Lawrence Berkeley National Laboratory

Title

Quantifying the effect of multiple load flexibility strategies on commercial building electricity demand and services via surrogate modeling

Permalink

Authors

Luo, Na
Langevin, Jared
Chandra-Putra, Handi
et al.

Publication Date

DOI

Quantifying the effect of multiple load flexibility strategies on commercial building electricity demand and services via surrogate modeling

Na Luo[a], Jared Langevin[a,*], Handi Chandra-Putra[a], Sang Hoon Lee[a]

[a]*Lawrence Berkeley National Laboratory, Building Technology and Urban Systems Division, Berkeley, CA, USA*

Abstract

The expansion of commercial building demand response as a demand-side management resource for the electric grid necessitates new decision support resources for customers seeking to assess the benefit-risk tradeoffs of possible strategies for energy flexible building operations. To address this need, we develop surrogate models that predict the impacts of several load flexibility strategies on commercial building electricity demand and indoor temperature, focusing on offices and retail buildings at multiple scales. The surrogate models are fit to a synthetic database generated via whole building simulations, which establish the relationships between the key operational features of a given strategy and potential changes in building demand and temperature in a wide variety of contexts. The surrogate models are translated to a Bayesian framework to allow straightforward communication of uncertainty and parameter updating given new evidence. We find strong predictive performance across the suite of models, underscoring the usefulness of the approach in guiding decisions about implementing load flexibility strategies under a particular set of operational and environmental conditions.

Keywords: commercial building demand response, surrogate modeling, Bayesian framework, demand-side management, grid-interactive efficient buildings

*Corresponding author, email: jared.langevin@lbl.gov

1. Introduction

Commercial building demand response (DR), which encompasses a set of time-dependent utility program activities and tariffs that seek to reduce electricity demand or shift demand across time periods [1], is likely to play an expanded role in the coming years as a demand-side management tool that facilitates grid reliability and resilience in the face of day-to-day stresses, increased renewable energy penetration, and emergency events [2]. Applications of DR have been reported and assessed across the individual building, building district, electricity system, and whole sector levels [3]. At the system level, previous studies suggest multiple power grid benefits from improved demand management through DR, including: lower peak demand, improved market efficiency, and coordination of efficient and flexible demand-side resources [4]; support of load shifting from peak to off-peak times [5]; and increased ability of grid operators to manage high penetration of renewable energy sources such as solar and wind energy that have highly variable generation profiles [6, 7].

Along with these grid benefits, however, expanded commercial DR introduces new risks on the utility customer side. Commercial building operators and owners have a primary interest in providing comfortable environmental conditions for building occupants, given the role of indoor environmental quality in facilitating workplace productivity, health, and well-being [8]; the potential for disruption to normal operating conditions constitutes a key concern about participation in DR programs [9]. Moreover, the prospect of further automating customer responses to DR calls from the grid engages additional customer concerns about a lack of transparency into or control over scheduled adjustments to normal building operations [10] — particularly if customers are not allowed to override or opt-out of responding [9, 11]. Looking ahead, widespread adoption of expanded commercial DR program offerings will require customer-facing resources that address such concerns by providing greater insight into the potential implications of flexible adjustments for building operations, including both

the likelihood of changes in electricity demand and changes to core building services like comfortable temperatures.

Existing reviews of the commercial DR context establish the range of load flexibility strategies that customers will typically consider, grouping these strategies into four areas: HVAC, lighting, miscellaneous equipment, and measures that work across components or end uses [12, 13]. Adjustments to commercial HVAC and lighting set point schedules are considered as particularly attractive strategies, due to the substantial portion of total loads that these strategies can affect and their comparatively lower risks to occupant comfort than strategies that make centralized adjustments to air distribution or cooling systems, for example. Thermal load flexibility strategies include pre-cooling and pre-heating [14, 15, 16, 17], which shift cooling and heating loads to the hours preceding those of peak demand on the grid, along with global temperature adjustments (GTA) [18, 19], which relax zone thermostat set points across all thermal zones in a building during peak periods. Lighting strategies include dimming or switching off lights for load shedding during peak demand periods; while such actions may be performed manually, previous studies demonstrate the ability to implement lighting load flexibility through a central energy management system (EMS) [18]. Outside of HVAC and lighting, commercial flexibility strategies may also target miscellaneous electric equipment (e.g., computers and office equipment, fountain pumps, industrial process loads, etc.) to reduce demand without influencing the basic activities of occupants [18]. Non-component-specific measures, which coordinate across the aforementioned types of DR strategies based on conditions such as outdoor temperature and electricity price, constitute more sophisticated ways of managing a building's demand dynamically, but also require a higher degree of programming [12].

Recent studies have attempted to represent the key types of commercial load flexibility strategies in dynamic building energy simulations [20, 21]. Such studies use physics-based energy modeling tools such as EnergyPlus, DOE-2, and TRNSYS to investigate changes in building demand, thermal dynamics, and changes to other building services from flexible operations during DR events.

3

Since the underlying modeling tools are able to capture both whole building and sub-system/component changes under DR at temporal increments that are suitable for grid planners (e.g., 5 minute-to-hourly changes in HVAC, lighting, and electric equipment schedules), these tools are well-suited to represent the various types of load flexibility strategies identified above. Other tools such as Demand Response Quick Assessment Tool (DRQAT) [22], and EnergyPLAN [23] have drawn from physics-based building simulation capabilities specifically for the purpose of assessing tradeoffs across candidate flexibility strategies, in terms of potential changes in building demand.

Data-driven models offer an alternative to physics-based simulations for characterizing a building's response to dynamic load adjustments [24]. Yin et al. [25] develop a surrogate modeling framework for estimating the theoretical demand flexibility of both commercial and residential buildings, in which regression models are trained to large EnergyPlus datasets on the simulated demand impacts of thermostat adjustment strategies. Kara et al. [26] use statistical autoregressive moving average models fit to measured data to quantify the flexibility of residential thermostatically controlled loads for demand response. The approach relates thermostat setpoint to changes in indoor temperature and instantaneous HVAC power consumption, estimating both the magnitude and duration of demand shedding potential. Afzalan and Jazizadeh [27] segment load data by user types to determine demand reduction potential, using a rough estimate of power reduction potential per degree change in thermostat set point to determine load shedding potential. Pinto et al. [28] combine a surrogate modeling approach with machine learning techniques, developing a deep reinforcement learning controller that manages the operation of heat pumps and chilled and domestic hot water storage at the district scale, minimizing peak demand across the district while maintaining indoor temperatures within a comfort band for individual buildings. Chen et al. [29] use a comfort temperature band to determine potential load flexibility in offices from thermostat adjustments and occupant temperature preferences, including the comfort band as an input to simple single zone energy balance equations; simple models of

4

demand reduction potential from lighting and appliance and thermal storage are also reported. Hurtado et al. [30] quantify the available demand flexibility of individual commercial buildings by fitting models of the conditional probability of various flexibility parameters, including ramp rates, power and energy capacity, and the time taken for indoor temperature to reach the edge of the set point band and subsequently recover to the normal temperature, to a synthetic database generated using whole building energy simulation of baseline and flexible operations. Contreras-Ocaña et al. [31] use clustered synthetic data generated via EnergyPlus reference buildings for offices and supermarkets to train linear models of feasible building load regions under HVAC flexibility, constraining these regions by acceptable temperature limits. The authors demonstrate the use of the models to support aggregators tasked with compensating deviations from forecasted wind generation with building load flexibility. Amara et al. [32] identify a hybrid physics-statistical modeling framework as most effective among data-driven approaches for managing building energy use, and other studies similarly highlight the use of hybrid models for building control and operations [33]. Finally, multiple studies use a resistance-capacitance circuit analogy to model thermal dynamics under load flexibility, requiring physical building information (e.g., zone volume, thermal envelope characteristics) to infer model parameters [34, 35, 36, 37].

Data-driven approaches require substantial measured or simulated data for model training – including building energy use, indoor and outdoor environmental variables, and system control data [38]. Accordingly, data-driven approaches remain a promising but uncommon basis for informing commercial load flexibility strategies in practice [39].

On the basis of existing literature, we identify the following key challenges to quantitative assessment of commercial building load flexibility strategies:

- Existing models of building load flexibility are not easily adapted to specific building instances; given the lack of modeling and computational resources and/or metered data for model training purposes, it is not feasi-

ble for building owners, operators, or consultants to build unique physics-based or data-driven models of the potential impacts of flexibility strategies in specific building applications.

- Few studies attempt to model the potential energy and service impacts of multiple load flexibility strategies at once across multiple commercial building types, with many studies focusing on individual flexibility strategies relating to HVAC adjustments in offices. This is despite the potential for flexibility packages to yield deeper demand reductions while distributing risks across building services, and despite interest in pursuing load flexibility in other commercial building contexts — particularly retail settings.

- Studies tend to focus on the implications of thermal load flexibility strategies in terms of potential changes in building demand; while multiple studies examine the associated implications of such strategies for indoor temperature — a key concern for building operators and owners – such studies tend to either equate indoor temperature changes with set point adjustment levels or rely on idealized models of thermal dynamics that do not account for context-specific factors such as HVAC system constraints or the effects of having multiple thermal zone types in the building.

- Few studies communicate the uncertainties in their predictions, imparting false confidence about potential flexibility impacts.

To address such limitations, we develop surrogate models of key load flexibility strategies for multiple commercial settings that are comprehensive, predictive, and adaptable to new information. Specifically, we use EnergyPlus to generate synthetic data on the potential demand and service impacts of an array of commercial load flexibility strategies and packages across office and retail contexts of multiple scales; train a series of multiple regression models of building demand and service changes under these strategies; and translate the regression-based surrogate models to a Bayesian framework to facilitate communication of

6

model prediction uncertainty and parameter updating given new evidence. The developed models are intended to serve as simple tools that commercial DR participants, consultants, and/or service providers can use to prospectively assess the relative benefits and risks of candidate load flexibility response strategies under a particular set of conditions (e.g., weather, occupancy, economic incentives).

2. Methodology

Surrogate models of commercial building demand and temperature service under load flexibility are developed through the following steps: 1) define commercial DR contexts (climate, building type, building vintage) and develop candidate load flexibility Measures in OpenStudio/EnergyPlus; 2) simulate load flexibility Measures across climate zones, building types, and vintages of interest; 3) compile results into synthetic database covering simulated electricity demand and indoor temperature outcomes under the various measure settings and contexts; 4) fit a series of multiple regression models of building demand and indoor temperature under load flexibility using the synthetic database; translate the regression models to a Bayesian inference framework to address model uncertainty and new parameter updating; and 5) assess model fit and predictive performance using quantitative metrics and qualitative graphical checks. Figure 1 summarizes the methodological steps used in this study, which are further enumerated below.

2.1. Definition of DR contexts and energy models

In this study, we model five different types of commercial buildings using the EnergyPlus engine [40]: medium offices, large offices, all-electric large offices, standalone retail, and big box retail. The medium and large office and standalone retail models are drawn from the U.S Department of Energy's commercial prototype models [41]; the all-electric large office model is based on the prototypical large office model, but is modified to reflect an all-electric HVAC
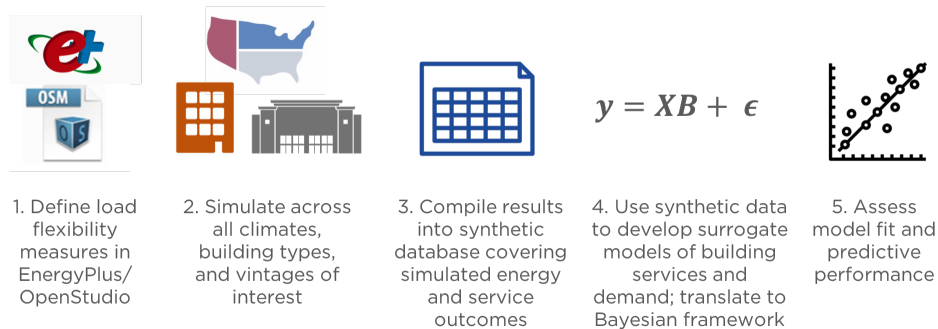
7

Figure 1: Summary of methodological steps for developing surrogate models of changes in commercial building demand and temperature service under load flexibility strategies.

system; the big box retail model draws from the base-case EnergyPlus model previously developed by the National Renewable Energy Laboratory (NREL) for the assessment of the ASHRAE Advanced Energy Design Guide (AEDG) for medium-to-big-box retail [42]. The characteristics of each building type are further described here.

2.1.1. Medium office

The prototypical medium office building is a three-storey office building of 4,900 m^2 with 13 space types. The building is conditioned by a central packaged air conditioning unit with a gas furnace. The building's air distribution is through variable air volume (VAV) terminal boxes with dampers and electric reheating coils. During weekdays, the building is occupied between 6 AM and midnight, with peak occupancy at a fraction of 0.95 for two periods of 8AM-12PM and 1PM-5PM. For lighting, the building is scheduled to a peak fractional intensity of 0.9 from 7AM-5PM, which is gradually reduced to zero by midnight. Equipment schedules (representing plug loads such as PCs and non-PC office type equipment) follow a peak fractional intensity of 0.9 from 8AM-5PM, which is sharply reduced to a baseline level of 0.4 for the remainder of each night.

2.1.2. Large office

The large office building corresponds to the version of the prototypical large office with more detailed thermal zoning; in total, the building is modeled with 100 zones and 14 space types, allowing a more finely-grained assessment of the impacts of load flexibility in different parts of the building. The detailed large office prototype version is available via the OpenStudio energy modeling environment [43]. The building has 13 floors in total including a basement level, ground floor, and 11 above-ground floors; the total conditioned building area is 46,321 m^2. The building is conditioned by a chiller for cooling and gas boiler for space heating with air handling units in each floor that distribute air to each zone via variable air volume (VAV) terminal units with a hot water reheat system. During weekdays, the building is occupied from 8AM-6PM, with peak occupancy from 10AM-12PM and 2PM-5PM. Occupancy, lighting, and electric equipment schedules for the large office are the same as the medium office building. The large office prototype has a main data center in the basement and small data center spaces on each floor; for the purposes of this study, we do not include data center spaces in the scope of electricity demand reductions from load flexibility measures.

2.1.3. All-electric large office

The all-electric large office building variant used in this study is derived directly from the large office detailed standard prototype, but we replace the prototypical gas boiler and chiller HVAC and VAV terminal reheat systems with a water-source heat pump (WSHP) and dedicated outdoor air system (DOAS) using the OpenStudio Building Component Library Measure, *WSHP with DOAS* [44]. With the application of this Measure, the all-electric large office model has centralized DOAS per floor with zone-level conditioning to all space types except storage, stair, and mechanical/electrical rooms. We model an all-electric large office variant for this study in anticipation that the current emphasis on building electrification [45, 46] will result in such buildings and heat pump systems becoming more common in the future, with associated implications for

the scope and applicability of our prototype-based surrogate models of load flexibility impacts.

2.1.4. Standalone retail

The prototypical standalone retail building is a one-storey building of 2,290 m^2 with 5 space types. The building is conditioned by a packaged air conditioning unit with a gas furnace. The building's air distribution is through a constant air volume (CAV) system with four rooftop units (RTUs) serving 4 thermal zones. During weekdays, building occupancy begins at 7 AM, with peak occupancy occurring from 11AM-5PM at a fraction of 0.95 of the maximum; the occupant fraction is then gradually reduced to zero by 9PM. The building is scheduled to a peak fractional lighting intensity of 0.9 from 9AM-6PM, which gradually reduces to zero by midnight. Equipment schedules follow a peak fractional intensity of 0.9 from 9AM-7PM, which is sharply reduced to a baseline level of 0.2 for the remainder of each night.

2.1.5. Big box retail

Finally, the big box retail building model is an adapted version of the baseline model developed originally by NREL [42] to represent larger merchandise stores under the ASHRAE 90.1-2004 code level. The model has sales floor and back-of-house areas – 11 zones and 9 space types in total with a conditioned building area of 9,218 m^2. The building is conditioned via unitary heating and cooling equipment, with a single duct air terminal, direct expansion (DX) air conditioner, and electric baseboard heating. During weekdays, building occupancy begins at 7 AM and lasts until midnight, with peak occupancy occurring between 11PM-1PM and 5PM-7PM at a fraction of 0.625 of the maximum; the occupant fraction is then gradually reduced to zero by 1 AM the following day. The building is scheduled to peak fractional lighting and electric equipment intensities of 0.9 from 7AM-9PM. For the purposes of this study, we set the building's humidistat, which controls indoor humidity, to 90% to avoid over-cooling issues observed with this control in place in the original model (the

original model uses a 60% humidistat set point). The original model is available for EnergyPlus version 6.0 on the Building Component Library website [47] and was updated to be compatible with EnergyPlus version 9.2.

2.1.6. Climate zones and building vintages

Office and retail building models are simulated across two building vintages, 1980-2004 and 2010, which are chosen to represent older and newer building construction characteristics, respectively. The adapted big box retail model is only available for the 2004 vintage and simulations of this building type are therefore limited to this vintage. All building types are simulated across a range of ASHRAE 90.1 climate zones [41] – 2A, 3C, 4A, and 6B – which are chosen to capture the influence of variation in weather as well as location-specific building design codes. In all, we simulate 36 unique contexts (4 building types * 2 vintages * 4 climate zones, plus 4 climate zones for the single big box retail vintage).

2.2. Definition of load flexibility strategies

Table 1 summarizes the load flexibility strategies implemented in the EnergyPlus simulations, presenting key characteristics assumed for each strategy. We represent the application of four types of flexibility strategies: 1) lighting dimming, 2) plug load reduction through low-priority computer/office electronic device switching, 3) global temperature adjustment (GTA), and 4) GTA + precooling in the hours directly preceding a DR event window. Each strategy is implemented in response to an assumed DR event window of 3-7 PM, which is chosen to reflect a net summer peak load period that is expected to be typical for utility systems in the coming years, as summarized across U.S. regions in the Supplemental Information of [48].[1] In the simulations, each strategy is represented as a modification of daily baseline operational schedules. The lighting and plug load strategies are assessed together exclusively across a full

---

[1]Net peak load is defined as the peak in total summer system load minus projected variable renewable energy generation.

year, while the GTA and pre-cooling strategies are assessed for the summer period only (May–Sep), in some instances alongside packaged lighting and/or plug loads strategies. Modified schedules that represent the strategies are generated and applied in EnergyPlus simulations using the OpenStudio Measures feature, which consists of small scripts that adjust an EnergyPlus model across various dimensions including the building and system characteristics as well as operational and occupancy schedules [49]. The modified hourly schedules are represented in "Compact:Schedule" objects in order to maintain the same system design settings as those in the baseline. Measure definitions for each flexibility strategy have been published on a GitHubrepository [2].

Table 1: Load flexibility strategy assumptions.

| Category | Measure | Magnitude of adjustment | | | Duration of adjustment |
		Low	(Uniformly distributed)	High	
HVAC	Global cooling temp. adjustment (GTA)	+1°F	to	+6°F	3-7 PM
	GTA and	+1°F	to	+6°F	3-7PM
	pre-cooling	-1°F	to	-4°F	1 to 6-hour ahead
					(Uniformly distributed)
Lighting	Dimming	0%	to	-100%	3-7 PM
Plug Loads	Low-priority device switching				

To ensure that a wide range of possible adjustment levels is represented for each flexibility strategy, we sample schedule values from uniform distributions that cover the feasible ranges of values shown in Table 1. Specifically, for both lighting and plug loads strategies, the reductions are generated using a continuous uniform distribution bounded from 1% to 100%. The GTA and pre-cooling strategy settings are also generated using a discrete uniform distribution; GTA cooling set point increases during the DR period are sampled between the range of 1°F and 6°F, while pre-cooling set point decreases are sampled between the range of 1°F and 4°F. Pre-cooling is represented for half of the simulated summer

---

[2]https://github.com/jtlangevin/flex-bldgs/tree/master/measures

days to ensure a balance in the data set between days with and without pre-cooling applied; pre-cooling application ranges between 1 and 6 hours directly preceding the DR event period, with duration drawn from a discrete uniform distribution. Sampled schedule draws are independent across the strategies.

Finally, to ensure that we capture the effects of individual strategies in isolation in our data, we force a single occurrence of 0% for the lighting strategy, plug loads strategy, and both lighting and plug loads on a weekly basis. As a result of this approach and the aforementioned restriction on pre-cooling instances, there are 22 occurrences of a GTA + pre-cooling-only strategy, 18 occurrences of a GTA-only strategy, 32 occurrences of a GTA + lighting + plug loads strategy, and 30 occurrences of a GTA + pre-cooling + lighting + plug loads strategy across simulated summer weekdays (May to September); there are 52 weekday occurrences of a lighting-only strategy, 48 weekday occurrences of a plug loads-only strategy, and 150 weekday occurrences of a lighting + plug loads strategy in the year-round simulations of those measures exclusively.

2.3. Whole building simulation of load flexibility strategies

Load flexibility strategies represented as OpenStudio Measures are executed via batch simulations of EnergyPlus that leverage the OpenStudio Workflow (OSW). Using the OSW, the office and standalone retail prototype models are seeded as OpenStudio Models, to which flexibility Measures and an additional set of Reporting Measures are applied. The big box retail runs are also performed in batches utilizing the JEPlus software platform [50], a tool for running EnergyPlus models parametrically. Flexibility Measure settings are applied via .csv files consisting of pre-defined schedules for each candidate strategy (based on settings in Table 1), while Reporting Measure arguments contain the type of reporting variable desired and the time resolution with which that variable should be reported. Given these arguments, the OSW executes the EnergyPlus engine and generates the required reporting files across the various DR contexts. All simulations are executed with an hourly time step.

Table 2 summarizes the full range of flexibility measure types, simulation

periods, building types and vintages, climate zones, and output metrics that are covered by the whole building simulations.

Table 2: Summary of load flexibility measure categories, simulation periods, building types and vintages, climate zones, and output metrics that are covered by this paper's whole building simulations.

Measure Category	Simulation Period	Building Types	Building Vintage(s)	Climate Zones [41]	Output Types
HVAC, HVAC and Lighting and/or Plug Loads	Summer (May-Sep)	Medium Office, Large Office, All-Electric Large Office, Standalone Retail	1980-2004, 2010	2A, 3C, 4A, 6B	Demand shed intensity (W/sf), Indoor temp. change (°F)
		Big Box Retail	2004		
Lighting and/or Plug Loads	Full Year	Medium Office, Large Office, All-Electric Large Office, Standalone Retail	1980-2004, 2010	2A, 6B	Demand shed intensity (W/sf)
		Big Box Retail	2004		

2.4. Development of synthetic database

After conducting the batch simulations of commercial load flexibility strategies across the dimensions listed in Table 2, hourly results are compiled into a synthetic database that includes the electricity demand and indoor temperature outputs and potential predictors of these outputs. Regarding indoor temperature results, the following adjustment is applied to the raw temperature data generated for each building thermal zone to yield a single occupant-weighted average indoor temperature variable to use in the dataset:

$$T_{ave} = \frac{\sum_{i=1}^{n}(T_i \times Occ_i)}{\sum_{i=1}^{n} Occ_i} \tag{1}$$

Where T_{ave} represents the occupant-weighted averaged indoor temperature; i represents the conditioned zone number, while n represents the total number

of conditioned zones; T_i represents the indoor temperature within zone i; and Occ_i represents the maximum potential occupant number within zone i.

Leveraging the authors' domain expertise and the literature review presented earlier, we identify three categories of potential predictors to report to the synthetic database and include in subsequent model development: 1) outdoor environmental conditions and occupancy, 2) load flexibility strategy characteristics (e.g., changes in cooling set point, lighting and plug load schedules) and – for cooling set point – its single time step lag, and 3) time duration indicators (e.g., hours since flexible operations have started and ended).

Raw simulation results are further filtered in the following ways to develop the final synthetic database:

- we restrict the data to weekdays only, under the expectation that commercial load shedding and shifting opportunities will be limited for this day type, when the buildings are typically occupied;

- within each weekday, we restrict hourly data points to those that fall in the DR event window, as well as any hours preceding the window in which pre-cooling was simulated and one hour following the event for any strategies with temperature adjustment, to capture a potential rebound in demand as the HVAC system recovers from this adjustment; and

- for any strategies with temperature adjustment (i.e., GTA and pre-cooling), we restrict the data to hours in which the outdoor air temperature is equal to or higher than 70°F, which is assumed to be a threshold at which substantial cooling loads are present and consistent with previous studies finding fundamentally different relationships between building demand and moderate to low outdoor temperatures [25].

The synthetic data are published on GitHub[3].

_____

[3] https://github.com/jtlangevin/flex-bldgs/tree/master/data

2.5. Surrogate model development using synthetic data

To develop surrogate models of building electricity demand and indoor temperature under load flexibility strategies, we fit multiple linear regressions to the synthetic data that predict these outcomes of interest in terms of their key correlates. In choosing the multiple regression approach, we prioritize model simplicity and interpretability, seeking to develop accurate predictions that rely on only a few readily quantified predictor variables for improved usability. A minimum set of model predictors is determined as outlined in section 2.4 based on expectations about the variables that are likely to have a real-world relationship to building demand and indoor temperature that is well-captured by a linear model. Given the multiple regression framework, we are able to assess the significance of individual variables to the predicted outputs alongside the determination of overall model predictive performance. Interaction terms are included in the multiple regressions under the expectation that the influence of certain model predictors on the output is conditioned on the values of other predictors in the model, such as the moderating effect of weather and occupancy on the impacts of changing a zone cooling setpoint.

We group our surrogate models by the five building types and two vintage bins described in section 2.1 under the expectation that differences in the operational schedules across building types and in construction characteristics across building vintages warrant the development of unique models for each building type and vintage combination in our synthetic data. Within each building type and vintage combination, we further segment the surrogate models of changes to building demand into models for strategies that drive changes in demand through changes in thermal loads (e.g., GTA and pre-cooling) and strategies that do not primarily influence demand through changing thermal loads (e.g., dimming lights and reducing plug load power). The former thermally-driven demand model is trained on synthetic data that includes the one hour rebound period, while the latter non-thermally-driven demand model is trained on synthetic data that excludes the rebound period – lighting and plug load strategies are observed to cause negligible recovery demand in the hours following the

16

DR period. We also fit a separate model for changes in demand during the pre-cooling period, given the unique thermal load dynamics of this period compared to the DR event window (e.g., pre-cooling *increases* in thermal load vs. the *decreases* during the DR event window). Finally, to predict changes to indoor temperature, we restrict the synthetic training dataset to exclude data points that reflect only lighting or plug load changes, as such changes were observed to have only small effects on zone temperature. The temperature model is trained on synthetic data with the one hour rebound period removed, as the indoor temperature tends to move quickly back to the thermostat set point during this period, leaving no temperature changes for the model to predict.

Table 3: Summary of demand and temperature surrogate model inputs and outputs for a given combination of building type and vintage. Shown are model types, outputs, and predictors, with interactive predictor variables highlighted in red.

	Whole Building Demand (DR, Non-thermally-driven)	Whole Building Demand (DR, Thermally-driven)	Whole Building Demand (Pre-cool)	Indoor Temperature (DR)
Output	Demand shed intensity (W/ft^2)	Demand shed intensity (W/ft^2))	Demand shed intensity (W/ft^2)	Indoor temperature change (°F)
Input	Lighting dimming (%)	Outdoor temperature (°F)	Outdoor temperature (°F)	Outdoor temperature (°F)
	Plug loads reduction (%)	Outdoor humidity	Outdoor humidity	Outdoor humidity
		Occupancy fraction	Occupancy fraction	Occupancy fraction
		Cooling set pt. change (°F)	Cooling set pt. change (°F)	Cooling set pt. change (°F)
		Lighting dimming (%)	Hours since pre-cool started	Cooling set pt. lag (°F)
		Plug loads reduction (%)	*Cooling change * Outdoor temp.*	Hours since DR started
		Cooling set pt. lag (°F)	*Cooling change * Occ. fraction*	Pre-cool set pt. change (°F)
		Hours since DR started	*Cooling change * Since Precool started*	Pre-cool duration
		Hours since DR ended		*Cooling change * Outdoor temp.*
		Pre-cool set pt. change (°F)		*Cooling change * Occ. fraction*
		*Cooling change * Outdoor temp.*		*Cooling change * Since DR started*
		*Cooling change * Occ. fraction*		*Pre-cool change * Pre-cool duration*
		*Cooling change * Since DR started*		

Table 3 summarizes the ultimate set of surrogate models that was developed for each combination of building type and vintage, including the outputs and set of predictor variables that were chosen for each model. The surrogate models for demand yield demand shed intensity (W/ft^2), while the surrogate model for indoor temperature yields the change in indoor temperature (°F). Predictor variables follow the three categories identified in section 2.4, and also include

interactive terms that capture the moderating effect of one predictor variable on another. Note that in the models of thermally-driven changes in demand and indoor temperature during the DR event window, inputs concerning the magnitude and duration of any pre-cooling before the event are included to account for the effects of pre-cooling during the DR event window.

Initial inference of model coefficient estimates is performed with frequentist regression methods using the full synthetic dataset for training purposes; subsequent Bayesian inference of coefficient distributions is performed using the full synthetic dataset as described in the following section. Use of the full synthetic dataset for model inference avoids the need to set aside potential training data for model testing; however, this approach introduces the possibility of overfitting the models to data that do not generalize well to the true distribution of observations. To mitigate this issue, we use k-fold cross validation [51] to re-estimate the models under multiple partitions of training/testing data and compare model performance under those alternate training/testing cases to those generated using the full synthetic database for model training and testing. Specifically, under k-fold cross validation, the original sample is randomly partitioned into k equal sized sub-samples. Of the k sub-samples, a single sub-sample is retained as the validation data for testing the model, while the remaining k - 1 sub-samples are used as training data. The cross validation process is then repeated k times, with each of the k sub-samples used once as the validation data. In this work, we use k=10; thus, model performance metrics (see 2.7) are recorded for each of the 10 assessments for each model and the resulting ranges and averages of these metrics are summarized alongside the estimates based on the use of the full synthetic dataset for training, to determine sensitivity to training and testing data choice and identify potential overfitting issues.

2.6. Translation of surrogate models to Bayesian framework

In practice, the input/output relationships that are mapped by the surrogate models may differ between prototypical and real offices. To mitigate this issue, uncertainty in surrogate model predictions should be represented in a straight-

forward manner and the models must be flexible to updating given new evidence collected in real building environments. Both of these criteria are well supported by translation of the surrogate models to a Bayesian inference framework.

Bayesian inference treats model parameters as random variables, deriving the *posterior* probability of model parameter values as a function of the *likelihood* of observed data given the parameters and *prior* parameter probabilities:

$$p(\theta|y, X, \alpha) \propto p(y|\theta, X)p(\theta|\alpha) \tag{2}$$

Where θ is a vector of model parameters, y and X are the observed response variable and covariates, respectively, and α is a vector of hyperparameters; $p(\theta|y, X, \alpha)$ is the posterior parameter probability given the observed data and hyperparameter values; $p(y|\theta, X)$ is the likelihood of observed data given the parameter values; and $p(\theta)$ is the prior parameter probability given the hyperparameter values.

Following from equation 2, the *posterior predictive distribution* of new data points $\hat{y}$ is generated by marginalizing over the posterior parameter distributions:

$$p(\hat{y}|y, X, a) = \int p(\hat{y}|\theta)p(\theta|y, X, \alpha) \, d\theta \tag{3}$$

The range of model predictions generated by equation 3 satisfies the requirement of communicating uncertainty in model outcomes, while the ability to weigh prior expectations about model parameters against new evidence as in equation 2 supports model updating with data collected in the field.

To implement the surrogate models in this framework, we use PyMC3 [52], a Python package for probabilistic programming. Models are initialized using the previously described synthetic data to populate the likelihood function for standard linear regression ($p(y|\theta, X) \sim \mathcal{N}(\mu = X^T\theta, \sigma)$) and assuming diffuse parameter prior distributions ($\theta \sim N(\mu = 0, \sigma = 10)$, $\sigma \sim \mathcal{H}(\sigma = 20)$) to reflect our lack of *a priori* beliefs about parameter values. Model inference is performed using the No U-Turn Sampler (NUTS) algorithm [53], a Markov chain Monte

Carlo (MCMC) sampling method that utilizes gradient information from the likelihood function to converge more quickly than traditional sampling methods for Bayesian inference, avoiding the random walk behavior and sensitivity to correlated parameters of the traditional methods while also requiring no no manual user tuning to converge. The posterior parameter distributions estimated through this process are published on GitHub[4] and may serve as informative prior distributions in subsequent rounds of model updating.

2.7. Model evaluation

The fit and predictive performance of the surrogate regression models were assessed via the following metrics:

1. *Adjusted R-squared (R^2)* is a goodness–of–fit metric; The R^2 metric indicates the proportion of the variability in the response data about its mean that is explained by one or more independent variables. The adjusted R^2 additionally accounts for the explanatory power of each variable that is added to the model, increasing only if the variable improves the model more than would be expected by chance; accordingly, the adjusted R^2 is always equal to or lower than the R^2 value. Previous building energy simulation studies tend to report R^2 values, finding R^2 up to 0.78 for prediction of hourly DR potential using regression models [25]; studies that predict annual energy use with linear models suggest that higher R^2 values above 0.90 are possible [54].

2. *Absolute Relative Error (ARE)* is a predictive accuracy metric; ARE subtracts observed from predicted values and normalizes each difference by the observed values [25]. Previous work focused on DR potential estimation for global temperature adjustments via linear models Yin et al. [25] demonstrates that more than 80% of data points can be predicted with an ARE of 20% or less.

---

[4]https://github.com/jtlangevin/flex-bldgs/tree/master/model_stored

3. *Mean Absolute Deviation Percentage (MADP)* is a second measure of accuracy; MADP is the ratio of the sum of deviations between observed and model-predicted values to the sum of observed values, and is one of the most commonly-used evaluation measures for models that predict building energy use [55]. MADP is preferred as a performance metric over Mean Absolute Percent Error (MAPE) because the latter misrepresents the overall error rate when actual values are close to zero [56]. A previous building energy simulation study by Fan et al. [57] documents MADP values less than 9% for prediction of next-day energy use and peak demand; more generally, MADP values less than 20% are considered acceptable in the building simulation context [56, 58].

4. *Variance inflation factor (VIF)* is a measure of variable multicollinearity [59]; VIF is the ratio of overall model variance to the variance of a model that includes only a given predictor variable [60]. VIF values higher than 10 are assumed to suggest the need for model adjustments to improve the stability of variable coefficient estimates. In Aghdaei et al. [54]'s work, for example, regression models that predict annual cooling and heating demand are evaluated with VIF values less than 1.6, indicating insignificant correlation between model parameters.

5. *Posterior predictive checks (PPCs)*, which are relevant specifically to the Bayesian model implementation, graphically compare observed and modeled data [61]. PPCs are a qualitative assessment method that can be used to check for systematic discrepancies between the modeled and observed outcomes.

3. Results

The following sub-sections summarize the predictive performance of each of the developed surrogate models in accordance with the aforementioned assessment metrics. Each sub-section presents two types of plots: 1) a bar plot showing the distribution of the absolute relative errors (ARE) between predicted

21

values and values observed in the synthetic training dataset, with the 20% error line denoted on the plots by a red dashed line; and 2) a scatter plot comparing predicted vs. simulated values for the modeled variable, with a red solid line denoting an ideal 1:1 relationship between the two and a red dashed line indicating the 20% error threshold.

Model coefficient estimates and their statistical significance levels are reported in Appendix A, Tables A1 - A4. We also demonstrate in Appendix B that the use of the full synthetic database for model training and assessment does not present overfitting issues, as model performance under the k-fold cross validation method described in section 2.5 is generally comparable to or somewhat better than that reported in the following sections, under the use of the full synthetic database for training and assessment.

3.1. Electricity demand models

3.1.1. Non-thermally-driven demand during DR event period

Figure 2 shows the performance of the surrogate model of the non-thermally-driven change in electricity demand under DR for the prototypical offices. The error distribution plots in Figures 2a–d show that 88% and 83% of the predictions fall within the 20% error threshold for the new and old medium offices, respectively, while 90% and 97% of predictions fall within this error threshold for the new and old large offices, respectively. R^2 values are similarly high at generally above 0.88 across models, while MADP values are at between 12–16% for medium offices and 3–8% for large offices, respectively. The scatter plots in (e–h) indicate that most of the predicted data points fall within 20% of the ground truth values from the synthetic database, though a slight tendency to over-predict changes in demand is observed in the medium office results. Demand shed intensities predicted for lighting and plug load strategies in the older vintage office buildings are generally higher than that of the new office buildings – a reflection of the lower efficiency of baseline lighting and electronic equipment in the older building vintages that is apparent across all building types examined in this analysis.

22

Figure 2: Histogram of the absolute relative error (a–d) and scatter plot of the predicted versus simulated values (e–h) in the model of non-thermally-driven changes in demand during the DR event window, for new medium office (a, e), old medium office (b, f), new large office (e, g), and old large office (d, h).
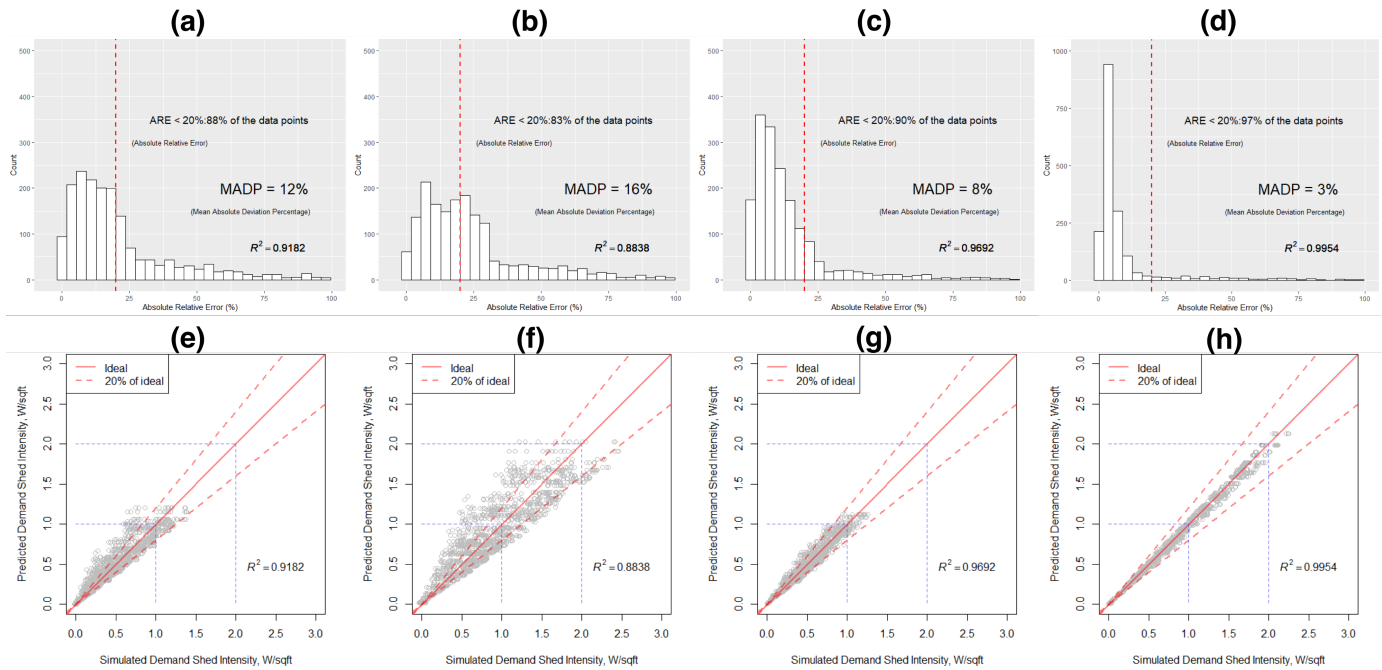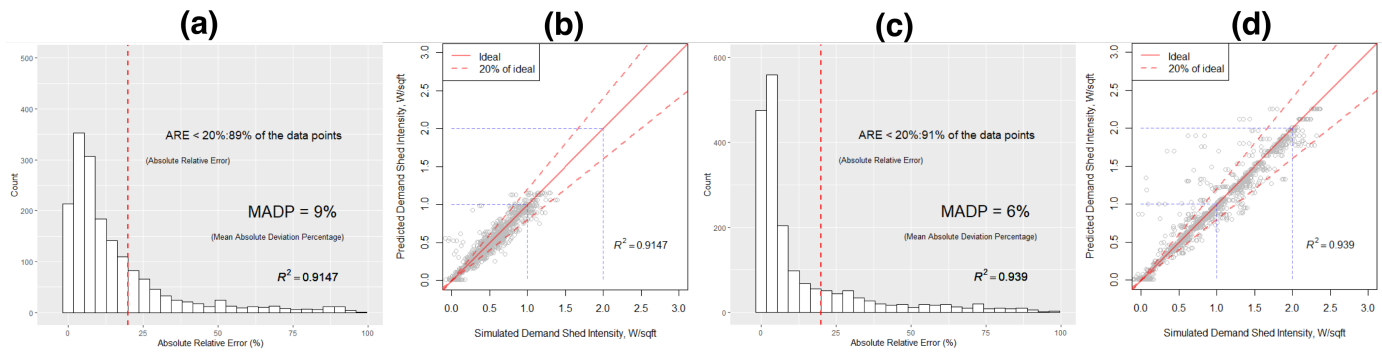

Figure 3: Histogram of the absolute relative error (a, c) and scatter plot of the predicted versus simulated values (b, d) in the model of non-thermally-driven changes in demand during the DR event window, for new all-electric large office (a, b), and old all-electric large office (c, d).

23

Figure 3 shows the same information as Figure 2 for new and old vintages of the all-electric large office, respectively. The error distribution plots in Figures 3a and c show that 89% and 91% of the predictions fall within the 20% error threshold for the new and old all-electric offices, respectively. R^2 values are generally above 0.91 for the models in both vintages, while MADP values are at between 6–9%. The scatter plots in (b) and (d) indicate that as for the prototypical large office, most of the predicted data points for the all-electric large office fall within 20% of the ground truth values from the synthetic database, though the distribution of points around the 1:1 reference line is wider in the all-electric case.

Figure 4 shows the same information as Figures 2-3 for the retail building types. Here again, predictive performance is high – 94% and 97% of the predictions fall under the 20% error threshold for new and old standalone retail, respectively, while 95% of predictions fall under the threshold for big box retail. R^2 values are consistently above 0.98 across models, while MADP values are comparably low to the office context at 4–6%. As in Figures 2-3, the scatter plots in Figure 4d–f indicate that almost all predicted data points fall within 20% of the ground truth values from the synthetic database. As for offices, the demand shed intensities predicted for lighting and plug load strategies in the older vintage standalone retail buildings are generally higher than that of the new standalone retail buildings.

3.1.2. Thermally-driven demand during DR event period

Surrogate models of thermally-driven changes in building demand under DR demonstrate comparable predictive capability to the non-thermally driven demand models, across building types. Per the error distribution plots in Figures 5a–d, more than 90% of the predictions fall under the 20% error threshold for prototypical medium offices and large offices across vintages. Across models, R^2 values are above 0.97, while MADP values are between 7–8%. The scatter plots in Figures 5e–h further indicate good agreement between predicted values and those observed in the underlying synthetic data. Models of thermally-driven
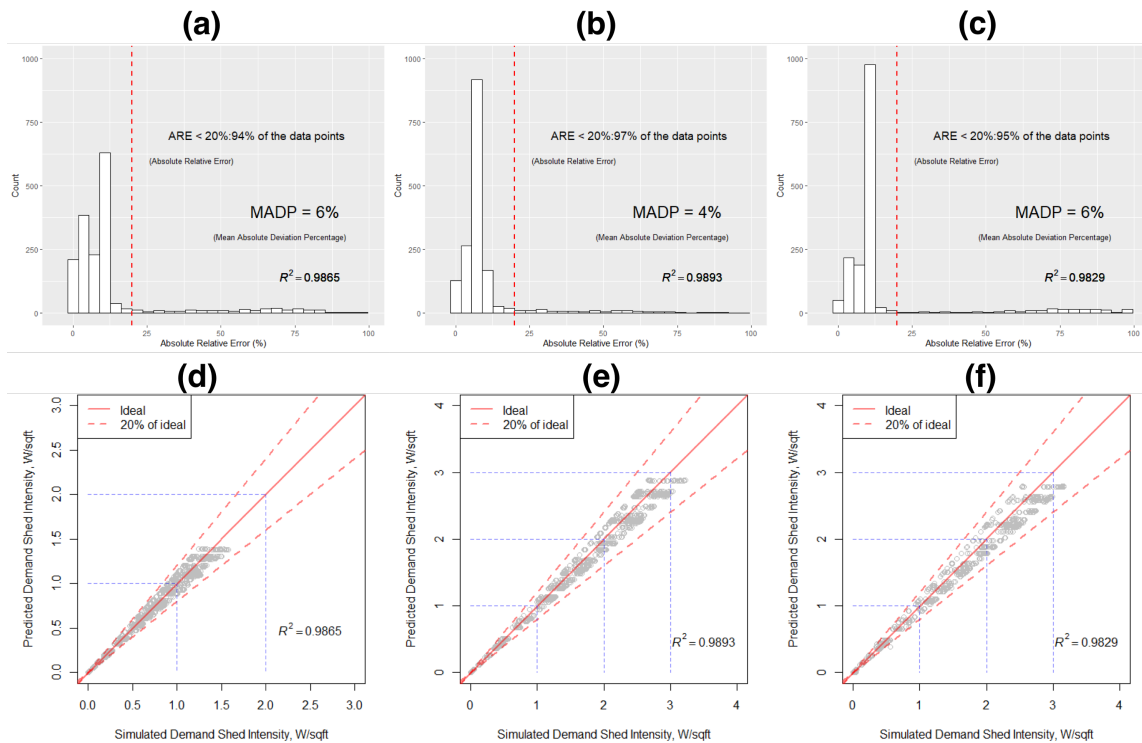
Figure 4: Histogram of the absolute relative error (a–c) and scatter plot of the predicted versus simulated values (d–f) in the model of non-thermally-driven changes in demand during the DR event window, for new standalone retail (a, d), old standalone retail (b, e), and 2004 vintage big box retail (c, f).

demand in the older large office vintage demonstrate a superior predictive performance to the other model types – likely reflecting the poorer envelope characteristics of this particular building type, which leads to a stronger relationship between outdoor air temperature, cooling set points, and cooling loads that favors higher predictive performance for the surrogate model of change in demand under DR. The poorer envelope and HVAC equipment performance characteristics of the older vintage models yield a larger range of demand shed intensities than the newer vintage models, mirroring the trend seen in the models of non-thermally-driven changes in building demand under DR.

Figure 6 shows the same information as Figure 5 for new and old vintages of

Figure 5: Histogram of the absolute relative error (a–d) and scatter plot of the predicted versus simulated values (e–h) in the model of thermally-driven changes in demand during the DR event window, for new medium office (a, e), old medium office (b, f), new large office (e, g), and old large office (d, h).

the all-electric large office, respectively. The error distribution plots in Figures 6a and c show that 92% of the predictions fall within the 20% error threshold for both vintages, while R^2 values are also high at above 0.97 and MADP values are low at between 6–7%. Accordingly, scatter plots for the all-electric office in Figure 6b and 6d indicate that most of the predicted data points fall within 20% of the ground truth values from the synthetic database.

In retail buildings, a strong relationship is observed in the synthetic data between outdoor air temperature, cooling set points, and cooling loads, which yields even higher performance for the thermally-driven demand model than is achieved in the office building contexts. As shown in Figure 7, in the retail context more than 94% of the thermally-driven demand model's predictions fall within the 20% error threshold; R^2 values are consistently above 0.98, and
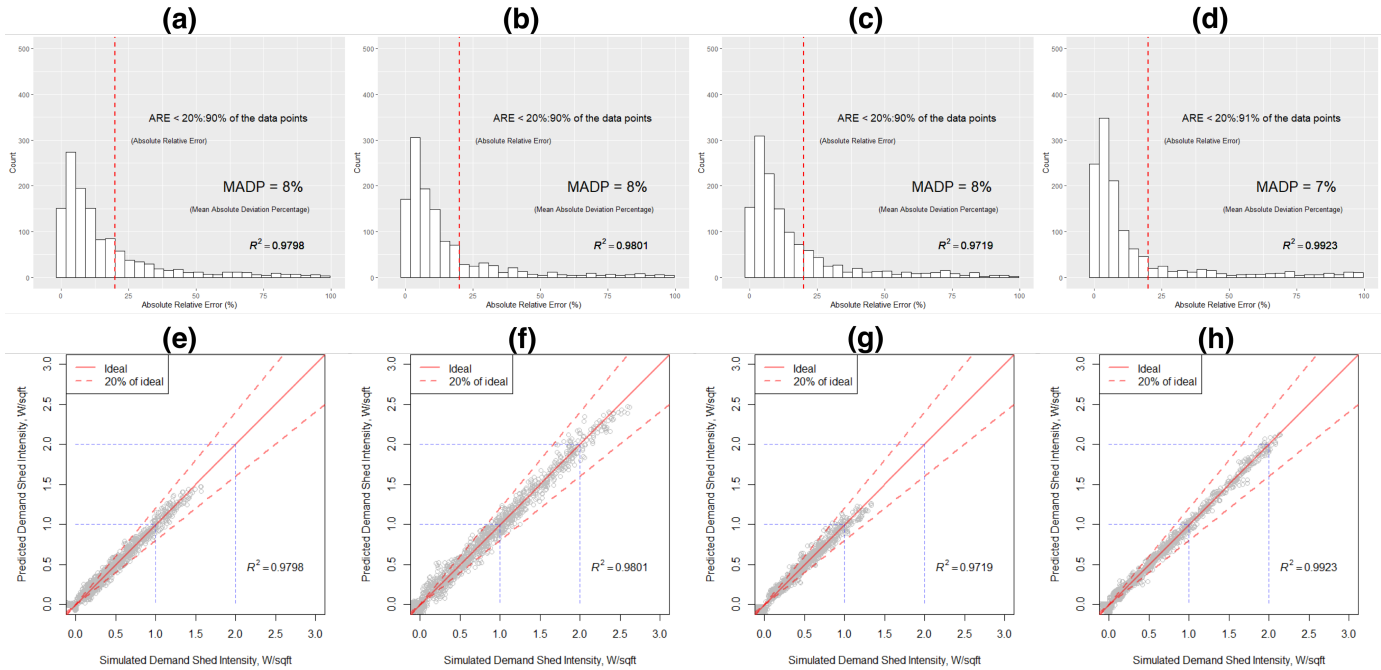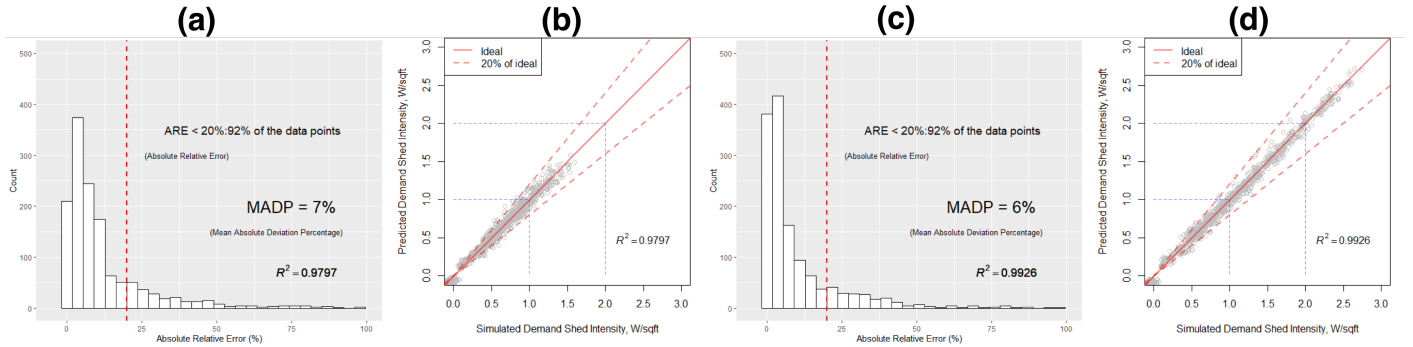
26

Figure 6: Histogram of the absolute relative error (a, c) and scatter plot of the predicted versus simulated values (b, d) in the model of thermally-driven changes in demand during the DR event window, for new all-electric large office (a, b), and old all-electric large office (c d).

MADP values are between 4–6%; data points in the scatter plots of Figures 7d–f are consistently well within the 20% error threshold.

3.1.3. Thermally-driven demand during pre-cooling period

Figures 8-10, which summarize the performance of models of the effects of pre-cooling on demand in prototypical and all-electric offices, and prototypical retail buildings, respectively, demonstrate somewhat lower but acceptable predictive performance than is observed for the models of changes in demand during the DR event period in Figures 2-7. Under pre-cooling, the range of simulated changes in demand observed in the scatter plots of Figures 8-10e–h is notably narrower than the range of simulated changes in demand during the DR event window, reflecting greater consistency in the need for cooling energy to maintain the lower pre-cooling set point temperature. Nevertheless, across building types, more than 85% of predictions are within 20% of the ground truth data for the prototypical offices; 91% of the predictions are within 20% of the ground truth data for all-electric large offices; while more than 83% of predictions are within 20% of the ground truth data for retail buildings. R^2 values are consistently above 0.88 for offices and above 0.82 for retail buildings, and scatter plots demonstrate good correspondence between predicted demand values and those observed in the synthetic database. MADP values range between 7–15%

27

Figure 7: Histogram of the absolute relative error (a–c) and scatter plot of the predicted versus simulated values (d–f) in the model of thermally-driven changes in demand during the DR event window, for new standalone retail (a, d), old standalone retail (b, e), and 2004 vintage big box retail (c, f).

across building types and models, with the largest errors observed for the big box retail building type. In this case, a wider range of occupancy fractions observed during the pre-cooling period – from 0.1-0.7 during the 8 hours preceding the DR period, compared to 0.3-0.6 for the office buildings – likely challenges the model's ability to capture hour-to-hour changes in the additional cooling demand that is needed to meet the pre-cooling set point, pushing up the model error.

3.2. Indoor temperature models

The final set of surrogate models predicts changes in indoor temperature under DR, given the expected need for building operators to trade off predicted

28

Figure 8: Histogram of the absolute relative error (a–d) and scatter plot of the predicted versus simulated values (e–h) in the model of changes in demand during the pre-cooling period, for new medium office (a, e), old medium office (b, f), new large office (e, g), and old large office (d, h).

Figure 9: Histogram of the absolute relative error (a, c) and scatter plot of the predicted versus simulated values (b, d) in the model of changes in demand during the pre-cooling period, for new all-electric large office (a, b), and old all-electric large office (c, d).

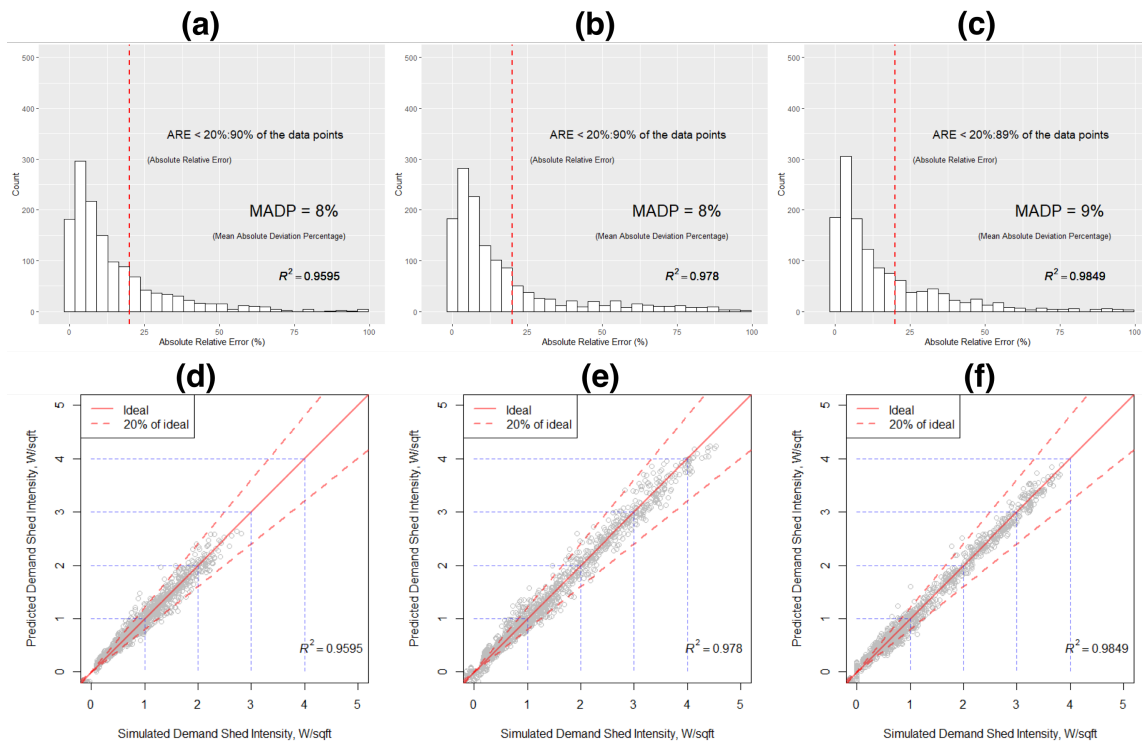changes in demand and associated economic benefits against the risk of building

Figure 10: Histogram of the absolute relative error (a–c) and scatter plot of the predicted versus simulated values (d–f) in the model of changes in demand during the pre-cooling period, for new standalone retail (a, d), old standalone retail (b, e), and 2004 vintage big box retail (c, f).

service losses. Here, model performance is highest across the all-electric large office (in Figure 12) and retail types (in Figure 13), and lower but still acceptable for the medium office and standard large office building types (in Figure 11).

For all-electric large offices and retail, more than 95% of predictions fall within the 20% error threshold; R^2 values are consistently above 0.93; and MADP values range from 3–6%. For prototypical medium and large offices, more than 81% of predictions fall within the 20% error threshold across building types and vintages; R^2 values are generally above 0.76; and MADP values range from 12–17%.

Examining the scatter plots in Figures 11-13, a notably fuller range of tem-

30
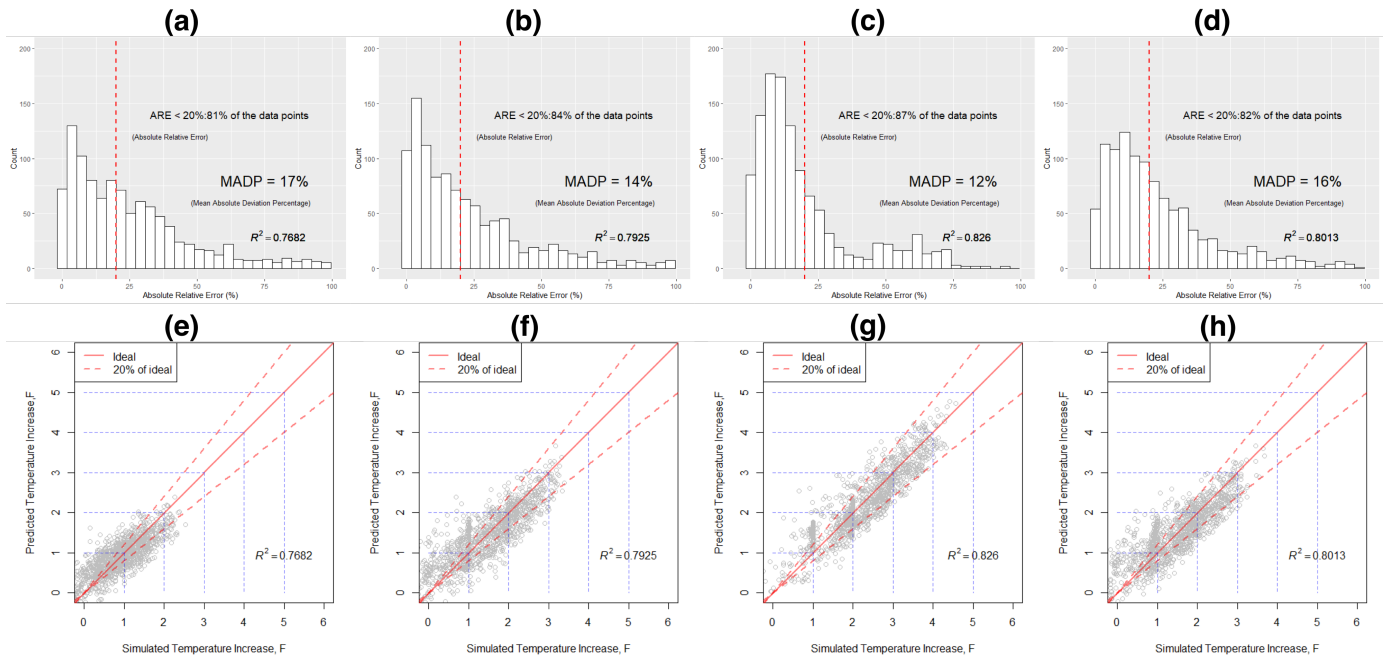
Figure 11: Histogram of the absolute relative error (a–d) and scatter plot of the predicted versus simulated values (e–h) in the model of changes in indoor temperature during the DR event window, for new medium office (a, e), old medium office (b, f), new large office (e, g), and old large office (d, h).
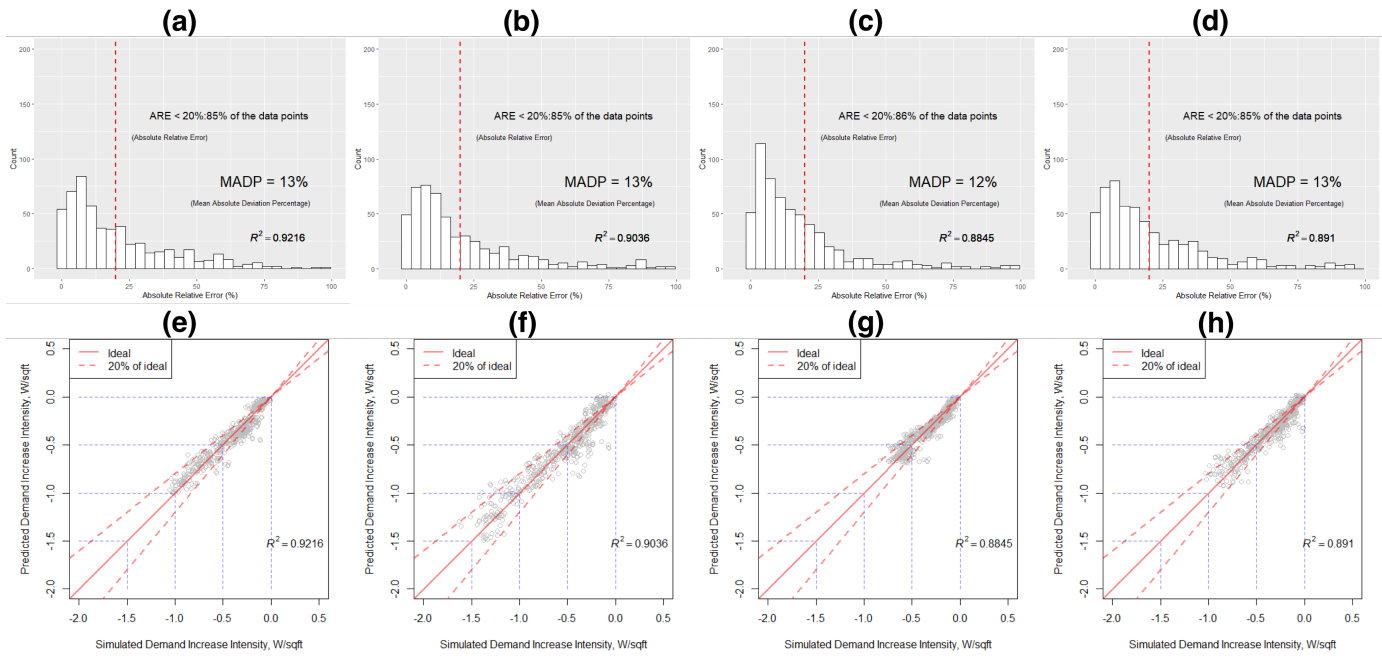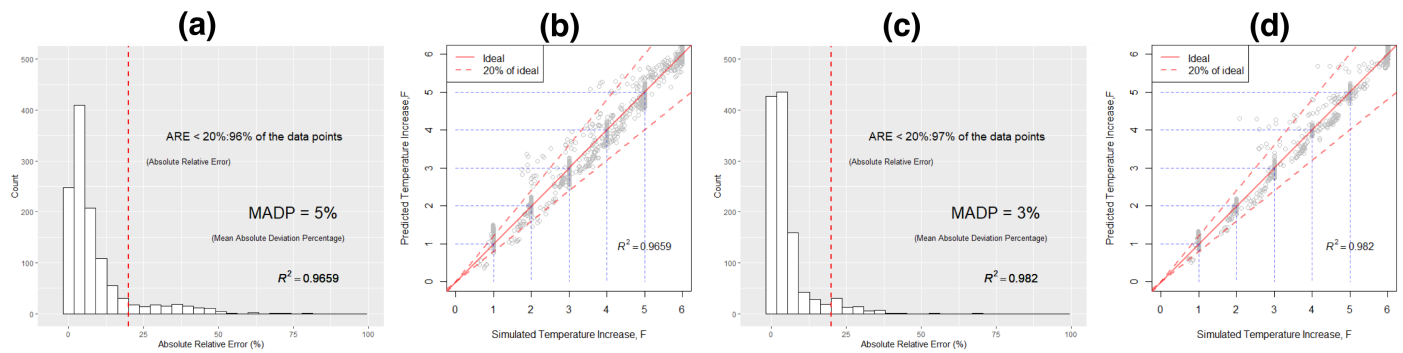

Figure 12: Histogram of the absolute relative error (a, c) and scatter plot of the predicted versus simulated values (b, d) in the model of changes in indoor temperature during the DR event window, for new all-electric large office (a, b), and old all-electric large office (c, d).
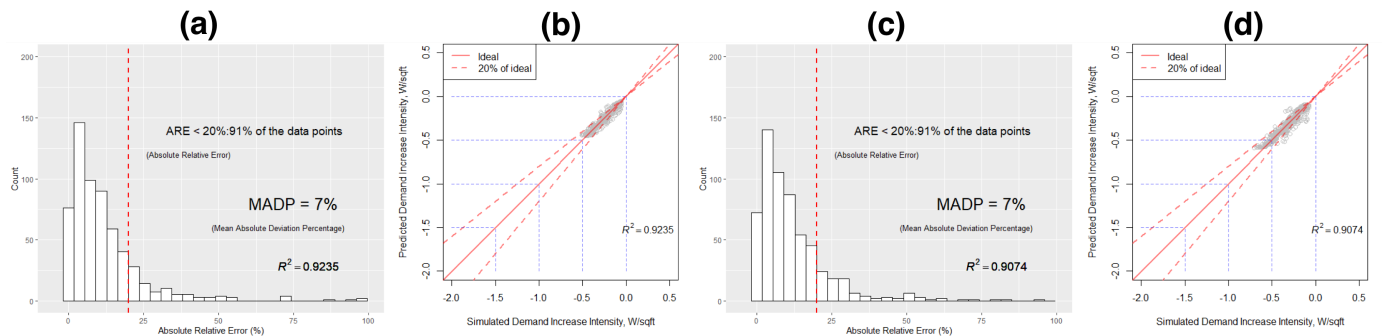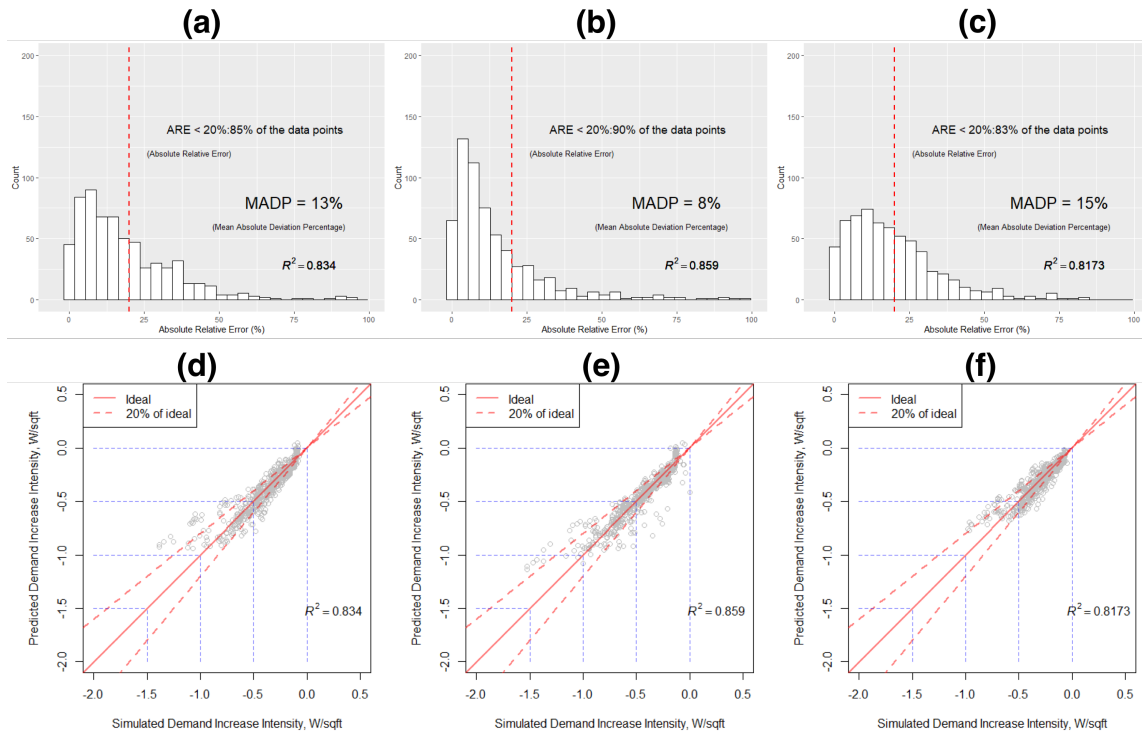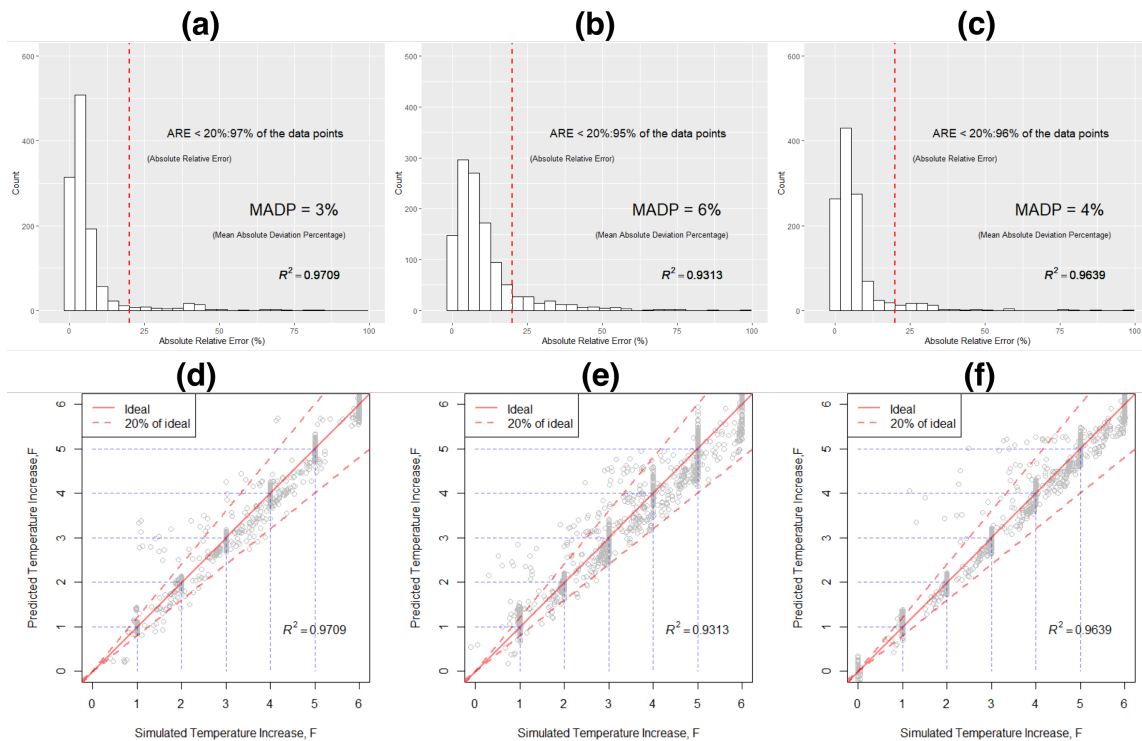
Figure 13: Histogram of the absolute relative error (a–c) and scatter plot of the predicted versus simulated values (d–f) in the model of changes in indoor temperature during the DR event window, for new standalone retail (a, d), old standalone retail (b, e), and 2004 vintage big box retail (c, f).

peratures is observed for the prototypical medium and large office building types (Figures 11e–h) than the all-electric (Figures 12b and 12d) and retail building types (Figures 13d–f), where more stratification in the temperature values is evident. This stratification in values reflects the tendency for indoor temperatures to quickly reach heightened cooling set points under DR in the all-electric office and retail contexts. In the all-electric office case, this tendency is attributable to the DOAS with zone-level ventilation control chosen for that model. This system provides more efficient introduction of outdoor air to zones, as opposed to the prototypical large office model with central chiller systems where ventilation needs are met with a global minimum outdoor air flow rate and a constant sup-

32

ply air temperature of 55°F; this leads to over-cooling under the higher cooling set points of the GTA strategy that prevents the indoor temperature from reaching the new set point. In the retail case, the performance of the prototypical building envelopes is generally inferior to that of the office prototypes, causing the indoor temperature to be more sensitive to warm outdoor conditions under a higher set point, and aforementioned changes to humidstat settings in the big box retail model prevent the over-cooling that is observed for the prototypical offices.

3.3. Bayesian model assessment

All models described in sections 3.1–3.2 were successfully re-estimated in a Bayesian framework and subjected to a series of posterior predictive checks (PPCs) as described in section 2.7. Figures 14a–b and 14d–e demonstrate example PPC results for the model of thermally-driven changes in demand during the DR event window in the two prototypical large office vintages. Figures 14a and 14d show that the distribution of mean demand reductions under the Bayesian modeling is well-centered on the observed mean demand reduction in underlying synthetic training data; Figures 14b and 14e further demonstrate that the full Bayesian posterior predictive distribution of demand reductions closely follows the distribution of demand reductions in the training data. Indeed, such posterior checks on the posterior distribution of all model outputs (including those not shown in Figure 14) reveal no systematic discrepancies between simulated and observed data.

Figures 14c and 14f compare the Bayesian posterior distribution of the set point change parameter coefficient in the thermally-driven demand model against the point estimate for this parameter generated via frequentist inference; these plots also demonstrate the potential to update model parameter distributions given new evidence. In both of the prototypical large office vintages, the initial set point parameter distributions — which reflect the full influence of the training data, given the use of uninformative prior distributions in their esti-
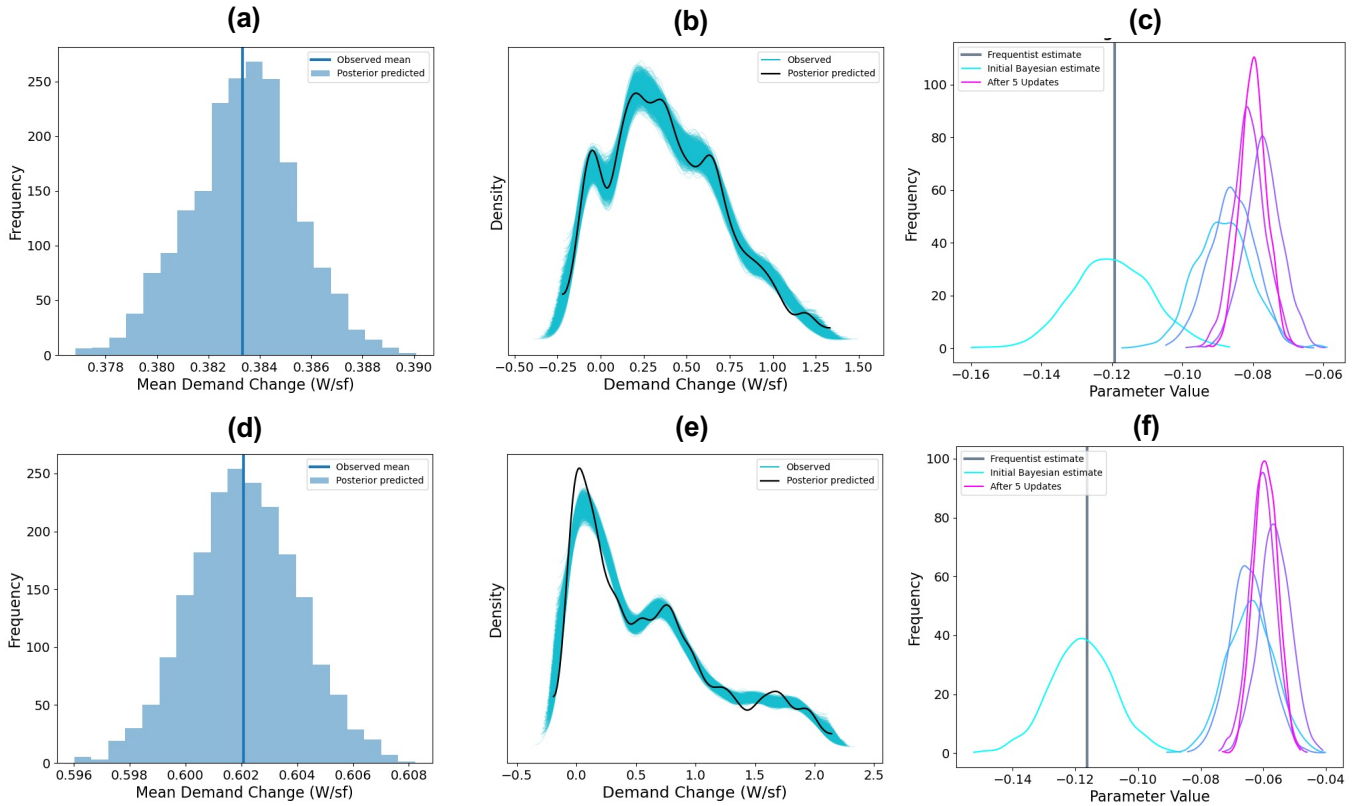
Figure 14: Example of Bayesian estimation and assessment of thermally-driven demand change model for newer (a–c) and older (d–f) prototypical large office vintages. (a, d, b, e) Posterior predictive checks against observed mean demand change and the distribution of observed demand change, including histogram (a, d) and kernel density estimation (b, e) of demand reduction output as observed in the underlying synthetic dataset and predicted by the Bayesian models. (c, f) Distribution of set point change model parameter estimate given full training datasets for the two large office vintages ("Initial Bayesian estimate") and after 5 successive updates with training data from 10 events in the all-electric large office analogues, climate zone 2A (hot-humid); the initial parameter estimate distributions are benchmarked against parameter point estimates generated via frequentist inference methods.

mation[5] — is well-centered on the frequentist estimate of approximately -0.12.

---

[5] As mentioned, we use Normal prior distributions on all model parameters ($\mu = 0$, $\sigma = 10$), and a Half Normal prior error distribution ($\sigma = 20$).

To demonstrate the parameter updating process, these initial distributions are used as priors for a subsequent round of parameter inference that introduces new evidence — in this case, a subset of the synthetic training data from the all-electric large office analogue for the same vintages, restricted to a hot-humid climate (2A). Each round of new evidence spans 10 DR events, and we carry out five successive parameter updating rounds. Across both vintages, Figures 14c and 14f show that the introduction of the new evidence from the all-electric large office setting shifts the initial set point parameter distribution in the positive direction, somewhat more strongly in the older than the newer vintage model, and that the variance of the parameter distribution is reduced with each update, implying greater certainty about the parameter estimate as more evidence is introduced.

4. Discussion

4.1. Implications of findings on model performance

Model assessment results generally show strong predictive performance for the developed surrogate models, which yield MADP values of less than 17% and R^2 values above 0.76 across the various model types and DR contexts explored; higher R^2 values above 0.95 and MADP values below 9% are observed for the models of thermally-driven changes in building demand under DR. This level of predictive accuracy compares favorably with the most directly comparable previous study of DR potential from Yin et al. [25], which reports 90% of predictions falling within a 20% prediction error for a surrogate regression model that is fit to simulation data aggregated across the medium and large office prototypes and a GTA strategy only. The high performance and wide scope of the investigated models underscores their practical usefulness in guiding decision-making regarding load flexibility strategies for office and retail contexts under a wide variety of conditions.

Moreover, reformulating the models in a Bayesian framework supports updating of parameter coefficients given new input data from real buildings, where

non-prototypical equipment scheduling and occupant behavior and/or faulty operations may lead to different input and output relationships than are initially expected by our models. The Bayesian implementation also facilitates the characterization of uncertainties in predictions for building operators, which is necessary to directly assess the risk of flexible load adjustments disrupting normal building operations.

4.2. Limitations and related work

While our findings on the potential performance of surrogate models for DR are encouraging, we note the following limitations with our approach:

- our models are trained against simulated, rather than metered data; while this approach allows examination of a wide range of flexibility strategies, settings, and building contexts, further validation of model performance against real-world case studies is needed to bolster the practical utility of our models,

- we investigate a limited set of load flexibility strategies and contexts, focusing on strategies that adjust control schedules rather than more fundamental modifications of equipment settings, examining set point adjustments in in summer only; future DR programs may also target commercial building participation in winter DR, particularly given high electrification that shifts regional grids towards winter peaks in grid system loads,

- our models are limited to producing decision-making insights at the hourly temporal resolution, in line with the resolution that outcomes are assessed under for many traditional DR programs and in wholesale energy markets; the models do not provide insights regarding faster DR services such as frequency regulation or load modulation, for which sub-hourly predictions are required,

- we do not extend our modeling to other building services such as control of humidity, illuminance, and CO_2, which are not as readily measured

or modeled by the physics-based simulations that underpin our synthetic database of load flexibility impacts, and

- while we predict the likely magnitude of impacts from load flexibility strategies on demand and indoor temperature, we do not assign operator valuations of these impacts; in practice, operators may weigh changes to demand (and associated economic benefits) more heavily than changes to temperature, or vice versa.

Parallel efforts to investigate commercial building operator and occupant load adjustment preferences have aimed to address the latter of these limitations. Specifically, discrete choice experiments were conducted to elicit the weightings that decision-makers use in trading off the potential benefits and drawbacks of possible responses to DR event calls – for example, the trade off between a certain level of economic benefit and a certain magnitude of temperature, lighting, or plug load reduction. Results have been integrated into an openly available decision tool, *FlexAssist* [62], that layers the weightings from the choice experiments on top of the predictions of the likely change in building demand and temperature from the surrogate models developed in this paper to enable a comprehensive assessment of the risks of implementing candidate load flexibility strategies in commercial office and retail settings. The Bayesian surrogate models developed in this paper are also separately available for use in the PyMC3 framework via the previously referenced GitHub repository.[6]

5. Conclusion

In this study, we used a surrogate modeling approach to predict changes in office and retail building electricity demand and indoor temperature under candidate strategies for flexible building operations under demand response (DR), including adjustments to HVAC, lighting, and plug load schedules. The surrogate models were fit to a large synthetic database generated via whole build-

---

[6]https://github.com/jtlangevin/flex-bldgs/tree/master/model_stored

ing simulations of the strategies under a variety of conditions; models were translated to a Bayesian framework to allow straightforward communication of uncertainty and parameter updating given new evidence. The surrogate models showed strong predictive performance, yielding overall prediction errors of less than 17% and R^2 values above 0.76. Surrogate models of thermally-driven changes in building demand under the various strategies display particularly high accuracy, yielding error values less than 9% and R^2 values above 0.95 across multiple building types and vintages. The strong predictive performance of the models underscores the potential for surrogate modeling approaches to support decision-making about commercial customer DR participation, by generating reliable predictions about possible changes in building demand and services under candidate DR strategies with an approach that is both computationally tractable and adaptable to a wide variety of building settings.

References

[1] P. Pinson, H. Madsen, Benefits and challenges of electrical demand response: A critical review, Renewable and Sustainable Energy Reviews 39 (2014) 686–699.

[2] B. Chew, B. Feldman, D. Ghosh, M. Surampudy, 2018 Utility Demand Response Market Snapshot, Technical Report, Smart Electric Power Alliance, Washington, DC, 2018.

[3] H. Li, Z. Wang, T. Hong, M. A. Piette, Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications, Advances in Applied Energy 3 (2021) 100054. doi:`10.1016/J.ADAPEN.2021.100054`.

[4] P. Siano, Demand response and smart grids - A survey, Renewable and Sustainable Energy Reviews 30 (2014) 461–478. doi:`10.1016/j.rser.2013.10.022`.

[5] Y. Liu, Demand response and energy efficiency in the capacity resource procurement: Case studies of forward capacity markets in ISO New England, PJM and Great Britain, Energy Policy 100 (2017) 271–282. doi:`10.1016/j.enpol.2016.10.029`.

[6] F. C. Robert, G. S. Sisodia, S. Gopalan, A critical review on the utilization of storage and demand response for the implementation of renewable energy microgrids, Sustainable Cities and Society 40 (2018) 735–745. doi:`10.1016/j.scs.2018.04.008`.

[7] X. Wang, A. Palazoglu, N. H. El-Farra, Operational optimization and demand response of hybrid renewable energy systems, Applied Energy 143 (2015) 324–335. doi:`10.1016/j.apenergy.2015.01.004`.

[8] O. A. Seppänen, W. Fisk, Some quantitative relations between indoor environmental quality and work performance or health, HVAC&R Research 12 (2006) 957–973.

[9] Demand Response Market Assessment, Technical Report, Michigan Public Service Commission, Lansing, MI, 2017.

[10] M. A. Piette, G. Ghatikar, S. Kiliccote, E. Koch, D. Hennage, P. Palensky, C. McParland, Open Automated Demand Response Communications Specification (Version 1.0), Technical Report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2009.

[11] K. Smith, Scaling Demand Response through Interoperability in Commercial Buildings, in: Grid Interop Forum, GridWise Architecture Council, Richland, WA, 2010.

[12] N. Motegi, M. A. Piette, D. S. Watson, S. Kiliccote, P. Xu, Introduction to Commercial Building Control Strategies and Techniques for Demand Response, Technical Report, Lawrence Berkeley National Laboratory, Berkeley, CA, 2007.

[13] Y. Chen, P. Xu, J. Gu, F. Schmidt, W. Li, Measures to improve energy demand flexibility in buildings for demand response (DR): A review, 2018. doi:10.1016/j.enbuild.2018.08.003.

[14] A. Rabl, L. K. Norford, Peak load reduction by preconditioning buildings at night, International Journal of Energy Research 15 (1991) 781–798. URL: http://doi.wiley.com/10.1002/er.4440150909. doi:10.1002/er.4440150909.

[15] K. R. Keeney, J. E. Braun, Application of building precooling to reduce peak cooling requirements, ASHRAE transactions 103 (1997) 463–469. URL: https://www.osti.gov/biblio/345252.

[16] W. J. Turner, I. S. Walker, J. Roux, Peak load reductions: Electric load shifting with mechanical pre-cooling of residential buildings with low thermal mass, Energy 82 (2015) 1057–1067. doi:10.1016/j.energy.2015.02.011.

[17] J. Sánchez Ramos, M. Pavón Moreno, M. Guerrero Delgado, S. Álvarez Domínguez, L. F. Cabeza, Potential of energy flexible buildings: Evaluation of dsm strategies using building thermal mass, Energy and Buildings 203 (2019). doi:10.1016/J.ENBUILD.2019.109442.

[18] F. Sehar, M. Pipattanasomporn, S. Rahman, An energy management model to study energy and peak power savings from PV and storage

in demand responsive buildings, Applied Energy 173 (2016) 406–417. doi:10.1016/j.apenergy.2016.04.039.

[19] 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 and Buildings 173 (2018) 19–27.

[20] T. Bossmann, E. J. Eser, Model-based assessment of demand-response measures—a comprehensive literature review, Renewable and Sustainable Energy Reviews 57 (2016) 1637–1656.

[21] Y. Chen, P. Xu, J. Gu, F. Schmidt, W. Li, Measures to improve energy demand flexibility in buildings for demand response (dr): A review, Energy and Buildings 177 (2018) 125–139.

[22] Demand Response Quick Assessment Tool (DRQAT), https://buildings.lbl.gov/demand-response-quick-assessment-tool-drqat (Accessed 11/10/21), 2021.

[23] D. Neves, A. Pina, C. A. Silva, Demand response modeling: A comparison between tools, Applied Energy 146 (2015) 288–297.

[24] Z. Wang, Y. Chen, Data-driven modeling of building thermal dynamics: Methodology and state of the art, Energy and Buildings 203 (2019) 109405.

[25] 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, Applied Energy 177 (2016) 149–164. doi:10.1016/j.apenergy.2016.05.090.

[26] E. C. Kara, M. D. Tabone, J. S. MacDonald, D. S. Callaway, S. Kiliccote, Quantifying flexibility of residential thermostatically controlled loads for demand response: A data-driven approach, in: Proceedings of the 1st ACM conference on embedded systems for energy-efficient buildings, 2014, pp. 140–147.

[27] M. Afzalan, F. Jazizadeh, Residential loads flexibility potential for demand response using energy consumption patterns and user segments, Applied Energy 254 (2019). doi:10.1016/J.APENERGY.2019.113693.

[28] G. Pinto, D. Deltetto, A. Capozzoli, Data-driven district energy management with surrogate models and deep reinforcement learning, Applied Energy 304 (2021) 117642. doi:10.1016/J.APENERGY.2021.117642.

[29] Y. Chen, Z. Chen, P. Xu, W. Li, H. Sha, Z. Yang, G. Li, C. Hu, Quantification of electricity flexibility in demand response: Office building case study, Energy 188 (2019) 116054.

[30] L. Hurtado, J. Rhodes, P. Nguyen, I. Kamphuis, M. Webber, Quantifying demand flexibility based on structural thermal storage and comfort management of non-residential buildings: A comparison between hot and cold climate zones, Applied Energy 195 (2017) 1047–1054.

[31] J. E. Contreras-Ocaña, M. A. Ortega-Vazquez, D. Kirschen, B. Zhang, Tractable and robust modeling of building flexibility using coarse data, IEEE Transactions on Power Systems 33 (2018) 5456–5468. doi:10.1109/TPWRS.2018.2808223.

[32] F. Amara, K. Agbossou, A. Cardenas, Y. Dubé, S. Kelouwani, et al., Comparison and simulation of building thermal models for effective energy management, Smart Grid and Renewable Energy 6 (2015) 95.

[33] X. Li, J. Wen, Review of building energy modeling for control and operation, Renewable and Sustainable Energy Reviews 37 (2014) 517–537. doi:10.1016/J.RSER.2014.05.056.

[34] J. E. Contreras-Ocaña, M. R. Sarker, M. A. Ortega-Vazquez, Decentralized coordination of a building manager and an electric vehicle aggregator, IEEE Transactions on Smart Grid 9 (2018) 2625–2637. doi:10.1109/TSG.2016.2614768.

[35] Y. Ma, A. Kelman, A. Daly, F. Borrelli, Predictive control for energy efficient buildings with thermal storage: Modeling, stimulation, and experiments, IEEE Control Systems Magazine 32 (2012) 44–64. doi:`10.1109/MCS.2011.2172532`.

[36] P. Radecki, B. Hencey, Online model estimation for predictive thermal control of buildings, IEEE Transactions on Control Systems Technology 25 (2017) 1414–1422. doi:`10.1109/TCST.2016.2587737`.

[37] X. Jin, K. Baker, D. Christensen, S. Isley, Foresee: A user-centric home energy management system for energy efficiency and demand response, Applied Energy 205 (2017) 1583–1595. URL: `https://www.sciencedirect.com/science/article/pii/S0306261917311856`. doi:`https://doi.org/10.1016/j.apenergy.2017.08.166`.

[38] I. Korolija, Y. Zhang, L. Marjanovic-Halburd, V. I. Hanby, Regression models for predicting uk office building energy consumption from heating and cooling demands, Energy and Buildings 59 (2013) 214–227.

[39] A. Kathirgamanathan, M. De Rosa, E. Mangina, D. P. Finn, Data-driven predictive control for unlocking building energy flexibility: A review, Renewable and Sustainable Energy Reviews 135 (2021) 1364–0321. doi:`"https://doi.org/10.1016/j.rser.2020.110120"`.

[40] R. K. Strand, D. B. Crawley, C. Pedersen, R. Liesen, L. Lawrie, F. Winkelmann, W. Buhl, Y. Huang, D. Fisher, Energyplus: A new-generation energy analysis and load calculation engine for building design, in: Association of Collegiate Schools of Architecture Technology Conference, 2000.

[41] DOE, Commercial Prototype Building Models, https://www.energycodes.gov/development/commercial/prototype_models (Accessed 01/30/21), 2021.

[42] Bonnema, Eric and Leach, Matt and Pless, Shanti, Technical Support Document: Development of the Advanced Energy Design Guide for Medium to

Big Box Retail Buildings - 50% Energy Savings, Technical Report, National Renewable Energy Laboratory, Golden, CO, 2013.

[43] NREL, Updated small/medium/large office models with detailed space types, https://github.com/NREL/openstudio-standards/pull/562 (Accessed 04/15/21), 2020.

[44] NREL, Building Component Library, https://bcl.nrel.gov/node/82301/ (Accessed 04/15/21), 2020.

[45] J. H. Williams, R. A. Jones, B. Haley, G. Kwok, J. Hargreaves, J. Farber, M. S. Torn, Carbon-Neutral Pathways for the United States, AGU Advances 2 (2021) 1–25.

[46] National Academies of Sciences, Engineering, and Medicine, Accelerating Decarbonization of the U.S. Energy System, The National Academies Press, Washington, DC, 2021. doi:10.17226/25932.

[47] NREL, Building component library, https://bcl.nrel.gov (Accessed 01/30/21), 2020.

[48] J. Langevin, C. B. Harris, A. Satre-Meloy, H. Chandra-Putra, A. Speake, E. Present, R. Adhikari, E. J. Wilson, A. J. Satchwell, US building energy efficiency and flexibility as an electric grid resource, Joule 5 (2021) 2102–2128.

[49] A. Roth, D. Goldwasser, A. Parker, There's a measure for that!, Energy and Buildings 117 (2016) 321–331. doi:10.1016/j.enbuild.2015.09.056.

[50] Y. Zhang, jeplus, http://www.jeplus.org/(Accessed 01/30/21), 2020.

[51] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, S. Ahrentzen, Random forest based hourly building energy prediction, Energy and Buildings 171 (2018) 11–25.

[52] J. Salvatier, T. V. Wiecki, C. Fonnesbeck, Probabilistic programming in python using pymc3, PeerJ Computer Science 2 (2016) e55.

[53] M. Hoffman, A. Gelman, The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo, The Journal of Machine Learning Research 15 (2014) 1593–1623.

[54] N. Aghdaei, G. Kokogiannakis, D. Daly, T. McCarthy, Linear regression models for prediction of annual heating and cooling demand in representative australian residential dwellings, Energy Procedia 121 (2017) 79–86.

[55] K. Amasyali, N. M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews 81 (2018) 1192–1205.

[56] Z. Petojević, R. Gospavić, G. Todorović, Estimation of thermal impulse response of a multi-layer building wall through in-situ experimental measurements in a dynamic regime with applications, Applied Energy 228 (2018) 468–486.

[57] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Applied Energy 127 (2014) 1–10.

[58] E. Lucas Segarra, H. Du, G. Ramos Ruiz, C. Fernández Bandera, Methodology for the quantification of the impact of weather forecasts in predictive simulation models, Energies 12 (2019) 1309.

[59] D. H. Vu, K. M. Muttaqi, A. P. Agalgaonkar, A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables, Applied Energy 140 (2015) 385–394.

[60] T. A. Craney, J. G. Surles, Model-dependent variance inflation factor cutoff values, Quality Engineering 14(3) (2002) 391–403.

[61] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian Data Analysis, 2nd ed., Chapman and Hall/CRC, 2004.

[62] J. Langevin, N. Luo, FlexAssist Model Documentation, https://flexible-buildings.readthedocs.io/en/latest/ (Accessed 11/02/21), 2021.

Appendix A. Model coefficient estimates

Tables A1 — A4 summarize the model coefficient values derived using frequentist inference (see section 2.5) for the full range of models covered in this paper — models of non-thermally-driven demand, thermally-driven demand (during DR event period), thermally-driven demand (during a pre-cooling period), and indoor temperature, respectively. The tables report mean model coefficient estimates and the statistical significance of these estimates, demonstrating that most input variables included in the models are highly significant across the wide range of building types and vintages covered in the study.

Appendix B. Assessment of model performance using k-fold cross validation

Table B1 summarizes the results of the analysis of the sensitivity of model performance results to choice of model training and testing data, which is conducted using the k-fold cross validation method (k=10) described in Methods section 2.7. In general, model assessment results under each of the 10 training/testing data subsets are comparable to or better than those reported with the use of the full synthetic dataset for training and testing in Results sections 3.1.1 to 3.2, suggesting the models estimated using the full dataset do not suffer from overfitting.

Table A1: Model coefficient values for the non-thermally-driven demand model (during DR event period) across all building types and vintages

Whole Building Demand (DR, Non-thermally-driven)							
	Building Type_Vintage	**Medium Office_New**		**Medium Office_Old**		**Large Office_New**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.9182		0.8838		0.9692	
	Relative Error (<20%):	88%		83%		90%	
	MADP:	12%		16%		8%	
	Input Variables	**Estimate**	**Signif.**	**Estimate**	**Signif.**	**Estimate**	**Signif.**
	(Intercept)	3.16E-03		0.0173502	*	0.0099978	***
	Lighting.Power.Diff.pct.	0.9126368	***	1.5574793	***	0.7226561	***
	MELs.Diff.pct.	0.6736572	***	0.8718433	***	0.655689	***
Office	**Building Type_Vintage**	**Large Office_Old**		**Large Office_allElec_New**		**Large Office_allElec_Old**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.9954		0.9147		0.939	
	Relative Error (<20%):	97%		89%		91%	
	MADP:	3%		9%		6%	
	Input Variables	**Estimate**	**Signif.**	**Estimate**	**Signif.**	**Estimate**	**Signif.**
	(Intercept)	0.0079855	***	0.0102969	**	0.0133642	*
	Lighting.Power.Diff.pct.	1.69239	***	0.7098196	***	1.7281562	***
	MELs.Diff.pct.	0.8524441	***	0.7175611	***	0.9800347	***
	Building Type_Vintage	**Standalone Retail_New**		**Standalone Retail_Old**		**Big Box Retail_New (2004)**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.9865		0.9893		0.9828	
	Relative Error (<20%):	94%		97%		95%	
Retail	**MADP:**	6%		4%		6%	
	Input Variables	**Estimate**	**Signif.**	**Estimate**	**Signif.**	**Estimate**	**Signif.**
	(Intercept)	0.0009071		0.002222		0.0006011	
	Lighting.Power.Diff.pct.	1.5563203	***	3.2269112	***	3.1366648	***
	MELs.Diff.pct.	/	/	/	/	/	/

*Signif. Codes: '***' p value < 0.001; '**' p value < 0.01; '*' p value < 0.05; '.' p value < 0.1; ' ' p value < 1*

Table A2: Model coefficient values for the thermally-driven demand model (during DR event period) across all building types and vintages

	Whole Building Demand Model (DR, Thermally-driven)						
Office	Building Type_Vintage	**Medium Office_New**		**Medium Office_Old**		**Large Office_New**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.9798		0.9801		0.9719	
	Relative Error (<20%):	90%		90%		90%	
	MADP:	8%		8%		8%	
	Input Variables	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
	(Intercept)	-0.164888	***	-0.608997	***	-0.319153	***
	Outdoor.Temp.F.	0.0038831	***	0.0102892	***	0.0041007	***
	Outdoor.Humid.	0.0003201	***	0.002008	***	0.0014681	***
	Occ.Fraction.	0.0157284		-0.108302		0.086439	**
	Cooling.Setpoint.Diff.	-0.124009	***	-0.237366	***	-0.11923	***
	Lighting.Power.Diff.pct.	1.0026444	***	1.7028471	***	0.6822556	***
	MELs.Diff.pct.	0.7064912	***	0.9568507	***	0.5959393	***
	Since.DR.Started.	-0.023026	***	-0.024639	***	-0.006586	
	Since.DR.Ended.	-0.093155	***	-0.25093	***	-0.126046	***
	Cooling.Setpoint.Diff.One.Step.	-0.005567	***	-0.005554	***	0.0013678	
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.001637	***	0.0029673	***	0.0015045	***
	Cooling.Setpoint.Diff.:Occ.Fraction.	0.0338338	*	0.0689113	*	0.019291	**
	Cooling.Setpoint.Diff.:Since.DR.Started.	-0.002621	**	-0.004302	**	0.0003022	
	Building Type_Vintage	**Large Office_Old**		**Large Office_allElec_New**		**Large Office_allElec_Old**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.9923		0.9797		0.9926	
	Relative Error (<20%):	91%		92%		92%	
	MADP:	7%		7%		6%	
	Input Variables	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
	(Intercept)	-0.406579	***	0.1320304	**	0.1316493	**
	Outdoor.Temp.F.	0.0050324	***	0.0001898		-0.000421	
	Outdoor.Humid.	0.0012449	***	0.0007217	***	0.0006437	***
	Occ.Fraction.	0.1224593	***	-0.095627	**	-0.016613	
	Cooling.Setpoint.Diff.	-0.116258	***	-0.046379	***	-0.021638	.
	Lighting.Power.Diff.pct.	1.6887175	***	0.6825002	***	1.7400303	***
	MELs.Diff.pct.	0.8032331	***	0.6793341	***	0.970081	***
	Since.DR.Started.	-0.005496		-0.026454	***	-0.021732	***
	Since.DR.Ended.	-0.072365	***	-0.081348	***	-0.06648	***
	Cooling.Setpoint.Diff.One.Step.	0.0005602		0.0161688	***	0.0235631	***
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.0014865	***	0.0010657	***	0.0012318	***
	Cooling.Setpoint.Diff.:Occ.Fraction.	0.0063284		0.0270481	***	0.0118631	
	Cooling.Setpoint.Diff.:Since.DR.Started.	-7.06E-05		0.0064706	***	0.002206	
Retail	Building Type_Vintage	**Standalone Retail_New**		**Standalone Retail_Old**		**Big Box Retail_New (2004)**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.9595		0.978		0.9849	
	Relative Error (<20%):	90%		90%		89%	
	MADP:	8%		8%		9%	
	Input Variables	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
	(Intercept)	-0.045085		-0.28212		-0.229538	*
	Outdoor.Temp.F.	0.0024819	*	0.0031292	*	0.0027492	**
	Outdoor.Humid.	0.003242	***	0.0064839	***	0.0029023	***
	Occ.Fraction.	-0.188465		-0.217494		0.0241719	
	Cooling.Setpoint.Diff.	-0.421545	***	-0.608274	***	0.0247687	
	Lighting.Power.Diff.pct.	1.6117979	***	3.3936214	***	3.3414764	***
	MELs.Diff.pct.	/	/	/	/	/	/
	Since.DR.Started.	-0.048772	***	-0.038504	*	-0.03394	***
	Since.DR.Ended.	-0.154235	***	-0.102278	**	-0.168946	***
	Cooling.Setpoint.Diff.One.Step.	0.0921623	***	0.0579895	***	0.0325032	***
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.0032656	***	0.0072248	***	0.0012336	***
	Cooling.Setpoint.Diff.:Occ.Fraction.	0.2767189	***	0.16794	***	-0.087192	*
	Cooling.Setpoint.Diff.:Since.DR.Started.	0.0297547	***	0.0230307	***	0.0082483	**

*Signif. Codes: '***' p value < 0.001; '**' p value < 0.01; '*' p value < 0.05; '.' p value < 0.1; ' ' p value < 1*

Table A3: Model coefficient values for the thermally-driven demand model (during pre-cooling period) across all building types and vintages

		Whole Building Demand Model (Pre-cool)					
	Building Type_Vintage	**Medium Office_New**		**Medium Office_Old**		**Large Office_New**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.9216		0.9036		0.8845	
	Relative Error (<20%):	85%		85%		86%	
	MADP:	13%		13%		12%	
	Input Variables	**Estimate**	**Signif.**	**Estimate**	**Signif.**	**Estimate**	**Signif.**
	(Intercept)	0.39113893	***	0.88485183	***	0.77729706	***
	Outdoor.Temp.F.	-0.0017126		-0.0060854	**	-0.0084933	***
	Outdoor.Humid.	-0.0001287		-0.0025149	***	-8.01E-04	***
	Occ.Fraction.	-0.1543354	.	-0.2720267	.	-0.2021857	***
	Cooling.Setpoint.Diff.	-0.0023863		-0.1544212	**	0.28371602	***
	Since.Pre.cooling.Started.	-0.0113895	**	-0.0070688	.	0.02670644	***
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.00354309	***	0.00682285	***	-0.0013774	***
	Cooling.Setpoint.Diff.:Occ.Fraction.	-0.0469055		-0.1291422	*	-0.0387751	.
	Cooling.Setpoint.Diff.:Since.Pre.cooling.Started.	-0.0095466	***	-0.0171153	***	0.00180534	
	Building Type_Vintage	**Large Office_Old**		**Large Office_allElec_New**		**Large Office_allElec_Old**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.891		0.9235		0.9074	
	Relative Error (<20%):	85%		91%		91%	
	MADP:	13%		7%		7%	
Office	**Input Variables**	**Estimate**	**Signif.**	**Estimate**	**Signif.**	**Estimate**	**Signif.**
	(Intercept)	0.51884927	***	0.01053278		-0.0118288	
	Outdoor.Temp.F.	-0.0033637	**	-0.0004778		-0.00034	
	Outdoor.Humid.	-1.54E-03	***	2.81E-04	***	-0.0001157	
	Occ.Fraction.	-0.0688629		-0.028865		0.00304142	
	Cooling.Setpoint.Diff.	-0.0722838	.	0.13439955	***	0.16243424	***
	Since.Pre.cooling.Started.	-0.0054437		0.0058963	***	0.00472763	*
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.00359616	***	-0.0002558		-0.0002391	
	Cooling.Setpoint.Diff.:Occ.Fraction.	0.02784841		-0.0078599		-0.0036593	
	Cooling.Setpoint.Diff.:Since.Pre.cooling.Started.	-0.014579	***	-0.0043676	***	-0.0065891	***
	Building Type_Vintage	**Standalone Retail_New**		**Standalone Retail_Old**		**Big Box Retail_New (2004)**	
	Output	Demand.Power.Diff.sf.		Demand.Power.Diff.sf.		Demand.Power.Diff.sf.	
	R-square:	0.834		0.859		0.8173	
	Relative Error (<20%):	85%		90%		83%	
	MADP:	13%		8%		15%	
Retail	**Input Variables**	**Estimate**	**Signif.**	**Estimate**	**Signif.**	**Estimate**	**Signif.**
	(Intercept)	0.67231969	***	0.80980755	***	0.28899129	**
	Outdoor.Temp.F.	-0.003479	*	-0.0042196	**	-0.0010988	
	Outdoor.Humid.	-0.0034245	***	-0.0046076	***	-0.0022483	***
	Occ.Fraction.	-0.4547411	***	-0.4711062	***	-0.2061352	**
	Cooling.Setpoint.Diff.	0.06208525		0.06996161		0.20132257	***
	Since.Pre.cooling.Started.	0.01623325	**	0.01249589	*	0.00126619	
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.00263248	***	0.00311365	***	0.00019944	
	Cooling.Setpoint.Diff.:Occ.Fraction.	-0.131634	***	-0.2100632	***	-0.1149814	***
	Cooling.Setpoint.Diff.:Since.Pre.cooling.Started.	-0.0111815	***	-0.0046761	*	-0.0118476	***

*Signif. Codes: '***' p value < 0.001; '**' p value < 0.01; '*' p value < 0.05; '.' p value < 0.1; ' ' p value < 1*

Table A4: Model coefficient values for the indoor temperature model across all building types and vintages

Indoor Temperature (DR)							
	Building Type_Vintage	**Medium Office_New**		**Medium Office_Old**		**Large Office_New**	
	Output	Indoor.Temp.Diff.F.		Indoor.Temp.Diff.F.		Indoor.Temp.Diff.F.	
	R-square:	0.7682		0.7925		0.826	
	Relative Error (<20%):	81%		84%		87%	
	MADP:	17%		14%		12%	
	Input Variables	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
	(Intercept)	-2.422466	***	-3.311087	***	1.9495424	***
	Outdoor.Temp.F.	0.0302677	***	0.0458428	***	-0.00689	
	Outdoor.Humid.	0.0048396	***	0.0147983	***	0.002794	***
	Occ.Fraction.	1.100022	*	-0.254312		-0.207709	
	Cooling.Setpoint.Diff.	-0.12383		-0.704968	***	-1.059504	***
	Since.DR.Started.	-0.016504		0.0259084		-0.067005	.
	Cooling.Setpoint.Diff.One.Step.	-0.162815	***	-0.143011	***	-0.178221	***
	Pre.cooling.Temp.Decrease.	-0.028785	*	-0.042564	*	-0.052081	**
	Pre.cooling.Duration.	0.0084109		0.0080855		0.000504	
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.0045025	***	0.010388	***	0.0145585	***
	Cooling.Setpoint.Diff.:Occ.Fraction.	-0.026256		0.2189828	.	0.3298304	***
	Cooling.Setpoint.Diff.:Since.DR.Started.	-0.048418	***	-0.030422	***	0.028129	**
Office	Pre.cooling.Temp.Decrease.:Pre.cooling.Duration.	-0.030377	***	-0.032117	***	-0.026045	***
	Building Type_Vintage	**Large Office_Old**		**Large Office_allElec_New**		**Large Office_allElec_Old**	
	Output	Indoor.Temp.Diff.F.		Indoor.Temp.Diff.F.		Indoor.Temp.Diff.F.	
	R-square:	0.8013		0.9659		0.982	
	Relative Error (<20%):	82%		96%		97%	
	MADP:	16%		5%		3%	
	Input Variables	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
	(Intercept)	-2.835725	***	3.2085779	***	1.7737865	***
	Outdoor.Temp.F.	0.0379978	***	-0.01579	***	-0.007056	**
	Outdoor.Humid.	0.0084079	***	0.0010608	*	0.0015663	***
	Occ.Fraction.	0.4212918	.	-1.18219	***	-0.743638	***
	Cooling.Setpoint.Diff.	-0.728801	***	-0.245135	***	0.3508159	***
	Since.DR.Started.	0.0792337	.	-0.364055	***	-0.267292	***
	Cooling.Setpoint.Diff.One.Step.	-0.124254	***	-0.189135	***	-0.146183	***
	Pre.cooling.Temp.Decrease.	-0.071235	***	-0.003198		-0.004595	
	Pre.cooling.Duration.	0.0126697		0.0141066	*	0.0094649	*
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.0104484	***	0.0089837	***	0.0047279	***
	Cooling.Setpoint.Diff.:Occ.Fraction.	0.1125076	*	0.3226307	***	0.1593479	***
	Cooling.Setpoint.Diff.:Since.DR.Started.	0.0030245		0.0907259	***	0.055168	***
	Pre.cooling.Temp.Decrease.:Pre.cooling.Duration.	-0.030343	***	-0.018987	***	-0.009151	***
	Building Type_Vintage	**Standalone Retail_New**		**Standalone Retail_Old**		**Big Box Retail_New (2004)**	
	Output	Indoor.Temp.Diff.F.		Indoor.Temp.Diff.F.		Indoor.Temp.Diff.F.	
	R-square:	0.9709		0.9313		0.9639	
	Relative Error (<20%):	97%		95%		96%	
	MADP:	3%		6%		4%	
	Input Variables	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
	(Intercept)	3.636189	***	4.2370523	***	1.8414063	***
	Outdoor.Temp.F.	-0.008051	**	-0.034792	***	-0.020613	***
	Outdoor.Humid.	0.0007086	.	0.0046837	***	0.0037167	***
Retail	Occ.Fraction.	-2.615499	***	-1.160373	**	0.2495808	*
	Cooling.Setpoint.Diff.	0.1322176		-1.363184	***	-0.1997	**
	Since.DR.Started.	-0.417573	***	-0.226894	***	-0.118512	***
	Cooling.Setpoint.Diff.One.Step.	-0.27709	***	-0.160252	***	-0.097888	***
	Pre.cooling.Temp.Decrease.	0.0081908		-0.034667	.	-0.031852	*
	Pre.cooling.Duration.	0.0100444	.	0.0104695		0.0104961	
	Cooling.Setpoint.Diff.:Outdoor.Temp.F.	0.006373	***	0.0237166	***	0.0129684	***
	Cooling.Setpoint.Diff.:Occ.Fraction.	0.291762	***	0.2469961	*	0.0145317	
	Cooling.Setpoint.Diff.:Since.DR.Started.	0.0513724	***	0.0637705	***	0.0312796	***
	Pre.cooling.Temp.Decrease.:Pre.cooling.Duration.	-0.009744	***	-7.42E-03	.	-0.005175	

Signif. Codes: '***' p value < 0.001; '**' p value < 0.01; '*' p value < 0.05; '.' p value < 0.1; ' ' p value < 1

Table B1: Summary of R^2 and MADP values for models estimated using the full synthetic database for training and assessment vs. estimated using k-fold cross validation (k=10), for all model types, building types and vintages

Model Type	Building Type	Vintage	Full Synthetic Database		K-fold Cross Validation (k=10)					
			R-squared	MADP	Averaged R-squared	Min R-squared	Max R-squared	Averaged MADP	Min MADP	Max MADP
Whole Building Demand (DR, Non-thermally-driven)	Medium office	New	0.92	12%	0.92	0.89	0.95	10%	8%	11%
		Old	0.88	16%	0.88	0.86	0.91	15%	12%	16%
	Large office	New	0.97	8%	0.97	0.96	0.98	5%	5%	6%
		Old	1.00	3%	1.00	0.99	1.00	3%	3%	4%
	Large office_allElec	New	0.91	9%	0.91	0.86	0.94	8%	7%	10%
		Old	0.94	6%	0.94	0.89	0.96	5%	2%	10%
	Standalone retail	New	0.99	6%	0.99	0.99	0.99	5%	5%	6%
		Old	0.99	4%	0.99	0.99	0.99	4%	3%	4%
	Big box retail	New	0.98	6%	0.98	0.98	0.98	5%	5%	6%
Whole Building Demand (DR, Thermally-driven)	Medium office	New	0.98	8%	0.98	0.97	0.98	5%	5%	6%
		Old	0.98	8%	0.98	0.98	0.99	6%	5%	7%
	Large office	New	0.97	8%	0.97	0.96	0.98	6%	5%	7%
		Old	0.99	7%	0.99	0.99	0.99	5%	5%	5%
	Large office_allElec	New	0.98	7%	0.98	0.98	0.98	6%	5%	6%
		Old	0.99	6%	0.99	0.99	0.99	5%	4%	6%
	Standalone retail	New	0.96	8%	0.96	0.95	0.97	7%	5%	9%
		Old	0.98	8%	0.98	0.97	0.98	7%	7%	8%
	Big box retail	New	0.98	9%	0.98	0.98	0.99	8%	7%	10%
Whole Building Demand (Pre-cool)	Medium office	New	0.92	13%	0.92	0.88	0.94	9%	8%	11%
		Old	0.90	13%	0.90	0.86	0.94	10%	8%	12%
	Large office	New	0.88	12%	0.89	0.83	0.93	8%	6%	9%
		Old	0.89	13%	0.89	0.86	0.93	8%	7%	9%
	Large office_allElec	New	0.92	7%	0.92	0.89	0.95	3%	3%	4%
		Old	0.91	7%	0.91	0.86	0.93	4%	4%	5%
	Standalone retail	New	0.83	13%	0.84	0.81	0.92	9%	6%	14%
		Old	0.86	8%	0.86	0.80	0.90	5%	2%	7%
	Big box retail	New	0.82	15%	0.82	0.78	0.88	11%	10%	12%
Indoor Temperature (DR)	Medium office	New	0.77	17%	0.77	0.68	0.87	16%	14%	20%
		Old	0.79	14%	0.79	0.73	0.82	13%	9%	18%
	Large office	New	0.83	12%	0.83	0.76	0.86	12%	7%	16%
		Old	0.80	16%	0.80	0.68	0.88	15%	11%	21%
	Large office_allElec	New	0.97	5%	0.97	0.95	0.98	4%	2%	9%
		Old	0.98	3%	0.98	0.96	0.99	3%	2%	6%
	Standalone retail	New	0.97	3%	0.97	0.96	0.98	4%	2%	7%
		Old	0.93	6%	0.93	0.91	0.95	6%	2%	11%
	Big box retail	New	0.96	4%	0.96	0.94	0.98	5%	2%	9%