When Bias Matters: An Economic Assessment of Demand Response Baselines for Residential Customers

Tri Kurniawan Wijaya, *Graduate Student Member, IEEE,* Matteo Vasirani, and Karl Aberer, *Member, IEEE*

***Abstract*—Demand response (DR) has been known to play an important role in the electricity sector to balance supply and demand. To this end, the DR baseline is a key factor in a successful DR program since it influences the incentive allocation mechanism and customer participation. Previous studies have investigated baseline accuracy and bias for large, industrial and commercial customers. However, the analysis of baseline performance for residential customers has received less attention. In this paper, we analyze DR baselines for residential customers. Our analysis goes beyond accuracy and bias by understanding the impact of baselines on all stakeholders' profit. Using our customer models, we successfully show how customer participation changes depending on the incentive actually received. We found that, in general, bias is more relevant than accuracy for determining which baseline provides the highest profit to stakeholders. Consequently, this result provides a valuable insight into designing effective DR incentive schemes.**

***Index Terms*—residential demand response, smart grid, demand-side management, demand response baseline, net benefit analysis.**

I. Introduction

MATCHING supply and demand is a key feature in the reliability of an electricity grid, since failure to ensure it could result in blackouts. Demand response (DR) can be seen as a demand side effort to match the available supply. This is essential, especially when there is nothing more that can be done from the supply side (particularly when energy sources are renewables).[1] The U.S. Federal Energy Regulatory Commission (FERC) gives a clear definition of DR: changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized [3]. In this paper, we focus on incentive-based DR, where the incentive can be in the form of bill rebates, redeemable vouchers, discounts or any other monetary incentive.

There are two key factors to ensure the success of a DR program, namely (i) how to operate DR resources, and (ii) how to measure DR performance. The first factor depends on customers, the energy market, devices, and the utility company (or *company*). In this paper, we focus on the second factor, and more specifically on the DR baseline (or simply *baseline*), which is an estimate of what customers would have consumed in the absence of a DR event. In an incentive-based DR program, the baseline is important because it determines the incentives allocated to customers, and thus influences customers' decisions and participation.

A. Baseline Analysis

Schnittger and Beare [4] provided a brief overview of a DR baseline and its importance in a DR mechanism. Coughlin et al. [5], [6], KEMA [7], and EnerNOC [8] analyzed the accuracy and bias of DR baselines, and suggested different adjustments to improve baseline accuracy. Our set of baselines is inspired by their work. However, we consider only the core baseline method, without adjustments. Their studies focused on accuracy and bias, and did not analyze how baselines affect stakeholders' profits. Mathieu et al. [9] analyzed DR baseline error (hence, accuracy). However, instead of using a set of baseline methods, they considered only a regression-based baseline. They characterized baseline error using several parameters, and aimed to compute the error associated with each parameter. As in previous work, this work focused on baseline error/accuracy, and did not analyze further how a baseline affects the stakeholder profit.

B. Residential Demand Response

Herter and Wayland [10] quantified the effect of residential DR using critical peak pricing. They compared customer's critical and normal weekday loads at the same temperature and found statistically significant average customer responses. In addition, a number of studies have proposed an automated response (aided by a software agent or energy management system) in reaction to the variability in energy prices [11]–[15]. Others have assumed the existence of small energy storage on consumers' premises [16]–[19]. Na Li et al. [20] modeled customer benefit when using a particular appliance at a particular time. This model was then used to provide a schedule that maximized customer profits in response to a dynamic pricing scheme. In contrast to previous works, which rely on the customer side taking to make an action when supply is short, Wijaya et al. [21] proposed that the supply side should explicitly announce the available load and let the customers bid for it. Overall, our literature review suggested that research in residential DR focused on dynamic pricing schemes, and automated response. Investigations about

T. K. Wijaya, M. Vasirani, and K. Aberer are with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: {tri-kurniawan.wijaya, matteo.vasirani, karl.aberer}@epfl.ch).

[1]For more information about the benefit of DR, see [1], [2].

incentive-scheme based DR for residential customers has largely been ignored so far.

C. Overview of Contributions

We summarize our contributions as follows. We study the impact of DR baselines applied to residential customers, whereas previous works focused on large, industrial and commercial customers [5]–[8]. While they concentrated on baseline accuracy and bias, we go beyond these by explaining how a baseline affects customer decision and participation in a DR event, and how it affects both customer and company profit. We evaluate existing methods as well as the new ones that we introduce here. We also develop three models of customer response during a DR event. We are able to model changes in customer participation as a response to the incentives they have actually received. We show that more positively biased baselines foster greater customer participation. Interestingly, while the idea of positively biased baselines does not work to the favor of utility companies, it does deliver the highest overall profits when profit sharing is low.

The rest of the paper is organized as follows. In Section II, we present different DR baseline methods. In Section III, we provide the necessary definitions needed for our analysis, including customer and company profit. We present our analysis in Section IV and V. In Section VI, we conclude and outline the further implications of our work.

II. Demand Response Baselines

Consider a set of (residential) customers $\mathcal{C}$. We divide a day into a set of timeslots $\mathcal{T} = \{t_0, \dots t_{|\mathcal{T}|}\}$. We define the actual load of customer i on day (or date) d at timeslot $t \in \mathcal{T}$ as $\ell_i(d,t)$. In the presence of a DR event, we define the load that a customer would have consumed in the absence of a DR event as *true baseline*, i.e., the customer's intended consumption.[2] However, in practice, during a DR event the true baseline is unknown. Thus, to calculate a customer's reduction in demand (and her incentive for participation) during a DR event, the utility company needs to establish a DR baseline, or predicted baseline (or simply a *baseline* when the context is clear). *Predicted baseline* is the load that the utility company estimates a customer would have consumed in the absence of a DR event, i.e., the prediction of a customer's true baseline. We denote the predicted baseline of customer $i \in \mathcal{C}$ on day d at timeslot $t \in \mathcal{T}$ as $b_i(d,t)$, and her true baseline as $b_i^*(d,t)$.

Several methods have been proposed in the literature and used in practice to compute a predicted baseline load for a DR event: these include HighXofY, MidXofY, exponential moving average, and regression baselines [7], [8]. For the completeness of our analysis, we also define a new baseline method: LowXofY. In addition, a DR baseline should be simple enough for all stakeholders to understand, calculate, and implement, including end-use customers [7], [8]. Thus, even though more sophisticated machine learning methods could deliver higher prediction accuracy, we do not consider them in this study.[3]

[2]In the absence of a DR event, true baseline is equal to actual load.

[3]We refer interested readers to [22] for performance analysis of more complex prediction methods.

We define *DR days*, or *target days*, as days when DR events occur, and others as *non-DR days*. Furthermore, we define two *day-types*: weekdays (Monday to Friday), and weekend (Saturday and Sunday). Let $D(Y,d)$ be a set of Y most recent non-DR days preceding the day d having the same day type as d. In addition, let $\ell_i(d) = \sum_{t \in \mathcal{T}} \ell_i(d,t)$ be the total load of customer i on day d.

A. HighXofY baseline

HighXofY baseline considers Y non-DR days preceding the DR event. The baseline is the average load of the X highest consumption days within those Y days. More formally, for customer i, we define her HighXofY days preceding a DR event day d as $High(X,Y,d) \subseteq D(Y,d)$, where:

(i) $|High(X,Y,d)| = X$, and
(ii) $\ell_i(\hat{d}) \geq \ell_i(d')$ where $\hat{d} \in High(X,Y,d)$ and $d' \in D(Y,d) \setminus High(X,Y,d)$.

The HighXofY baseline of customer i for timeslot t on day d is

$$b_i(d,t) = \frac{1}{X} \sum_{d \in High(X,Y,d)} \ell_i(d,t). \tag{1}$$

Examples of HighXofY baseline are [7]:

- PJM Economic: High4of5 for a weekday, and High2of3 for a weekend DR event.
- NYISO: High5of10 for a weekday, and High2of3 for a weekend DR event.
- CAISO: High10of10 for a weekday, and High4of4 for a weekend DR event.

B. LowXofY

We propose this new, yet relatively simple baseline method. Similar to the HighXofY baseline, the LowXofY baseline for day d is calculated using the X lowest consumption days of Y non-DR days (of the same day type) preceding d. More formally, for customer i, we define her LowXofY days preceding a DR event day d as $Low(X,Y,d) \subseteq D(Y,d)$, where:

(i) $|Low(X,Y,d)| = X$, and
(ii) $\ell_i(\hat{d}) \leq \ell_i(d')$ where $\hat{d} \in Low(X,Y,d)$ and $d' \in D(Y,d) \setminus Low(X,Y,d)\}$.

The LowXofY baseline of customer i for timeslot t on day d is

$$b_i(d,t) = \frac{1}{X} \sum_{d' \in Low(X,Y,d)} \ell_i(d',t). \tag{2}$$

From this baseline method, we use Low4of5, Low5of10, and Low10of20. Unlike previous baselines in HighXofY, for these three baselines we use the same configuration for X and Y for weekday and weekend DR events. As we will show later, this baseline method has higher accuracy, but has more negative bias than the others.

C. MidXofY

The MidXofY baseline for day d is calculated using X of Y non-DR days preceding d by dropping some of the lowest and highest consumption days, retaining only the X middle consumption days. Let $X, Y \in \mathbb{N}$, $X \leq Y$, and $(Y - X) \bmod 2 = 0$. In addition, let $Z = (Y - X)/2$. The MidXofY baseline is calculated using $D(Y, d)$ by dropping the Z-lowest and Z-highest consumption days. More formally, for customer i, we define her MidXofY days preceding a DR event d as $Mid(X, Y, d) = D(Y, d) \backslash (Low(Z, Y, d) \cup High(Z, Y, d))$. The MidXofY baseline of customer i for timeslot t on day d is

$$b_i(d, t) = \frac{1}{X} \sum_{d \in Mid(X,Y,d)} \ell_i(d, t). \tag{3}$$

For our analysis, we consider Mid4of6 (which has also been considered in [7]).

D. Exponential moving average

The exponential moving average baseline is a weighted average of a customer's historical load, where the weight decreases exponentially with time. This baseline is computed using historical load data from the beginning of the measurement day up to the day preceding the target day d.

Let $D(\infty, d) = \{d_1, \ldots, d_k\}$ be the set of all measured days preceding the target day d having the same day type as d. In addition, let $1 \leq \tau \leq k$ be a constant. We define $s_i(d_\tau, t)$ as the *initial* average load of customer i on timeslot t, i.e.,

$$s_i(d_\tau, t) = \frac{1}{\tau} \sum_{j=1}^{\tau} \ell_i(d_j, t) \tag{4}$$

The exponential moving average for $\tau < j \leq k$ is

$$s_i(d_j, t) = \big(\lambda \cdot s_i(d_{j-1}, t)\big) + \big((1 - \lambda) \cdot \ell_i(d_j, t)\big) \tag{5}$$

where $\lambda \in [0, 1]$. Then, we define the exponential moving average baseline for customer i on day d at timeslot t as:

$$b_i(d, t) = s_i(d_k, t). \tag{6}$$

The baseline for days earlier than $d_{\tau+1}$ is undefined.

For this baseline method, we consider the ISONE baseline [7] where $\tau = 5$ and $\lambda = 0.9$. The ISONE baseline is undefined for a customer who joined the DR program for less than 5 days. Even though in practice the ISONE baseline is not applied to the weekend, in this paper, we also compute the ISONE baseline for the weekend.

E. Regression

The baseline of day d is computed using linear regressions whose parameters are inferred on the basis of historical data taken from $D(Y, d)$. This method uses one linear regression predictor for each timeslot during the day, i.e., the baseline of customer i on day d at timeslot t is computed by:

$$b_i(d, t) = (\boldsymbol{\theta}_{i,t})^T \boldsymbol{x}_{i,t} + \epsilon_{i,t} \tag{7}$$

where $\boldsymbol{x}_{i,t}$ is the feature vector, $\boldsymbol{\theta}_{i,t}$ is the (vector of) regression coefficient, and $\epsilon_{i,t}$ is the error term. The feature vector is a vector of explanatory variables such as historical consumption, temperature, or sunrise/sunset time. Then, we estimated the regression coefficient and the error term using ridge regression, although other estimation methods can also be used.

Because our dataset, as we will explain later, does not contain temperature or other measurements which could potentially be explanatory variables, we use historical consumption at the same hour of the day as feature vectors. In order to capture the weekly trend, we set the length of the feature vector to 7 for weekday estimation, and to 5 for weekend days.[4] We consider two regression baselines:

- Reg1, where $D(Y, d)$ contains all historical data available,
- Reg2, where $Y = 150$.

Table I summarizes the baseline methods explained in this section.

TABLE I: Summary of the baseline methods.

Baseline	Short description
HighXofY	average of the highest X of Y days
LowXofY	average of the lowest X of Y days
MidXofY	average of the middle X of Y days
Exp. moving avg.	weighted average of customer's consumption
Regression	linear regression of customer's consumption

note: all historical data considering only non-DR days preceding the DR event

III. Residential Demand Response

With large, industrial and commercial customers, DR is often contract-based. For example, customers agree to respond to a fixed number of DR events per year. However, DR among residential customers can be very dynamic and need not be contract-based. One of the common execution scenarios, which we also consider in this paper, is as follows:

1) The company sends a DR signal to the customers.
2) Each customer decides whether she would like to respond to the signal or not.
3) Using customer's smart meter data, the company reads her actual load and calculates her incentive.

A. DR Signal

Below are two possible types of DR signal for residential customers:

1) A signal which communicates the DR event start/end times and the amount of kWh to be reduced.
2) A signal which communicates the DR event start/end times and lets the customer decide how much she is willing to reduce.

In order to understand how baselines and incentive allocation influence customers' decisions to reduce their consumption, we focus on DR signals of type 2.

[4]Note that from a specific weekday (or weekend day), to reach that same day one week ago, we need to go back at least 5 previous weekdays (or 2 previous weekend days).

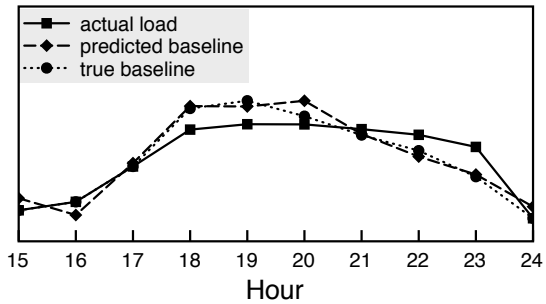


Fig. 1: An illustration of the true baseline, predicted baseline, and actual load where a DR event occurs from 17:00 to 20:00 to curtail the evening peak.

B. DR Event

We define a DR event as a tuple $\delta = (\delta_{start}, \delta_{end})$ where δ_{start} is the start time and δ_{end} is the end time of the event, and we denote $d(\delta)$ as the day/date of the event. In addition, we define the total actual load, predicted baseline, and true baseline of customer i during event δ as:

$$l_i(\delta) = \sum_{t=\delta_{start}}^{\delta_{end}} l_i(d(\delta), t), \tag{8}$$

$$b_i(\delta) = \sum_{t=\delta_{start}}^{\delta_{end}} b_i(d(\delta), t), \tag{9}$$

$$b_i^*(\delta) = \sum_{t=\delta_{start}}^{\delta_{end}} b_i^*(d(\delta), t). \tag{10}$$

For our analysis (later in Section V), we obtained customers' true baselines from the real-world dataset, and we model customers' responses during a DR event to generate the actual load. Figure 1 provides a simple illustration of the customer's actual load, predicted baseline, and true baseline when there is a DR event from 17:00 to 20:00.

Then, for an event δ, we define the aggregate actual load, predicted baseline, and true baseline over all customers as:

$$L(\delta) = \sum_{i \in \mathcal{C}} l_i(\delta). \tag{11}$$

$$B(\delta) = \sum_{i \in \mathcal{C}} b_i(\delta), \tag{12}$$

$$B^*(\delta) = \sum_{i \in \mathcal{C}} b_i^*(\delta), \tag{13}$$

In practice, while $B^*(\delta)$ is not known, $B(\delta)$ and $L(\delta)$ are known. Publishing this information does not violate customer privacy (since both of them describe information aggregated over all customers). In addition DR performance feedback can also be useful to foster customer participation [23]. Thus, we assume that the utility company publish $B(\delta)$ and $L(\delta)$.

C. Cost and Profit Functions

Cost We denote $c(L)$ as the total cost of meeting load demand L. We assume that c is monotonically increasing and strictly convex. An example of a real energy cost function that satisfies both the above assumptions is the quadratic cost function for thermal generators [11], [24], [25]:

$$c(L) = a_1 L^2 + a_2 L + a_3. \tag{14}$$

where a_1, a_2, and a_3 are constants.

Note that our assumption about the monotonically increasing cost function might not hold in cases of renewable energy generation, such as solar or wind power when there is an abundance of sunlight or wind. In this case, we consider a typical situation where there is a lack of supply, and more expensive generator or power sources need to be activated. This may involve advanced buying from the wholesale energy market at a few hours' notice.

A DR event typically happens when there is not enough supply to meet demand or when the spot market price is higher than the retail price. Therefore, the lower the customers' consumption, the greater the company's saving. We introduce the notion of a *company's true saving* as the differences between the cost of generating the customers' true baseline and the cost of generating the customers' actual load, i.e.,

$$c(B^*(\delta)) - c(L(\delta)). \tag{15}$$

However, in practice, what we can compute during a DR event is the predicted baseline, and not the true baseline. Under this condition, the company's saving is computed as the difference between the cost of generating the predicted baseline and the cost of generating the actual load. Hence, we define the notion of the *company's perceived saving*:

$$c(B(\delta)) - c(L(\delta)). \tag{16}$$

Customer's profit function Because of the customer's own efforts to reduce the load, and so as to give the customers further incentives, we define $\alpha \in [0, 1]$ as the proportion of the saving that the company would be willing to share with its customers. A customer only receives an incentive if she reduces her consumption in comparison to the predicted baseline. The incentive received is proportional to the aggregate load. We define the *received incentive* of customer i as:

$$rv_i(\delta) = \begin{cases} \alpha \cdot \left(\frac{b_i(\delta)}{B(\delta)} c(B(\delta)) - \frac{l_i(\delta)}{L(\delta)} c(L(\delta)) \right), & \text{if } l_i(\delta) < b_i(\delta) \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

We can see from Eq. 17, that customer's received incentive depends on the predicted baseline established by the company. From the customer's perspective, if we know what she would have consumed (if there were no DR event), then we can also compute her *true incentive*, which is defined by replacing the predicted baseline in Eq. 17 with customer's true baseline:

$$tv_i(\delta) = \begin{cases} \alpha \cdot \left(\frac{b_i^*(\delta)}{B(\delta)} c(B(\delta)) - \frac{l_i(\delta)}{L(\delta)} c(L(\delta)) \right), & \text{if } l_i(\delta) < b_i^*(\delta) \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

In this case, we assume that the customer is able to estimate her own intended consumption (true baseline). This could be made possible with the help of software agents for example, since the customer is the stakeholder with the most knowledge about the residence: the number and type of inhabitants, the number and type of appliances, and access to personal agendas to know when someone is at home or away. Because the company publishes only $B(\delta)$ and $L(\delta)$, $B^*(\delta)$ remains unknown to the customer. Thus, in Eq. 18 we use $B(\delta)$ as the approximation of $B^*(\delta)$. A customer's true incentive can also be thought of as the customer's received incentive when the predicted baseline established by the company perfectly estimates customer's true baseline.

In addition, for customer i, we define the difference between her received incentive and her true incentive for a DR event δ as her *additional profit*, i.e.,

$$rv_i(\delta) - tv_i(\delta). \tag{19}$$

Positive additional profit means that customer i receives more incentive than she deserves.

Company's profit function The company's profit can be calculated by subtracting the amount of incentives allocated to customers from the true or perceived savings. Note that in practice what it is possible to calculate is the company's perceived savings. However, the perceived savings depend on the chosen baseline method and does not reflect the company's true savings or losses after a DR event. Therefore, if possible, analyzing the company's true profit using its true savings is highly desirable. Below, we specify the company's true profit by its proportion to the cost of generating true baseline:

$$\frac{c(B^*(\delta)) - c(L(\delta)) - \sum_{i\in\mathcal{C}}(rv_i)}{c(B^*(\delta))}. \tag{20}$$

In our analysis, we compute the company's true profit by maintaining the customers' true baseline (obtained from a real-world dataset), and modeling customer responses during a DR event to obtain the actual load.

D. Customer Model

For our analysis, we propose three customer models.

Naïve This first model introduces a naïve customer model whose fixed parameter $\gamma \in [0,1]$ determines by how much she reduces her intended load (true baseline). For customer i, for each DR event δ, we have:

$$l_i(\delta) = (1-\gamma_i)\cdot b_i^*(\delta). \tag{21}$$

This parameter γ remains constant over time, regardless of the chosen baseline or the incentive given by the company.

Rational From Eq. 17 and 18, we can see that when the predicted baseline underestimates a customer's true consumption she receives less incentive than she deserves (and *vice versa*). In this model, a customer responds to a DR signal only if the predicted baseline established by the company does not underestimate her true consumption. More formally, for customer i, during a DR event δ:

$$l_i(\delta) = \begin{cases} (1-\gamma_i)\cdot b_i^*(\delta), & \text{if } b_i(\delta) \geq b_i^*(\delta) \\ b_i^*(\delta), & \text{otherwise} \end{cases} \tag{22}$$

where $\gamma_i \in [0,1]$ is the proportion of customer i's reduction compared to her true consumption.

Adaptive In this model, we introduce a customer who learns to make her decisions with regards to past experiences. The greater incentive the customer actually receives, in relation to the incentive she actually deserves, the more eager she will be to participate in the next event. That is, the customer's decision to reduce load evolves from one DR event to the next, influenced by the ratio between her received incentive (what she receives) and her true incentive (what she should have received).

Let δ_j be the jth DR event. From a predefined γ_i^0, this parameter evolves as the exponential moving average ratio between customer i's received incentive and her true incentive. Moreover, let

$$\begin{cases} \Delta^\omega = \frac{1}{\omega}\cdot\sum_{j=1}^{\omega}\frac{rv_i(\delta_j)}{tv_i(\delta_j)} \\ \Delta^j = \rho\cdot\Delta^{j-1} + (1-\rho)\cdot\frac{rv_i(\delta_j)}{tv_i(\delta_j)}, \quad \text{for } j > \omega \end{cases} \tag{23}$$

where $\omega \in \mathbb{N}$ is the initial learning length parameter and $\rho \in [0,1]$ is the decaying parameter to discount previous observations. Then, we define:

$$\begin{cases} \gamma_i^j = \gamma_i^0, & \text{for } j \leq \omega \\ \gamma_i^j = \Delta^{j-1}\cdot\gamma_i^{j-1}, & \text{for } j > \omega \end{cases} \tag{24}$$

We restrict the minimum value of γ_i^j to 0 and its maximum value to 1. In this model, customer i reduces γ_i^j of her true consumption during DR event δ_j, i.e.,

$$l_i(\delta_j) = (1-\gamma_i^j)\cdot b_i^*(\delta_j). \tag{25}$$

The larger the ratio between customer's received incentive and her true incentive, the higher her γ for the next DR event.

IV. Accuracy and Bias

A. Setup

For our analysis, we use the Irish CER smart metering trial dataset [26]. This dataset contains measurements of around 5,000 customers over 1.5 years. The customers consist of residential houses and small and medium-sized enterprises. The measurements started in July 2009 and ended in December 2010. Since the trial was about dynamic pricing, we use only the data from the control group, composed of customers who are not affected by the different pricing schemes. More specifically, we choose residential customers that belong to the control group and have no missing values. This results in the selection of 782 customers. In order to take into account the seasonal variation in customers' loads, we use a full year of measurement data, from January 1st to December 31st 2010.

We analyze the hourly accuracy and bias of each baseline. Let $\mathcal{C}$ be the set of our 782 customers, $\mathcal{D}$ be the set of all days in 2010, and $\mathcal{T}$ be the set of hourly timeslots in a day. We measure baseline accuracy in terms of Mean Absolute Error (MAE):

$$\frac{\sum_{i\in\mathcal{C}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}|b_i(d,t) - \ell_i(d,t)|}{|\mathcal{C}|\cdot|\mathcal{D}|\cdot|\mathcal{T}|}. \tag{26}$$

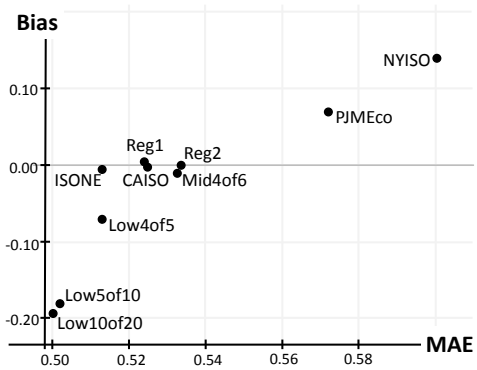


Fig. 2: Mean Average Error (MAE) and bias of different baselines in kWh. Average hourly load over all customers is 0.97 kWh.

The lower the MAE, the higher the accuracy. And we define baseline bias as:

$$\frac{\sum_{i\in\mathcal{C}}\sum_{d\in\mathcal{D}}\sum_{t\in\mathcal{T}}\left(b_i(d,t)-\ell_i(d,t)\right)}{|\mathcal{C}|\cdot|\mathcal{D}|\cdot|\mathcal{T}|}. \tag{27}$$

Baseline methods which have positive bias tend to overestimate customers' actual consumption (or true baseline when DR events occur), and *vice versa*.

B. Analysis

Figure 2 shows the accuracy and bias of each baseline in kWh. As expected, the HighXofY baselines whose $X < Y$, have a more positive bias than the others, whereas our LowXofY baselines have more negative bias than the others.

One interesting point here is that our LowXofY baselines are able to provide better accuracy than more sophisticated baselines such as ISONE (exponential moving average) and Reg1/Reg2 (regression).[5] The accuracy gained by our LowXofY baselines is driven by the days when a day with unusually high consumption is followed by days with normal (and thus much lower) consumption. In this case, LowXofY baselines will most likely exclude this unusually high consumption day from the next day's (or days') baseline computation(s). However, other baselines take this unusual day into account. For example, using CAISO, this unusual day will be carried through in the next 10 baseline computations (if the unusual day happens on a weekday, or 4 if it happens on the weekend). This is also the case with the exponential moving average and regression baselines, which both take into account this unusual day in their models.

V. Net Benefit Analysis

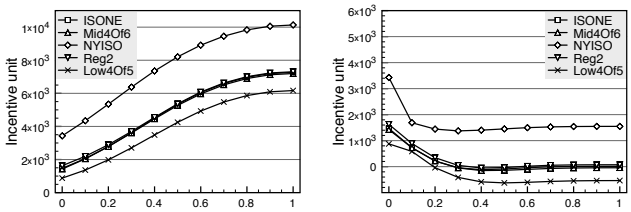
A. Setup

Baselines To analyze customer and company profit we focus on five representative baseline methods: ISONE, Low4of5, Mid4of6, Reg2, and NYISO. We chose them such that they use different baseline methods, and have different accuracy and bias profiles. Table II shows the details of their profiles.

[5]However, one should not conclude that LowXofY is better than regression models in general. Adding some additional explanatory variables to Reg1/Reg2 could increase their predictive power and improve their accuracy.

TABLE II: Detailed characteristics of ISONE, Low4of5, Mid4of6, Reg2, and NYISO (ordered by MAE).

Baseline	Baseline family	MAE (kWh)	Bias (kWh)
ISONE	Exponential moving average	0.51362	-0.00259
Low4of5	LowXofY	0.51645	-0.07293
Mid4of6	MidXofY	0.53163	-0.01416
Reg2	Regression	0.53368	0.00893
NYISO	HighXofY	0.60611	0.14796



(a) Received incentive (b) Additional profit

Fig. 3: Received incentive and additional profit in the naive customer model. Both are calculated as the sum of customers over all 52 DR events.

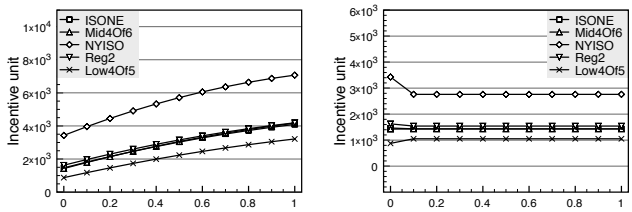
DR event We set the DR event to occur once a week on a random day. In total, we have 52 DR events for a year, between January 1st and December 31st 2010. The events happen evening during peak hours, starting at 17:00 and ending at 20:00. As described in Section III-A, we use a type 2 DR signal, which allows us to analyze how the baseline and the incentive affect a customer's decision to reduce her load.

We use a simple cost function as described Eq. 14 with $a_1 = 0.0001$, $a_2 = 0$, and $a_3 = 0$. Any other cost function could be used as long as it satisfies the assumption stated in Section III-C. It will not affect our analysis (the end result might have different exact numbers, but would show the same trends). In addition, we assume that the imbalance between supply and demand occurs during the DR events. Hence, we focus our analysis exclusively on the time of the DR events.

B. Customers Profit

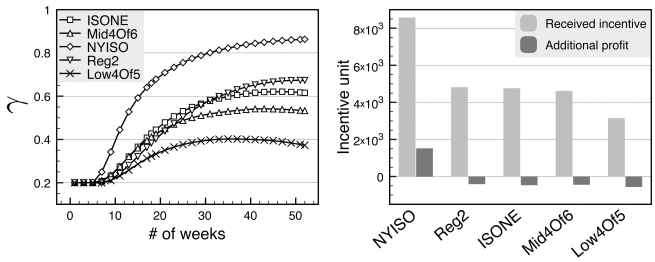
For each customer model, we analyze customers' received incentive and their additional profit when a particular baseline is used. Note that our cost function is not associated to a currency. Thus we define the unit of measurement of the customers' incentive as an *incentive unit*.

Received incentive in the naïve and rational customer models Figure 3a shows the received incentive in the naïve customer model, and Figure 4a shows the received incentive in the rational customer model. Both are shown as the sum of all customers over all 52 DR events with $\alpha = 0.1$ (defined in Section III-C). We have the same trends, i.e., the larger the γ, the larger the received incentive. This is expected since γ represents the proportion of the intended consumption (true baseline) that customers reduce. A larger gamma means a lower actual load (Eq. 21 and 22), thus higher incentives (Eq. 17).



(a) Received incentive (b) Additional profit

Fig. 4: Received incentive and additional profit in the rational customer model. Both are calculated as the sum of all customers over all 52 DR events.



(a) Weekly average of γ (b) Received incentive and additional profit

Fig. 5: The evolution of γ, received incentive, and additional profit of adaptive customer model with $\gamma^0 = 0.2$ and $\alpha = 0.1$ (defined in Section III-C). Received incentives and additional profit shown are the sum over all customers and 52 DR events.

The figures show that irrespective of customers' γ, NYISO delivers them the highest incentives; Reg2, ISONE, and Mid4of6 come second; and Low4of5 delivers the lowest incentives. We highlight that this ranking is ordered by their bias (not accuracy). The more positive the bias, the higher we set the customers' predicted baseline loads. As a result, the customers received higher incentives (Eq. 17). When different α is used, our analysis does not change. Different α only introduces a constant shift to the current plots upward or downward (Eq. 17).

Additional profit of naïve and rational customer model Figure 3b shows the additional profit of the naïve customer model, and Figure 4b shows the additional profit of the rational customer model. See Section III-C for the definition of additional profit. Similar to the received incentive case, baseline methods with more positive bias give higher additional profit to the customers. This can be understood since more positive bias baseline methods tend to overestimate the customers' true baseline, thus they give higher additional profit to the customers.

It is interesting to note how customers receive the highest additional profit when γ is low, especially when $\gamma = 0$ (except using the Low4of5 baseline on rational customers). Let $\mathcal{C}^+$ be the set of customers whose true consumption is overestimated and $\mathcal{C}^-$ be the set whose consumption is underestimated by a particular baseline for a particular DR event. When $\gamma = 0$, $\mathcal{C}^+$ has positive additional profit, whereas $\mathcal{C}^-$ has 0 additional profit. When $\gamma > 0$, $\mathcal{C}^+$ still has positive additional profit but $\mathcal{C}^-$ experiences negative additional profit, and the overall customers' additional profit decreases. In addition, when $\gamma = 0$ the true incentive of $\mathcal{C}^+$ is 0, whereas when $\gamma > 0$ the true incentive of $\mathcal{C}^+$ is > 0. This also causes the trend of additional profit going down when $\gamma > 0$. Additional profit can also be thought of as a "free lunch" for the customers.

Customers with $\gamma = 0$ have a true incentive equal to 0, because they do not carry out any reduction in consumption. However, there are some customers whose loads are overestimated by the baseline methods, and others whose loads are not. While the non-overestimated customers do not receive any incentives, the overestimated customers do receive some incentives (free riders). This is why the total received incentive (and additional profit) over all customers is positive when $\gamma = 0$. For these customers, in order to receive bigger incentives, they need to increase their γ, i.e., they cannot be free riders any more. Even though their additional profits decrease, their received incentives increase (as the payoff for reducing their load).

Adaptive customers Figure 5 shows the evolution of customers' γ over time (weekly), their received incentives, and additional profits, given different baselines with initial gamma, $\gamma^0 = 0.2$. We recall that the evolution of a customer's γ depends on the ratio between her received incentive and her true incentive. The higher the ratio, the higher the customer's γ for the next DR event.

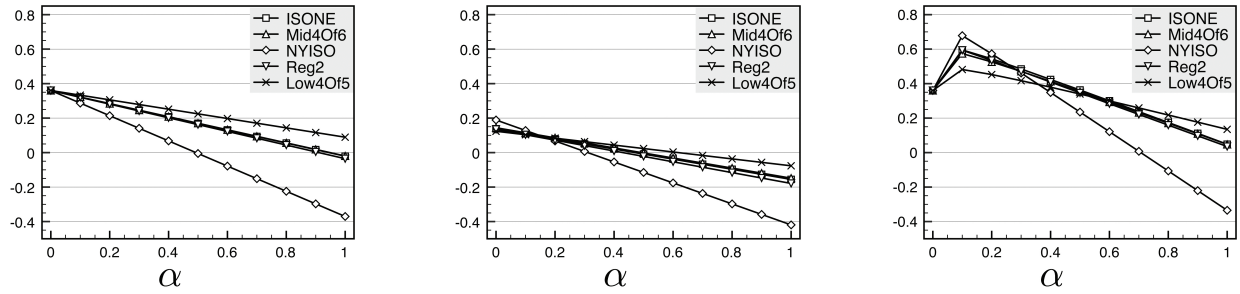
The primary motivation for the development of the NYISO baseline was to encourage customer participation. Using an adaptive customer model, we successfully realized this phenomenon. Figure 5a shows that using a positive bias baseline increases customers' γ more than using a negative bias baseline (see Table II for the bias of the baseline methods). This is due to the fact that more positively biased baselines provide a higher received incentives compared to the true incentives.

The baseline bias and customer incentive trends seen in both the naïve and rational customer model can also be found in this customer models. A baseline with a more positive bias results in higher received incentive and additional profit (which encourages customer participation, as we mentioned earlier).

C. Company's Profit

Figure 6 shows the company's profit using different customer models. We discussed earlier that more positively biased baselines deliver higher customer profit. However, this is not the case with the company's profit. More negatively biased baselines provide higher company profit because they lower the amount of incentives allocated to the customers.

In the naïve customer model, and in the relatively high α of the rational and adaptive customer models, Low4of5 delivers the highest company profit compared to the other baselines. Moreover, we can see that the more negative the baseline's bias, the higher the company's profit. This is understandable since having more negatively biased baseline methods means that the company tends to set lower predicted baseline load,



(a) Naive customer model (b) Rational customer model (c) Adaptive customer model with $\gamma^0 = 0.2$

Fig. 6: Company's profit under different customer models and various α.

and hence distributes lower incentives to the customers. These lower total incentives lead to higher company profit. (see Eq. 20).

More interesting facts are shown in Figures 6b and 6c, which present some cases in which more positively biased baselines provide higher profit to the company. This is interesting because positively biased baselines (and, of course, larger α) are in line with customer preference. In general, negatively biased baselines, which tend to underestimate customers' true consumption, ought to deliver higher profit to the company. If this were always the case, then there would be a conflict of interest with the customers. Therefore, a case where more positively biased baselines deliver higher profit to the company is attractive because it opens the possibility of satisfying both stakeholders – the customers and the company.

However, determining the right α becomes crucial. Larger α results in a higher overall incentive allocated, thus, lower company profit. Figures 6b and 6c show that NYISO is the best baseline to use with the rational customer model with $\alpha < 0.15$ and with the adaptive customer model with $\alpha < 0.2$. In these two cases, for some small α, more positive bias baselines provide better profit for the company. This can be understood because a more positively biased baseline encourages more customer participation. Thus, it reduces the overall actual load, but does not give away too much in terms of incentives, which potentially increases the company's profit (see Eq. 20).

VI. Conclusion

In this paper, we analyzed the performance of DR baselines in the context of residential demand response, whereas previous works on baseline analysis focused only on large customers and commercial buildings. Furthermore, the baseline analyses performed to date were limited to "classic" analyses, i.e., baseline accuracy and bias. In this paper, we went beyond these classic analyses by explaining the impact of DR baselines on the stakeholders' benefits, i.e., the profit of both the customers and the company. These are all essential elements for making residential DR a reality in the future.

As a supplement to the current baseline methods found in the literature, we proposed a novel yet relatively simple baseline method: LowXofY. This method has a more negative bias than other baselines, but it is more accurate. We also successfully confirmed the fact that positively biased baselines increase customer participation. While the motivation for using more positively biased baselines is indeed to encourage customer participation, little is known about whether overestimating customer consumption could benefit the company. We showed that when the company shares a small portion of its profit with its customers (i.e., small α), it can actually increase its profit (and deliver higher profit compared to the other baselines) due to the increased customer participation. This opens up the possibility of a win-win solution for both the customers and the company. In addition, our result provides a valuable insight for the design of future incentive schemes for DR, i.e., even though the company can perfectly estimate a customer's baseline (which is useful to assess DR success rate), it might want to increase the baseline's bias a little to encourage her participation

As a future study, we plan to incorporate social interaction into the customer models, since this has been shown to have an impact on energy conservation awareness [27]–[29] and more environmentally friendly behavior in general [30]. In addition, while baseline methods described in this paper use only non-DR days, designing a DR baseline which take into account customer's consumption on DR days could be another interesting direction. This would be particularly important, in the future, when DR events are set to run almost everyday. We also plan to quantify baseline integrity, i.e., how far customers can manipulate a baseline to their own advantage. In addition to its economic impact, baseline integrity is another important aspect that needs to be assessed for a successful and sustainable DR program.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 288322 (Wattalyst) and 288021 (EINS).

References

[1] "Benefit of Demand Response in Electricity Markets and Recommendations for Achieving Them," U.S. Department of Energy, Feb. 2006.
[2] G. Strbac, "Demand Side Management: Benefits and Challenges," *Energy Policy*, vol. 36, no. 12, pp. 4419 – 4426, 2008.
[3] "Assessment of Demand Response & Advanced Metering," Federal Energy Regulatory Commission, Dec. 2010.

[4] S. Schnittger and S. Beare, "Economic implications of the proposed demand response mechanism," Australian Energy Market Commission, Tech. Rep., 2012.

[5] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote, "Statistical Analysis of Baseline Load Models for Non-Residential Buildings," *Energy and Buildings*, vol. 41, no. 4, pp. 374 – 381, 2009.

[6] K. Coughlin, M. A. Piette, C. A. Goldman, and S. Kiliccote, "Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California," Demand Response Research Center, CA, Tech. Rep. LBNL-63728, Jan. 2008.

[7] "PJM Empirical Analysis of Demand Response Baseline Methods," White Paper, KEMA, Inc., Apr. 2011.

[8] "The Demand Response Baseline," White Paper, EnerNOC, 2011.

[9] J. Mathieu, D. Callaway, and S. Kiliccote, "Examining uncertainty in demand response baseline models and variability in automated responses to dynamic pricing," in *The 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, 2011, pp. 4332–4339.

[10] K. Herter, P. McAuliffe, and A. Rosenfeld, "An exploratory analysis of California residential customer response to critical peak pricing of electricity," *Energy*, vol. 32, no. 1, pp. 25 – 34, 2007.

[11] A. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 320 –331, Dec. 2010.

[12] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential Demand Response Using Reinforcement Learning," in *The First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 409–414.

[13] T. K. Wijaya, T. Papaioannou, X. Liu, and K. Aberer, "Effective consumption scheduling for demand-side management in the smart grid using non-uniform participation rate," in *Sustainable Internet and ICT for Sustainability (SustainIT), 2013*, Oct 2013, pp. 1–8.

[14] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. Jennings, "Agent-based control for decentralised demand side management in the smart grid," in *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, ser. AAMAS '11. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 5–12.

[15] M. Vasirani and S. Ossowski, "Smart consumer load balancing: state of the art and an empirical evaluation in the Spanish electricity market," *Artificial Intelligence Review*, vol. 39, no. 1, pp. 81–95, 2013.

[16] P. Vytelingum, T. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings, "Theoretical and practical foundations of large-scale agent-based micro-storage in the smart grid," *J. Artif. Intell. Res. (JAIR)*, vol. 42, pp. 765–813, 2011.

[17] T. Carpenter, S. Singla, P. Azimzadeh, and S. Keshav, "The impact of electricity pricing schemes on storage adoption in ontario," in *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, ser. e-Energy '12. New York, NY, USA: ACM, 2012, pp. 18:1–18:10.

[18] J.-Y. Le Boudec and D.-C. Tomozei, "A demand-response calculus with perfect batteries," in *Proceedings of the 16th International GI/ITG Conference on Measurement, Modelling, and Evaluation of Computing Systems and Dependability and Fault Tolerance*, ser. MMB'12/DFT'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 273–287.

[19] A. Mishra, D. Irwin, P. Shenoy, and T. Zhu, "Scaling distributed energy storage for grid peak reduction," in *Proceedings of the 4th International Conference on Future Energy Systems*, ser. e-Energy '13. New York, NY, USA: ACM, 2013, pp. 3–14.

[20] N. Li, L. Chen, and S. H. Low, "Optimal Demand Response Based on Utility Maximization in Power Networks," in *Proceedings of the 2011 IEEE Power & Energy Society General Meeting*, Jul. 2011.

[21] T. K. Wijaya, K. Larson, and K. Aberer, "Matching demand with supply in the smart grid using agent-based multiunit auction," in *The Fifth International Conference on Communication Systems and Networks (COMSNETS)*, 2013, pp. 1–6.

[22] S. Humeau, T. K. Wijaya, M. Vasirani, and K. Aberer, "Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households," in *The 3rd IFIP Conference on Sustainable Internet and ICT for Sustainability*, Oct. 2013.

[23] A. Faruqui, S. Sergici, and A. Sharif, "The impact of informational feedback on energy consumptiona survey of the experimental evidence," *Energy*, vol. 35, no. 4, pp. 1598 – 1608, 2010.

[24] M. Gonzalez Chapa and J. Vega Galaz, "An economic dispatch algorithm for cogeneration systems," in *IEEE Power Engineering Society General Meeting*, 2004, pp. 989–994 Vol.1.

[25] P. Rao and S. Nagendra, "Combined Heat and Power Economic Dispatch: A Direct Solution," *Electric Power Components and Systems*, vol. 34, no. 9, pp. 1043–1056, 2006.

[26] "Smart Metering Trial Data Publication," Irish CER, 2012.

[27] R. Brewer, G. Lee, and P. Johnson, "The Kukui Cup: A Dorm Energy Competition Focused on Sustainable Behavior Change and Energy Literacy," in *The 44th Hawaii International Conference on System Sciences (HICSS)*, 2011, pp. 1–10.

[28] P. Petkov, F. Köbler, M. Foth, R. Medland, and H. Krcmar, "Engaging energy saving through motivation-specific social comparison," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '11. New York, NY, USA: ACM, 2011, pp. 1945–1950.

[29] A. Laskey, "How behavioral science can lower your energy bill," Video file, TED, Feb. 2013. [Online]. Available: http://www.ted.com/talks/alex_laskey_how_behavioral_science_can_lower_your_energy_bill.html

[30] J. Mankoff, S. R. Fussell, T. Dillahunt, R. Glaves, C. Grevet, M. Johnson, D. Matthews, H. S. Matthews, R. McGuire, R. Thompson, A. Shick, and L. D. Setlock, "Stepgreen.org: Increasing Energy Saving Behaviors via Social Networks," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, May 2010.

Tri Kurniawan Wijaya is a PhD student in the Distributed Information System Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. His research interest is about energy consumption behavior analysis and mechanism design for demand-side management in the Smart Grid. Before joining EPFL, he received his M.Sc. degree in Computational Logic from the Dresden University of Technology and the Vienna University of Technology with distinction. He has received Erasmus Mundus scholarship from the European Commission and IBM PhD Fellowship Award 2013-2014.

Matteo Vasirani received the M.Sc. degree in Computer Engineering from the University of Modena, Reggio Emilia, Italy, in 2004 and the Ph.D. degree in computer science from the University Rey Juan Carlos, Madrid, Spain, in 2009. He is a Research Associate in the Distributed Information Systems Laboratory, EPFL, Switzerland. Previously, he was an Assistant Professor and Associate Director for Training and Promotion at the Centre for Intelligent Information Technologies (CETINIA), University Rey Juan Carlos, Madrid, Spain. He is the author of several publications in journals and international conferences and currently work on agent-based techniques and data analytics for demand-response.

Karl Aberer received the PhD degree in mathematics from the ETH Zurich in 1991. He has been a full professor of distributed information systems at the EPFL, Switzerland, since 2000. His research interests are on highly distributed information systems with applications in peer-to-peer search, semantic interoperability, trust management, mobile, social and sensor networks, and smart energy systems. From 1991 to 1992, he was postdoctoral fellow at the International Computer Science Institute, University of California, Berkeley. In 1992, he joined the Integrated Publication and Information Systems institute at GMD in Germany, where he was leading the research division Open Adaptive Information Management Systems. From 2005 till 2012 he was the director of the Swiss National Research Center for Mobile Information and Communication Systems (NCCR-MICS, www.mics.ch). He is member of the editorial boards of the VLDB Journal, the ACM Transaction on Autonomous and Adaptive Systems, and the World Wide Web Journal. He has also been consulting for the Swiss government in research and science policy as a member of the Swiss Research and Technology Council (SWTR) from 2004 to 2011 He is a member of the IEEE and the IEEE Computer Society.