Energy & Buildings 278 (2023) 112599

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Unlocking demand response in commercial buildings: Empirical response of commercial buildings to daily cooling set point adjustments

Jacques A. de Chalendar [a,*], Caitlin McMahon [a], Lucas Fuentes Valenzuela [b], Peter W. Glynn [c], Sally M. Benson [a]

[a] Department of Energy Resources Engineering, Stanford University, Stanford 94305, California, United States
[b] Department of Electrical Engineering, Stanford University, Stanford 94305, California, United States
[c] Department of Management Science and Engineering, Stanford University, Stanford 94305, California, United States

ARTICLE INFO

Article history:
Received 7 August 2022
Revised 16 October 2022
Accepted 20 October 2022
Available online 02 November 2022

Keywords:
Commercial building flexibility experiments
Data-driven building energy models
Thermal demand response
Temperature set point experiments
Commercial HVAC flexibility potential
Grid-interactive efficient buildings

ABSTRACT

Buildings represent over half of global electricity demand. Cooling buildings already accounts for over 9% of global electricity demand and is expected to grow rapidly due to climate-change induced hot-spells and increasing prosperity in developing economies. In the US, commercial buildings represent 35% of nationwide electricity consumption. Increased electricity demand for cooling services will challenge already stressed power grids, particularly during times of peak demand. This work explores the flexibility and demand response potential of large Heating, Ventilation and Air Conditioning (HVAC) systems based on an extensive set of measurements from six commercial buildings in a Warm-summer Mediterranean climate. Over a three-month summer period, zone-level temperature set points were adjusted daily in six commercial buildings to determine the effect on chilled water and electricity loads, as well as on zone-level temperatures. External weather conditions were measured continuously during the testing period. The experimental data that were collected are published with this article. These experiments confirmed the potential to provide flexibility by reducing energy demands based on modest zone-level temperature set point adjustments. A two-degree Fahrenheit increase of the cooling set point resulted in a 13–28% reduction in daily building-level cooling loads on average for four office buildings and 3–4% for two laboratory buildings. The impact on electric loads was less than 2% (excluding for cooling water but including for ventilation). Zone-level temperature increases were measurable but temperatures remained within the target ranges. By collecting 385 experiment-days of experiment data, we were able to parameterize statistical models for the response of the buildings. These models provide statistical guarantees on the reliability of thermal demand response. This work provides a blueprint for constructing building and zone-level energy-response functions and highlights the value of testing buildings repeatedly and across a range of weather conditions. Providing statistical performance guarantees will be critical for widespread adoption of demand response technology to provide the flexibility needed to meet peak electricity demands. Combined with thermal storage, the daily flexibility studied here would also unlock daily and sub-daily electrical flexibility, and can also be integrated with sub-daily flexibility from building-level electrical loads.

1. Introduction

Energy for buildings represented 30% of global energy end-use in 2018 and 28% of the corresponding emissions (excluding construction) [1]. Despite efficiency gains in the past decade, global building energy use continues to grow, driven by population, heat waves, a growing middle-class, and floor area expansion. This is especially true for space cooling, that already represented 9% of world electricity use in 2018 [2,3]. This work focuses on the flexibility potential of the US commercial sector, that represented 35% of 2020 US electricity sales [4]. Active management of energy consumption in buildings can reduce the costs and the carbon impact of our energy systems and improve their resiliency and their efficiency. In commercial buildings, e.g. offices, retail sites, supermarkets, schools, laboratories, or data centers, Heating, Ventilation and Air Conditioning (HVAC) systems have been a prime target for energy flexibility for over a decade [5]. Accurate measurement tools are now needed to unlock mass deployment of Demand Response (DR) technologies. Beyond engineering calculations and

* Corresponding author.

https://doi.org/10.1016/j.enbuild.2022.112599
0378-7788/© 2022 The Author(s). Published by Elsevier B.V.

simulations, energy system operators require measurement-based models for the response of building energy systems to integrate them into their decision-making.

This paper presents and demonstrates practical and scalable experimental methods to construct quantitative, data-driven flexibility models for the energy behavior of commercial buildings in response to daily temperature set point adjustments, while controlling for weather conditions and occupancy. The flexibility model we propose is simple to interpret, calibrate, and update through repeated testing. It provides statistical information that is directly relevant to electric grid or district energy system operators calling on DR. More regular testing of buildings will also lead to energy performance benefits by stress testing the response of a building's overall energy system to controlled perturbations.

The paper reports on three important empirical findings from analyzing data generated during 385 experiment-days during the cooling season in six large office and laboratory buildings located on a university campus in Northern California (Warm-summer Mediterranean climate). First, buildings don't all respond identically to temperature set point changes nor to outside weather conditions, which further motivates the need for scalable modeling and testing methods. Second, a building's response to a set of given temperature set points, weather conditions and occupancy levels is uncertain. However, building HVAC systems can still be expected to provide a reliable DR resource in future energy systems, because that uncertainty can be quantified through repeated testing. Third, in contrast with several previous field studies where HVAC system flexibility was mainly provided by reduced electricity consumption from the ventilation fans [6] (note that those studies consider shorter time scales), in our experiments the HVAC system response due to reduced building cooling loads was much larger than from changes in fan loads. A possible explanation for this difference is the more recent building control system logic that is operational on the buildings in our testbed. A key implication from this third finding is that building flexibility models will need to capture the *full cyber-physical response* of buildings.

2. Context and related work

2.1. Motivation

As electricity grids progressively decarbonize, electrifying heating and cooling will increasingly become an option to decarbonize building energy systems. In 2018, electricity already represented 34% of world building energy use [1]. Buildings represented 53% of world electricity consumption.

In fast-changing electricity grids, existing flexibility needs are exacerbated by the integration of non-dispatchable wind and solar generation. Options include short and long duration electricity storage, transmission expansion, and demand-side management. Flexibility on durations of 10 h and longer is sorely needed [7]. The daily timescale could be a natural fit for energy systems serving building Heating, Ventilation and Air Conditioning (HVAC), where relevant timescales are typically slower than for other electricity-consuming energy assets.

Active management of building energy consumption offers a lower-cost alternative to sizing systems based on peak loads. It can reduce consumption, achieve environmental benefits, and increase resilience in both new and old buildings. A first set of opportunities exists through continuous optimization of a building's operation, without impacting occupant comfort. Second, deeper energy flexibility needs to be made available, whether it is to respond to DR events in a market or to enhance resiliency and adaptability during supply shortages or demand surges.

The U.S. Department of Energy estimates the value of the untapped opportunity for Grid-interactive Efficient Buildings (GEBs) to be between $8 billion and $18 billion annually by 2030, or 2–6% of total U.S. electricity generation and transmission costs ($100–200 billion cumulative benefits from 2021 to 2040, in 2019 dollars) [8]. Through demand flexibility and energy efficiency, GEBs could decrease US CO_2 emissions by 80 Mtons per year by 2030, 6% of total US power sector CO_2 emissions. US buildings currently account for over 70% of electricity use [4] and 33% of CO_2 emissions [9].

In electrified commercial buildings, flexibility from HVAC systems provides value for the electric grid in two coupled ways. First, DR improves resilience and lowers infrastructure requirements both for the building energy system and for the wider electric grid. DR is available both from the heating and cooling equipment (chillers, heat pumps, boilers) and from the ventilation fans, which typically represent a significant but smaller fraction of total electric load. Second, when heating and cooling infrastructure is shared, e.g. in a district energy system, building-level thermal DR also provides value to manage the portfolio of buildings and more efficiently schedule district-level thermal energy operations. In an electrified district energy system with thermal storage like this study's testbed [10], building-level thermal DR unlocks district-level electrical DR.

2.2. Temperature set point adjustment strategies

This work focuses on flexibility strategies based on temperature set point adjustments. The design and implementation of commercial building HVAC systems varies significantly. Their primary control objective does not and is to maintain occupant thermal comfort. This is typically treated as equivalent to maintaining the different zones within temperature "deadbands", defined by heating and cooling temperature set points [11].

In 22% of US commercial floorspace, the thermostats that control these deadbands are programmable [12] and are therefore a frequent target for flexibility strategies [5]. Temperature set point strategies do not require significant reprogramming of internal building energy management control systems, which is why they are sometimes referred to as "supervisory control". They are likely to be a scalable, easily implementable, and lower cost alternative to methods that issue commands to HVAC equipment directly. Numerical and field studies have also shown that they would be acceptable from the perspective of occupant comfort, especially if occupant comfort is explicitly considered in the control strategy [13].

The simplest strategy uses Global Thermostat Adjustments (GTA), in which one master thermostat controls the entire building [5]. A review of commercial buildings enrolled in mostly GTA-based DR programs in California from 2003 to 2010 concluded that they were able to provide 13% reductions in peak afternoon demand on average (from 5 to 15%) [14].

Physics-based digital twins for buildings (EnergyPlus, TRNSYS, or eQuest) can be used to simulate many different scenarios and to characterize the potential for demand flexibility. On the annual timescale, it was found that increasing the cooling set point from 72°F to 76°F reduces energy consumption by 10% in a reference medium-sized office building with strong ventilation constraints [15]. When changes in the zone-level minimum air flow rates were allowed, savings grew to 20% at 76°F, 30% at 78°F, and 40% at 80°F. Reported savings were averaged over seven US climate zones, and were highest for San Francisco, the simulated city that is closest to our testbed. In another study with reference office buildings of different sizes and construction types, re-optimizing the center of the deadband daily in the range of 22.5±3°C resulted in average annual savings of 10–37% versus maintaining the setpoint fixed at 22.5°C (these simulations used a 3°C-wide deadband) [16]. In both cases, the ranges reflect simulations for several climate zones and building sizes. While both studies targeted energy efficiency from more flexible temperature setpoint settings, their findings also suggest strong flexibility on a daily or hourly timescale with a similar strategy. On the hourly timescale, simulations suggest that the response of HVAC systems to faster changes is dynamic and varies with time of day, day of week, and outside weather conditions [17,18].

While physics-based numerical simulations are most often chosen to evaluate the impact of different control strategies, they cannot replace the value of data from field tests, especially as technologies near deployment. Comparisons of simulations to experiments show that it is often difficult to fully capture the physical and operating characteristics of real-world commercial buildings with simulations only [17,6].

Also, simulation models typically rely on inputs such as geometric and physical parameters from buildings. Generating and maintaining models for real buildings is time-consuming, both in cases where a small team manages a large portfolio of buildings and in cases where the building doesn't have a dedicated operator.

While simulations point to credible potential from temperature set point adjustments, they need to be validated on site. Of course, testing real buildings is challenging. Earlier field tests suffered from small sample sizes that made statistically significant results challenging to obtain [5]. Still today, experiments yield discrepancies and sometimes contradictory results between tested buildings, even when procedures are automated. Often cited drivers for discrepancies include 1) the selection of the baseline method and its accuracy, 2) limited site specific information, and 3) control accuracy [19,6].

The widespread adoption of programmable actuators and sensors makes testing cheaper, easier to automate, and to replicate. Driven by experimentation capabilities, experiments reported on in the literature increasingly involve more samples, supporting more quantitative statements, such as the ones that are possible in this paper. This is needed to build confidence in DR technologies. The variability and discrepancies that are often observed between field studies reinforce the need for standardized testing methods to enable mass DR deployment.

2.3. Relation to model predictive control approaches

Direct control of building equipment and actuators based on Model Predictive Control (MPC) technologies has been proposed by several researchers as a more energy efficient and robust alternative to the HVAC industry's current standard of using Ruled-Based Control (RBC) [11,20]. Beyond energy efficiency improvements, MPC-equipped buildings would also be easier to enroll in DR programs, because controllers could now be reprogrammed to account for the benefits of responding to DR events. But, despite over two decades of significant research activity, including several full-scale demonstrations [21–23], these approaches have not been widely adopted by industry [20].

In many of these prior studies, authors note that while MPC approaches have been shown to dramatically outperform naive control strategies (e.g. fixed controls), the performance gains are typically mild when compared to well-implemented RBC. Implementing MPC strategies also requires significant upfront engineering time and cost. Since building HVAC systems are not standardized, deploying the technology typically requires customized and non-scalable integration work to instrument a new unique set of actuators in each new building and tailoring models and control designs. Poor information and communications infrastructure is a strong limiting factor in commercial buildings and will remain so for the foreseeable future [12]. Beyond upfront costs, MPC strategies also require significant maintenance. Since the usage of buildings changes over time, e.g. due to tenant improvements or HVAC system retrofits, the models that MPC relies on will also need to be updated to remain accurate.

Our conclusion from the lack of adoption by industry is that a new research approach is now needed. There is strong value in researching less invasive control strategies that do not seek to replace the industry standard RBC strategies but instead to augment them, e.g., through distributed temperature set point strategies like those that are experimented with in this work. These less invasive strategies remain compatible with the general MPC framework. MPC would now be used for controlling higher level and more generic set points, possibly across a large number of buildings at once, in an approach sometimes referred to as supervisory control. These less invasive strategies will often be cruder, so they can and should use prior, equipment-level MPC results as an optimal benchmark. A key challenge to enabling MPC for supervisory control is of course the development of suitable models, which is a main goal of this work.

3. Materials and methods

3.1. Experiment design

Cooling set points were adjusted every two days in the morning and cycled repeatedly between a low (74°F, 23.3°C) and a high (76°F, 24.4°C) value. A DR scenario consistent with this simple experimental setup is one where occupants need to be notified one day in advance that tomorrow will be a DR day, so that they have the option not to come to work. Zone-level heating set points were fixed to 68°F (20°C) for the entire duration of the experiments, to minimize heating loads. Prior to the experiments, a number of sensitive zones were identified in each building and excluded from the experiments. All other zones received the same set point commands throughout the experiments. This set point strategy is similar to what is sometimes referred to as Global Temperature Adjustments (GTA) [5], but with the option to enforce the adjustment for only a fraction of the zones.

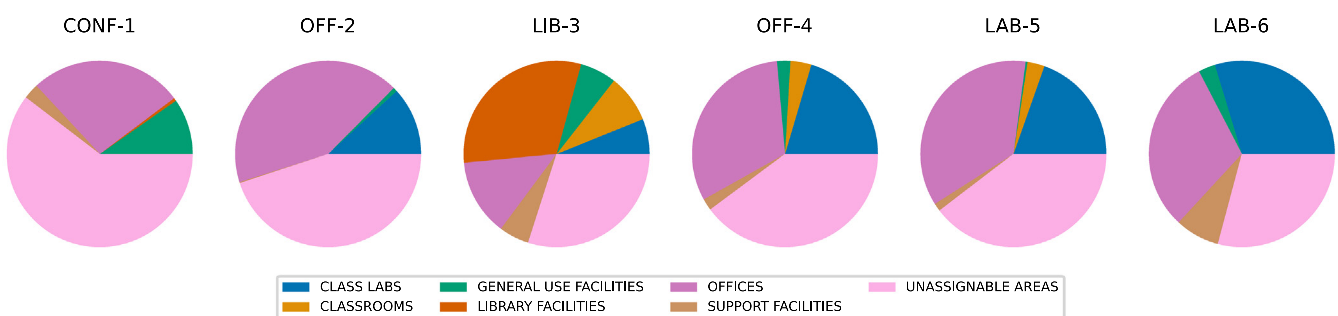
3.2. Experiment testbed

3.2.1. Buildings

The experiments were conducted on six commercial buildings on a university campus in Northern California (Warm-summer Mediterranean climate) during the summer of 2021. Summary characteristics for the buildings are presented in Table 1. Overall, the testbed covers 56,000 m^2. Three buildings are mainly comprised of office and classroom space, two have a significant fraction of laboratory space and the last building houses a conference center along with offices. 45 to 76 days of experiment data were collected for each building. Note that while the age of the buildings varies by over a century, all buildings underwent significant retrofits since their construction date. The oldest building was constructed in 1893 but its HVAC system underwent major upgrades in 2015. These buildings were compliant with regulations specified by California's Title 24 at the date of their construction or last major retrofit, so building envelopes and mechanical systems can be considered comparable to modern commercial buildings in California. The two laboratory buildings are much more energy intensive. Process chilled water loops are expected to represent a significant fraction of the cooling consumed by these buildings (not sub-metered). Temperature and ventilation requirements are also stricter in these buildings.

Table 1
Summary characteristics of the buildings in the experimentation testbed. OFF: office, CONF: Conference, LIB: Library, CLS: Classroom, LAB: Research laboratory. Building cooling and electricity loads are measured separately. Building electricity loads do not include energy required to produce cooling. Table S1 provides an indicative breakdown of floor space by usage.

	CONF-1	OFF-2	LIB-3	OFF-4	LAB-5	LAB-6	Total
Type	OFF/CONF	OFF/CLS	OFF/CLS	OFF/CLS	LAB/CLS	LAB/CLS	–
Year of construction	2000	1893	1996	1998	1965	1963	–
Year of last retrofit	2021	2015	2021	–	2020	2018	–
Average cooling (MJ/m^2/day)	0.68	0.69	0.37	1.24	4.84	2.30	1.35
Average cooling (kWh/m^2/day)	0.19	0.19	0.10	0.35	1.34	0.64	0.38
Average electricity (kWh/m^2/day)	0.10	0.09	0.15	0.28	0.81	0.59	0.28
Floors	4	3	7	4	5	5	–
Air Handling Units	5	2	13	4	2	4	17
Variable Air Volume systems	136	33	143	217	0	0	529
Fan Coil Units	0	50	0	12	117	166	345
Controlled zones	136	73	139	223	117	133	821
Excluded zones	0	10	4	6	0	33	53
Measured zones	136	83	143	229	117	166	874
Daily samples	45	63	54	75	72	76	385
Area (1,000 m^2)	13.5	2.6	15.8	9.8	7.0	7.2	55.7



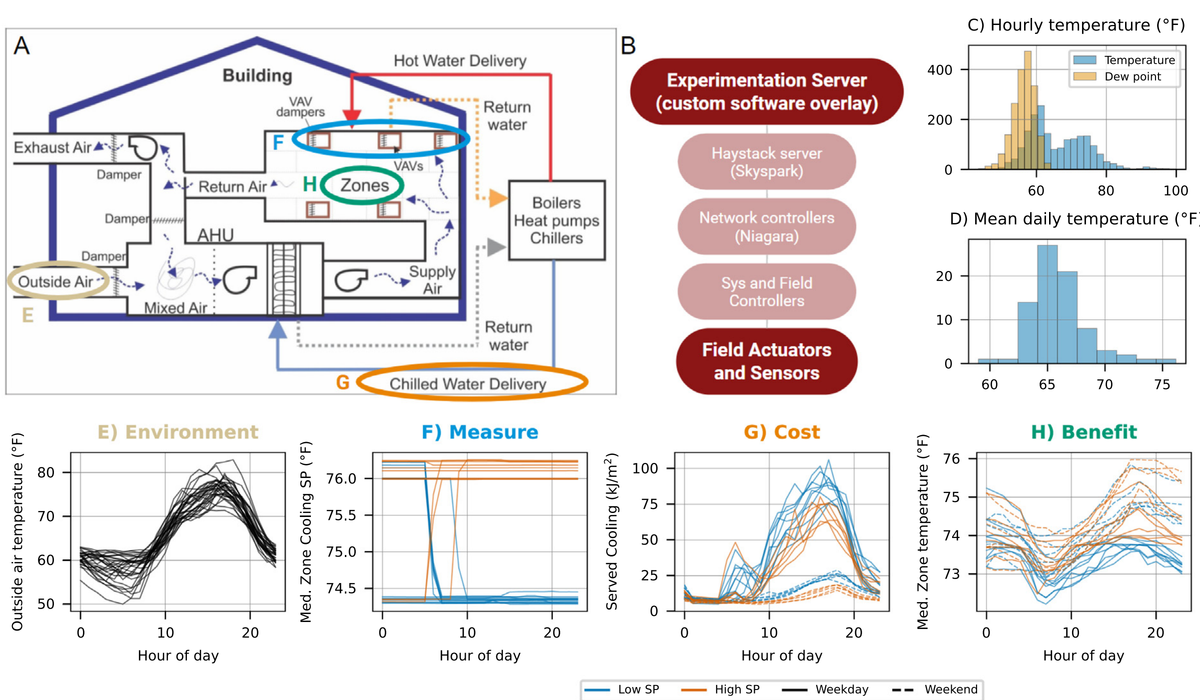


Fig. 1. Experiment testbed and illustration of experimentation protocols. (A) Building schematic. (B) Software overlay. (C-D) Histograms for hourly temperature and daily averages measured at the campus weather station. (E-H) Daily profiles for key variables in OFF-2 where mean daily outside air temperatures are between 65°F and 67°F (11 days of high set points and 14 days of low set points). A dew point below 60°F is typically associated with dry conditions.

3.2.2. HVAC equipment and controls

In the experiment testbed, cooling is supplied to the buildings by a Central Energy Facility through a chilled water loop system. Two types of HVAC systems extract cooling from the chilled water loop in the buildings: (1) centralized systems comprised of Air Handling Units (AHU) and Variable Air Volume systems (VAVs), and (2) distributed systems comprised of Fan Coil Units (FCUs). A recent survey estimated that 30% of commercial US floorspace is served by AHU-VAV systems [12]. A schematic for the AHU-VAV system is provided in Fig. 1A. In the AHUs, fans powered by electrical motors blow air over a chilled water coil to extract cooling from the chilled water loop and maintain the air leaving the AHU at the AHU's target pressure and temperature. The cooled air is blown through a central duct to the zones, where the VAVs control the flow rate of air that is sent to each zone and can also reheat the air. Equipment-level controllers were not modified for this study. The AHUs and the VAV systems each have their own controllers, typically programmed according to the same general logic, but often implemented differently from one building to the next or even within a building. The VAV system controller attempts to maintain the temperature measured inside the zone between the cooling and the heating set points (the temperature "deadband") and the flow rate of air entering the zone within design boundaries, using a mechanical damper to adjust the flow rate of air to the zone. When the VAV is unable to maintain the associated zone's temperature within the specified deadband, it issues a request for a higher discharge pressure or a lower discharge temperature to the AHU. The AHU controller sets the AHU's fan speeds and the position of a chilled water valve that determines the flow rate of the chilled water that is exposed to the air inside the AHU. Combined, these two parameters determine the pressure and temperature of the supply air. We note that in most AHUs in this study's testbed, supply air pressure and temperature reset strategies were implemented, which is in contrast with many prior field studies where the supply air temperature is maintained fixed, typically at 55°F [5,6]. These reset strategies, that use a logic referred to in the industry as Trim-and-Respond, are now part of the sequence of operations recommended by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) [24]. Finally, the FCUs extract cooling directly from the chilled water loop at the zone-level. They are more common in laboratory buildings in the testbed. The physical and controls configuration of the different HVAC systems varies from zone to zone and building to building. However, all zones attempt to control their temperature to a specified deadband, defined by heating and cooling set points, which is why temperature set point strategies are a naturally scalable supervisory control strategy. In the experiments reported on in this paper, the controls logic for the different HVAC systems were not modified for the experiments, and correspond to the industry standard of rule-based controls. The behavior of these control systems is considered to be an integral part of the building response function that is estimated through the collection of data during the experiments. The building response functions that were estimated capture the entire cyber-physical response of the buildings. They are directly usable by an energy system operator, without needing to communicate any specifics about the building's cyber-physical parameters. In this framework, an update to the building control system is treated similarly to an energy retrofit and triggers new stress tests to assess whether the building model needs to be updated.

3.2.3. Experimentation software systems

Several different software layers interact to control and measure the behavior of the HVAC systems in the test buildings, as shown in Fig. 1B. Rather than communicate with the zone-level actuators and sensors directly, our approach is to use a custom-built software overlay to automatically schedule and send commands, as well as collect and visualize data from the building energy management systems. This custom software overlay was written in Python and communicates with pre-existing building energy management systems that relay information to the zone-level. To connect to zone-level actuators and sensors, the software overlay leverages functionality from the open-source pyhaystack module [25], that allows users to connect to a server implementing the Haystack semantic model. Project Haystack is an open-source initiative to standardize semantic models for working with IoT data [26]. The Haystack server connects to network controllers (typically, one per floor) that then relay information to and from the zone-level controllers. The software overlay is also capable of connecting to a separate data historian used by the university to manage utility meter data. This system is used to retrieve records of chilled water and electricity usage per building. The custom software overlay system was installed on a central server on the university's internal IT system from which it could securely communicate with the different buildings in the testbed. A major advantage to this approach is that it is naturally scalable. The underlying building management systems in the tested buildings were procured from different vendors and installed by different contractors, so that each building has its own unique system of software layers and often customized controls. After an integration process however, each new building could be controlled in a similar fashion by our central software overlay.

3.2.4. Experiment monitoring and data collection

Data collection began on June 16th in all buildings except LAB-5 (began June 18th) and ended on August 30th in all buildings except CONF-1 and OFF-2 (ended August 17th). During the experiments, weather conditions were measured continuously at the campus weather station. Temperatures and humidity levels recorded at the campus weather station are shown by the histograms for temperatures and dew point temperatures in Figs. 1C-D. Summary data for additional measured weather data, on solar irradiation, wind speeds and relative humidity can be found in Fig. S12. In Figs. 1E-H, daily profiles for key variables in OFF-2 on days with mean daily outside air temperatures between 65°F and 67°F provide a visual summary of experiment protocols. On each experiment day, either a high or low cooling set point was broadcast to all of the zones that were included in the experiment. Fig. 1F shows the median of these set points, computed across zones. Chilled water meters measured the building-level chilled water usage, or load. Reductions in chilled water load from the low (blue) to high (red) set point days shown in Fig. 1G correspond to the benefit from DR. Building-level cooling loads were measured from the flow rate of chilled water going through the building and the difference in temperature between the building supply and return water. Fig. 1H shows the median of the temperatures that were measured by sensors in the different zones. They were higher on the high (red) than on the low (blue) set point days. Higher temperatures correspond to the cost from DR (occupant discomfort). For zone-level data, the median is preferred over the average as a summary statistic that is more robust to the data outliers that can be very large in some buildings. Other physical variables are measured throughout the buildings during the experiments and used to interpret results, including the operating state of different components of the HVAC systems and aggregate building electrical load. During the experiments, writing of zone-level set point adjustments was not perfect. Communication delays and errors were routinely observed in some buildings and are thought to be due to overloading of network and zone-level controllers. To address this, our custom software system also includes monitoring capabilities and can automatically verify which zones responded to a command followed by rescheduling that command for those zones that failed their attempt, due e.g. to network overloading. This is further detailed in the Supplementary Information (SI) to this paper, where Figs. S10 and S11 show the target and effective set points that were written during the experiments. Treatment effects presented in the results section are computed with respect to the target set points, rather than the effective ones.

4. Results

This section presents results on the impact of increasing the temperature set point by 2°F in the tested buildings on 1) building cooling load, 2) building electricity load (excluding for cooling), 3) overall building electricity load (including an estimate of electricity for cooling) and 4) room temperatures. To estimate the "treatment effect" associated with the set point change, functional forms are specified for the relationship between the different *dependent variables* (building energy loads, zone temperature percentiles) and a set of *explanatory variables* (temperature set point, weather variables, building operational schedule). For each of the modeled dependent variables, ordinary least squares (OLS) estimates are computed for the model parameters using experimental data collected during the summer of 2021. Summary modeling results are presented for each variable in this section. Modeling choices are discussed in more detail in the SI.

4.1. Cooling load

To study the impact of increasing temperature set points on building-level energy for cooling, we compute the ordinary least squares (OLS) estimate for the parameters of a linear model expressing the logarithm of energy for cooling as a function of mean daily Outside Air Temperature (OAT), an indicator variable for whether the scheduled set point was 74°F or 76°F, and an indicator variable for weekdays,

$$\log y = \alpha T + \beta I_{SP} + \gamma I_W + \delta + \epsilon. \tag{1}$$

In this equation, y is energy for cooling, T is the mean daily temperature measured at the campus weather station, I_{SP} is the indicator function taking a value of 1 if the scheduled zone cooling set point is 76°F and 0 if it is 74°F. I_W is the indicator function taking a value of 1 for weekends and 0 otherwise. $\alpha, \beta, \gamma, \delta$ are the parameters to be estimated, and ϵ is an error term. The main assumption underlying Eq. 1 is that the impact of changing the set point on energy for cooling is a fixed percentage, independent of outside air temperatures. It also assumes that one degree of outside air temperature increases the energy for cooling by a fixed percentage, independent of set point change. Similarly, whether the day is a weekday or a weekend is assumed to impact served cooling by a fixed percentage. In Section 5 we discuss the assumptions underlying this specification and alternative functional forms. As a reminder, served cooling loads are measured at the building level from the flow rate of chilled water going through the building and the difference between the chilled water temperature in the supply and return loops. Data are also presented in the SI for cooling loads measured in different units, including estimates for the electricity required to produce the chilled water to meet these loads from chillers (see Figs. S1 and S2).

Figs. 2 A-F show how daily energy for cooling increases with the daily average of OAT, using scatter plots generated from experimental data along with superimposed lines corresponding to the estimated models for energy for cooling in each building. The full blue, respectively red, line shows the estimated model for cooling load on weekdays with a set point of 74°F, respectively 76°F. The dashed lines similarly show the estimated models on weekends.

$$y = e^{\hat{\alpha}T+\hat{\delta}} \quad \text{full blue line}, \tag{2}$$

$$y = e^{\hat{\alpha}T+\hat{\beta}+\hat{\delta}} \quad \text{full red line}, \tag{3}$$

$$y = e^{\hat{\alpha}T+\hat{\gamma}+\hat{\delta}} \quad \text{dashed blue line}, \tag{4}$$

$$y = e^{\hat{\alpha}T+\hat{\beta}+\hat{\gamma}+\hat{\delta}} \quad \text{dashed red line}, \tag{5}$$

where $\hat{\cdot}$ denotes the OLS estimate of a parameter. Numerical values for the estimated parameters and the corresponding p-values are reported in Table S2. The R^2 statistics in that table indicate that the predictors collectively explain from 80 to 96% of the variance in cooling load for OFF-2, LIB-3, OFF-4 and LAB-6, 66% for CONF-1, and 74% for LAB-5.

The effect on energy for cooling of average OAT and of increasing the zone temperature set point by two degrees Fahrenheit is summarized in Figs. 2 G-H. Fig. 2 G shows the percent increase of served cooling per degree Fahrenheit increase in average OAT, $e^{\hat{\alpha}} - 1$. For each building, the expected value estimate is shown as a dot, and the vertical line denotes the 95% confidence interval estimated during the OLS procedure. The right panel in Fig. 2 H similarly shows the percent decrease of served cooling from increasing the temperature set point, $e^{\hat{\beta}} - 1$. While the impact of the temperature set point change is more differentiated across buildings than the impact of average OAT, office-type buildings are overall more responsive to both predictors than the laboratory-type buildings (13–28% versus 3–4% in Fig. 2 G). Fig. 2 H indicates that a 10°F increase in mean daily OAT results in a 2.0 to 2.3 factor increase in served cooling in offices and classrooms, and a 1.5 to 1.7 factor increase in labs.

The sensitivity of served cooling to average OAT, measured by $\hat{\alpha}$ in Table S2, is consistent with building type. In the office and classroom buildings (CONF-1, OFF-2, LIB-3, OFF-4) a 1°F increase of the average OAT is associated with a 7.5 to 8.9% increase in served cooling. In the laboratory buildings (LAB-5 and LAB-6) a 1°F increase of the average OAT is associated with a 4.1 to 5.8% increase in served cooling. Schedules vary from building to building, as confirmed by the numerical values found for $\hat{\gamma}$. Served cooling drops very significantly on the weekends in OFF-2, significantly in OFF-4, noticeably in LAB-6, and very little in LAB-5. For CONF-1 and LIB-3, the building HVAC systems are turned off during the weekends and weekend data are not considered.

Several reasons could explain why the laboratory buildings are less responsive to the set point increase. These two buildings have a larger number of sensitive zones that are excluded from the experiments. Laboratories have stricter air ventilation requirements than office buildings, which tends to reduce the impact of set point changes on chilled water usage. Although sub-metering is not available, it is expected that a significant fraction of the total chilled water usage in the laboratory buildings is associated with the process chilled water loop (e.g. to cool equipment), while the HVAC system air handlers are the main consumers of chilled water in the office-type buildings. The two laboratory buildings have different HVAC equipment and control logic than the office-type buildings, as they rely on fixed set point AHUs and variable set point FCUs while the office-type buildings rely on variable set point AHUs and VAVs. Finally, writing set points was not as reliable as in the other buildings (see Figs. S10 and S11).

4.2. Electricity loads

4.2.1. Electricity loads excluding for cooling water

A similar approach is taken to estimate the impact of increasing room temperature set points on building-level electrical energy consumption. We compute OLS estimates for a model corresponding to Eq. 1 where y is now the building's electrical consumption. These measurements exclude electricity for cooling water, but

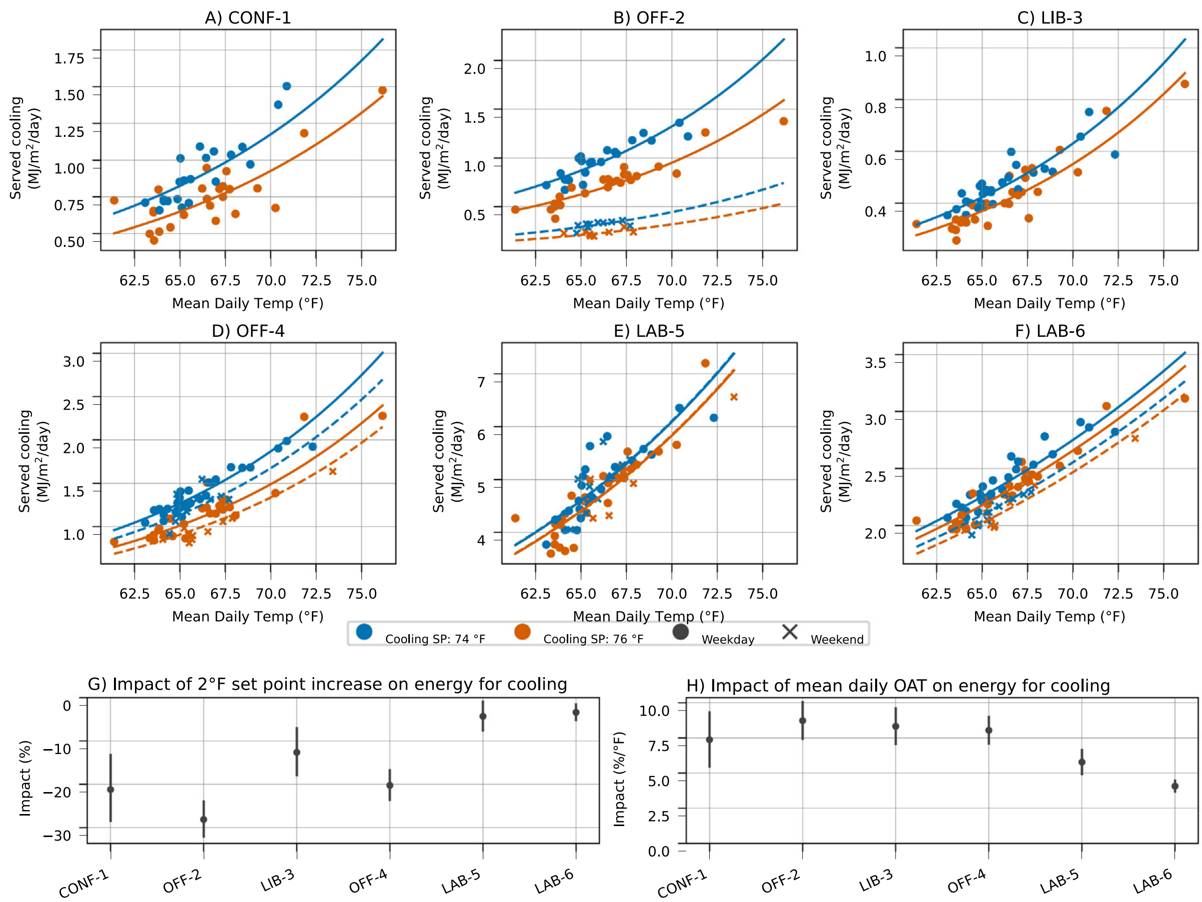


Fig. 2. Experiment results: energy for cooling. A-F) Daily energy for cooling increases with mean daily temperatures and decreases when the cooling set points are increased. Empirical data collected during the summer of 2021 and log model from Eq. 1. G-H) Measured effect of OAT and set point change on energy for cooling. Additional numerical results are reported in Table S2 and data in different units are shown in Figs. 1 and 2. Model structure and predictors are discussed in Section 5.

include electricity for moving air throughout the building with the AHU fans. The AHU fans are not sub-metered. Fig. S3 reports results for electricity consumption in a similar format to Fig. 2 and shows a small (less than 2%) reduction in building electricity consumption when the temperature set points are increased. For most buildings in the testbed, the 95% confidence intervals on the estimated effects are also quite large. The p-values in Table S3 indicate the effect is significant only for OFF-2 and OFF-4 at the 5% level. OAT is also observed to have an overall limited impact on building electricity consumption. Together, these results suggest either that the setpoint changes and OAT have a limited impact on fan energy consumption and/or that changes in fan energy consumption are small compared to other electricity uses inside the buildings.

4.2.2. Electricity loads including for cooling water

Since most DR applications focus on electricity demands, we also generate impact estimates for overall building electricity loads. Electricity for cooling cannot be measured directly in the testbed, since the buildings are supplied by a district cooling system. 4.4% of US commercial floorspace is currently served by district cooling systems and 19% by central chillers [12]. To generalize our results to both categories, we estimate electricity for cooling assuming a constant chiller Coefficient Of Performance (COP). In reality, the COP will not be constant and will depend on OAT, relative humidity, and in a district cooling system, on the possibly time-dependent combination of chillers that are producing. For example, the effective COP of the central chiller plant that supplies this testbed typically varies from 3.5 to 4.5. To assess the sensitivity of results to chiller COP, we use a representative range for the industry: a COP of 3 to represent in-building cooling systems, 5.5 to represent electrified district cooling systems, and 8 to represent more efficient technologies than what is currently deployed [27]. Fig. 3 presents results for the overall daily building electricity loads, including these calculations of electricity for cooling. Estimates are computed for specifications corresponding to Eq. 1 and numerical results are reported in Table S4.

Assuming a coefficient of performance of 5.5 for the electric chillers, the impact of the 2°F set point increase is measured to be 4.6 to 9.4% for CONF-1, OFF-2, and OFF-4, 2.5% for LIB-3, and less than.8% for LAB-5 and LAB-6. Consistent with our cooling results, the buildings that are most responsive to the set point changes are also the ones that are most responsive to OAT.

These results are sensitive to the efficiency assumption for the electric chillers supplying the cooling. Increasing the efficiency of the chillers reduces the responsiveness (and vulnerability) of overall electric load to higher temperatures but also decreases the overall electric flexibility potential, since a large fraction of the flexibility is attributable to space cooling.

4.3. Room temperatures

To study the impact of increasing room temperature set points on room-level temperatures inside the building, we compute OLS estimates for the parameters of a linear model expressing different

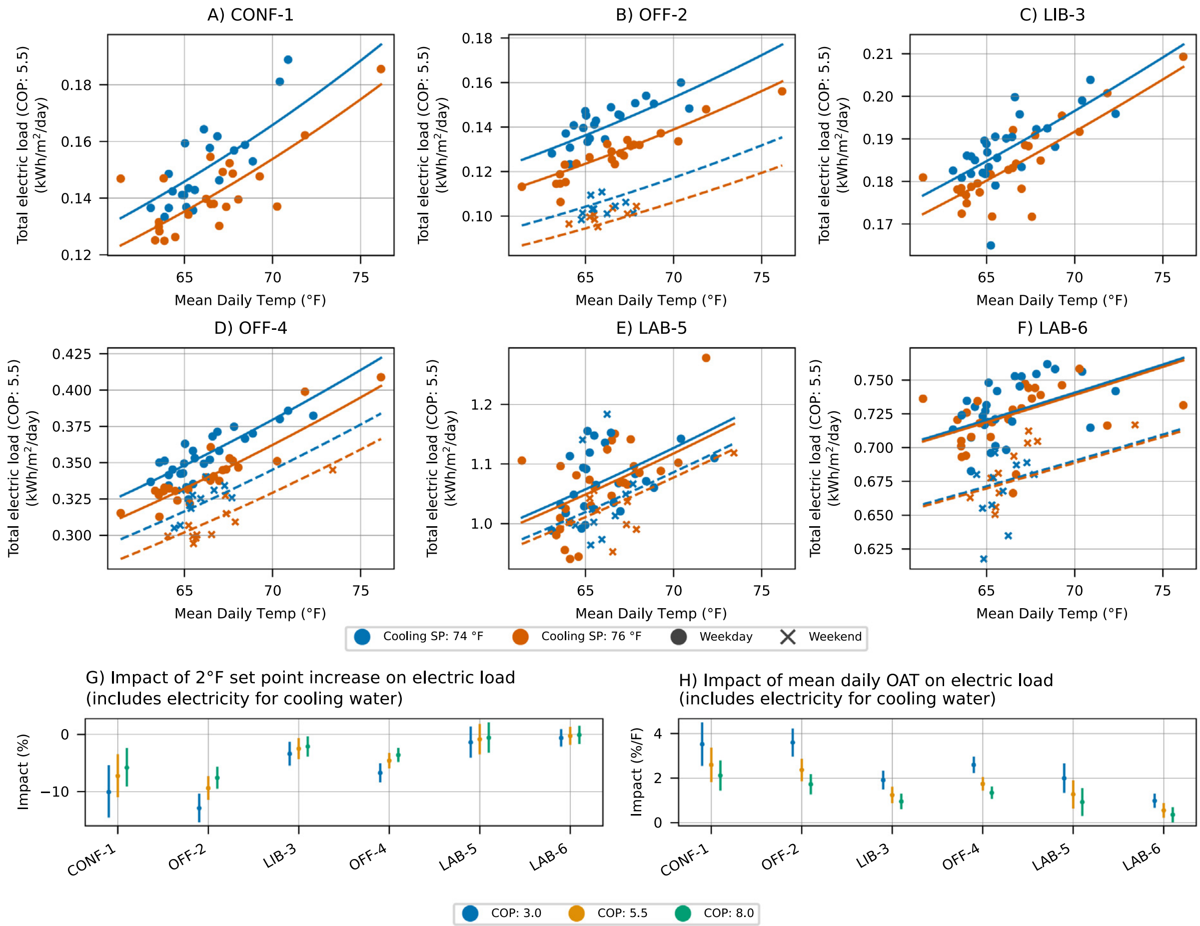


Fig. 3. Experiment results: electricity loads (including for cooling water). (a) Daily electricity as a function of mean daily temperatures of the cooling set points. Empirical data collected during the summer of 2021 and log model from Eq. 1. (b) Measured effect of OAT and set point change on electricity. To estimate the electricity required for cooling water, different assumptions are made for chiller Coefficient Of Performance (COP). Numerical results for a COP of 5.5 are report.ed in Table S4.

zone temperature percentiles as a function of mean daily OAT, an indicator variable for the scheduled zone set point, and an indicator variable for weekdays,

$$T_p = \alpha T + \beta I_{SP} + \gamma I_W + \delta + \epsilon. \tag{6}$$

Here T_p is a summary statistic for room temperatures throughout the building: the 8am to 8 pm average of the pth percentile for room temperatures, where the percentile is taken over zones. The main assumption underlying Eq. 6 is that T_p changes by a fixed offset (in degrees F) for every degree of mean daily OAT, due to the treatment (set point change) and as a function of whether the day is a weekday. Importantly, for OAT we use the average over the full 24 h of data, because OAT is a dependent variable and nighttime temperatures could have an effect because of thermal inertia. For room temperatures on the other hand, we only use data from 8 am to 8 pm, because temperatures during the unoccupied periods at night are not important for the thermal comfort of the occupants.

Scatter plots for experimental data are shown in Figs. 4A-F along with superimposed lines corresponding to the estimated models for the 8am-8 pm average for T_{50} (similar conventions as Fig. 2). Additional numerical data are reported in Table S5. Figs. 4-G-H show the effects of the treatment and OAT on $T_{10}, T_{25}, T_{50}, T_{75}$ and T_{90}, respectively.

A 2°F set point increase resulted in an increase of less than 1.5°F in our summary statistics for room temperatures (T_p's). We also measure a small but statistically significant effect of average daily OAT. A 10°F increase in mean daily OAT resulted in a 0.1 to 2.1°F increase in the T_p's. The effect is larger for the two laboratory buildings and the library than for the office buildings. In some buildings, the effect of the treatment and of the OAT grows with the percentile, but not in others. In OFF-2 the 8am-8 pm average for the median zone temperature (T_{50}) is on average 0.88°F hotter on the weekend which is more than the impact of increasing the set point by 2°F (0.04°F).

The indoor temperature data in Figs. 4A-F are all below the corresponding cooling set points, sometimes substantially. Additional scatter plots for the 50th, 75th, and 90th percentiles for room temperatures at 3 pm in Figs. S4-S6 show that room temperatures vary significantly within the buildings. They also vary depending on the time of day. The observation that the temperatures in many rooms are often below the cooling set point helps explain why the T_p's are found to increase with OAT in Fig. 4 H: this is likely the symptom of over-cooling from ventilation constraints in multi-zone HVAC systems, as discussed by several previous authors [28–30]. Ventilation constraints were also shown to have a strong limiting effect on the impact of temperature set point changes using simulated data by Hoyt et al. [15]. As OAT increases, overcooling decreases, and the T_p's increase. Another complicating factor is that occupants have

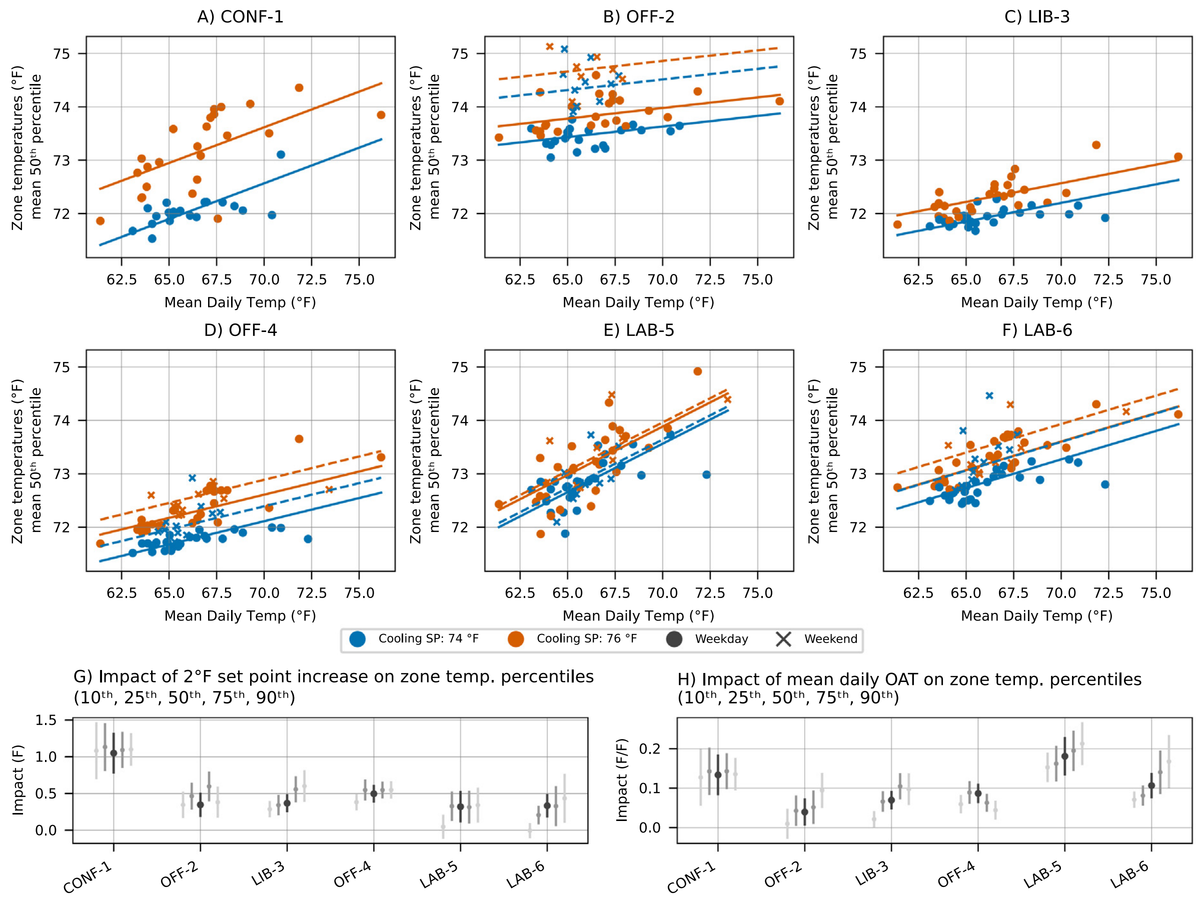


Fig. 4. Experiment results: room temperatures. (a) The 8am-8 pm average for the 50th percentile of room temperatures increases with mean daily OAT and as the cooling set points are increased. Numerical results are reported in Table S5.

the option to adjust their thermostat by up to 2°F in most rooms, which could contribute to temperature heterogeneity.

Fig. S7 shows the measured effects of OAT and the temperature set point increase on zone temperatures at different hours of the day. Unsurprisingly, the effect of both predictors is very limited at 6 am. At the other hours, the impact of raising the set point is less than 1°F for all buildings expect CONF-1. CONF-1's temperatures are the most responsive and all T_p's increase the most at 3 pm, on average by between 1.1 and 1.3°F.

Finally, some buildings are warmer than others, despite receiving the same set points. Differences across buildings are often larger than the impact of the set point change or of OAT.

Overall, we find limited impacts on measured room temperatures from increasing the cooling set point by 2°F. In the building where temperatures responded the most, CONF-1, room temperatures increased by at most 1.5°F at 3 pm (Fig. S7). In other buildings and at other times of the day, a response of 1°F or less was typical. These results suggest potential for additional flexibility and efficiency gains in the tested buildings.

5. Discussion

5.1. Model specification

The models in Section 4 use a specification for the logarithm of energy loads. As previously noted, the major underlying assumption is that the predictors have a fixed percentage effect on the predicted variable. Results for an alternative specification, that is linear in energy loads, are presented in Fig. S8 and Table S6. While a linear dependence of energy loads on OAT is more consistent with a heat balance at the building level, the model in Eq. 1 has several advantages. First, the coefficients α, β and γ lead to a natural, unit-less interpretation for the response of energy loads to their respective predictors that can easily be compared across buildings and extrapolated: they measure the relative change in consumption associated with the predictors, e.g. $100(e^{\hat{\beta}} - 1)$ is the estimate for the percent change in energy from the set point change. Also, while in Section 4.1 we measure the impact of the set point change on chilled water consumption, in many applications electricity will be the main focus. Assuming a constant conversion efficiency, the estimates remain directly applicable. Second, Eq. 1 predicts that increasing the set point will have a larger impact on energy loads at higher OAT, consistent with previous work [19]. Our results on room temperatures in Section 4.3 also offer a possible physical explanation for why the response to the set point change might grow with OAT. Those results show that the indoor air temperature is below the cooling set point in a significant fraction of rooms. As OAT increases, over-cooling decreases, a greater number of rooms is closer to their cooling set point, and will therefore be directly affected by a change in that set point. In contrast, the model used for Fig. S8 assumes the impact of the set point adjustment is constant with respect to OAT. Third, Eq. 1 has a statistical advantage over a model that is linear in energy loads in that it reduces heteroskedasticity. In the data, we observe that the magnitude of the residuals from the linear equation grows with OAT. By reducing heteroskedasticity, Eq. 1 leads to better estimates for the standard errors on the model coefficients. Eq. 1's main drawback is in its predictions at high temperatures, which is also where it significantly deviates from the linear model (Fig. S8). But for most of the OAT range for which data are recorded (specifically, from 62.5 to 72.5°F), Fig. S8 shows that the linear and logarithmic specifications produce very similar results, consistent with a first-order Taylor series expansion of the exponential function. The R^2 statistics reported in the SI also differ by at most 3%. Finally, we note that while our results show that simple linear models are adequate for modeling daily loads, more complex models will likely be required to capture hourly dynamics (e.g., to capture thermal inertia effects).

Additional predictors. The sensitivity of the results in Section 4 was also tested by including the maximum daily OAT, minimum daily OAT, wind, solar irradiation, and relative humidity as predictors in Eq. 1. A summary for the cooling models is shown in Fig. S9. Including these additional predictors did not significantly change our estimate for the impact of the 2°F set point increase on energy for cooling nor goodness-of-fit. The same is true for the impact of mean daily OAT on energy for cooling, with the exception of the models with maximum and minimum daily temperatures. In that case, including those predictors produced larger estimates, but also larger confidence intervals, which can be explained by the strong correlation between the average, maximum and minimum daily temperatures. An advantage of using the average rather than the maximum or minimum temperature is that this predictor is less sensitive to faulty data measurements.

The number of occupants was not measured, but is expected to have been fairly constant in most buildings apart from CONF-1, where variability in occupancy could be the explanation for the comparatively lower performance of the models.

Choice of treatment variable. Special care is taken in choosing the treatment variable. In the models presented in Section 4, the treatment variable corresponds to the set point command that was scheduled to be sent to the building, not the actual set point command that was adopted by the different rooms inside the building. Consequently, the variability in the outcome data are attributable not only to the physical response of the buildings, but also to the response of their software systems. Figs. S10 and S11 show how the software systems did not respond perfectly. Two possible explaining factors are 1) occupants have the option to re-adjust temperature set points locally; and 2) network and controller overloading can cause communication delays and data loss between the centralized software system and the rooms. Our choice of treatment variable yields models that encompass the entire cyber-physical response of a building energy system, which is more relevant to electric grid or district energy system operators.

5.2. Informing operator flexibility decisions

Data-driven models for flexibility. Historically, most field tests were used to complement and validate simulations by providing case studies and demonstrations. Ubiquitous sensors and actuators promise a new paradigm under which field tests can play a much more ambitious role in developing data-driven flexibility models for buildings. The work presented here provides a blueprint for developing these models. The findings in this paper suggest that if more ambitious experimentation programs are conducted, a more ambitious role can be played by field data. The data-driven models for building response functions developed in this paper are directly relevant to electric grid and district energy system operators. The response of buildings under the change remains uncertain and variable, but the uncertainty is quantified in the models, by providing statistical information and guarantees that can be used by consumers of the DR service when making their decisions, e.g. confidence intervals. Past field tests were invaluable to demonstrate potential from providing DR services through building energy systems when the technology was in its infancy. As the technology nears deployment, new, standardized field tests supported by rigorous statistical analyses are now needed to build confidence.

Providing flexibility on other timescales through integrated energy systems. In contrast with most prior works reporting field test data (see Section 2), the experiments here evaluate daily rather than hourly flexibility. Indeed in many present applications, DR is provided on shorter timescales, e.g. to reduce peak afternoon loads. However, daily flexibility strategies will often be simpler to implement. What's more, we argue that providing daily flexibility from building-level loads may turn out to be just as valuable as hourly flexibility for two reasons.

First, there are expectations that longer term flexibility needs will arise in future electricity grids [7]. Second, it is very likely that in most real-life applications DR services from flexible loads will be aggregated. When aggregated with other energy systems, daily flexibility from building HVAC systems can unlock additional hourly flexibility from the integrated energy system, e.g. at the electric distribution system level. In an electrified district energy system with city-scale thermal storage like the one that provides heating, cooling and electricity to the buildings that are tested here [10], daily building-level flexibility unlocks more opportunities to provide hourly DR services measured at the integrated system level [31]. Other integration options include an electrochemical storage system at the distribution level, or an electric vehicle fleet.

5.3. Value of stress tests for building energy systems

The experiments in this work highlight the value of developing and standardizing stress tests for building energy systems. One of the most important benefits is to enable the development of data-driven flexibility models that account for the full cyber-physical response of buildings. Training and updating such models will be cheaper and more scalable than for computational models that heavily rely on building design parameters rather than on measured parameters. HVAC control logic is a good example of an operational parameter that is non-trivial to capture with computational models such as EnergyPlus, but naturally captured when generating training data for data-driven models through stress tests.

Baselines for DR. One specific area where the collection of data through stress tests offers an opportunity to rethink prior research is in the development of the baselines that are typically used to estimate a building's actual contribution during an event. Most prior works, reviewed in Weng et al. (2018) [32], use deterministic estimates, either computed using a (weighted) average of the previous ten days or a linear regression based on prior data, possibly adjusted based on information from the morning of the event. The data collected through stress tests can be used to calibrate more robust probabilistic baseline models. The experiments reported here also suggest including information from after the DR event may be just as relevant as information prior the event. The only constraint to implement such a *centered* baseline scheme is to wait until the end of the DR program to distribute rewards and penalties to participants.

Updating DR models. Energy infrastructure and large buildings are designed to last half a century, but the energy consumption patterns of a building change over time. In a research facility, for instance, new laboratory equipment has the potential to dramatically change the building's energy profile. The same is true in a large administrative building after a retrofit of the HVAC system or a tenant improvement. These changes typically render previously developed digital twins based on simulation models (such as EnergyPlus, eQUEST, TRNSYS) obsolete. In contrast, the methods

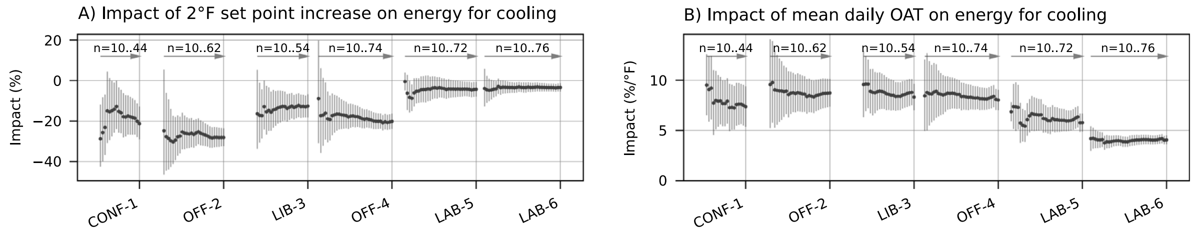


Fig. 5. Collecting a growing number of observations gradually improves parameter estimates. Measured effect of set point change and OAT on energy for cooling using the model in Section 4.1. The dots represent the average estimates and the vertical bars show the 95% confidence intervals. The number of observations (n) is reported for each building, e.g. 10 to 44 observations are .used for CONF-1.

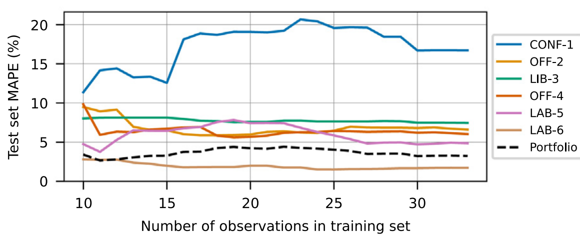


Fig. 6. Aggregating buildings improves out-of-sample predictive power. Performance is reported as Mean Absolute Percentage Error (MAPE) for a growing number of observations, using ten days as an out-of-sample test set.

developed in this work can be used to regularly and rapidly evaluate and update a building's thermal flexibility model. Fig. 5 shows the sensitivity of the impact of the 2°F set point increase and of mean daily OAT, as estimated from models trained on a growing number of observations. Observations were ordered chronologically in this analysis. As expected, the confidence intervals grow smaller as more data is collected. For most buildings, the value of collecting additional data points decreases after a few weeks.

Out-of-sample predictive power. To assess the predictive power of the models, the last ten observations common to each data set were excluded to form a "test set". Models trained on a growing number of observations were then evaluated against the test set. Performance is reported in Fig. 6. For all buildings except CONF-1, the test set error gradually decreases as the number of observations that are included grows. We also compute the out-of-sample performance of the sum of the buildings (labeled "Portfolio") and observe that aggregating buildings improves predictive power.

Energy efficiency and commissioning. Beyond flexibility, stress tests provide an empirical measurement tool for building managers to evaluate the benefits and costs of operational changes and a continuous assessment of opportunities for improved energy efficiency. The experiments here show that cooling loads are significantly reduced from increasing cooling set points by 2°F, with overall limited impacts to room temperatures. Where acceptable to occupants, cooling set points could be raised at all times and similar stress tests conducted to assess the flexibility benefit from increasing set points from this new baseline. Stress tests are a natural complement and extension to the existing commissioning tools that can be used to dramatically improve the energy efficiency of existing commercial building stocks [33,34]. In contrast with most existing commissioning methods that test the behavior of individual actuators and sensors, the experiments in this work test the overall cyber-physical response of a building's energy system. Control systems play a large role in building energy consumption: one simulation-based study found that upgrading control systems could save 23–30% energy in half of the US commercial building stock [35]. Additionally, as flexibility services become more used, flexibility-oriented commissioning tools will be needed. Stress testing buildings will help maintain a state of operational readiness that improve the reliability of the DR service.

5.4. Building response: cooling versus electricity

Our results provide an interesting perspective on previous field tests where the main target for flexibility was the electricity consumption associated with the AHUs [36,6]. The flexibility source relied on in many of those experiments is a reduction in the fan speed of the AHUs, resulting in a reduction of the overall building electricity consumption [37–39]. In contrast, in these experiments the daily response of the electrical loads is minimal, while that of the cooling loads is significant.

A first important difference is in the experiment time scale. In many prior experiments, flexibility is provided on an hourly time scale, typically in the middle of the day. In contrast, in the experiments reported on in this paper, set points are adjusted on a daily time scale. Transient temperature effects can be expected to play a larger role in hourly rather than daily experiments. A second important difference is the AHU control logic. While this is not reported in all of the prior studies referenced above, for the major part it appears that they considered HVAC systems where the temperature of the air leaving the AHUs to supply the zones was fixed, whereas most AHUs in this testbed used the more recent industry best practice of supply air temperature resets [24]. AHU controllers in this testbed leverage two main degrees of freedom to control the cooling power of the air that is sent to the zones: the flow rate of air blowing over the chilled water coils, controlled by the speed of the fans; and the flow rate of water through the chilled water coils, controlled by the position of the chilled water valves. A third setting is the fraction of building air that is re-circulated, but that was maintained fixed during the experiments so it is not considered here. Together, these settings control the temperature and pressure of the supply air. Finally, a third source of possible response heterogeneity also noted in previous work [6] is the presence of ventilation requirements that can enter in competition with cooling requirements.

In summary, our results on cooling and electricity loads show that new building control sequences could alter conclusions from previous DR field experiments. In the experiments reported on in this paper, flexibility was mainly provided from the chilled water valves rather than from the fan speeds, in contrast to previous field studies. However, the three possible causes for the difference that were discussed also indicate that this will not always be the case. This further motivates the need for scalable modeling and testing methods like the ones discussed in this work.

6. Conclusion

A scalable empirical method based on the idea of stress tests was developed to assess the cyber-physical response of commercial buildings to temperature set point adjustments. The method is generally implementable, and doing so will be of value to building researchers and industry practitioners in at least two ways. Previous literature reports large variability in building response during DR events. This work shows how empirical methods can be used to systematically measure, quantify and explain that variability by also measuring explanatory variables such as the outside air temperature. These methods can also be used during commissioning of a building as a system-level diagnostic tool, or stress test, that represents a significant departure from existing commissioning processes that overwhelmingly rely on component-level diagnostic tools. Diagnostic tools that rely on systems-level stress tests or tests to assess the overall responsiveness of buildings to perturbations in operating set points will be valuable to demonstrate continued DR readiness to consumers of DR services such as electric grid operators. The flexibility model that was proposed is simple to calibrate and to interpret. It provides statistical guarantees that can directly be used by electric grid or district energy system operators calling on DR. The application of the empirical method provides invaluable information to construct, maintain and update a building's weather-dependent flexibility model, including quantitative statistical information about the building's responsiveness. The response of six buildings to daily cooling set point adjustments was tested on a university campus in California (Warm-summer Mediterranean climate). The buildings do not all respond identically, and their response is uncertain. But, repeated testing can be used to build data-driven models to quantify the uncertainty, e.g. through confidence intervals. Data generated through stress tests enables the development of models that also capture the full cyber-physical response of buildings, which will be required for energy operators to integrate DR resources. The potential for and constraints on providing thermal and electrical demand response from temperature set point strategies was discussed. Also, the daily timescale considered here is in contrast with the majority of prior works on building DR that consider hourly or sub-hourly timescales. In integrated, electrified energy systems such as the one we consider, chilled water storage tanks can be used as a buffer between the electric chillers used to produce cooling and the buildings that consume it. Enabling daily flexibility in the served cooling loads at the building level in turn unlocks both daily and/or hourly flexibility at the integrated energy system level. This work provides a blueprint for developing reliable, scalable, experimental and testing methods with building energy systems. These methods will lead to a better scientific understanding of energy loads in large, modern buildings, and their flexibility in response to controlled perturbations. As distributed sensors and actuators continue to be more widespread, these methods will become increasingly cheaper, scalable, and attractive.

7. Data and code

Daily data from the experiments reported on in this paper are released as part of the supplemental material, along with code to reproduce Figs. 2 through 6. A detailed description of the data set that is released is provided in the SI.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding for this research was supported by Stanford Land, Buildings and Real Estate, and TotalEnergies SE. The authors thank several past and current members of the Stanford Cooler project for valuable discussions and contributions that made the experiments reported on in this paper possible, including Philippe Cordier, Austin Diaz, Gerry Hamilton, Scott Jespersen, Leslie Kramer, Jackie Lum, Jim Perrizo, Gabe Sandoval, and Tim Troxell.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.enbuild.2022.112599.

References

[1] IEA, Global status report for buildings and construction 2019, Tech. rep., IEA, Paris, 2019.
[2] IEA, The future of cooling – opportunities for energy-efficient air conditioning, Tech. rep., IEA, Paris, 2018.
[3] International Energy Agency, IEA Sankey diagram - World Final Consumption, available at: URL: https://www.iea.org/sankey/, 2022.
[4] EIA, Electric power annual 2020, Tech. rep, US Energy Information Administration (2021).
[5] N. Motegi, M.A. Piette, D.S. Watson, S. Kiliccote, P. Xu, Introduction to commercial building control strategies and techniques for demand response, Lawrence Berkeley National Laboratory LBNL-59975 4, 2007.
[6] J.S. MacDonald, E. Vrettos, D.S. Callaway, A critical exploration of the efficiency impacts of demand response from hvac in commercial buildings, Proc. IEEE 108 (9) (2020) 1623–1639.
[7] P. Albertus, J.S. Manser, S. Litzelman, Long-Duration Electricity Storage Applications, Econ., Technol., Joule 4 (1) (2020) 21–32, https://doi.org/10.1016/j.joule.2019.11.009.
[8] A. Satchwell, M.A. Piette, A. Khandekar, J. Granderson, N.M. Frick, R. Hledik, A. Faruqui, L. Lam, S. Ross, J. Cohen, et al., A national roadmap for grid-interactive efficient buildings, Tech. rep, US Department of Energy Building Technologies Office, 2021.
[9] J. Langevin, C.B. Harris, J.L. Reyna, Assessing the potential to reduce us building co2 emissions 80% by 2050, Joule 3 (10) (2019) 2403–2424.
[10] J.A. de Chalendar, P.W. Glynn, S.M. Benson, City-scale decarbonization experiments with integrated energy systems, Energy Environ. Sci. 12 (5) (2019) 1695–1707.
[11] Y. Ma, A. Kelman, A. Daly, F. Borrelli, Predictive control for energy efficient buildings with thermal storage: Modeling, stimulation, and experiments, IEEE Control Syst. Mag. 32 (1) (2012) 44–64.
[12] EIA, 2018 commercial buildings energy consumption survey (cbecs), Tech. rep., Energy Information Administration, 2021.
[13] S. Aghniaey, T.M. Lawrence, The impact of increased cooling setpoint temperature during demand response events on occupant thermal comfort in commercial buildings: A review, Energy Build. 173 (2018) 19–27, https://doi.org/10.1016/j.enbuild.2018.04.068, URL:https://doi.org/10.1016/j.enbuild.2018.04.068.
[14] S. Kiliccote, M.A. Piette, J. Mathieu, K. Parrish, Findings from seven years of field performance data for automated demand response in commercial buildings, Tech. rep., Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States), 2010.
[15] T. Hoyt, E. Arens, H. Zhang, Extending air temperature setpoints: Simulated energy savings and design considerations for new and retrofit buildings, Build. Environ. 88 (2015) 89–96.
[16] A. Ghahramani, K. Zhang, K. Dutta, Z. Yang, B. Becerik-Gerber, Energy savings from temperature setpoints and deadband: Quantifying the influence of building and system properties on savings, Appl. Energy 165 (2016) 930–942.
[17] R. Yin, E.C. Kara, Y. Li, N. DeForest, K. Wang, T. Yong, M. Stadler, Quantifying flexibility of commercial and residential loads for demand response using setpoint changes, Appl. Energy 177 (2016) 149–164.
[18] M. Cai, S. Ramdaspalli, M. Pipattanasomporn, S. Rahman, A. Malekpour, S.R. Kothandaraman, Impact of hvac set point adjustment on energy savings and peak load reductions in buildings, 2018 IEEE International Smart Cities Conference (ISC2), IEEE 2018 (2018) 1–6.
[19] J. Liu, R. Yin, M. Pritoni, M.A. Piette, M. Neukomm, Developing and evaluating metrics for demand flexibility in buildings: Comparing simulations and field data, in: ACEEE Summer Study on Energy Efficiency in Buildings, 2020, pp. 12-267–12-280.
[20] J. Drgoňa, J. Arroyo, I.C. Figueroa, D. Blum, K. Arendt, D. Kim, E.P. Ollé, J. Oravec, M. Wetter, D.L. Vrabie, et al., All you need to know about model predictive control for buildings, Annu. Rev. Control (2020).
[21] D. Sturzenegger, D. Gyalistras, M. Morari, R.S. Smith, Model Predictive Climate Control of a Swiss Office Building: Implementation, Results, and Cost-Benefit Analysis, IEEE Trans. Control Syst. Technol. 24 (1) (2016) 1–12, https://doi.org/10.1109/TCST.2015.2415411.
[22] P. Li, D. Vrabie, D. Li, S.C. Bengea, S. Mijanovic, Z.D. O'Neill, Simulation and experimental demonstration of model predictive control in a building hvac system, Sci. Technol. Built Environ. 21 (6) (2015) 721–732.

[23] J. Granderson, G. Lin, R. Singla, S. Fernandes, S. Touzani, Field evaluation of performance of hvac optimization system in commercial buildings, Energy Build. 173 (2018) 577–586.
[24] ASHRAE, Ashrae guideline 36–2021: High-performance sequences of operation for hvac systems, 2021.
[25] Tremblay Christian and Longland, Stuart, pyhaystack, available at: URL: https://github.com/ChristianTremblay/pyhaystack, 2020.
[26] Project Haystack, About Project Haystack, available at: URL: https://www.project-haystack.org/about (2022).
[27] ASHRAE, District cooling guide (2013).
[28] S. Derrible, M. Reeder, The cost of over-cooling commercial buildings in the united states, Energy Build. 108 (2015) 304–306.
[29] D. Yan, Y. Jiang, X. Shi, et al., Influence of asynchronous demand behavior on overcooling in multiple zone ac systems, Build. Environ. 110 (2016) 65–75.
[30] C. Zhuang, K. Shan, S. Wang, Coordinated demand-controlled ventilation strategy for energy-efficient operation in multi-zone cleanroom air-conditioning systems, Build. Environ. 191 (2021).
[31] J. de Chalendar, P. Glynn, S. Benson, Experimental investigation of a capacity-based demand response mechanism for district-scale applications, in: Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019, pp. 3709–3718.
[32] Y. Weng, J. Yu, R. Rajagopal, Probabilistic baseline estimation based on load patterns for better residential customer rewards, Int. J. Electr. Power Energy Syst. 100 (2018) 508–516.
[33] E. Mills, Building commissioning: a golden opportunity for reducing energy costs and greenhouse gas emissions in the united states, Energy Effi. 4 (2) (2011) 145–173.
[34] S. Katipamula, K. Gowri, G. Hernandez, An open-source automated continuous condition-based maintenance platform for commercial buildings, Sci. Technol. Built Environ. 23 (4) (2017) 546–556.
[35] N.E. Fernandez, S. Katipamula, W. Wang, Y. Xie, M. Zhao, C.D. Corbin, Impacts of commercial building controls on energy savings and peak load reduction, Tech. rep., Pacific Northwest National Lab. (PNNL), Richland, WA (United States) (2017).
[36] M.A. Piette, S. Kiliccote, G. Ghatikar, A. McKane, N. Matson, J. Page, J. MacDonald, A. Aghajanzadeh, D. Black, R. Yin, Demand response research center, in: California energy commission. publication number, Cec500-2015-xxx, Tech. rep., Lawrence Berkeley National Laboratory, 2015.
[37] I. Beil, I. Hiskens, S. Backhaus, Round-trip efficiency of fast demand response in a large commercial air conditioner, Energy Build. 97 (2015) 47–55.
[38] S. Afshari, J. Wolfe, M.S. Nazir, I.A. Hiskens, J.X. Johnson, J.L. Mathieu, Y. Lin, A.K. Barnes, D.A. Geller, S.N. Backhaus, An experimental study of energy consumption in buildings providing ancillary services, 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE 2017 (2017) 1–5.
[39] A. Keskar, S. Lei, T. Webb, S. Nagy, I.A. Hiskens, J.L. Mathieu, J.X. Johnson, Assessing the performance of global thermostat adjustment in commercial buildings for load shifting demand response, Environ. Res.: Infrastruct. Sustain. (2022).