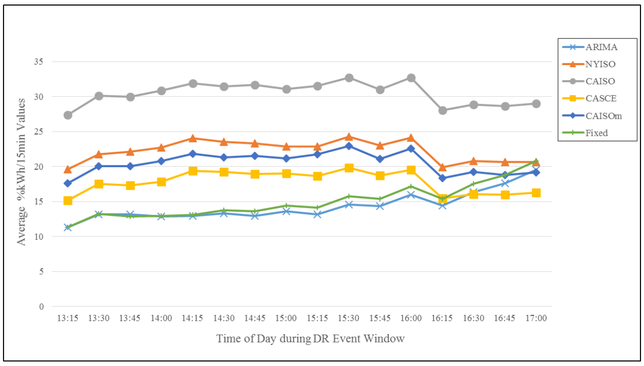


Figure 6: Comparison of average curtailment achieved by using different baselines

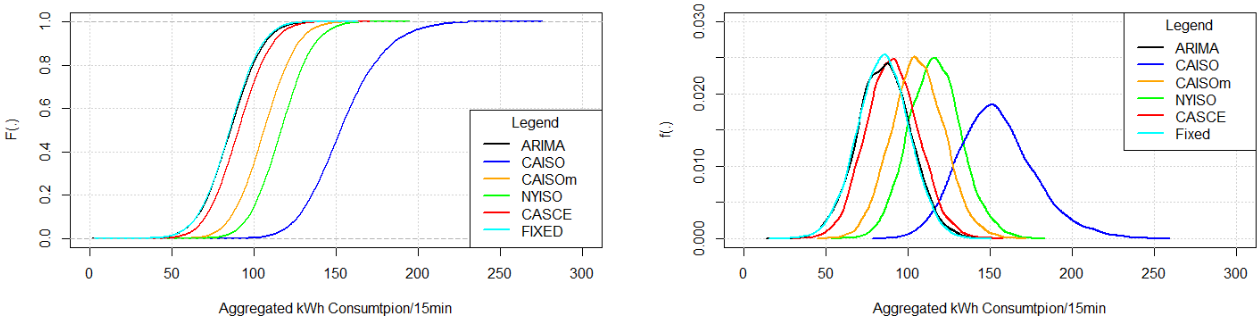
good MAPE values while at the same time being the least biased. ARIMA and CASCEISO constitute conservative predictors of baseline consumption, i.e., they are closest to the ideal behavior of equally over/under estimating the baseline load. Instead, we found CAISO to consistently over predict baseline load profile. Specifically, CASCE has the smallest MAPE (7.4%) and bias (-0.18). Fixed performs surprisingly well, outperforming CAISO and NYISO. Further, CAISOm, our modified version of CAISO, exhibits small variance but high error values in some cases. Finally, ARIMA achieves small MAPE values on average, however it exhibits great variance. We conjecture that CASCE is the best baseline prediction method. In what follows, we use CASCE for baseline consumption prediction during DR.

Finally, we examine the average estimated consumption shedding (cf. Fig. 1) calculated by the various baseline methods. We compute the total values by aggregating results across all buildings for all DR events in our dataset. Fig. 6 summarizes the results. Since there is no actual total consumption reduction to compare against, we can only speculate about the results. Fig. 6 suggests that on average, CAISO predicts the highest reduction, whereas Fixed and ARIMA estimate minimal consumption shedding. CASCE prediction is lower than the rest ISO methods, yet higher than Fixed and ARIMA. Interestingly, estimated reduction increases slightly for most baselines up to one hour before the end of the DR window, where a sharp drop is evident. Fixed and ARIMA do not exhibit this trend. We assume such an effect to be a result of control measures been taken to ensure that temperature does not exceed occupants' comfort limits.

5. ACHIEVING A CURTAILMENT GOAL

While the analysis in Section 4.3 can help establish a general understanding of the "behavior" of popular baseline models, it does not provide clear supporting evidence in support of one model over the others. To shed light on the effect of baseline selection on the interpretation of consumption reduction estimation and evaluation due to DR programs, we consider DR events in which all buildings participate, each following a random strategy (cf. Section 3). We measure campus wide estimated curtailment for the various baselines described in Section 4.1 (cf. Fig. 1 in Section 1) and report our findings.

The probability density function and the cummulative density function of "achieved" curtailment values are presented in Figure 7a and Figure 7b respectively. Table 1 summarizes the results. The last column in the table shows the probability of achieving a certain curtailment target. Clearly, the probability of successfully reaching the target of 500kWh/hr which we have set depends on the baseline which is used to analyze past DR events. As expected, when CAISO, which



(a) Probability density function

(b) Cummulative density function

Figure 7: Aggregate curtailment over all buildings

Table 1: CDF Analysis of Different ISO models with the random strategy selection

Baseline	Min	Median	Max	P(c=500kWh/hr)
CAISO	84.15	152.40	256.90	90.78%
CAISOm	55.88	105.80	165.80	12.11%
NYISO	57.11	116.80	181.50	30.72%
CASEISO	32.53	89.77	149.80	1.6%
ARIMA	21.28	85.99	141.40	0.91%
Fixed	27.71	8522	139.40	0.41%

Table 2: CDF Analysis of Different ISO models with the best strategy selection

Baseline	Min	Median	Max	P(c=500kWh/hr)
CAISO	134.60	191.30	265.00	100%
CAISOm	95.08	138.60	186.40	85.56%
NYISO	100.30	147.60	191.50	96.16%
CASEISO	78.15	121.50	167.60	39.34%
ARIMA	41.03	106.10	155.20	9.6%
Fixed	48.56	103.10	149.00	6.26%

we argued in Section 4.3 that overestimates the achieved curtailment, is used, the target curtailment of 500kWh/hr is achieved with high probability (90.78%). Instead, when a "conservative" baseline, such as Fixed, or a "pessimistic" baseline such as ARIMA, is used, the target curtailment of 500kWh/hr is almost impossible to achieve (0.41% and 0.91% accordingly).

In order to ensure that randomly selecting a strategy for each building does not affect our findings, we repeat the experiment mentioned above, with the difference that we now use the "best" strategy per building. We determine the best strategy per building by examining the outcome of historical DR events as detailed in the following paragraph. Table 2 summarizes the results. Even though the probability of achieving the curtailment target is higher (as compared to the values reported in Table 1) for all baselines, our findings are consistent in both cases.

Next, we evaluate the correlation between the ADR strategy selected and a building's ability to reduce electricity consumption. As a side effect, the "best" strategy per building can be identified. We begin by computing the %kWh difference achieved by different strategies for each building. One such comparison is can be seen in Figure 8 for three buildings and two baselines. Figure 8 suggests that different buildings respond to different strategies differently. For example, according to Figure 8, VFD is the best strategy for buildings BLDG16 and BLDG17, and the combination

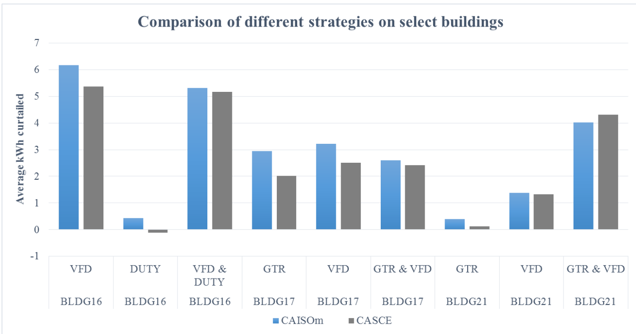


Figure 8: Strategy comparison for sample buildings

of GTR and VFD is the best for building BLDG21. As a result, we can impose a per building ranking of strategies according to expected kWh curtailment. Once again, baseline appropriateness becomes crucial, especially for ADR policy optimization where a subset of buildings needs to be automatically determined for participation in order to achieve a goal determined by the utility [16]. Specifically, a distribution based on historical observations of consumption reduction can be constructed for each building-strategy pair. Doing so, the probability of curtailing a specific amount by selecting any given building-strategy pair can be computed. An example is shown for building BLDG26 in Figure 9. The distribution appears as a discrete step function due to the limited number of DR events per building in our dataset.

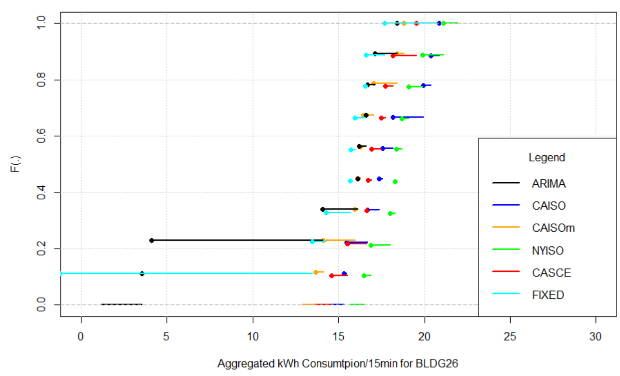


Figure 9: Distribution of kWh amount reduction values for BLDG26

According to NYISO, the probability of achieving curtailment greater than 18.3kWh/15min is 50%. The probability decreases (e.g., 16.19kWh/15min) if a more conservative baseline (e.g., ARIMA) is used instead.

6. DISCUSSION

We have evaluated a wide range of popular methods for household electricity consumption forecasting in the absence of DR based on actual data. We discuss some of the implications of our results here. We start with pointing out that selecting a baseline is non trivial, is error prone, and may lead to misinterpretation of the results. Our analysis shows that curtailment (i.e., electricity consumption reduction) estimation is highly correlated to the baseline selected for analysis. It is therefore straightforward that more effort needs to be done in the following areas of research. First, coming up with better baseline methods that can be applied to all customers without exhibiting the volatility to external factors, such is the building type, which we report here would be highly desirable. If a "one solution fits all" is not possible, developing a framework that would adapt to individual household attributes so as to select the "best" performing baseline method for each individual customer would be advisable. Learning to switch between baselines as time progresses to adapt to customers (changing) behavior would also be beneficial, but at the same time computationally expensive.

While much work has been done on electricity consumption forecasting, reduced consumption prediction is so far an open problem that is under-studied. Instead of estimating what the consumption would have been in the absence of DR (i.e., baseline consumption), and then calculating the difference between such estimate and the actual consumption during DR (as in Fig. 1), computational methods for reduced consumption prediction would be beneficial. The advantage of such an approach is twofold. First, reduced consumption prediction does not require a baseline calculation. Instead, observed curtailed consumption from past events could be used to predict future curtailed consumption. Second, predicted values would be directly comparable against observed consumption during DR for a fair performance evaluation. Some works [5] have already proposed solutions towards this direction. Our findings reinforce this research direction motivating an exploration of promising future work.

The drawback of our work is that it only considers a single regional scenario, even though our analysis involves a heterogeneous collection of buildings with diverse functions and purpose, covering a wide percentage of consumer demographics. Considering scenarios on a per-household basis, as well as including more diverse customer types (e.g. industrial or residential) would be beneficial to the strength of our study. As we could not get access to such datasets, we leave such large-scale in-depth study for future work.

7. CONCLUSIONS

In this paper, we studied the effect of baseline prediction models on the accuracy of estimation and evaluation of consumption reduction achieved by participants during Demand Response motivating an exploration of promising directions for future work. Various methods have been proposed in literature to determine the extent of consumption shedding by estimating what the consumption would have been in the absence of curtailment, and comparing such predictions with observed electric consumption during curtailment. We argued that evaluating a baseline is not straightforward, and more importantly, that not careful consideration of appropriate baseline selection can result in erroneous measurements and misinterpretation of DR results. To address this limitation, we performed an in-depth study of popular methods for estimation of counterfactual electricity demand during DR. Our analysis, which was based on a real-world Smart Grid dataset from the University of Southern California campus microgrid, entails very serious implications for automated Demand Response programs. Particularly, we have shown that curtailment estimation is highly correlated to the baseline of choice. Specifically, over- or under- prediction can result in erroneous assessment of DR programs' outcomes. Our findings can act as guidelines for utilities engaging in post Demand Response analysis to determine the extent of curtailment undertaken by consumers so that appropriate insensitive mechanisms can be designed and the amount of "true" curtailment being properly calculated.

Acknowledgments

This material is based upon work supported by the United States Department of Energy under Award Number number DE-OE0000192, and the Los Angeles Department of Water and Power (LA DWP). The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, the LA DWP, nor any of their employees.

8. REFERENCES

[1] S. Aman, M. Frincu, C. Chelmis, M. U. Noor, Y. Simmhan, and V. Prasanna. Empirical comparison of prediction methods for electricity consumption forecasting. Technical report, 2014. http://www.cs.usc.edu/assets/007/89887.pdf.

[2] S. Aman, Y. Simmhan, and V. K. Prasanna. Energy management systems: state of the art and emerging trends. *IEEE Communications Magazine*, 51(1):114–119, 2013.

[3] CaliforniaISO. Caiso demand response resource user guide. Technical report, 2007. http://www.caiso.com/1ca6/1ca67a5816ee0.pdf.

[4] S. Chan, K. M. Tsui, H. C. Wu, Y. Hou, Y.-C. Wu, and F. Wu. Load/price forecasting and managing demand response for smart grids: Methodologies and challenges. *Signal Processing Magazine, IEEE*, 29(5):68–85, 2012.

[5] C. Chelmis, S. Aman, M. R. Saeed, M. Frincu, and V. K. Prasanna. Estimating reduced consumption for dynamic demand response. In *Proceedings of the Twenty-Ninght AAAI Conference on Artificial Intelligence*. AAAI Press, 2015.

[6] J. Contreras, R. Espinola, F. Nogales, and A. Conejo. Arima models to predict next-day electricity prices. *Power Systems, IEEE Transactions on*, 18(3):1014–1020, 2003.

[7] S. Fan and H. R.J. Short-term load forecasting based on a semi-parametric additive model. *Power Systems, IEEE Transactions on*, 27(1):134–141, 2012.

[8] B. George E. P. and J. Gwilym M. *Time series analysis, forecasting and control*. Holden-Day, 1970.

[9] Y. Goude, R. Nedellec, and N. Kong. Local short and middle term electricity load forecasting with semi-parametric additive models. *Smart Grid, IEEE Transactions on*, 5(1):440–446, Jan 2014.

[10] J. Granderson, M. Piette, G. Ghatikar, and P. Price. Building energy information systems: State of the technology and user case studies. Technical report, Lawrence Berkeley National Laboratory, 2009. http://eis.lbl.gov/pubs/lbnl-2899e.pdf.

[11] T. Hong, M. Gui, B. M.E., and W. H.L. Modeling and forecasting hourly electric load by multiple linear regression with interactions. In *Power and Energy Society General Meeting, 2010 IEEE*, pages 1–8, 2010.

[12] T. Hong, J. Wilson, and J. Xie. Long term probabilistic load forecasting and normalization with hourly information. *Smart Grid, IEEE Transactions on*, 5(1):456–462, Jan 2014.

[13] K. J. Kelly and E. Carl. Measuring industrial energy savings. *Applied Energy*, 85(5):347–361, 2008.

[14] C. Katie, P. Mary Ann, G. Charles A., and K. Sila. Statistical analysis of baseline load models for non-residential buildings. *Energy and Buildings*, 41(4):374–381, 2009.

[15] C. Kuo, M. Lee, C. Fu, Y. Ho, and L. Chen. An in-depth study of forecasting household electricity demand using realistic datasets. In *The Fifth International Conference on Future Energy Systems, e-Energy '14, Cambridge, United Kingdom - June 11 - 13, 2014*, pages 145–155, 2014.

[16] J. Kwac and R. Rajagopal. Demand response targeting using big data analytics. In *Proceedings of the 2013 IEEE International Conference on Big Data, 6-9 October 2013, Santa Clara, CA, USA*, pages 683–690. IEEE, 2013.

[17] X. Lou, D. Yau, H. H. Nguyen, and B. Chen. Profit-optimal and stability-aware load curtailment in smart grids. *Smart Grid, IEEE Transactions on*, 4(3):1411–1420, Sept 2013.

[18] J. Mathieu, P. Price, S. Kiliccote, and M. Piette. Quantifying changes in building electricity use, with application to demand response. *Smart Grid, IEEE Transactions on*, 2(3):507–518, 2011.

[19] C. Morana. A semiparametric approach to short-term oil price forecasting. *Energy Economics*, 23(3):325–338, May 2001.

[20] N. Motegi, M. A. Piette, D. S. Watson, S. Kiliccote, and P. Xu. Introduction to commercial building control strategies and techniques for demand response. Technical report, Lawrence Berkely National Laboratory, 2006. http://gaia.lbl.gov/btech/papers/59975.pdf.

[21] NewYorkISO. Emergency demand response program manual. Technical report, 2013. http://www.nyiso.com/public/webdocs/markets_operations/docu

[22] P.-F. Pai and C.-S. Lin. A hybrid arima and support vector machines model in stock price forecasting. *Omega*, 33(6):497–505, December 2005.

[23] M. A. Piette, G. Ghatikar, S. Kiliccote, E. Koch, D. Hennage, P. Palensky, and C. McParland. Open automated demand response communications specification (version 1.0). Technical Report CEC-500-2009-063, California Energy Commission, PIER program, 2009.

[24] M. A. Piette, O. Sezgen, D. S. Watson, N. Motegi, C. Shockman, and L. ten Hope. Development and evaluation of fully automated demand response in large facilities. Technical report, Lawrence Berkely National Laboratory. http://drrc.lbl.gov/system/files/cec-500-2005-013.pdf.

[25] D. Rahayu, B. Narayanaswamy, S. Krishnaswamy, C. Labbé, and D. P. Seetharam. Learning to be energy-wise: discriminative methods for load disaggregation. In *Future Energy Systems: Where Energy, Computing and Communication Meet (e-Energy), 2012 Third International Conference on*, pages 1–4. IEEE, 2012.

[26] H. Simao, H. Jeong, B. Defourny, W. Powell, A. Boulanger, A. Gagneja, L. Wu, and R. Anderson. A robust solution to the load curtailment problem. *Smart Grid, IEEE Transactions on*, 4(4):2209–2219, Dec 2013.

[27] SouthernCaliforniaEdison. 10 day average baseline and "day-off" adjustment. Technical report, 2011. http://asset.sce.com/Documents/Business - Energy Management Solutions/10DayAvgBaselineFS.pdf.

[28] A. Veit, C. Goebel, R. Tidke, C. Doblander, and H. Jacobsen. Household electricity demand forecasting: benchmarking state-of-the-art methods. In *The Fifth International Conference on Future Energy Systems, e-Energy '14, Cambridge, United Kingdom - June 11 - 13, 2014*, pages 233–234. ACM, 2014.