Occupant Plugload Management for Demand Response in Commercial Buildings: Field Experimentation and Statistical Characterization

Chaitanya Poolla$^{1, *}$, Abraham K. Ishihara$^2$, Dan Liddell$^3$, Rodney Martin$^4$, Steven Rosenberg$^5$

Abstract

Commercial buildings account for approximately 35% of total US electricity consumption, of which nearly two-thirds is met by fossil fuels resulting in an adverse impact on the environment. This adverse impact can be mitigated by lowering energy consumption via control of occupant plugload usage in a closed-loop building environment. In this work, we conducted multiple experiments to analyze changes in occupant plugload energy consumption due to incentives and/or visual feedback. The incentives entailed daily monetary values between $5 and $50 administered in a randomized order and the visual feedback consisted of a web-based dashboard aimed at increasing the energy awareness of participants. Experiments were performed in government office and university buildings at NASA Ames Research Park located in Moffett Field, CA. Autoregressive models were constructed to predict expected plugload savings in the presence of exogenous variables. Analysis of the data revealed modulation of plugload energy consumption can be achieved via visual feedback and incentive mechanisms suggesting that occupant-in-the-loop control architectures may be effective in the commercial building environment. Our findings indicate that the mean energy reduction due to visual feedback in office and university environments were $\approx 9.52\%$ and $\approx 21.61\%$, respectively. By augmenting the visual feedback in the university environment with a monetary incentive, the mean energy reduction was found to be $\approx 24.22\%$.

Keywords: Commercial buildings, Plugload, Demand Response, Energy Efficiency, Sustainability, Experiment design, Dashboard, Incentives, Carbon footprint, Occupants, Building facilities

1. Introduction

Buildings account for more than 55% of electricity consumption and 28% of energy-related $CO_2$ emissions worldwide [1]. In 2019, these emissions amounted to 10 Gigatonnes, reaching their highest level ever recorded [2]. Considering the rapid growth in building energy demand and related $CO_2$ emissions, there is a need to improve energy efficiency measures in buildings for achieving a Sustainable Development Scenario (SDS) in the society [2][3][4]. Within the US, electricity consumed by residential, commercial, and industrial buildings account for 37%, 35%, and 28%, respectively [5].

$^*$Corresponding author: cpolla@alumnii.cmu.edu

$^1$Chaitanya Poolla is with Intel Corporation. This work was done when he was with Carnegie Mellon University (SV).

$^2$Abraham K. Ishihara is with KBR. This work was done when he was with Carnegie Mellon University (SV).

$^3$Dan Liddell contributed to this work when he was with Carnegie Mellon University (SV).

$^4$Rodney Martin is with NASA Ames Research Center.

$^5$Steven Rosenberg is with Carnegie Mellon University (SV).
Electricity consumption in buildings is classified into three categories: Heating, Ventilation, and Air Conditioning (HVAC); Lighting; and Plug and Process Loads (PPLs) [5]. HVAC and lighting systems contribute to nearly two-thirds of the electricity consumption and PPLs account for the remainder [7]. Building electricity consumption is reduced through methods such as Demand Response (DR) programs [8] or retrofitting [9]. While there have been several studies on reducing the energy consumption incurred by HVAC and lighting systems [10] [11], the problem of plugload energy reduction has received considerably less attention despite being recognized as the next big hurdle toward improving energy efficiency in buildings [12] [13].

The potential for energy savings due to plugload energy reduction in office buildings is estimated to be in the range of 15 – 40% [14] [15]. Given that plugloads are controlled by occupants, it is of interest to determine how interventions motivate occupants to vary their plugload energy consumption. Previous studies indicate that promoting energy efficiency among occupants by feedback and/or incentives results in significant energy savings: Jain et al. [16] studied the role of interventions in motivating energy efficient behavior among 43 participants over a period of six weeks. They concluded that feedback via historical comparisons and incentives are statistically significant in motivating energy reduction. Based on techniques from human-computer interaction, psychology, and energy efficiency, Yun et al. [17] found that 12% – 20% energy savings can occur by behavioral modification. Petersen et al. [18] conducted a five-week study involving feedback and incentives across 18 dormitories and found a significant energy conservation of 32%. The potential of online games for plugload energy conservation was studied by Gandhi et al. [19]. Online games were found effective to reinforce conservation behavior, while suggesting the need for further studies on non-monetary commercial plugload reduction.

Despite its significance toward building energy efficiency, the problem of plugload management is challenging due to its dependence on occupant usage [12] [13]. This dependence also results in the difficulty to model plugload consumption. Nevertheless, models are essential to forecast energy consumption and to assess the efficacy of interventions [20]. In most of the published literature on plugload energy consumption, either field experimentation or statistical modeling are absent. This absence limits the scope for demand management and sustainable development in cities. Thus, there exists a need for plugload studies that not only involve experiments but also infer models from the resulting data. To the best of our knowledge, this is the first work to statistically characterize and model occupant plugload energy consumption in commercial buildings based on designed experiments involving incentives and/or visual feedback.

In this work, two field experiments were designed to study the effect of incentives and/or dashboard feedback on occupant plugload energy consumption. These experiments were carried out in government office and educational (university) buildings inside the NASA Ames Research Park Complex. The experiment related to the government office building was conducted within NASA Sustainability Base (SB), a 50,000 square foot LEED platinum certified commercial building [21] in the presence of a visual (dashboard) feedback intervention. The other experiment was conducted within two Carnegie Mellon University Silicon Valley (CMU SV) campus buildings (NASA Ames Research Park buildings 19 and 23) in the presence of incentives and/or dashboard feedback. The data from these experiments was used to construct
models for predicting the effects of incentive and feedback interventions on plugload energy consumption. These occupant plugload models enable the integration of occupants-in-the-loop within existing building optimization frameworks. The integration of the occupant plugload allows for holistic load modeling, thereby enabling optimal energy management in buildings. For example, a model-based optimal policy could be designed and updated in real-time based on system identified parameters. [22]. The major contributions of this work are:

1. A paired experiment design to study the effects of dashboard-enabled feedback and/or incentives on occupant plugload energy consumption.

2. A statistical characterization of the data where hypothesis tests are conducted and confidence intervals are estimated to determine the efficacy of interventions. Autoregressive models with exogenous inputs are proposed to model occupant plugload energy consumption. This can enable occupants-in-the-loop control strategies within a demand response framework.

3. Visual feedback via informative dashboards offers statistically significant plugload reduction in both government office and university environments.

The rest of this paper is organized as follows: Section 2 describes the design and execution of the experiments at NASA SB and CMU SV. A statistical analysis of the data is presented in Section 3 along with the respective results and discussion. Concluding remarks are presented in Section 5.

2. Experiment design and execution

We designed and conducted experiments to study the influence of incentives and/or feedback interventions on occupant plugload energy consumption. Our research hypothesis is:

\textit{Providing incentives and/or dashboard-based feedback to occupants in office and university buildings reduces occupant plugload energy consumption.}

Consequently, we examine the claim that the average occupant plugload energy consumption in the presence of an incentive and/or feedback is less than the energy consumption in the absence of incentive and/or feedback based on data from the experiments. In the rest of this section, we present the experiment design and implementation.

2.1. Location and duration

Two experiments were conducted within the NASA Ames Research Center, one within a government office environment (NASA SB) and the other within a university environment (CMU SV - buildings 19 and 23). Let the symbols $E_N$ and $E_C$ denote the experiments at NASA SB and CMU SV, respectively. Each experiment consisted of multiple phases classified into a baseline or an experiment phase depending on the absence or presence of an intervention, respectively.
The interventions employed during experiment phases consisted of dashboard feedback, incentives, or both dashboard feedback and incentives. The feedback-based intervention was applicable to both experiments $E_{NI}$ and $E_{CI}$ whereas the incentive-based intervention was applicable only to experiment $E_{CI}$ (at CMU SV). Thus, experiment $E_{CI}$ was designed to study both the individual and interaction effects of the interventions using a two-factor factorial design. The duration of both of the experiments were based on findings from habit modeling [23]. The properties for each phase of $E_{NI}$ and $E_{CI}$ are provided in table 1.

| Expt. Property | Phase       | Baseline Phase | Incentive Phase | Feedback Phase | Feedback & Incentive Phase |
|----------------|-------------|----------------|-----------------|----------------|----------------------------|
| $E_{NI}$ (NASA SB): Applicable? | Yes         | No             | Yes             | No             |
| $E_{NI}$ (NASA SB): Notation | $p^{31}_{NI}$ | N/A            | $p^{31}_{NI}$   | N/A            |
| $E_{NI}$ (NASA SB): Duration | Five weeks  | N/A            | Four weeks      | N/A            |
| $E_{CI}$ (CMU SV): Applicable? | Yes         | Yes            | Yes             | Yes            |
| $E_{CI}$ (CMU SV): Notation | $p^{1}_{CI}$ | $p^{2}_{CI}$   | $p^{3}_{CI}$    | $p^{4}_{CI}$   |
| $E_{CI}$ (CMU SV): Duration | Five weeks  | Two weeks      | Two weeks       | Two weeks      |

Table 1: Description of properties in experiments $E_{NI}$ and $E_{CI}$

### 2.2. Variables

We discuss the response variables and interventions in both experiments here. The response variable was defined as the time-averaged power consumption of the participant. Its value was computed based on data from smart powerstrips [24]. The interventions employed are described in Table 1. The incentive interventions were administered as daily monetary rewards aimed at promoting energy conservation among the participants. The feedback intervention was administered by a web browser-based dashboard tool which was designed to raise awareness about the participant’s plugload energy consumption. It is important to note that the feedback provided by the experimenters and the feedback received by the participants differed since each participant does not necessarily use all features of the dashboard at all times. The time spent by each participant on their dashboard is used to quantify the feedback received.

### 2.3. Design principles and implementation

The ideal experiment design allows us to attribute any changes in energy consumption exclusively to the interventions employed and not to nuisance factors such as differences in participants’ workloads or preferences. For example, one participant may consume more energy than another due to the difference in a device workload or setting. In such a case, the device workload and setting are the inevitable nuisance factors that the experiment design needs to account for. By accounting for nuisance factors, we can ensure a strong connection between the interventions employed and the response recorded. To this end, we employ a combination of the recommended design techniques consisting of blocking, randomization, and replication [25] as described below:
1. The experiments at the office building (E₁) and the university (E₂) were regarded as separate experiments instead of a single experiment with larger sample size. This separation mitigates the influence of nuisance factors such as those arising from heterogeneous work environments or device setups.

2. The participants in each experiment were randomly selected without preferences toward age, gender, work function, or any other nuisance factors except for their willingness to participate. Such randomization is necessary to average out the effect of such nuisance factors that cannot be blocked.

3. The participants in each experiment phase were paired to their baseline selves as the control counterparts. Such a matched paired design blocks the effect of inter-participant variation due to difference in workloads and/or energy consumption preferences.

In addition to blocking and randomization, the sample size was so determined to allow for the study of statistical differences in the experimental responses. In this manner, the experiment was designed to ensure well-treated, random samples for the statistical analyses described in Section 3.

2.4. Feedback intervention design: Dashboard application

A dashboard was designed to provide the participants with information relevant to their plugload energy consumption. The elements of the dashboard were defined based on analytics that were previously found effective in motivating energy conservation among occupants in commercial buildings [17, 26, 27]. These analytics were represented by easily comprehensible elements with minimal cognitive and visual load [28]. The back end of the dashboard was implemented in PHP and the front end was implemented as a browser-independent webpage in HTML, JavaScript, jQuery, and HighCharts. The servers for the webpage were hosted on Amazon AWS EC2 and RDS instances. An image of the dashboard is shown in Figure 1. Each feature of the dashboard is described in section 2.4.1.

2.4.1. Dashboard features

1. Comfort feature (upper left): The comfort feature was represented by mutually exclusive radio buttons that allow participants to report their comfort levels. The options represented an ASHRAE 7-point scale [29]. This feature motivated participant engagement based on their historical interest in communicating their comfort levels to building facilities. Along with this feature, the dashboard also displayed energy consumption-related features.

2. Individual power feature (center): The instantaneous power consumed by the individual participant was pointed to by the needle in the dial. Similar visualizations were found effective for energy reduction in households [30]. The dial’s needle was set to saturate beyond the dial’s maximum reading. The dial was calibrated using data collected...
during the baseline phase. The average baseline power usage was computed by considering data points above 5 W. This average was chosen to represent the zenith of the dial for the participant under consideration. The 5 W threshold was chosen to avoid a participant’s inactivity from lowering the average value. The calibration also provided the color-coded context within which the current usage was positioned.

3. Scoreboard feature (upper right): The scoreboard feature provided participants with the score and relative position in the participant pool. When an incentive was provided, the participant with the highest score (rank 1) was declared the winner of the day. The scoring mechanism was designed to measure the improvement of the participant compared to his/her baseline, and is described in Section 2.4.2.

4. Serial power feature (lower left): The power series of an individual (in orange) relative to the pool (in green) was depicted by line charts in the serial power feature. Such social comparisons have proven successful in motivating energy reduction among participants [31, 32]. The vertical axis depicting power usage was scaled based on the individual and pool values during the time the dashboard window was active in the corresponding session.

5. Socket split feature (lower right): The instantaneous power consumed via the individual sockets in the powerstrip was represented by bar charts. While other features represented the participant’s cumulative power consumption across sockets, these bars provided actionable feedback by corresponding to the device plugged in the socket.

6. Notification feature (top right): A notification feature was provided in the dashboard to notify winners, if applicable.
2.4.2. Score computation

The scoreboard described above represents the participant’s score along with the relative position in the competition against other participants to win the incentive. The steps involved in the scoring mechanism are described below:

1. The time-averaged power consumption across each powerstrip socket was computed for the baseline phase by excluding data points below an inactivity threshold (5 W). The threshold served as a measure of inactivity.

2. The socket-specific averages computed above were aggregated over all the sockets assigned to a participant to obtain the average active baseline power consumption of a participant.

3. The above steps were repeated across all participants to obtain baselines for the score computation described below.

4. During each day of the incentive competition, each participant’s average active power consumption was determined similar to determining the baseline. The only procedural difference between the experiment and baseline computations was that the average power during the experiment was computed using data from local midnight till the scoring instant unlike the baseline computation which was performed using data from midnight to the next midnight.

5. The participant score was computed by the percentage improvement during the experiment compared to his/her baseline. That is, \( \text{score} = 900 + 100 \times \frac{\text{baseline\_average} - \text{expt\_average}}{\text{baseline\_average}} \), where \( \text{baseline\_average} \) and \( \text{expt\_average} \) represent the participant’s baseline average power (step 2) and the experiment day average power (step 4), respectively.

The choice of the score range in the interval [900, 1000] was motivated by previous findings that suggest a relationship between round numbers and human goals [33].

In this manner, the scoring mechanism assigns scores to the participants based on power reduction relative to their baseline consumption. We also note that an already energy-conscious participant may receive a low score due to a low baseline consumption and lack of ability for further energy reduction. This is expected since the study is designed to examine energy reduction irrespective of the baseline usage.

2.4.3. Inactivity detection

The inactivity threshold (5 W) mentioned above was unknown to the participants to ensure that no participant was declared to be the winner either due to inactivity or absence. However, the participants were informed that the scoring mechanism only rewards reducing power consumption via active changes as opposed to reducing the power consumption via passive changes such as turning off devices or being absent. While turning off unused devices could be an active change, such a consideration could also encourage unwanted effects such as leaving devices off or working from other locations during the experiment just for the sake of incentives. Thus, we consider powering off devices to be passive and

---

7The inactivity threshold was employed both in the scoring algorithm and the testing, estimation across all phases, thereby deeming any inferred power reduction conservative.
detect it using the inactivity threshold. Despite such inactivity measures, it was also possible that a participant could win
due to apparent activity such as leaving a single unused device on while turning off all other devices. In such cases, a
metric based on sliding time windows was used to detect participant inactivity. In this manner, the scoring algorithm was
designed to guard against winning strategies driven by inactivity.

2.5. Incentive intervention design

For the experiment phases involving incentives (P2C and P4C), a fixed monetary value was announced at the start of
each workday for participants to compete by changing their energy behavior compared to respective baselines. The values
of the incentives ranged between $5 and $50 in multiples of 5 over a duration of ten working days or two weeks. Based on
a random number generator, these values were randomly sampled without replacement to determine the incentive value
for the day. The random ordering of incentives ensured that any changes in the consumption could not be attributed to the
order in which the incentives were provided.

2.6. Data collection

The power consumption of devices associated with each participant were monitored in real-time by smart powerstrips
from Enmetric systems [24]. The monetary value associated with incentive inputs were recorded on a daily basis. As
noted in 2.2 the amount of feedback received by each participant was quantified by their time spent on the dashboard.
In what follows, we refer to this time spent on the dashboard as the screen time. The screen time was measured as the
aggregate duration spanned by the active sessions in the participant’s browser. This duration was obtained by software
running alongside the dashboard application.

2.7. Execution of the experiment

We describe the setup and implementation details for executing the experiment below.

2.7.1. Experiment setup

With the proposed design and permissions for experiments E_N and E_C, the participants were recruited. At NASA SB,
sixteen full-time employees were recruited for the experiment E_N. At CMU SV (buildings 19 and 23), a mix of faculty,
staff, and students totalling sixteen in number were recruited for experiment E_C. Smart powerstrip(s) were installed in
each participant’s workspace for collecting data during baseline and experiment phases. The devices plugged into the
powerstrips were noted as shown in Table 2.

2.7.2. Experiment E_N: NASA Sustainability Base

This experiment was conducted in two phases, a baseline phase and a feedback intervention phase. The baseline
phase (P1E) spanned five weeks from September 12, 2016 to 17 October, 2016, during which no interventions were
administered. Thereafter, the feedback intervention phase (P3E) was conducted during which the participants were
Table 2: List of devices associated with experiments in NASA SB (E_N) and CMU SV (E_C)

| Experiment | #Monitors | #Laptops | #Docking stations | #Desktops | #Headsets | #Landline telephones |
|------------|-----------|----------|-------------------|-----------|-----------|---------------------|
| E_N        | 21        | 11       | 9                 | 2         | 7         | 9                   |
| E_C        | 23        | 13       | 0                 | 8         | 0         | 6                   |

provided with dashboard feedback described in Section 2.4.1. The participants were provided with relevant explanation as shown in Figure 2(c). This phase was conducted for four weeks from October 18, 2016 to November 11, 2016.

2.7.3. Experiment E_C: CMU SV - buildings 19 and 23

The experiment E_C was conducted in four phases. The first phase was the baseline phase (P_1^C) during which no intervention was administered. This phase was conducted for five weeks from September 12, 2016 to October 17, 2016. The second phase of the experiment was the incentive-only phase (P_2^C) wherein monetary incentives were provided for participants to compete with the objective of winning the incentive. The individual with the highest score at local midnight was the winner of the day considered. The participants were also provided access to dashboards containing only the scoreboard, which showed their near real-time scores. An explanation of the relevant elements received by the participants during this phase is shown in Figure 2(a). The incentive-only phase was conducted for two weeks from October 18, 2016 to October 30, 2016. The third phase was the dashboard feedback only phase (P_3^C) during which each participant was provided with a dashboard depicting comparisons relative to their corresponding baseline and to the participant pool. All the dashboard features described in Section 2.4.1 except the scoreboard were provided to the participants. An explanation of the features shown in Figure 2(b) were provided to the participants. This phase was conducted for two weeks from October 31, 2016 to November 13, 2016. Finally, the both incentive and dashboard feedback phase was conducted during which the participants were provided with both the incentive and dashboard feedback. All the features of the dashboard were made available to the participants during this phase. The participants were provided with explanations of each feature as shown in Figure 2(c). This phase was conducted for two weeks from November 14, 2016 to November 25, 2016.

2.7.4. Energy conservation information

At the beginning of every experiment phase, namely P_3^N, P_2^C, P_3^C, P_4^C, the participants were provided with information on possible practices to reduce plugload energy consumption as shown in Figure 2(d). The specific choices on the appropriate implementations of these practices and compliance to safety regulations were left to the discretion of the participants. In this manner, the experimenters ensured that any absence of behavioral changes during the intervention phases could not be attributed to lack of information. These suggestions were compiled after surveying and classifying the devices used by each participant. The list of devices associated with experiments E_N and E_C are shown in Table 2.
3. Statistical analysis and modeling

In this section, we analyze the results from experiments $E_N$ and $E_C$. This analysis involves performing hypothesis tests, estimating confidence intervals, and developing statistical models from the data. Given that each of the baseline
and experiment phases for $\mathbb{E}_\mathbb{H}$ and $\mathbb{E}_\mathbb{C}$ were conducted over several days, we consider the following temporal context for analysis. Let the duration of each phase be represented as follows: For experiment $\mathbb{E}_\mathbb{H}$, let the set of days corresponding to $\mathcal{P}j^\mathbb{H}$ be denoted by $D^\mathbb{H}_j$, where $j \in \{1, 3\}$. Similarly, let the set of days corresponding to $\mathcal{P}k^\mathbb{C}$ be denoted by $D^\mathbb{C}_k$, where $k \in \{1, 2, 3, 4\}$. In addition, let $\mathbb{D}_W = \{\text{Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday}\}$ represent the days of a week. Further, in a phase $\mathcal{P}X^\mathbb{C}$ where $X \in \{\mathbb{H}, \mathbb{C}\}$, let the set of participants be represented by $\mathcal{P}f^X$. Let the $i^{th}$ participant of this set be denoted by $\mathcal{P}f^X_i$.

### 3.1. Data analysis for experiment $\mathbb{E}_\mathbb{H}$ at NASA SB

During the baseline phase $\mathcal{P}1^\mathbb{H}$, let the power consumption of the $i^{th}$ participant $\mathcal{P}3^\mathbb{H}_i$ on day $d \in D^\mathbb{H}_1$ at time instant $t$ be denoted by $y_{\mathcal{P}1^\mathbb{H}_i}(d, t)$ and let the time-averaged power consumption during $[t_0, t_f]$ be denoted by $\bar{y}_{\mathcal{P}1^\mathbb{H}_i}(d, t_0, t_f)$. Similarly, let the instantaneous and time-averaged power consumption during the feedback intervention phase $\mathcal{P}3^\mathbb{H}_i$ be denoted by $y_{\mathcal{P}3^\mathbb{H}_i}(d, t)$ and $\bar{y}_{\mathcal{P}3^\mathbb{H}_i}(d, t_0, t_f)$, respectively. Further, let this participant’s screentime during the time interval $[t_0, t_f]$ on day $d$ be denoted by $x_{\mathcal{P}3^\mathbb{H}_i}(d, t_0, t_f)$. Let the random variable corresponding to the baseline response of the $i^{th}$ participant during the time interval $[t_0, t_f]$ on day $d \in D^\mathbb{H}_1$ be denoted by the random variable $Y_{\mathcal{P}1^\mathbb{H}_i}(d, t_0, t_f)$. Similarly, let the random variables associated with the response and screentime input during the feedback phase be represented by $Y_{\mathcal{P}3^\mathbb{H}_i}(d, t_0, t_f)$ and $X_{\mathcal{P}3^\mathbb{H}_i}(d, t_0, t_f)$, respectively.

#### 3.1.1. Statistical assumptions

Given the baseline and experiment conditions, we represent the day of the week ($d \in \mathbb{D}_W$) statistics corresponding to $Y_{\mathcal{P}1^\mathbb{H}_i}(d, t_0, t_f)$ and $Y_{\mathcal{P}3^\mathbb{H}_i}(d, t_0, t_f)$ by $\mu_{\mathcal{P}1^\mathbb{H}_i}(d, t_0, t_f)$ and $\mu_{\mathcal{P}3^\mathbb{H}_i}(d, t_0, t_f)$, respectively. For hypothesis testing and interval estimation, we consider the sample constituted by the differences in the daily-averaged experiment response and the corresponding baseline response, sampled across participants and days of the week $\mathbb{D}_W$. The sample elements are assumed to be independent across the days of the week and the participant pool. Let this response differential for the $i^{th}$ participant on a day $d \in \mathbb{D}_W$ be represented by the random sample $\mu_{\mathcal{P}1^\mathbb{H}_i}(d, 0, 86400) - \mu_{\mathcal{P}3^\mathbb{H}_i}(d, 0, 86400)$. The response differential across participants and days of the week mitigates the corresponding nuisance factors stated in Section 2.3.

Given these matched pairs, statistical testing allows us to attribute any significant changes between the baseline and experiment responses to the intervention administered rather than to nuisance factors such as differences in individual energy needs or differences due to daily work schedules.

#### 3.1.2. Hypothesis testing and confidence interval estimation

We employ a paired difference test to examine the differential population sampled across participants and days of the week. Given the matched pairs $\mu_{\mathcal{P}1^\mathbb{H}_i}(d, 0, 86400)$ and $\mu_{\mathcal{P}3^\mathbb{H}_i}(d, 0, 86400)$, the paired difference t-test checks if the mean differential sample is significantly different from zero. The null and the alternative hypotheses are presented below:

1. $H_0^\mathbb{H}_0$: $\mu_{\mathcal{P}1^\mathbb{H}_i}(d, 0, 86400) - \mu_{\mathcal{P}3^\mathbb{H}_i}(d, 0, 86400)$ is sampled from a population with zero mean
2. $H^N_{A^{-}}: \mu_{P_1^N}(d, 0, 86400) - \mu_{P_3^N}(d, 0, 86400)$ is sampled from a population with non-zero mean.

3.1.3. Regression-based modeling

Given the statistical significance, it is of interest to predict the experiment phase power consumption based on a model. To model the hourly power consumption of an average participant, we employ an autoregressive model with an exogenous input consisting of the average screen time associated with the dashboard during the past hour. In place of specifying the start time ($t_0$) and end time ($t_f$) arguments, let the single argument $h \in \{1, ..., 24\}$ specify the hour of the day $d$. Therefore, we can write the experiment and baseline hourly power consumption of the $i^{th}$ participant during hour $h$ on day $d$ as $Y_{P_3^N}(d, h)$ and $Y_{P_1^N}(d, h)$, respectively. Similarly, the intervention variable can be written as $X_{A}^{A}(d, h)$. Also, let ':' denote an index representing the sample statistic when used in place of 'i', the index corresponding to the $i^{th}$ participant. Instead of explicitly modeling the experiment hourly power consumption of an average participant $Y_{P_3^N}(d, h)$, we model the difference between the averaged baseline and experiment responses, $\mu_{P_3^N}(d, h) - Y_{P_3^N}(d, h)$. The paired difference mitigates subjective variation due to individual energy consumption arising from varied energy preferences or work schedules, thereby enabling better prediction. Let this mean differential response be represented by $\Delta Y_{\bar{i}}(d, h)$. To model this response, we considered various model families differing in structural complexity. These included timeseries regression using linear, polynomial, and logarithmic functions; neural networks; kernel regression; and Gaussian process regression. It was observed that increasing the model complexity did not necessarily translate to tangible improvements in performance (< 5%). Based on the evaluation, the autoregressive model structure with exogenous inputs (ARX) shown below is chosen:

$$\Delta Y_{\bar{i}}(d, h) = \alpha^N + \beta^N \Delta Y_{\bar{i}}(d, h - 1) + \delta^N X_{A}^{A}(d, h - 1) + \epsilon^N(d, h)$$

(1)

where, $\epsilon^N(d, h)$ represents the error process following a Gaussian distribution $N(0, \sigma^N_\epsilon)$. The introduction of the lagged variable $\Delta Y_{\bar{i}}(d, h - 1)$ is instrumental in weakening the residual serial correlation and thus mitigates systematic factors in the error process as shown in Figure 4(a). It depicts the impact of adding time-lagged dependent variables on the serial correlation of the residuals. It is evident that the first order lag significantly reduces the correlation and the introduction of further lags do not contribute toward reducing the correlation further. From an experiment perspective, the time-lagged dependent variable enables us to account for changes between experiment conditions with respect to the baseline conditions. For example, any change in workload between the baseline and the experiment conditions can be captured by the introduction of the time-lagged dependent term in the model. This allows us to strengthen the assumption that the residuals corresponding to consecutive hours are a result of random factors and hence uncorrelated given the inputs. Given the justified model in Equation 1 we train and test it against the data collected from experiment $E_{\bar{i}}$ and the results are discussed in Section 4.1.2.

---

8The difference is constructed by considering the baseline week day $d \in D_W$ corresponding to the experiment day $d \in D_3^N$. 

---

12
3.2. Data analysis for experiment $E_C$ at CMU SV

The experiment $E_C$ was conducted in four phases: A baseline phase and three experiment phases. The experiment phases $P^2_C$, $P^3_C$, and $P^4_C$ consisted of interventions in the form of incentives, feedback, and both incentives and feedback, respectively. Similar to the experiment $E_{n1}$, during phase $P^k_C$ ($k \in \{1, 2, 3, 4\}$), let the instantaneous power consumption of the $i^{th}$ participant $P^k_{C,i}$ on the day $d \in D^k_C$ at instant $t$ be denoted by $y_{P^k_{C,i}}(d,t)$, and let the average power consumption during $[t_0,t_f)$ be denoted by $\bar{y}_{P^k_{C,i}}(d,t_0,t_f)$. Also, let the incentive and feedback provided during the time interval $[t_0,t_f)$ for the respective phases be denoted by $x^I_{P^k_{C,i}}(d,t_0,t_f)$, $(k \in \{3,4\})$ and $x^A_{P^k_{C,i}}(d,t_0,t_f)$, respectively.

For inference, we regard the observations as the realizations of a random sample from the occupant population. Similar to the experiment $E_{n1}$, we use $Y_{P^k_{C,i}}(d,t)$, $Y_{P^k_{C,i}}(d,t_0,t_f)$, $X^A_{P^k_{C,i}}(d,t_0,t_f)$, and $X^I_{P^k_{C,i}}(d,t_0,t_f)$ to denote random variables corresponding to the $i^{th}$ participant and the experiment day $d \in D^k_C$. Thus, the random variables pertaining to the response of the $i^{th}$ participant on day $d$ during the time interval $[t_0,t_f)$ corresponding to each of the phases $P^1_C$, $P^2_C$, $P^3_C$, and $P^4_C$, by convention, become $Y_{P^1_{C,i}}(d,t_0,t_f)$, $Y_{P^2_{C,i}}(d,t_0,t_f)$, $Y_{P^3_{C,i}}(d,t_0,t_f)$, and $Y_{P^4_{C,i}}(d,t_0,t_f)$, respectively. Similarly, the corresponding random variables representing the interventions during each of the three experiment phases $P^2_C$, $P^3_C$, and $P^4_C$ become $(X^A_{P^2_{C,i}}(d,t_0,t_f),X^I_{P^2_{C,i}}(d,t_0,t_f))$, $(X^A_{P^3_{C,i}}(d,t_0,t_f))$, and $(X^A_{P^4_{C,i}}(d,t_0,t_f),X^I_{P^4_{C,i}}(d,t_0,t_f))$, respectively.

3.2.1. Statistical assumptions

Given the baseline and experimental conditions, let the day of the week $d \in D_W$ statistics corresponding to $Y_{P^1_{C,i}}(d,t_0,t_f)$, $Y_{P^2_{C,i}}(d,t_0,t_f)$, $Y_{P^3_{C,i}}(d,t_0,t_f)$, and $Y_{P^4_{C,i}}(d,t_0,t_f)$ by $\mu_{P^1_{C,i}}(d,t_0,t_f)$, $\mu_{P^2_{C,i}}(d,t_0,t_f)$, $\mu_{P^3_{C,i}}(d,t_0,t_f)$, and $\mu_{P^4_{C,i}}(d,t_0,t_f)$. For performing inference, similar to experiment $E_{n1}$, we consider the sample constituted by the differences in the daily-averaged experiment response and the corresponding baseline response, sampled across participants and days of the week. For a day $d \in D_W$ in the experiment phase $j \in \{2,3,4\}$, let this averaged response differential for the $i^{th}$ participant be represented by the random sample $\mu_{P^j_{C,i}}(d,0,86400) - \mu_{P^j_{C,i}}(d,0,86400)$. The sample elements are assumed to be independent across the days of the week and the participant pool. Given the matched pairs design similar to Section 3.1.1 any inferences from the differential sample can be attributed to the intervention(s) administered during $P^j_C$ instead of the nuisance factors stated in Section 2.3.

3.2.2. Hypothesis testing and confidence interval estimation

Given the assumptions about the population consisting of the differential responses we employ a paired difference t-test to draw inferences about the underlying population. For each of the experiment phases $P^j_C$, $j \in \{2,3,4\}$, hypothesis tests and confidence interval estimation are performed on the mean of the differential response $\mu_{P^j_{C,i}}(d,0,86400) - \mu_{P^j_{C,i}}(d,0,86400)$. Thus, the null and alternate hypothesis for $P^j_C$ become:

---

9The scoreboard feature of the dashboard was made visible to the participants even during the incentive phase ($k = 2$) to let them monitor their position in competing for the incentive. Hence, in $E_C$, the screen-time variable is applicable to all the experiment phases $P^k_C$ ($k \in \{2,3,4\}$).
1. $H_0^C : \mu_{P_1}(d, 0, 86400) - \mu_{P_1'}(d, 0, 86400)$ is sampled from a population with zero mean

2. $H_A^C : \mu_{P_1}(d, 0, 86400) - \mu_{P_1'}(d, 0, 86400)$ is sampled from a population with non-zero mean

3.2.3. Regression-based modeling

Given the interval estimates, we are interested in a predictive model akin to the one in Section 3.1.3. We employ a similar notation here. In case of experiment $E_N$, dashboard feedback was the only intervention used and hence screen time was the only exogenous variable considered. However, in experiment $E_C$, each phase consists of either an incentive intervention (in phase $P_2^C$) or a dashboard feedback intervention (in phase $P_3^C$) or both (in phase $P_4^C$). For modeling purposes, we note that each observation in phase $\kappa$ can have a non-negative value for each of the intervention variables $x_A^I_{P_k^C}(d, h)$ and $x_I^I_{P_k^C}(d, h)$, thereby simultaneously accommodating both exogenous inputs into the model structure. Further, let the baseline and experiment hourly power consumption of the $i^{th}$ participant during hour $h$ on the experiment day $d$ be denoted by $Y_{P_1^C}(d, h)$ and $Y_{P_e^C}(d, h)$, respectively. We then model the mean differential response $\Delta Y_C(d, h) := \mu_{P_1^C}(d, h) - Y_{P_e^C}(d, h)$ by an ARX model consisting of screen time, the incentive, and their interaction as the inputs. Similar to the model selection process described in Section 3.1.3, we evaluated various model structures prior to selecting the following ARX model:

$$\Delta Y_C(d, h) = \alpha_C + \beta_C \Delta Y_C(d, h-1) + \gamma_C x_A^I_{P_e^C}(d, h-1) + \delta_C x_I^I_{P_e^C}(d, h-1) + \tau_C x_{A_e^C}(d, h-1) + x_{A_e^C}(d, h-1) + \epsilon^C(d, h) \quad (2)$$

where, $\epsilon^C(d, h)$ represents the error process following a Gaussian distribution $N(0, \sigma^C_e)$. The introduction of the lagged dependent term $\Delta Y_C(d, h-1)$ is instrumental in weakening the residual serial correlation. Figure 4(b) depicts the relationship between the number of added lags and the residual correlation. It is evident that the additional lags do not add further systematic information about the predicted variable and hence do not significantly contribute toward weakening the residual serial correlation. From an experiment standpoint, these lags capture the change in experiment conditions as compared to the baseline conditions, thereby strengthening the assumption that the residuals corresponding to consecutive hours are uncorrelated given the model inputs. The justified model in Equation 2 is trained and tested against the data collected from experiment $E_C$, and the results are discussed in Section 4.2.2.

4. Results and discussion

The results from hypotheses testing, interval estimation, and regression modeling for the experiments $E_N$ and $E_C$ are presented in Sections 4.1 and 4.2 respectively. These results are discussed in Section 4.3.

4.1. Results from experiment $E_N$

As per the design formulated in Section 3.1, we present the results from the government office experiment below.

14
4.1.1. Hypothesis testing and confidence interval estimation: Consider the hypotheses in Section 3.1.2. The mean baseline power consumption and the mean feedback phase power consumption are 51.51 W and 48.86 W, respectively. The t-statistic is found to be t(86)=3.64, and the corresponding p-value is $p = 4.61 \times 10^{-4}$. Therefore, the evidence against the null hypothesis is statistically significant at $\alpha = 0.05$. Further, we note that the 95% confidence interval of the mean of the differential sample $\mu_{P_{1N}^d}(d, 0, 86400) - \mu_{P_{3N}^d}(d, 0, 86400)$ is [2.22, 7.57] W. Thus, we conclude that the mean power consumption during the feedback phase $P_{3N}$ is (statistically) significantly less than that of the baseline phase $P_{1N}$. The summary data and test results are presented in Figure 3(a) and Table 3, respectively.

4.1.2. Autoregressive model: Based on the model designed in Section 3.1.3, we estimate the parameters on the training set using 70% of the data. These parameters are estimated by Ordinary Least Squares (OLS) and are provided in Table 5. The coefficients $\beta_{3N} = 0.8042$ and $\delta_{3N} = 0.0019$ represent the effect of last hour power differential in Watts and the effect of change in screen time in seconds, respectively. While both coefficients are positively correlated with the predicted variable, the difference in their magnitudes stems from the differences in the units of the predictors and their predictive
contributions. The intercept $\alpha_N = -0.0298$ represents the constant residual effect after adjusting for the predictor variables given the zero-mean gaussian process $\varepsilon^N(d, h)$. The model performance on the test dataset is shown in Figure 5(a). The RMS residuals on the training and test sets are 3.51W and 3.53W, respectively. The RMS accuracy of the model on the test set is $77.39\%$.

4.2. Results from experiment $E_C$

As per the design formulated in Section 3.2, we present the results from the university experiment below.

4.2.1. Hypothesis testing and confidence interval estimation: Similar to experiment $E_N$, we consider the hypotheses for each of the intervention phases as outlined in Section 3.2.2.

4.2.1.1. Incentive phase ($P2$): We find the mean baseline power consumption and the mean incentive phase power consumption to be 60.90 W and 53.91 W, respectively. The t-statistic is found to be $t(74) = 1.62$, and the corresponding p-value is $p = 0.11$. The 95% confidence interval of the mean of the differential response $\mu_{P2}^C(d, 0, 86400) -$
\[ \mu_{P_2}(d, 0, 86400) = [-1.84, 17.63] \text{ W}. \] Thus, the mean power consumption during the incentive phase is not statistically different from the baseline power consumption at \( \alpha = 0.05 \). The summary data and test results are presented in Figure 3(b) and Table 4 respectively.
4.2.1.2. Feedback phase (P3): We find that the mean baseline power consumption and the mean feedback phase power consumption are 60.90 W and 49.27 W, respectively. The t-statistic is found to be $t(75) = 2.26$, and the corresponding p-value is $p = 0.03$. The 95% confidence interval of the mean of the differential response $\mu_{p^3_1}(d, 0, 86400) - \mu_{p^3_1}(d, 0, 86400)$ is $[1.58, 24.82]$ W. Thus, the mean power consumption during the feedback phase is statistically significantly less than that of the baseline phase at $\alpha = 0.05$. The summary data and test results are presented in Figure 3(c) and Table 4, respectively.

4.2.1.3. Feedback and Incentive phase (P4): We find that the mean baseline power consumption and the mean feedback & incentive phase power consumption are 60.90 W and 50.33 W, respectively. The t-statistic is found to be $t(67) = 2.30$, and the corresponding p-value of $p = 0.02$. The 95% confidence interval of the mean of the differential response $\mu_{p^4_1}(d, 0, 86400) - \mu_{p^4_1}(d, 0, 86400)$ is $[1.96, 27.63]$ W. Thus, the mean power consumption during the feedback & incentive phase is statistically significantly less than that of the baseline phase at $\alpha = 0.05$. The summary data and test results are presented in Figure 3(d) and Table 4, respectively.

4.2.2. Autoregressive model: Consider the model designed in Section 3.2.3. We estimate the parameters with 70% data via OLS similar to that of Section 4.1.2. These estimates are listed in Table 6. The coefficient $\beta^c$ is 0.7679 represents the effect of the last hour power differential in Watts. Similarly, the predictive effects of the incentive in $\$, the dashboard feedback in seconds, and their interaction in $\$-$seconds are represented by $\gamma^c = -0.0048$, $\delta^c = 0.0045$, and $\tau^c = 0.0001$, respectively. The switch of signs between the incentive and the interaction coefficients suggests a crossover interaction between the feedback and the incentive for purposes of prediction. The intercept term $\alpha^c = 2.4667$ represents the constant residual effect after accounting for the predictors given the zero-mean gausian process $\epsilon^c(d, h)$. The model performance on the test dataset is shown in Figure 5(b). The RMS residuals on the training and test sets are 3.95 W and 4.09 W, respectively. The RMS accuracy of the model on the test set is $\approx 76.38\%$. 

| Property | Coeff. | $\alpha^c$ | $\beta^c$ | $\gamma^c$ | $\delta^c$ | $\tau^c$ | $\sigma^c_\epsilon$ |
|----------|--------|------------|-----------|------------|-----------|----------|------------------|
| Point estimate | 2.4667 | 0.7679 | -0.0048 | 0.0045 | 0.0001 | 3.9678 |
| 95% CI [lower] | 2.4399 | 0.7658 | -0.0056 | 0.0024 | -0.0001 | 3.7441 |
| 95% CI [upper] | 2.4935 | 0.7700 | -0.0040 | 0.0066 | 0.0003 | 4.1916 |

Table 5: Regression model parameter estimates from experiment $\mathbb{E}_N$

Table 6: Regression model parameter estimates from experiment $\mathbb{E}_C$
4.3. Discussion

The findings from experiment $E_N$ reveal that the dashboard feedback offers a statistically significant ($p = 4.16 \times 10^{-4}$) reduction in office plugload power consumption. The average reduction is 9.52%, and the 95% confidence interval is [4.32%, 14.71%]. This reduction is lower than the 15 – 40% commercial office energy savings estimated in [14, 15] or the 12% – 20% reduction deemed possible by behavioral modification in [17]. Similarly, the data from experiment $E_C$ also suggests that dashboard feedback results in a statistically significant ($p = 0.03$) power reduction in the university environment. The average reduction is 21.61%, and the 95% confidence interval is [2.59%, 40.63%]. This reduction is lower than the 30% reduction reported by a previous study within a university environment [26]. Our estimates may be considered conservative since they reflect power reduction only through active behavioral changes as noted in Section 2.4.3. We also note that the mean reduction in the government office (9.52%) is tangibly different than that of the university environment (21.61%). This difference could be the result of differing power reduction potentials driven by the device lists in both environments as shown in Table 2. The office environment consisted of two desktops compared to eight at the university. Desktop computers are known to be among the largest occupant plugloads with a substantial reduction potential [14]. Consequently, the power reduction realized at the university environment is larger than that of the office environment. In this manner, the compressibility of energy demand, or lack thereof, determines the scope for power reduction.

The findings from experiment $E_C$ suggest that the incentive, dashboard feedback, and their combination resulted in a mean reduction of 12.93% ($p = 0.11$), 21.61% ($p = 0.03$), and 24.22% ($p = 0.02$), respectively. It is noteworthy that the incentive intervention corresponds to a larger p-value and hence less significant than the dashboard or the combined intervention. A possible explanation is to consider the order of interventions. The first experiment phase $P_{2C}$ consisted of the incentive and the later phases $P_{3C}$ and $P_{4C}$ consisted of the feedback and the combination, respectively. The growth of practical and statistical significance in the order of phases is suggestive of the effect of time on plugload power consumption behavior. This suggestion is consistent with the finding that behavioral changes require adaptation time prior to the formation of habits [23]. From the point of sustainable energy reduction, it is desirable that the interventions are instrumental in cultivating habits that persist over time [34]. To account for the temporal effect on energy behavior, future studies could include an adaption or settle-in time during the experiment design. Further, to examine persistence and fade-out effects from one phase to another, a sufficient washout period and re-baselining need to be considered. While these considerations increase the experiment duration, they allow for a systematic investigation of the temporal effects of interventions on occupant power consumption behavior.

The ARX model coefficients for both experiments are provided in tables 5 and 6. In both models, the coefficients $\beta_N = 0.8042$ and $\beta_C = 0.7679$ indicate significant predictive information in the lagged power differential. Given the models’ accuracies, the lower effect size magnitudes of $\gamma_C$, $\delta_N$, $\delta_C$, and $\tau_C$ suggest the need to track other variables to model intervention effects better. Ideally, these variables would offer insights into participant attention, perception, and
behavior pertaining to plugload energy consumption. The difference in magnitudes of the constant terms $\alpha^N = -0.0298$ and $\alpha^C = 2.4667$ can be understood by considering the underlying AutoRegressive (AR(1)) processes in the absence of interventions. Let $Q \in \{N, C\}$. Given that $|\beta^Q| < 1$, we can treat these AR(1) processes as Wide Sense Stationary (WSS) and write $\alpha^Q = (1 - \beta^Q)\mu_{\Delta Y_Q}$, where $\beta^Q$ and $\mu_{\Delta Y_Q}$ are respectively the first order autocorrelation and mean of the hourly differential $\Delta Y_Q$. Thus, the magnitudes of $\alpha^N$ and $\alpha^C$ differ due to the differences between participants’ aggregate energy behavior in both environments.

The training and test RMS residuals of both ARX models are similar as noted in sections 4.1.2 and 4.2.2. This similarity suggests that the models are not overfitting. Further, the test set residual plots are shown in figures 4(c) and 4(d). The residual behavior does not suggest heteroscedasticity. From figures 4(a) and 4(b), we note that the residual process is not autocorrelated. To the extent each of the underlying Gaussian error processes $e^N$ and $e^C$ are uncorrelated and homoscedastic, the respective OLS estimators can be regarded as the Best Linear Unbiased Estimator (BLUE) based on the Gauss-Markov theorem. The RMS accuracy of the office and university models on the test set are 77.39% and 76.38%, respectively. The prediction error can be a product of factors related to modeling, estimation, or observation errors [36]. The observed significant reduction in plugload power consumption could be the result of behavioral changes induced either by the dashboards or cognitive factors such as the Hawthorne effect where the awareness of the participants about their energy consumption being monitored changes their energy behavior [37].

While the experiments are carried out in different buildings, the methodology to elicit the influence of visual and incentive interventions on occupant plugload usage is generalizable across buildings. We suggest the future studies to consider, (1) larger timescale and sample size with appropriate settle-in and washout periods, and (2) multimodal tracking of participant behavior to obtain further insights into occupant plugload energy consumption, including the persistence and fade-out effects of interventions. Such insights, apart from verifying the extent of generalizability, enable more accurate models comprising significant predictive information.

5. Conclusion

In this work, the problem of improving building energy efficiency through plugload management is considered. The changes in occupant plugload energy consumption due to monetary incentives and/or visual feedback were investigated by conducting experiments in office and university buildings. These experiments employed a matched pairs design to strengthen the causal connection between plugload consumption and the corresponding intervention used. During different phases of the experiments, interventions in the form of monetary incentives and/or dashboard feedback were provided. The incentives were offered in a random order and the dashboard was constructed with regard to occupant engagement and plugload consumption awareness. The experiment in the office environment was conducted at NASA Sustainability Base

---

10In the strict sense, this inequation is an assumption about the population parameter $\beta^Q$ by observing that its 95% CI estimates in tables 5 and 6 are less than 1 in magnitude. In general, the numbers associated with the regression coefficients (parameters) in this work represent their sample estimates.
in the presence of dashboard feedback. The average plugload reduction was found to be 9.52\% (\( p = 4.61 \times 10^{-4} \)) and the regression model RMS accuracy was found to be 77.39\%. The experiment in the university environment was conducted at the CMU Silicon Valley campus in the presence of incentives and/or dashboard feedback. The average plugload reduction in the presence of incentives, dashboard feedback, and their combination was found to be 12.93\% (\( p = 0.11 \)), 21.61\% (\( p = 0.03 \)), and 24.22\% (\( p = 0.02 \)), respectively. The regression model RMS accuracy for the university experiment was found to be 76.38\%. Findings from both experiments indicate that feedback intervention can be effective in both office and university environments with an estimated mean reduction of 9.52\% and 21.61\%, respectively. The proposed models potentially enable the integration of occupant plugload consumption control into demand response paradigms for achieving a low-carbon society. Future studies should investigate experiment designs with larger sample sizes, persistence of effects, load shifting mechanisms, sustainable interventions, and generalizability across building types and cities.

Acknowledgments

The authors thank the NASA Ames Research Center, Carnegie Mellon University (CMU), and their Institutional Review Boards (IRBs) for supporting this research under the cooperative agreement NNX13AD49A.

References

[1] IEA, Buildings, The critical role of buildings, International Energy Agency (2019).
[2] IEA, Paris, Tracking buildings 2020, International Energy Agency (2020).
[3] N. Hashempour, et al., Energy performance optimization of existing buildings: A literature review, Sustainable Cities and Society (2020).
[4] S. Longo, F. Montana, E. R. Sanseverino, A review on optimization and cost-optimal methodologies in low-energy buildings design and environmental considerations, Sustainable cities and society 45 (2019) 87–104.
[5] US Energy Information Administration, Annual energy outlook 2020: With projections to 2050 (2020).
[6] NREL, Assessing and Reducing Plug and Process Loads in Office Buildings, US Department of Energy (2010).
[7] K. McKenney, et al., Commercial miscellaneous electric loads: Energy consumption characterization and savings potential in 2008 by building type, TIAX LLC, Lexington, MA, Tech. Rep.D (2010).
[8] F. Sehar, M. Pipattanasomporn, S. Rahman, Integrated automation for optimal demand management in commercial buildings considering occupant comfort, Sustainable cities and society 28 (2017) 16–29.
[9] H. F. Castleton, et al., Green roofs; building energy savings and the potential for retrofit, Energy and buildings (2010).
[10] T. I. Salsbury, A survey of control technologies in the building automation industry, IFAC Proceedings Volumes 38 (1) (2005) 90–100.
[11] Y.-J. Wen, et al., Personalized dynamic design of networked lighting for energy-efficiency in open-plan offices, Energy and Buildings (2011).
[12] D. Kaneda, B. Jacobson, P. Rumsey, R. Engineers, Plug load reduction: The next big hurdle for net zero energy building design, in: ACEEE Summer Study on Energy Efficiency in Buildings, 2010, pp. 120–130.
[13] D. Agdas, R. S. Srinivasan, K. Frost, F. J. Masters, Energy use assessment of educational buildings: Toward a campus-wide sustainable energy policy, Sustainable Cities and Society 17 (2015) 15–21.
[14] C. Mercier, L. Moorefield, Commercial Office Plug Load Savings and Assessment: Executive Summary, California Energy Commission (2011).

[15] B. Acker, C. Duarte, K. Van Den Wymelenberg, Office space plug load profiles and energy saving interventions, Proc. of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA (2012).

[16] R. K. Jain, et al., Assessing eco-feedback interface usage and design to drive energy efficiency in buildings, Energy and buildings (2012).

[17] R. Yun, et al., Sustainability in the workplace: nine intervention techniques for behavior change, in: Persuasive Technology, Springer, 2013.

[18] J. E. Petersen, et al., Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives, International Journal of Sustainability in Higher Education 8 (1) (2007) 16–33.

[19] P. Gandhi, G. S. Brager, Commercial office plug load energy consumption trends and the role of occupant behavior, Energy and Buildings (2016).

[20] M. Bourdeau, X. qiang Zhai, E. Neftaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: A review of data-driven techniques, Sustainable Cities and Society 48 (2019) 101533.

[21] NASA, NASA Sustainability base, http://www.nasa.gov/ames/facilities/sustainabilitybase (2016).

[22] C. Poolla, A. K. Ishihara, R. Milito, Designing near-optimal policies for energy management in a stochastic environment, Applied Energy (2019).

[23] P. Lally, C. H. Van Jaarsveld, H. W. Potts, J. Wardle, How are habits formed: Modelling habit formation in the real world, European journal of social psychology 40 (6) (2010) 998–1009.

[24] Enmetric Plug Load Management System, https://www.enmetric.com/platform, accessed: 2016-10-29 (2016).

[25] D. C. Montgomery, Design and analysis of experiments, John Wiley & Sons, 2008.

[26] R. Yun, et al., Toward the design of a dashboard to promote environmentally sustainable behavior among office workers, in: Persuasive Technology, Springer, 2013, pp. 246–252.

[27] R. Gubina, J. E. Taylor, Effects of real-time eco-feedback and organizational network dynamics on energy efficient behavior in commercial buildings, Energy and Buildings 84 (2014) 493–500.

[28] R. Brath, M. Peters, Dashboard design: Why design is important, DM Direct (2004).

[29] R. de Dear, et al., Thermal comfort in naturally ventilated buildings: revisions to ashrce standard 55, Energy buildings (2002).

[30] P. Petkov, F. Köbler, M. Foth, H. Krcmar, Motivating domestic energy conservation through comparative, community-based feedback in mobile and social media, in: Proc. 5th Intl. Conf. on Communities and Technologies, ACM, 2011, pp. 21–30.

[31] H. Allcott, Social norms and energy conservation, Journal of Public Economics 95 (9) (2011) 1082–1095.

[32] I. Ayres, S. Raseman, A. Shih, Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage, Journal of Law, Economics, and Organization (2012) ewso20.

[33] D. Pope, U. Simonsohn, Round numbers as goals: Evidence from baseball, sat takers, and the lab, Psychological science 22 (1) (2011) 71–79.

[34] E. Frey, et al., Persistence: How treatment effects persist after interventions stop, Policy Insights from the Behavioral and Brain Sciences (2014).

[35] J. D. Hamilton, Time series analysis, Princeton university press, 2020.

[36] G. Shmueli, To explain or to predict?, Statistical science 25 (3) (2010) 289–310.

[37] H. Kosonen, A. Kim, Quantifying plug load energy use in a leed gold building—lessons learned in the installation phase, in: Construction Research Congress 2016, 2016, pp. 1234–1243.