Using Machine Learning to Predict Retrofit Effects for a Commercial Building Portfolio

Yujie Xu [1], Vivian Loftness [1,*] and Edson Severnini [2,3,4]

1   Center for Building Performance & Diagnostics, School of Architecture, Carnegie Mellon University, Pittsburgh, PA 15213, USA; yujiex@andrew.cmu.edu
2   Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, USA; edsons@andrew.cmu.edu
3   Institute of Labor Economics (IZA), 53113 Bonn, Germany
4   National Bureau of Economic Research (NBER), Cambridge, MA 02138, USA
*   Correspondence: loftness@cmu.edu

Abstract: Buildings account for 40% of the energy consumption and 31% of the CO_2 emissions in the United States. Energy retrofits of existing buildings provide an effective means to reduce building consumption and carbon footprints. A key step in retrofit planning is to predict the effect of various potential retrofits on energy consumption. Decision-makers currently look to simulation-based tools for detailed assessments of a large range of retrofit options. However, simulations often require detailed building characteristic inputs, high expertise, and extensive computational power, presenting challenges for considering portfolios of buildings or evaluating large-scale policy proposals. Data-driven methods offer an alternative approach to retrofit analysis that could be more easily applied to portfolio-wide retrofit plans. However, current applications focus heavily on evaluating past retrofits, providing little decision support for future retrofits. This paper uses data from a portfolio of 550 federal buildings and demonstrates a data-driven approach to generalizing the heterogeneous treatment effect of past retrofits to predict future savings potential for assisting retrofit planning. The main findings include the following: (1) There is high variation in the predicted savings across retrofitted buildings, (2) GSALink, a dashboard tool and fault detection system, commissioning, and HVAC investments had the highest average savings among the six actions analyzed; and (3) by targeting high savers, there is a 110–300 billion Btu improvement potential for the portfolio in site energy savings (the equivalent of 12–32% of the portfolio-total site energy consumption).

Keywords: building energy retrofits; energy savings evaluation; data-driven energy analysis; causal forest; heterogeneous treatment effect

1. Introduction

1.1. Motivation

In 2018, buildings consumed 40,000 trillion Btu of energy in the U.S., accounting for 40% of U.S. total energy consumption [1]. Buildings also contributed to over 2000 million metric tons of CO_2 equivalent greenhouse gas (GHG) emissions in the U.S., which constitutes up to 31% of U.S. total GHG emissions [2]. Reducing energy consumption and GHG emissions in the building sector is an important step to reducing energy-induced pollution and global warming.

Retrofits of existing buildings are recognized as an effective means to reducing building consumption and carbon footprints [3–5]. According to a report by Granade et al. [6], over 25% of primary energy consumption and the associated carbon footprint could be reduced by energy efficiency improvements in existing buildings [6]. Energy conservation measures such as upgraded building envelopes, lighting, and efficient appliances are projected to not only be effective in saving energy and reducing carbon footprint but also to easily pay for themselves with the avoided energy expenses [6,7].

Simulation-based tools and methods have been widely used in the prediction of retrofit savings. A growing number of studies use field experiments or observational data to measure the real effect of energy retrofits, finding discrepancies between the realized and the projected savings. For example, in a large field experiment of 30,000 households in Michigan, Fowlie et al. [8] showed that the Weatherization Assistance Program reduced energy expenses by an average of USD 2349 per home over a 16-year horizon (17 MMBtu/year), but only about a quarter of the projected USD 9810 (56 MMBtu/year). In a quasi-experimental study of 636 commercial buildings in Phoenix, Liang et al. [9] showed that the Energize Phoenix program, from which buildings receive single or combinations of retrofit actions among HVAC, lighting, windows, refrigeration, pumps, and motors, led to energy savings of 12%, one-fifth of the projected 60%. Table 1 summarizes such discrepancies between empirical evaluations and retrofit effects predicted with simulation-based tools. The commonly used experimental and quasi-experimental approaches include randomized control trials (RCT), randomized encouragement design (RED), instrumental variable (IV), difference-in-differences (DD), event study, matching, etc. Explanations about these methods can be found in [10].

Table 1. Comparison between measured savings with empirical methods and predicted savings with simulation-based tools.

Engineering Model or Projection	Empirical Approach	n	Realization Rate [1]	Building Use	References
NEAT	Experimental (RED)	around 30,000	30% in total energy	Residential	[8]
TREAT	DD, event study	101,881	35% in total energy 58% in energy expenses	Residential	[11]
DEER	DD, IV	275	79% in electricity for AC homes 0% in electricity for non-AC homes	Residential	[12]
[13]	Experimental + engineering [2]	504	87% in cooling 88–92% in heating	Residential	[14]
World Bank	DD, event study, matching	1,162,775	About 25% in electricity reduction from fridge replacement	Residential	[15]
Dwelling Energy Assessment Procedure (DEAP)	DD	640,000	$64 \pm 8\%$	Residential	[16]
Net Benefit Model	DD	12,000	1/3	Residential	[17]
Not specified	Time-series approach	2094	24% overall, 49% for HVAC, 42% for lighting	K-12 school	[18]
Not specified	DD	847	30–50%	Commercial and residential	[9]
[19]	Repeated cross-sectional comparison controlling for observables	Around 7000	<32% [3]	Residential	[20]

[1] Measured savings divided by simulation projected savings. [2] This study is a bit different from others, as it did not compare a pure engineering model vs. a pure empirical model. It attributed the deviation to the engineering model as the rebound effects (increase in usage as a result of higher appliance efficiency). [3] The realized electricity savings are no more than 15%, and gas savings no more than 15%. The total realized savings will be less than 25%, as compared to the projected 77%.

There could be a variety of reasons for such under-delivery. First, some simulation-based models in program projections are simplified and not calibrated to actual building

consumption or did not consider behaviors that impact consumption. A second reason could be retrofit-induced behavioral changes, where higher efficiency could stimulate higher consumption. For example, in a poorly insulated house, some building occupants might not heat or cool the house to the most comfortable temperature since too much energy expenditure is needed to achieve the desired indoor temperature. After the retrofit, these homeowners might need less energy expenditure to achieve the desired temperature, which could result in less savings in energy consumption, even if higher thermal comfort is achieved. This is called the "rebound effect"; however, other studies have found that the rebound effect might not explain all of the discrepancy [8,11]. A third reason could be the quality of installation and worker incentives. For example, energy savings are significantly lower when retrofit actions are carried out on a Friday—a day particularly prone to negative shocks on workers' productivity—than on any other weekday [21–23]. A fourth reason could be the large heterogeneity in the retrofit effects. This means that a retrofit program or action might have different levels of effectiveness in different buildings, and the retrofit tools did not successfully target the group of buildings that have the most substantial savings potential [24]. This paper focuses on understanding the heterogeneity of retrofit effects, i.e., the effects as a function of a series of characteristics related to buildings, energy-use level, weather, etc.

This paper presents a data-driven approach to predicting the average and distribution effects of six types of retrofit actions, including three capital retrofits with hardware installation and three operational retrofits. The research applies recently developed machine learning tools in the prediction of heterogeneous treatment effects of energy retrofits across a portfolio of commercial buildings, using a list of building characteristics and climate conditions [25]. In companion research to be published in the future, we examine how climate change mitigation scenarios could be incorporated into savings prediction and retrofit planning.

1.2. Related Literature

The main goals of retrofit analysis tools include evaluating the savings of past retrofits, predicting the savings potential of future retrofits, and/or optimizing retrofit decisions based on the predicted savings. According to ASHRAE Guideline 14 [26], past retrofit savings for whole buildings can be evaluated with either building energy simulations (the calibrated simulation approach) or data-driven methods (the inverse modeling approach). Today, the energy-savings prediction for retrofit planning purposes is dominated by building energy simulations [27]. Data-driven methods are mainly used in the decision optimization phase in retrofit planning given predicted savings from simulations [28]. This research fills the gap by presenting a data-driven method to predict retrofit savings.

1.2.1. The Applications of Data-Driven Methods in Existing Building Retrofit Studies

Data-driven methods are currently used to evaluate past retrofits, retrofit decision optimization, and pre- or post-processing of simulation inputs or outputs to reduce its complexity. In this section, each of the above applications will be briefly discussed.

One main application of data-driven models in building retrofits is the measurement and verification (M&V) inverse model evaluating savings of past retrofits, based on the M&V principles in ASHRAE Guideline 2014 [26] and IPMVP (International Performance Measurement and Verification Protocol) [29]. The effect of an implemented retrofit in a specific building is evaluated by creating a baseline model (counterfactual) with a regression model fitted to its pre-retrofit energy and a set of observed covariates, which usually include weather and occupancy. The regression model can take on many forms (see Table 2), from simple linear regression models to advanced techniques like deep neural networks.

Decision optimization is another main application of data-driven models in retrofit planning. Such optimization solvers can provide rankings of alternative design options or encode decision-makers' preferences in the selection of optimal strategies, based on predicted savings from simulation tools and other objectives. Some studies bypass the

retrofit savings prediction step and directly learn a model to predict the best retrofit actions from building characteristics [30].

Due to the intense computation requirements in building energy simulations (BES), some data-driven methods are deployed to reduce the set of buildings to be simulated. In this application, various clustering methods are used in grouping buildings with similar characteristics and identifying representative buildings of each group. Some studies cluster buildings based on their characteristics (age, window ratio, etc.) [31], whereas others do so based on the predicted retrofit effects [32].

To reduce the simulation time, some data-driven models are trained to approximate the simulation model results of energy consumption or occupant comfort with a series of building characteristics. The results are commonly fed to optimization schemes to produce ECM (energy conservation measure) recommendations.

Data-driven approaches modeling the conditional average treatment effect of behavioral energy interventions with detailed characteristics of occupants and buildings appear in some recent studies on residential buildings [33,34]. These studies model household energy savings potential as a function of the occupant and building characteristics and target high potential savers to reduce implementation costs while maintaining high overall savings. This study extends this branch of research to the commercial building sector, with a different set of retrofit effect predictors that focuses less on occupant characteristics and more on weather and climate.

Table 2. Data-driven methods used in retrofit evaluation or planning.

Role of the Model	Model	Source
M&V inverse modeling	Linear regression	[35–38]
	Piecewise linear regression	[35,39,40]
	Neural network (shallow)	[39,41,42]
	Deep learning	[43]
	Support vector machine regression (SVR)	[44,45]
	Kernel regression	[42]
	Gaussian process regression	[39,46]
	Gaussian mixture regression	[39]
	Decision tree	[47]
	Random forest	[48]
Multi-objective optimization	Genetic algorithm (GA)	[49–52]
	Nondominated sorting genetic algorithm II (NSGA-II)	[53–55]
	Particle swarm optimization	[52]
	Sequential search	[52]
Predict retrofit decisions	Falling rule list	[30]
Approximate BES results	Neural network (shallow)	[49,55,56]
Generate inputs for BES parametric runs	Genetic algorithm (GA)	[57]
Identify representative buildings	Clustering	[31,58]
Estimate retrofit effect	Elastic nets	[34]
	Gradient forest	[34]
	Causal forest	[33]

1.2.2. Advantage of Data-Driven Methods in Retrofit Savings Prediction

Simulation tools allow full control of physical and environmental parameter settings and have the potential to analyze a wide range of retrofit scenarios, including prototyping new retrofits. However, they require more intensive data input, higher expertise in building energy modeling, and long computation time. To reduce the computation time, many tools use a pre-simulated database, usually with prototype buildings. This could speed up the analysis and provide a quick screening of retrofit alternatives, but many such tools do not

have model calibration. As a result, whether the prototype building matches the reality is left unverified. Even for those with calibration, how close the model parameters are to the actual building is still unclear, as the goodness of the calibration is mainly measured by the closeness of simulated energy and the actual energy, not the model parameters. How this uncertainty propagates when the ECM-related model parameters change is unclear [59]. On the other hand, the use of measured data with data-driven approaches has the potential to better reflect the savings actually achieved. Equally important, most simulation-based tools cannot assess operational or behavioral efficiency measures [59], whereas the data-driven approaches established in this study can evaluate operational improvements (for example, dashboard tools) as well as hardware retrofits.

The following paper is organized into four sections: database and methodology, results, discussion, and conclusion. In Section 2, the paper will first introduce the dataset, the assumptions of the causal model, the inputs, and the machine learning model. In Section 3, the results of the average electricity and gas savings of the six main retrofit actions and their sub-actions will be presented, followed by a demonstration of the benefit targeting buildings with high savings potential. To assist the interpretation of the model, the association between the weather inputs and the energy savings is also presented. Section 4 first identifies the most important retrofit effect predictors, then compares the magnitude of the savings of the current study against other similar studies. Section 5 concludes by summarizing the key results, providing guidelines on the application of the method to portfolio retrofit planning, and pointing out some limitations and directions of future developments.

2. Database and Methodology

This section describes the data set and some methodological choices. The diagram in Figure 1 shows the methodology workflow. The input data has three main components: the effect predictors, the retrofit action, and the pre- to post-retrofit energy change. Key effect predictors are identified through the causal mechanism analysis. With these, the machine learning model could learn a function that predicts the retrofit effect of various retrofit actions based on the predictor values of a target building. By interpreting the model, the most informative predictors could be identified. This result is shown in Section 4.1. The model also directly predicts retrofit effects for building–action pairs. With this information, one can compare the average effectiveness of different actions. This result is presented in Section 3.1. With the predicted savings, portfolio owners can target high-energy-saving buildings. The benefit of targeting is evaluated in Section 3.3 by comparing the portfolio total savings with retrofit decisions made with targeting and those with retrofit decisions made in reality.

2.1. Data and Summary Statistics

The building energy data and retrofit records in this study are from the U.S. General Service Administration (GSA) portfolio. GSA is a government agency providing office space, services, and goods to government agencies. Executive Order 13,423 required government agencies to reduce their energy-use intensity by 30% by the fiscal year 2015, compared with its usage in the fiscal year 2003. With USD 5.5 billion in funding from the American Recovery and Reinvestment Act (ARRA) [60], a series of retrofit actions were undertaken in buildings in the GSA portfolio between 2010 and 2015. Follow-up Executive Order 13,693 (revoked in 2018) required an annual energy use intensity reduction of 2.5% per year from 2016 to 2025 [61].

This study quantifies the retrofit effect of actions taken during the 2010–2015 period in the GSA portfolio and uses this information to provide decision support for future retrofits. A subset of 552 buildings in the GSA building portfolio is used in the retrofit effect analysis (270 with recorded retrofits, 282 with no recorded retrofits).
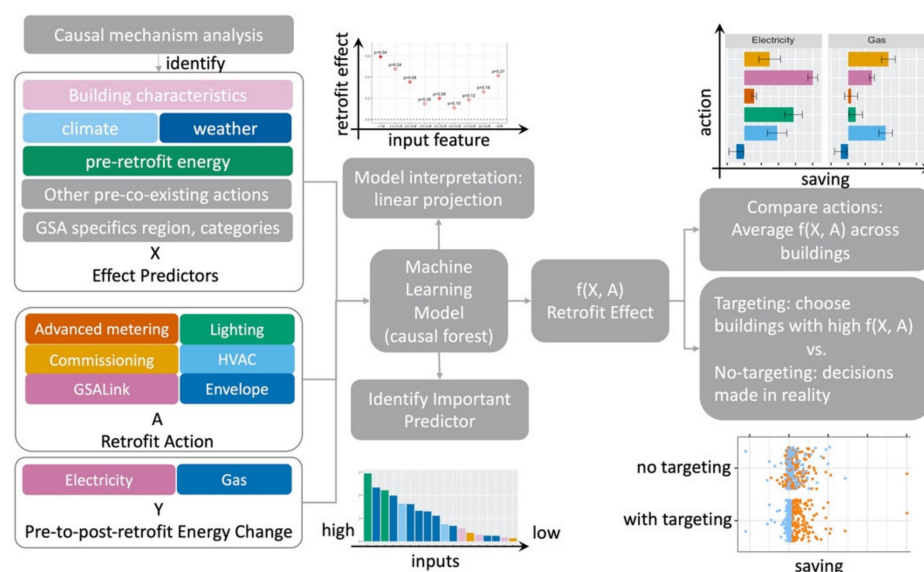
Figure 1. Methodology diagram.

The study quantifies the retrofit effect of six groups of retrofit actions, with building counts shown in Figure 2. Advanced metering refers to the installation of a system of smart meters to "monitor and/or store energy consumption data for specific building systems or the entire building" [62]. The commissioning retrofit involves seasonal system testing, identifying and fixing system issues, post-occupancy evaluation, etc. [63]. The GSALink action is a combined energy-use dashboard and fault-detection tool. These three actions involve relatively few hardware installations and are thus categorized as operational investments. The "building envelope" retrofits include new or repaired building roofs, facades, or windows. HVAC retrofits consist of repairing or replacing components in the heating, cooling, and ventilation systems such as chillers, boilers, and cooling towers. Lighting retrofits include daylighting strategies and the installation of indoor and outdoor LED lighting, occupant response lighting, integrated lighting control, etc. [64]. These three groups of actions involve substantial hardware changes and are thus classified as capital investments.
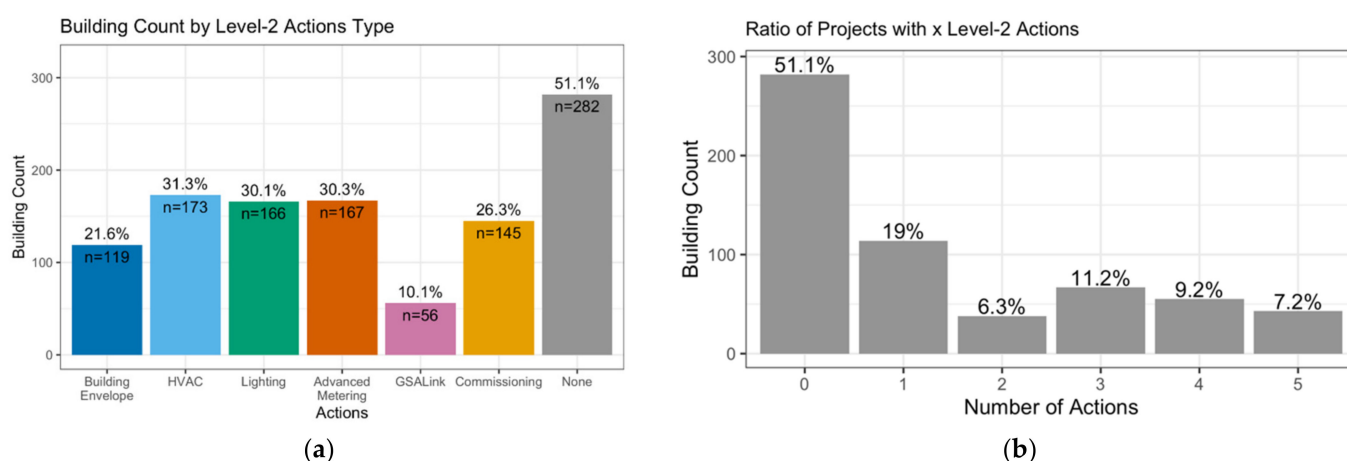

(a)

(b)

Figure 2. Retrofit building count: (**a**) percentage and number of buildings with each type of retrofit action, and (**b**) percent and number of buildings with 0 through 5 types of retrofit actions.

Among the 552 buildings in the study, around 30% had investments in advanced metering, 26% conducted commissioning, 10% had GSALink monitoring, 31% had investments in HVAC, 30% in lighting retrofits, and 22% in building envelopes. Since these

percentages exceed the 49% of buildings with investments, it should be clear that over 30% of all buildings (nearly 70% among the retrofitted buildings) had 2 or more retrofit actions at the same time (Figure 2).

Table 3 summarizes a few pre-retrofit numeric features of the retrofitted buildings and un-retrofitted buildings in the study. For an un-retrofitted building, a dummy retrofit time was selected by randomly sampling the retrofit times of the retrofitted buildings. Note that the retrofitted buildings had slightly larger gas consumption and larger building size. The differences between the retrofitted and the un-retrofitted buildings could have caused biases in the results. To account for this difference, a doubly robust causal inference model—the causal forest—was selected. The propensity score weighting component in the model "balanced out" the difference between the two groups and thus reduced biases in the predicted retrofit effects.

Table 3. Summary of the pre-retrofit characteristics of the retrofitted buildings and the un-retrofitted buildings.

|	Retrofit Effect Predictor	Min	Mean	Median	Max	St. Dev.
withretrofit	Electricity (kBtu/sqft/year)	1.85	41.67	38.72	287.25	20.89
	Gas (kBtu/sqft/year)	0.01	21.96	21.14	132.22	17.80
	Cooling degree day (CDD)	58.6	1409.3	1160.2	4628.2	989.5
	Heating degree day (HDD)	71.5	4484.1	4667.1	9172.7	2005.2
	Gross square footage	5912	402,742	269,946	3993,881	456,913
without retrofit	Electricity (kBtu/sqft/year)	2.11	45.21	40.21	182.01	23.93
	Gas (kBtu/sqft/year)	0.00	29.94	21.07	283.68	30.77
	Cooling degree day (CDD)	20.9	1398.5	1363.7	4589.0	961.2
	Heating degree day (HDD)	184.3	4460.4	4188.3	10,580.2	2389.8
	Gross square footage	1161	216,610	65,055	3456,919	428,105

2.2. Assumptions and Confounding Variables

To identify the retrofit effect from non-experimental data, it was crucial to understand the causal mechanisms and identify and control for the confounding variables, those factors affecting both the retrofit decisions (treatment) and building energy consumption (outcome).

Building energy consumption could be affected by building size due to the difference in surface-to-volume ratio, the different properties of the building systems, etc. Building type and ownership could influence the occupants' energy-use behaviors, and in turn, could impact energy consumption. Building vintage and retrofit history is related to construction property and equipment efficiency, and thus could influence the energy consumption of a building. Local climate affects the amount of heating or cooling needs, which in turn affects both the total energy consumption and heating/cooling end uses. Figure 3 summarizes the list of variables affecting the energy consumption of a building.

According to [65], two main determinants of retrofit decisions in commercial buildings are technical feasibility and financial consideration. Technical feasibility is affected by factors including building vintage, energy types, and construction properties [65]. Building type and local climate affect financial considerations through the building resiliency requirements and pre-retrofit energy consumption levels. The geographical location of a GSA region is related to climate, which in turn influences the financial benefits of retrofit decisions. Building size affects retrofit implementation cost, which then shapes retrofit decisions. Ownership could also influence retrofit decisions, as split incentives or regulatory hurdles might make tenants less willing or able to conduct retrofits. This can be seen in the current GSA building data, where all retrofitted buildings are owned buildings, and policies have been set for energy savings, all of which affect retrofit decisions. Figure 4 outlines the factors affecting building retrofit decisions.
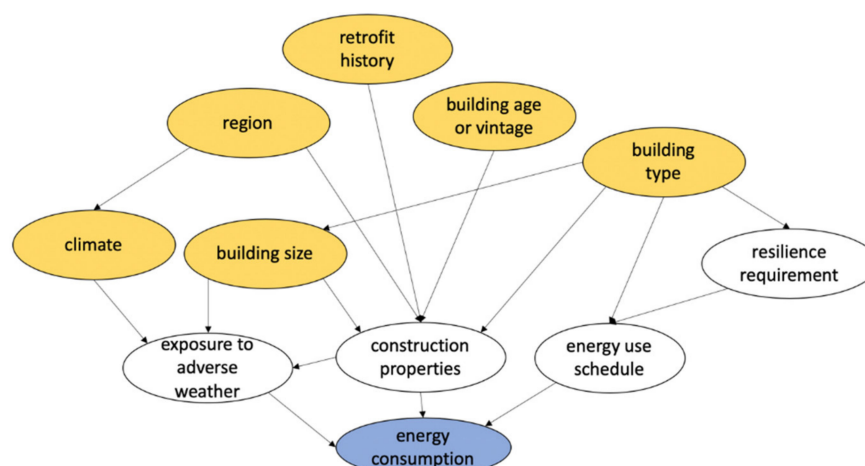
Figure 3. Variables affecting energy consumption (variables in yellow are the control variables).
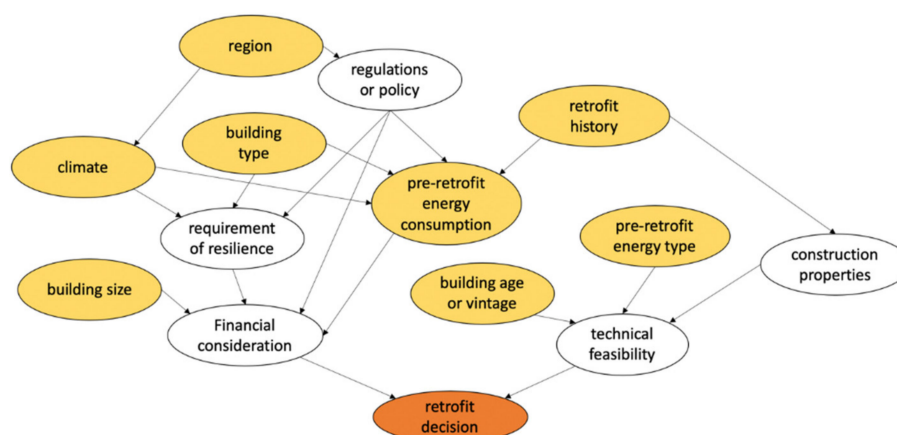
Figure 4. Retrofit decision determinants (variables in yellow are the control variables).

Based on the reasoning above, a simplified causal diagram was drawn incorporating the factors affecting building energy consumption and factors impacting retrofit decisions (Figure 5). Note that there could have been some un-measurable confounders, for example, whether the building manager is environmentally conscious. Buildings with an environmentally conscious manager could have been more likely to receive retrofit investments and more likely to use less energy in the absence of retrofits. This is illustrated by the dashed arrows in the diagram. The absence of these un-measurable confounders could have biased the results. The study assumes that the strength of the relationship encoded with the dashed arrows is weak.

Seven groups of input features were selected for the prediction of the retrofit effect, based on the causal mechanism discussed above and data availability. Table 4 lists the retrofit effect predictors. Building characteristics included 4 variables: building size, type, LEED (Leadership in Energy and Environmental Design) certification status, and indicator for historic buildings (a proxy for building age). Short-term weather was used in the form of a binned representation of daily average temperature three years before the retrofit, following [66–68]. The 30-year annual heating and cooling degree day climate normal from NOAA (National Oceanic and Atmospheric Administration) was used to reflect the long-term local climate condition. The pre-retrofit annual energy consumption category included the consumption of four fuel types. GSA sustainability practices were organized under the regional offices, each with its unique sustainability activities. Such policy regions are controlled with indicators of regional membership. GSA buildings were classified into a GSA-designated category, which denotes the building ownership and the energy-use

intensity. The last predictor class encoded the type of retrofit actions that took place before or at the same time as the current action under analysis.
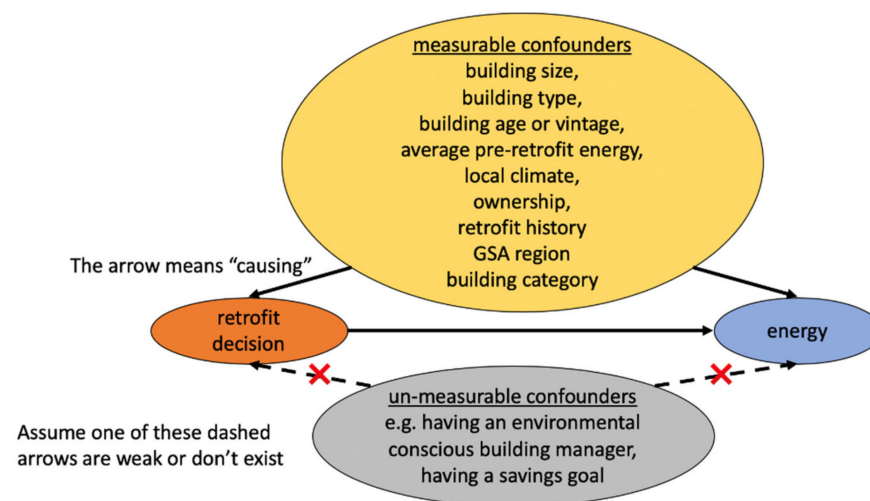

Figure 5. Simplified causal diagram.

Table 4. The classification of retrofit effect predictors.

| Retrofit Effect Predictor Class	Retrofit Effect Predictor
Building characteristics	Building size in gross square footage. Whether a building is a historic building. Whether a building had a LEED certificate before the retrofit project Whether a building is an office building or a courthouse
Weather (average annual number of days with daily mean temperature within a certain range)	A total of 8 variables, the kth variable indicating the annual average number of days with daily temperature in the kth temperature bin. The temperature bins are below 10 °F, between 10 °F and 20 °F, . . . , between 80 and 90 °F, and above 90 °F. The 60 °F to 70 °F bin is left out as a reference.
Climate	1981–2010 climate normal of annual cooling degree day (CDD), and annual heating degree day (HDD)
Pre-retrofit energy	Average monthly electricity, natural gas, chilled water, and steam consumption in kBtu/sqft/year.
Region	A total of 10 indicator variables, the kth indicator variable representing whether a building is in GSA region k. Region 1 is left out as the reference level.
Category	A total of 3 indicator variables, each corresponding to a GSA building category designation of B, E, or I. Category A is left out as the reference level [1].
Previous actions	Indicator variables of pre- or co-existing action categories.

[1] A: Government-owned buildings subject to the energy reduction goals, B: government-owned buildings exempt from the goals, E: reimbursable and non-re-portable, and I: energy-intensive buildings.

2.3. The Causal Forest Model

The study applied the causal forest estimator [69] to predict the retrofit effect as a function of building characteristics, climate, energy-use level, etc. The causal forest model was selected due to its better predictive accuracy and high confidence interval coverage, compared with traditional non-adaptive neighborhood matching methods [69].

A causal forest is an ensemble of many causal trees. Each causal tree is built with a random subsample of the whole data set. This subsample is randomly split into two halves. One half is used in learning the tree partition by maximizing the variance of the heterogeneous treatment effect, and the other half is used in computing the effect estimates. This approach both reduces the bias and minimizes the mean squared error.

In this study, a causal forest was fitted for each type of retrofit action and fuel type (electricity and gas) separately, using buildings with that retrofit type and buildings with no retrofit. The schematic diagram in Figure 6 illustrates how the HVAC causal forest model estimates the retrofit effect on the gas consumption of a target building. Suppose the target building is an office and non-historic building of 30,000 square feet. During the three years before the retrofit, on average there are 100 days in a year with temperature between 70 °F and 80 °F, and 30 days with a temperature between 30 °F and 40 °F. The causal forest predicts that investing in HVAC retrofits is estimated to reduce the average annual natural gas consumption of this target building by 1.5 kBtu/sqft. This savings is estimated by contrasting the pre to post-natural gas consumption changes of the retrofitted and the un-retrofitted buildings that share similar characteristics with this building.
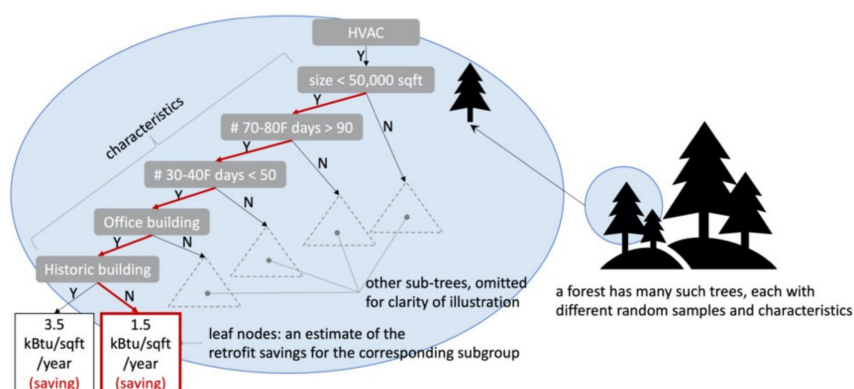
Figure 6. Illustration of a regression tree in the causal forest estimating effects of HVAC retrofits.

The causal forest model uses the implementation in the generalized random forest (grf) R package developed by Tibshirani et al. [70]. In the first round of model fitting, the propensity score (treatment probability) is estimated, then the model is re-fitted with the propensity score bounded within the range of 0.05 to 0.95 so that the overlapping assumption required in the identification of the causal effect is satisfied. The training process uses cross-validation to tune the following hyper-parameters: the proportion of data used in building a causal tree, the number of candidate features considered in a split, the minimum number of buildings in each leaf node, the proportion of data in the sub-sample used in computing the split, the accepted level of imbalance of a split, and the penalty of an imbalance split. All the data pre-processing and model fitting was conducted using R on a Mac laptop machine with a Quad-Core Intel Core i7 processor and 16 GB memory.

3. Results

3.1. The Average and Distribution of Retrofit Effect

Among all six retrofit actions, on average, the energy reporting and dashboards in GSALink had the highest average savings in annual electricity consumption per square foot (electric site EUI (EUI: energy use intensity)), whereas commissioning and HVAC

capital improvements had the highest average savings in natural gas consumption per square foot (gas site EUI). Figure 7 presents the mean and the 95% confidence interval of the predicted retrofit effect of past retrofits, restricted to the set of buildings with positive electricity or gas consumption three years before the retrofits. The summary statistics of the distribution of the estimated savings among the set of retrofitted buildings are shown in Table 5. For most actions, there were some buildings with negative savings, even negative median savings. This means the same retrofit action saved energy for some buildings but increased energy usage for others. It is thus crucial to target buildings with high savings potential in retrofit planning.
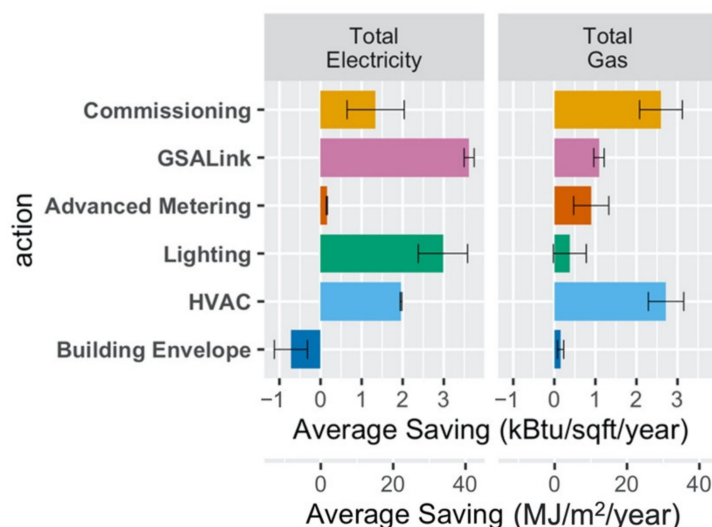
Figure 7. Bar chart of the mean and 95% C.I. of the estimated retrofit effects of past retrofit actions.

Table 5. The distribution of the estimated site electricity or gas EUI in kBtu/sqft/year savings among retrofitted buildings. 1 kBtu/sqft/year = 3.15 kWh/m^2 = 11.36 MJ/m^2.

Fuel	Action Type	Action	Summary Statistics					
			Min	Q1	Median	Q3	Max	Mean
Electricity	Capital	Building envelope	−5.09	−2.84	−0.82	1.58	4.47	−0.66
		HVAC	−5.02	−1.86	1.53	4.76	11.71	1.90
		Lighting	−4.36	0.52	2.20	4.53	14.20	2.55
	Operational	Advanced metering	−1.41	−0.09	0.36	0.93	2.86	0.47
		Commissioning	−7.15	−2.75	1.56	3.57	11.00	0.94
		GSALink	1.04	3.40	4.26	5.57	10.60	**4.87**
Gas	Capital	Building envelope	−4.81	−2.47	−1.10	0.97	1.95	−0.89
		HVAC	−1.93	1.04	2.75	3.83	8.50	**2.55**
		Lighting	−4.30	−1.15	0.06	1.61	6.78	0.35
	Operational	Advanced metering	−3.63	−1.15	−0.24	1.51	6.86	0.49
		Commissioning	−1.24	0.88	1.97	3.76	8.56	**2.62**
		GSALink	0.41	0.69	0.89	1.24	1.81	0.98

In addition to the six large groups of operational and capital ECMs, the data set enabled the analysis of a set of sub-actions in the capital ECMs. Among the HVAC sub-actions, new cooling towers and new air handlers had the highest average site electricity EUI savings. Repairing controls and repairing boilers had the highest average site gas EUI

savings. Surprisingly, new cooling towers also had noticeable average site natural gas EUI savings. This might have been due to the reduction in gas used in the reheat of cooling supply air. Another possibility might have been due to commissioning retrofits co-existing with HVAC capital investments, which is rather effective in saving gas consumption. Even though such co-existing actions were controlled with indicator variables of past retrofit categories, the interactions between actions might have been more complicated and not adequately accounted for. A future development of this data-driven analysis would be to gather more retrofit data sets so that the interactions of different retrofit investments can be estimated. Figure 8 visualizes the mean and 95% confidence interval of the predicted retrofit effect of HVAC sub-actions on the site electricity EUI and site gas EUI.
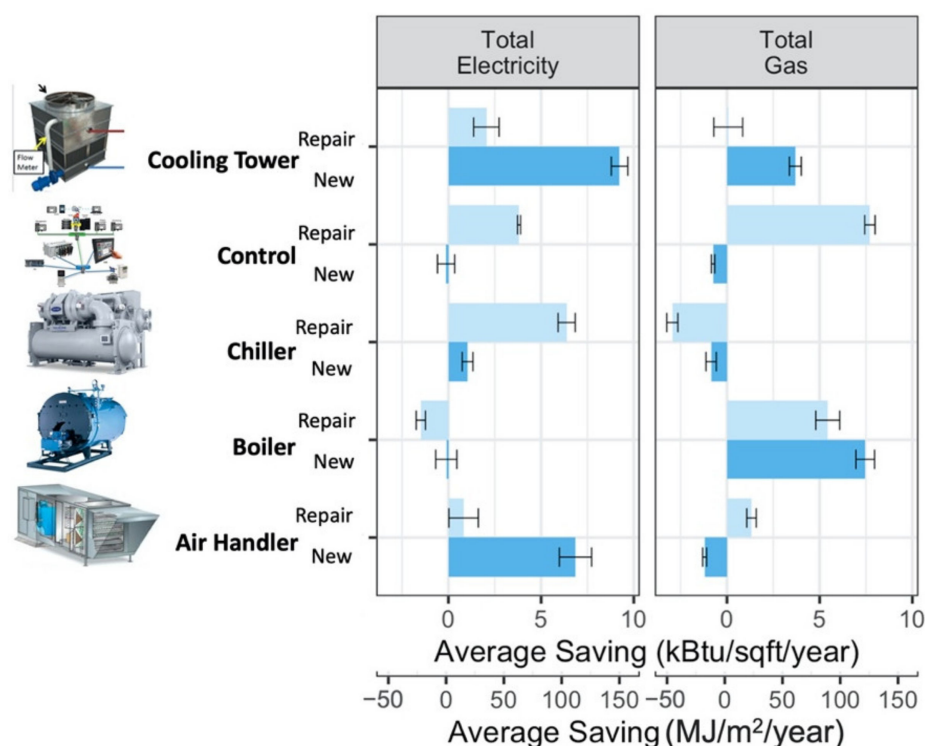
Figure 8. Bar chart of the mean and 95% C.I. of the estimated retrofit effects of past HVAC retrofit sub-actions on the site electricity EUI and site gas EUI.

Among the building envelope sub-actions, new windows had the highest site electricity EUI savings. Across all building envelope sub-actions, new windows were the best at reducing site electricity EUI, followed by repairing façades and repairing roofs. Repairing façades was the most effective in reducing site gas EUI. The negative gas EUI savings in new windows and new facades might have been due to the increased solar heat gain coefficient (SHGC) that reduces the contribution of passive solar heat in the winter. The negative site gas EUI savings of repairing roofs might have been a result of the reduction of solar gain in winter from adding more reflective roof coatings to reduce cooling loads. These negative values suggest that the goal of most envelope retrofits is reducing the cooling load rather than improving thermal insulation. Figure 9 presents the summary statistics of the predicted retrofit effects of six building envelope sub-actions on the site electricity EUI and site gas EUI.

As is anticipated, lighting sub-actions were more effective at saving site electricity EUI than site gas EUI (Figure 10). On average, daylighting and outdoor lighting controls were the most effective at saving site electricity EUI. Daylighting retrofits also had some site gas EUI savings, which might have been due to the improved windows and the increased solar heat gain from daylighting.
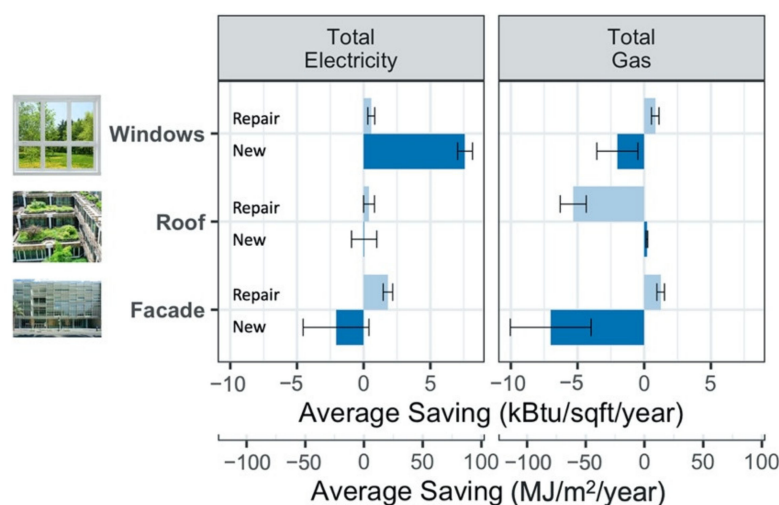
Figure 9. Bar chart of the mean and 95% C.I. of the estimated retrofit effects of past building envelope retrofit sub-actions on the site electricity EUI and site gas EUI.
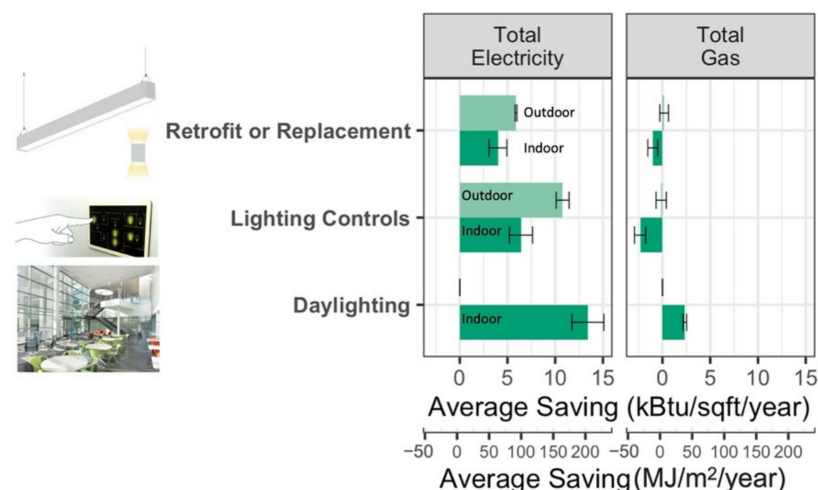
Figure 10. Bar chart of the mean and 95% C.I. of the estimated retrofit effects of past lighting retrofit sub-actions.

3.2. Model Performance

Currently, there is no established methodology to test the true performance of data-driven causal models using observational data, where the treatment, retrofit in this case, is not randomly assigned. This is because the retrofit effect is not observable. For data from experiments, the observed difference between the treated and untreated units equals the actual treatment effect, but this is not the case for non-experiment data.

There are a few heuristics measures provided in the grf package to evaluate the relative performance of the model. The mean forest prediction score evaluates the goodness of the prediction of the average effect. The closer the score is to 1, the better. GSALink and lighting had the most accurate mean prediction in electricity savings. HVAC, building envelope, and commissioning had the most accurate mean prediction in gas savings (Table 6). Another metric is the differential forest prediction score. The *p*-value of the score, also shown in Table 6, reflects whether the treatment effect heterogeneity is statistically significant. The effect heterogeneity was statistically significant for electricity savings of HVAC and commissioning, and marginally significant for building envelope. The effect heterogeneity was statistically significant for the gas savings of advanced metering.

Table 6. Model performance in the prediction of the mean and effect heterogeneity of the six retrofit actions.

Retrofit Action	The Absolute Value of the Mean Forest Prediction Score—1		The *p*-Value of the Differential Forest Prediction Score	
	Electricity	Gas	Electricity	Gas
Advanced metering	1.78	0.87	0.94	**0.04 ***
Building envelope	0.37	**0.13**	0.07	0.49
Commissioning	0.98	**0.15**	**0.04 ***	0.34
GSALink	**0.19**	1.55	0.95	0.93
HVAC	0.45	**0.01**	**0.00 ***	0.27
Lighting	**0.13**	1.41	0.21	0.73

*, *** in the last two columns indicate statistical significance at $p < 0.05$ and $p < 0.001$.

3.3. The Benefit of Targeting Buildings with Higher Savings

This section illustrates how the knowledge of predicted savings could help portfolio owners achieve higher portfolio-wide savings. The scatter plot in Figure 11 explains why targeting and prioritization could improve portfolio-wide savings on HVAC retrofits. The "reality" case reflects implemented retrofit decisions in reality. The "optimal" case represents the scenario where only buildings with high savings potential for that action are retrofitted. Compared to the "optimal" scenario, the implemented retrofit decisions were suboptimal in two aspects. First, it wasted resources retrofitting buildings with zero or negative savings potential and it missed out on buildings with high savings potential. Targeting and prioritization could correct both mistakes.
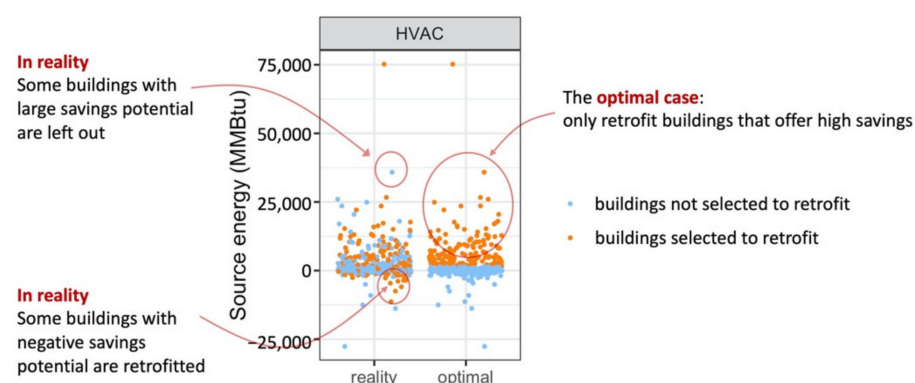

Figure 11. Illustration of why targeting works in improving portfolio-wide source energy savings of HVAC retrofits.

To evaluate the benefit of prioritization, a hypothetical scenario is examined where the portfolio manager could go back in time and use the knowledge of the predicted retrofit savings to re-assign retrofits to the buildings with the highest savings potential while maintaining the same number of retrofitted buildings. As is shown in Figure 12, by targeting the buildings with high predicted savings, the portfolio-wide total savings for 552 federal buildings could improve by about 110–300 billion Btu in site energy, 170–680 billion Btu in source energy, and 10,000–50,000 metric tons of CO_2 emissions reduction.

Figure 13 compares the benefit of prioritizing these HVAC sub-actions to the buildings with the greatest savings potential compared to the implemented retrofits, keeping the number of retrofitted buildings the same. Repairing controls had the highest portfolio savings improvement in site energy. New cooling towers improved most in source energy savings and CO_2 emissions reduction. New air handlers improved the two energy expense metrics the most by prioritization. The highest benefits of HVAC sub-actions were about 41% to 42% equivalent to the pre-retrofit energy consumption or GHG emissions among all sub-actions and objectives (site energy, source energy, energy expense, energy expense considering environmental externalities, and CO_2 emissions).
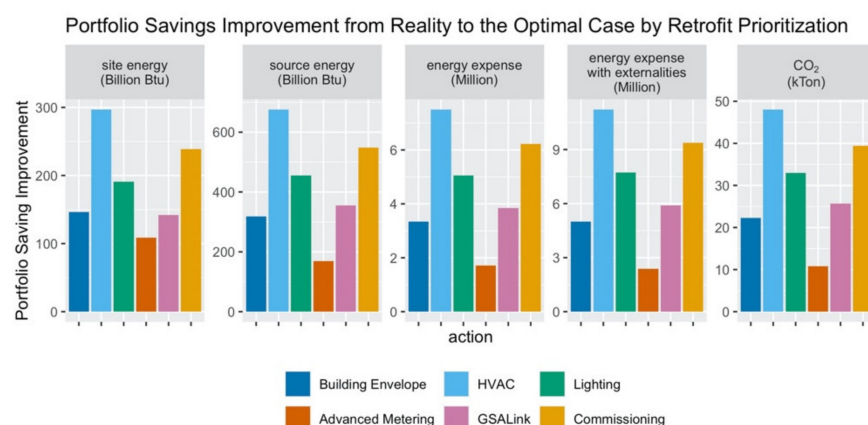
Figure 12. The improvement of portfolio-wide savings from the "reality" case to the "optimal" case when re-assigning retrofits to high-saving buildings (1 billion Btu = 1055 billion joules).
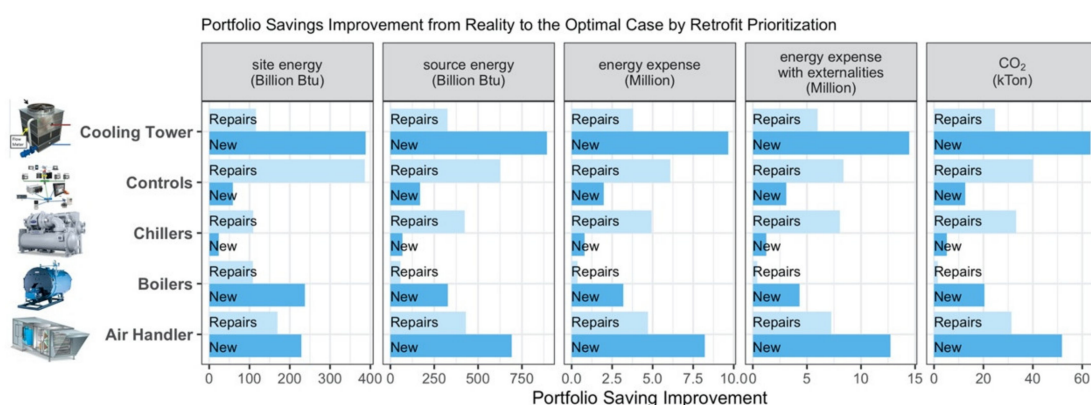
**Figure 13.** The benefit of targeting for HVAC retrofit sub-actions (1 billion Btu = 1055 billion joules).

Prioritizing envelope sub-actions for the buildings with the greatest savings potential revealed that new windows achieved the highest improvement of 456 billion Btu in source energy, 5 million in energy expenses considering externalities, and 35,000 metric tons in CO_2 emissions reduction. These improvements constitute 15% to 23% of the pre-retrofit energy consumption or GHG emissions for each of the five objectives. New roofs ranked second in prioritized benefits, with savings equivalent to 12% to 14% of the pre-retrofit quantities. Figure 14 summarizes the benefits of prioritizing building envelope sub-actions.
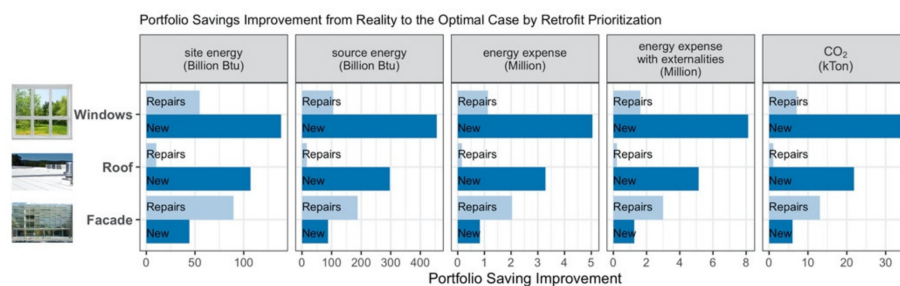
Figure 14. The benefits of prioritizing building envelope sub-actions (1 billion Btu = 1055 billion joules).

Prioritizing lighting sub-actions for the buildings with the greatest savings potential revealed that indoor daylighting benefitted the most from prioritization across all objectives, achieving an improvement of 1175 billion Btu in source energy, USD 13 million in energy expenses considering externalities, and 86,000 metric tons in CO_2 reductions. These improvements are as large as 50–56% of the pre-retrofit energy consumption or GHG emissions for each of the five objectives. The outdoor actions did not benefit as much as the

indoor actions, but they still achieved an improvement of close to 30% of the pre-retrofit quantities. Figure 15 illustrates the benefits of prioritizing these lighting sub-actions for the top saving buildings.
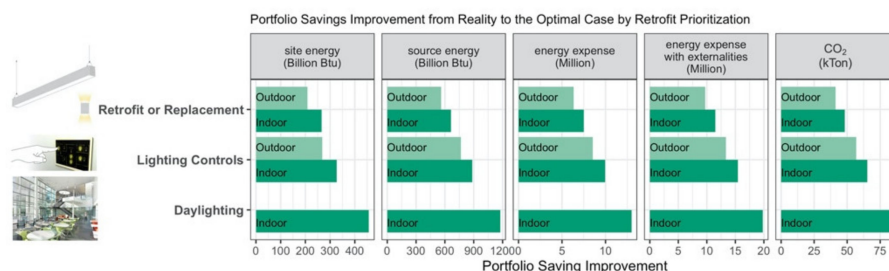
Figure 15. The benefits of prioritizing lighting sub-actions (1 billion Btu = 1055 billion joules).

3.4. The Association between Predicted Savings and Weather

To interpret the causal forest (CF) predictions and evaluate the association between the input features and the predicted savings, a linear regression model was fitted by regressing the CF-estimated retrofit effect (technically, the doubly robust scores [69]) onto a subset of the most informative predictors (the variable with importance scores at least 10 times the average importance scores across all variables was selected). A schematic flow diagram is shown in Figure 16. The coefficients of the linear model reflect how much difference in savings prediction could be associated with a one-unit difference in the value of the retrofit effect predictor.
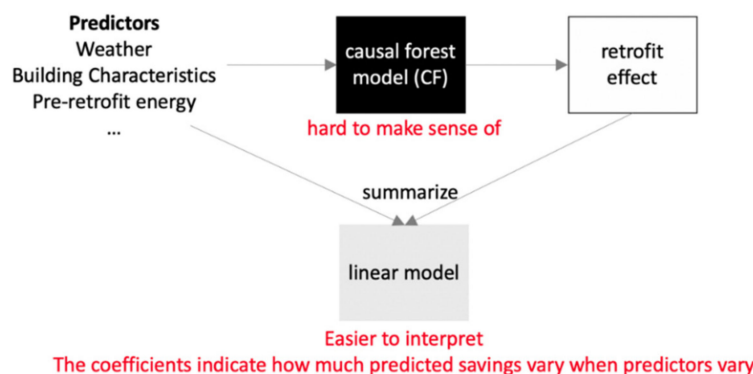
Figure 16. Schematic diagram of summarizing a causal forest with a linear model (linear model).

As an example, Figure 17 visualizes the linear model coefficients summarizing the CF prediction of HVAC on electricity savings. Hotter and colder weather was associated with increased HVAC electricity savings; specifically, one additional day with a temperature below 10 °F was associated with a 0.6 kBtu/sqft/year electricity savings increase by investment in HVAC retrofits ($p = 0.03$).

The previous U-shaped pattern appeared in the electricity savings of most actions, where total electricity savings increased with additional cold or hot days in a year. GSALink had the highest increase in hot-day electricity savings, at about 1.05 kBtu/sqft/year per one additional hot day in a year with a daily temperature above 90 °F. More cold days were associated with more gas savings in building envelope, HVAC, and lighting retrofits. Gas savings from lighting retrofits might have been due to the correlation between the outdoor daylighting level and outdoor temperature. For example, buildings at a high-altitude location could have cold and dark winters, resulting in increased consumption of both heating and lighting. These findings suggest that building envelope, HVAC, and lighting retrofits should be pursued in climates with more extreme weather, whereas advanced metering may have similar savings in extreme and mild climates. Figure 18 visualizes the weather variable coefficients of the linear model approximations of the causal forest models for each action–fuel combination.
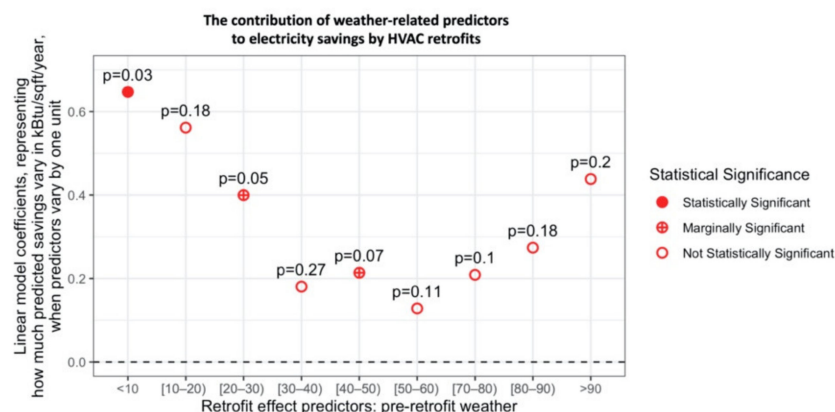
Figure 17. The contribution of weather predictors to total electricity savings by HVAC retrofits.
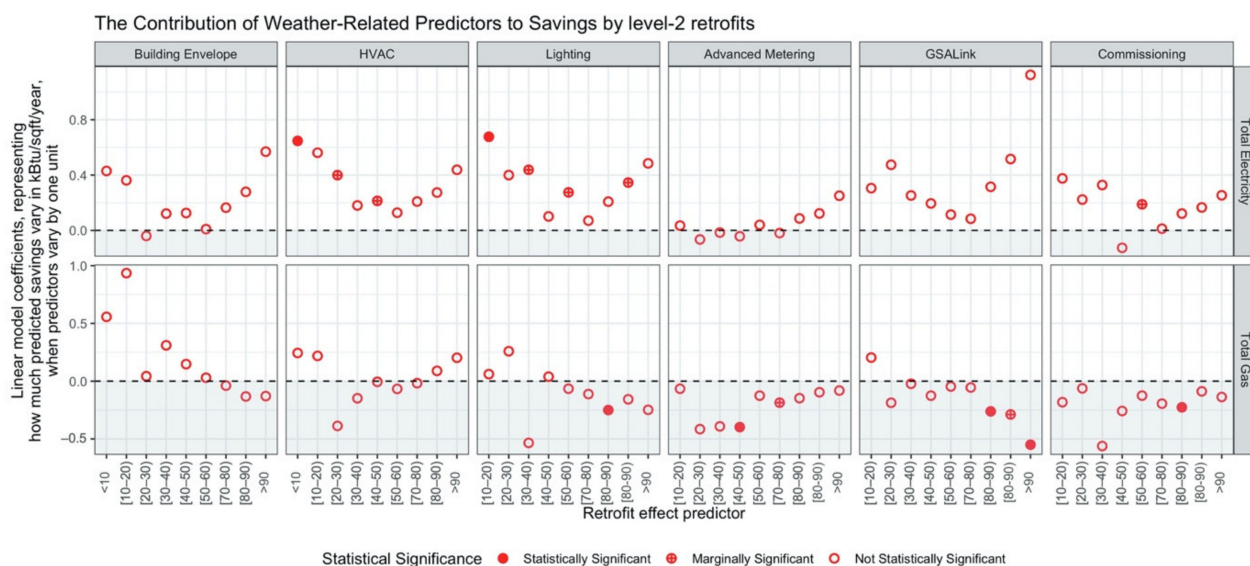

Figure 18. The contribution of weather-related predictors to the savings prediction of electricity and gas savings of six actions.

## 4. Discussion

4.1. Variable Importance of Input Features

A variable importance evaluation was conducted for each action–outcome combination based on how often each variable was used in top-level splits when building the causal trees [70].

As an example, Figure 19 shows the importance ranking of each retrofit effect predictor in the estimation of the effect of HVAC retrofits on annual electricity use (kBtu/sqft/year). The retrofit effect predictors with close to zero importance were omitted from the plot for clarity of presentation. The analysis revealed that for predicting the electricity savings of HVAC retrofits, pre-retrofit gas and electricity use ranked as the first and third most important variables in the prediction of the electricity savings of HVAC retrofits. Although it may be unexpected that pre-retrofit gas matters for electricity savings, the combination of the two pre-retrofit conditions determines whether space heating is all-electric, all-gas, or a combination of the two. This could determine the magnitude and seasonal pattern of electricity savings. Furthermore, the building with high gas usage could have had high electricity use as well, possibly due to the low thermal insulation level. The number of days within 30–40 °F or 80–90 °F ranked second and fourth in the prediction of the effect of HVAC retrofits on annual electricity use, as they were related to the heating and cooling load, and thus useful in predicting electricity savings.
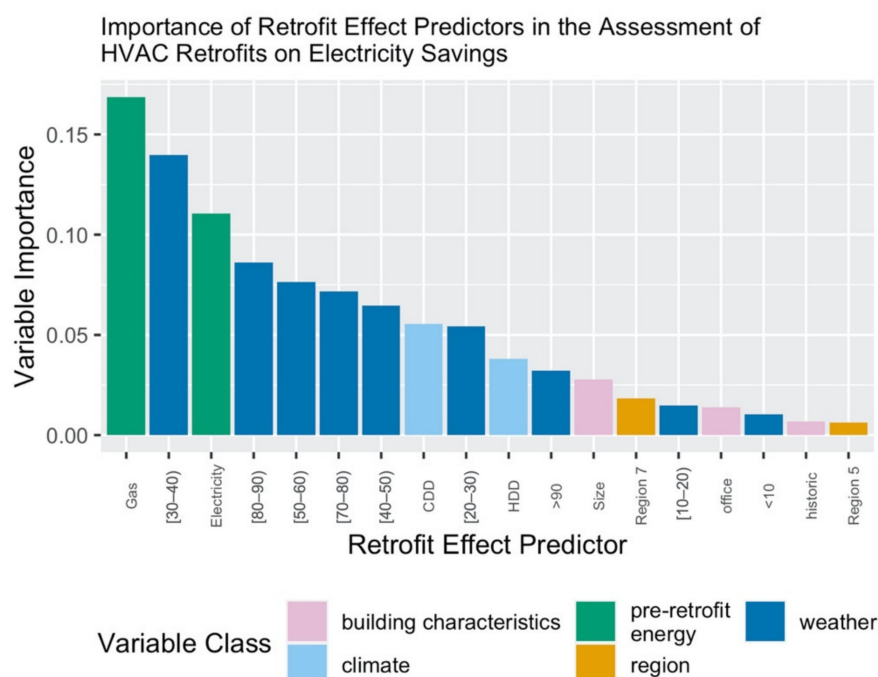
Figure 19. Variable importance of retrofit effect predictors predicting the effect of HVAC retrofits on annual electricity use.

Due to a large number of action–outcome combinations, the variable importance of each action–outcome combination was aggregated into one value, indicating the most important predictor class among the top five most important predictors, identified by selecting the class with the highest average importance score. Figure 20 visualizes the most important retrofit effect predictor classes for the effect of six retrofit actions on electricity and gas savings. Variable to variable-class mapping is shown in Table 4. Pre-retrofit gas and electric energy use and weather were the most important variable classes for estimating the effect of most actions on both electricity and gas savings. It would be most important to consider short-term weather (dark blue) for the prediction of the electricity savings of commissioning and lighting actions and long-term climate (light blue) for the prediction of the gas savings of GSALink and advanced metering.
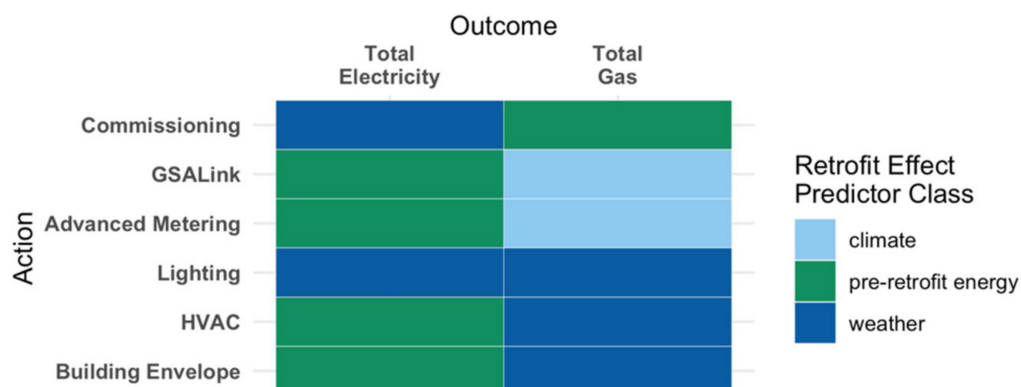
Figure 20. Most important retrofit effect predictor categories for level-2 actions.

4.2. Results Comparison with Similar Studies

In this sub-section, the retrofit effects estimated in this study are compared to several data-driven building studies evaluating similar retrofit actions.

Among the three operational actions, commissioning is the most well studied. According to a meta-analysis by Mills et al. [71], the median commissioning savings is 5.8 kBtu/sqft/year for

electricity and 6.5 kBtu/sqft/year for gas, substantially larger than the savings estimated in this study, at 0.9 kBtu/sqft/year for electricity and 2.6 kBtu/sqft/year for gas. The lower estimates in this study could have been due to the fact that it frequently co-exists with other retrofit actions. Even if multi-actions are controlled with an indicator variable of a co-existing action category in the model, this analysis assumed the additive of retrofit effects, which might not be the case, according to Chidiac et al. [72]. The effect of commissioning could have been overshadowed by co-existing actions.

Although commissioning is evaluated in commercial buildings, the energy savings impact of advanced metering is mostly studied in residential buildings with large variations in estimated electricity savings, from 1.6–1.7% [73] to 6.1–6.4% [74]. In this study, the installation and use of advanced metering was estimated to save an average of 2.1% (0.47 kBtu/sqft/year) electricity, which is within the range of savings magnitude for residential buildings.

The advantages of building automation, energy-usage visualization, and fault-detection programs such as GSALink have been quantified to have a median energy savings of 8% (ranging from −1% to 30%) [75]. In this analysis, the estimated median energy savings was 10.6% (ranging from 2% to 38%), which is slightly higher than the previously quantified median savings and its range.

For the three capital actions analyzed in this study, the average electricity savings demonstrated was 6.5% for lighting, −2.7% for enclosure retrofits, and 4.6% for HVAC retrofits. These values are lower than two previous commercial building studies of 10.2% for lighting and 18% for HVAC in [9], and 42% for lighting and 49% for HVAC in [18]. Among retrofit sub-actions, the average of window repairs and replacements at 22% is similar to the 17.4% electricity savings in [9]. In the current study, new controls reduced the electricity consumption by 0.13% on average, whereas repairing controls achieved 9.3% electricity savings, larger than the magnitude of savings of 2.1% reported in [9].

As these comparisons suggest, the energy savings quantified in this study were generally on the lower end of the savings estimated in the literature. In addition to the impacts of multiple investments, these lower estimates could have been due to a portfolio-wide energy reduction goal in the federal sector. All savings were estimated by contrasting the pre- to post-retrofit energy changes of the retrofitted and the un-retrofitted. For the federal portfolio, the no-retrofit buildings might also have had energy reductions due to Executive Order 13,423, which sets goals for annual reductions in energy.

5. Conclusions

This paper presents the application of a machine learning method with causal forest for the prediction of retrofit savings from six broad retrofit actions and sub-actions using a commercial building portfolio. The paper fills in a gap in the power of using data-driven methods to predict retrofit savings, which is currently dominated by simulation methods. Simulation-based retrofit planning methods allow full control of physical and environmental parameter settings and have the potential to analyze a wide range of retrofit scenarios, including prototyping new retrofits. However, they require intensive data input, high expertise, and long computation time. In addition, most simulation-based tools do not adequately assess operational or behavioral changes or their impacts on energy efficiency. The data-driven method proposed in this study captures capital and operational variables through actual energy records, is less computation- and data-intensive, and can be more easily extend to the evaluation of large portfolio projects. The savings predictions are most likely closer to reality due to the use of real measured data. The study also contributes to the limited literature on large-scale empirical evaluation of energy efficiency interventions in commercial buildings.

Based on measured data on energy use and weather history, retrofit records, building characteristics, and other control variables identified through the causal mechanism analysis, portfolio owners and policymakers can train models that estimate energy savings for past retrofits, predict energy savings for future retrofits, and design retrofit plans that

maximize portfolio savings or cost-effectiveness by targeting building sub-groups with high predicted savings. To save energy, portfolio owners can first rank buildings based on the model-predicted energy savings, then select buildings with positive savings. This ensures the largest portfolio-wide energy savings. When a budget constraint is added, the portfolio owners could rank buildings by energy savings per dollar spent in the retrofit implementation and select buildings from the top of the list until the budget runs out. When there are more layered savings objectives, multi-objective optimization methods such as [76,77] can be incorporated into the decision support process.

This study presents an initial demonstration of how to use a data-driven approach to predict the retrofit effect as a function of a series of characteristics. It is far from complete, with the following limitations.

- The sample size is still rather small, given the desire to evaluate six classes of energy conservation actions (ECM) and 21 sets of sub-actions.
- The ECM action classifications are broad with limited or conflicting indications of the sub-actions taken.
- The retrofit records only have a completion date, not a start date, such that before and after data sets may not be accurately defined. In addition, many actions share the same completion date, which could lead to the contamination of "pre-retrofit" data with "during-retrofit" data, potentially resulting in biased savings estimates.
- As previously discussed, Executive Orders 13,423 and 13,693 are strong policy drivers for buildings in the GSA portfolio to reduce their energy consumption, whether they are retrofitted or not. This might make the retrofit savings estimates more conservative and the same action might reach higher savings in buildings without such a policy driver.
- Due to some un-documented retrofit actions, some retrofitted buildings might be categorized as un-retrofitted in the study. This could also lead to a more conservative savings estimation. In the future, such uncertainty should be reduced by improving the retrofit action documentation or including a sensitivity analysis.
- Even though co-existing or past actions are controlled for with indicator variables, the interactions between actions might be more complicated and not adequately accounted for.

The following future improvements are identified to enhance the performance and increase the usability of the proposed method.

- Acquire larger data sets with more thorough documentation of retrofit details and building characteristics and covering a larger variety of retrofit actions. With this improvement, savings could be predicted more accurately for a larger set of more specific actions.
- The decision support section is relatively simple in this study, as the focus is more on the savings prediction, rather than decision optimization. More complicated multi-objective optimization methods in Section 1.2.1 could be incorporated.
- The effect of the retrofit action sequences could be studied in the future, possibly using methods from the time-varying treatment effect literature, such as [78–80].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

References

1. U.S. Energy Information Administration (EIA), Annual Energy Review. 2019. Available online: https://www.eia.gov/totalenergy/data/annual/index.php (accessed on 27 December 2019).
2. United States Environmental Protection Agency (EPA). *US Energy Use Intensity by Property Type*; Energy Star: Washington, DC, USA, 2018.
3. Luddeni, G.; Krarti, M.; Pernigotto, G.; Gasparella, A. An analysis methodology for large-scale deep energy retrofits of existing building stocks: Case study of the Italian office building. *Sustain. Cities Soc.* **2018**, *41*, 296–311. [CrossRef]
4. Zhai, J.; LeClaire, N.; Bendewald, M. Deep energy retrofit of commercial buildings: A key pathway toward low-carbon cities. *Carbon Manag.* **2011**, *2*, 425–430. [CrossRef]
5. Zuhaib, S.; Goggins, J. Assessing evidence-based single-step and staged deep retrofit towards nearly zero-energy buildings (nZEB) using multi-objective optimisation. *Energy Effic.* **2019**, *12*, 1891–1920. [CrossRef]
6. Granade, H.C.; Creyts, J.; Derkach, A.; Farese, P.; Nyquist, S.; Ostrowski, K. Unlocking Energy Efficiency in the US Economy; McKinsey Co. 2009. Available online: https://www.mckinsey.com/~{}/media/mckinsey/dotcom/client_service/epng/pdfs/unlocking%20energy%20efficiency/us_energy_efficiency_exc_summary.ashx (accessed on 22 October 2018).
7. Brainerd, J.G.; Stobaugh, R.; Yergin, D.; Phillips, O. Energy Future: Report of the Energy Project at the Harvard Business School. *Technol. Cult.* **1981**, *22*, 217. [CrossRef]
8. Fowlie, M.; Greenstone, M.; Wolfram, C. Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program. *Q. J. Econ.* **2018**, *133*, 1597–1644. [CrossRef]
9. Liang, J.; Qiu, Y.; James, T.; Ruddell, B.; Dalrymple, M.; Earl, S.; Castelazo, A. Do energy retrofits work? Evidence from commercial and residential buildings in Phoenix. *J. Environ. Econ. Manag.* **2018**, *92*, 726–743. [CrossRef]
10. Angrist, J.D.; Pischke, J.-S. *Mastering'metrics: The Path from Cause to Effect*; Princeton University Press: Princeton, NJ, USA, 2014.
11. Allcott, H.; Greenstone, M. *Measuring the Welfare Effects of Residential Energy Efficiency Programs*; National Bureau of Economic Research: Cambridge, MA, USA, 2017.
12. Zivin, J.G.; Novan, K. Upgrading Efficiency and Behavior: Electricity Savings from Residential Weatherization Programs. *Energy J.* **2016**, *37*, 1–24. [CrossRef]
13. Dubin, J.A.; McFadden, D.L. *A Heating and Cooling Load Model for Single-Family Detached Dwellings in Energy Survey Data*; California Institute of Technology: Pasadena, CA, USA, 1983.
14. Dubin, J.A.; Miedema, A.K.; Chandran, R.V. Price Effects of Energy-Efficient Technologies: A Study of Residential Demand for Heating and Cooling. *RAND J. Econ.* **1986**, *17*, 310. [CrossRef]
15. Davis, L.W.; Fuchs, A.; Gertler, P. Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico. *Am. Econ. J. Econ. Policy* **2014**, *6*, 207–238. [CrossRef]
16. Scheer, J.; Clancy, M.; Hógáin, S.N. Quantification of energy savings from Ireland's Home Energy Saving scheme: An ex post billing analysis. *Energy Effic.* **2012**, *6*, 35–48. [CrossRef]
17. Grimes, A.; Preval, N.; Young, C.; Arnold, R.; Denne, T.; Howden-Chapman, P.; Telfar-Barnard, L. Does Retrofitted Insulation Reduce Household Energy Use? Theory and Practice. *Energy J.* **2016**, *37*, 165–186. [CrossRef]
18. Burlig, F.; Knittel, C.; Rapson, D.; Reguant, M.; Wolfram, C. Machine Learning from Schools about Energy Efficiency. *J. Assoc. Environ. Resour. Econ.* **2020**, *23908*, 1181–1217. [CrossRef]
19. Horn, T.D.M.; Borden, M. *Cost Effectiveness*; California Energy Commission: Sacramento, CA, USA, 1980.
20. Levinson, A. How Much Energy Do Building Energy Codes Save? Evidence from California Houses. *Am. Econ. Rev.* **2016**, *106*, 2867–2894. [CrossRef]
21. Blonz, J.A. The Welfare Costs of Misaligned Incentives: Energy Inefficiency and the Principal-Agent Problem. *Financ. Econ. Discuss. Ser.* **2019**, *2019*, 1–39. [CrossRef]
22. Christensen, P.; Francisco, P.; Myers, E.; Souza, M. Decomposing the wedge between projected and realized returns in energy efficiency programs. *E2e Work. Pap.* **2020**, *46*, 1–37.

23. Giraudet, L.-G.; Houde, S.; Maher, J.A. Moral Hazard and the Energy Efficiency Gap: Theory and Evidence. *J. Assoc. Environ. Resour. Econ.* **2018**, *5*, 755–790. [CrossRef]

24. Allcott, H.; Greenstone, M. Is There an Energy Efficiency Gap? *J. Econ. Perspect.* **2012**, *26*, 3–28. [CrossRef]

25. Xu, Y. *Using Machine Learning to Target Retrofits in Commercial Buildings under Alternative Climate Change Scenarios*; Carnegie Mellon University: Pittsburgh, PA, USA, 2020.

26. ASHRAE. Ashrae Guideline 14-2014: Measurement of Energy, Demand and Water Savings. 2014. Available online: https://books.google.com/books?id=zlJkAQAACAAJ (accessed on 28 March 2018).

27. Ma, Z.; Cooper, P.; Daly, D.; Ledo, L. Existing building retrofits: Methodology and state-of-the-art. *Energy Build.* **2012**, *55*, 889–902. [CrossRef]

28. Grillone, B.; Danov, S.; Sumper, A.; Cipriano, J.; Mor, G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. *Renew. Sustain. Energy Rev.* **2020**, *131*, 110027. [CrossRef]

29. Efficiency Valuation Organisation. *International Performance Measurement and Verification Protocol. Concepts and Options for Determining Energy and Water Savings*; International Performance Measuremnt & Verification Protocol Committee: Golden, CO, USA, 2012.

30. Marasco, D.E.; Kontokosta, C.E. Applications of machine learning methods to identifying and predicting building retrofit opportunities. *Energy Build.* **2016**, *128*, 431–441. [CrossRef]

31. Salvalai, G.; Malighetti, L.E.; Luchini, L.; Girola, S. Analysis of different energy conservation strategies on existing school buildings in a Pre-Alpine Region. *Energy Build.* **2017**, *145*, 92–106. [CrossRef]

32. Geyer, P.; Schlüter, A.; Cisar, S. Application of clustering for the development of retrofit strategies for large building stocks. *Adv. Eng. Inform.* **2017**, *31*, 32–47. [CrossRef]

33. Knittel, C.; Stolper, S. Using Machine Learning to Target Treatment: The Case of Household Energy Use. *Nat. Bur. Econ. Res.* **2019**, *26531*, 1–47. [CrossRef]

34. Allcott, H.; Kessler, J.B. The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *Am. Econ. J.* **2015**, *11*, 236–276. [CrossRef]

35. Reddy, T.A.; Saman, N.F.; Claridge, D.E.; Haberl, J.S.; Turner, H.W.D.; Chalifoux, A.T. Baselining Methodology for Facility-Level Monthly Energy Use-part 1: Theoretical aspects. *ASHRAE Trans.* **1997**, *103*, 336–347.

36. Fels, M.F. PRISM: An introduction. *Energy Build.* **1986**, *9*, 5–18. [CrossRef]

37. Reddy, T.A.; Claridge, D.; Wu, J. Statistical Modeling of Daily Energy Consumption in Commercial Buildings Using Multiple Regression and Principal Component Analysis. 1992. Available online: https://www.researchgate.net/publication/4761 2918_Statistical_Modeling_of_Daily_Energy_Consumption_in_Commercial_Buildings_Using_Multiple_Regression_and_ Principal_Component_Analysis (accessed on 3 November 2016).

38. Thamilseran, S.; Haberl, J.S. A Bin Method for Calculating Energy Conservation Retrofit Savings in Commercial Buildings. 1994. Available online: http://oaktrust.library.tamu.edu/handle/1969.1/6640 (accessed on 21 October 2016).

39. Zhang, Y.; O'Neill, Z.; Dong, B.; Augenbroe, G. Comparisons of inverse modeling approaches for predicting building energy performance. *Build. Environ.* **2015**, *86*, 177–190. [CrossRef]

40. Granderson, J.; Price, P. *Evaluation of the Predictive Accuracy of Five Whole Building Baseline Models*; LBNL-5886E; USDOE Office of Science (SC): Berkeley, CA, USA, 2012.

41. Mackay, D.J.C. Bayesian Non-Linear Modeling for the Prediction Competition. In *Maximum Entropy and Bayesian Methods*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 1996; pp. 221–234.

42. Brown, M.; Barrington-Leigh, C.; Brown, Z. Kernel regression for real-time building energy analysis. *J. Build. Perform. Simul.* **2012**, *5*, 263–276. [CrossRef]

43. Mocanu, E.; Nguyen, P.; Gibescu, M.; Kling, W.L. Deep learning for estimating building energy consumption. *Sustain. Energy Grids Netw.* **2016**, *6*, 91–99. [CrossRef]

44. Dong, B.; Cao, C.; Lee, S.E. Applying support vector machines to predict building energy consumption in tropical region. *Energy Build.* **2005**, *37*, 545–553. [CrossRef]

45. Solomon, D.M.; Winter, R.L.; Boulanger, A.G.; Anderson, R.N.; Wu, L.L. *Forecasting Energy Demand in Large Commercial Buildings Using Support Vector Machine Regression*; Columbia University Libraries: New York, NY, USA, 2011.

46. Heo, Y.; Zavala, V.M. Gaussian process modeling for measurement and verification of building energy savings. *Energy Build.* **2012**, *53*, 7–18. [CrossRef]

47. Yu, Z.; Haghighat, F.; Fung, B.C.M.; Yoshino, H. A decision tree method for building energy demand modeling. *Energy Build.* **2010**, *42*, 1637–1646. [CrossRef]

48. Behl, M.; Mangharam, R. Evaluation of DR-Advisor on the ASHRAE Great Energy Predictor Shootout Challenge. 2014. Available online: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1093&context=mlab_papers (accessed on 16 May 2017).

49. Asadi, E.; da Silva, M.C.G.; Antunes, C.H.; Dias, L.; Glicksman, L. Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application. *Energy Build.* **2014**, *81*, 444–456. [CrossRef]

50. Ascione, F.; Bianco, N.; De Masi, R.F.; De Stasio, C.; Mauro, G.M.; Vanoli, G.P. Multi-objective optimization of the renewable energy mix for a building. *Appl. Therm. Eng.* **2016**, *101*, 612–621. [CrossRef]

51. Ascione, F.; Bianco, N.; Mauro, G.M.; Napolitano, D.F.; Vanoli, G.P. A Multi-Criteria Approach to Achieve Constrained Cost-Optimal Energy Retrofits of Buildings by Mitigating Climate Change and Urban Overheating. *Climate* **2018**, *6*, 37. [CrossRef]

52. Bichiou, Y.; Krarti, M. Optimization of envelope and HVAC systems selection for residential buildings. *Energy Build.* **2011**, *43*, 3373–3382. [CrossRef]

53. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]

54. Delgarm, N.; Sajadi, B.; Delgarm, S.; Kowsary, F. A novel approach for the simulation-based optimization of the buildings energy consumption using NSGA-II: Case study in Iran. *Energy Build.* **2016**, *127*, 552–560. [CrossRef]

55. Magnier, L.; Haghighat, F. Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. *Build. Environ.* **2010**, *45*, 739–746. [CrossRef]

56. Ascione, F.; Bianco, N.; De Stasio, C.; Mauro, G.M.; Vanoli, G.P. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. *Energy* **2017**, *118*, 999–1017. [CrossRef]

57. Siddharth, V.; Ramakrishna, P.; Geetha, T.; Sivasubramaniam, A. Automatic generation of energy conservation measures in buildings using genetic algorithms. *Energy Build.* **2011**, *43*, 2718–2726. [CrossRef]

58. Lara, R.A.; Pernigotto, G.; Cappelletti, F.; Romagnoni, P.; Gasparella, A. Energy audit of schools by means of cluster analysis. *Energy Build.* **2015**, *95*, 160–171. [CrossRef]

59. Lee, S.H.; Hong, T.; Piette, M.A.; Taylor-Lange, S.C. Energy retrofit analysis toolkits for commercial buildings: A review. *Energy* **2015**, *89*, 1087–1100. [CrossRef]

60. Burrell, M. GSA Looks Back at the American Recovery and Reinvestment Act. 2016. Available online: https://www.gsa.gov/blog/2016/02/18/gsa-looks-back-at-the-american-recovery-and-reinvestment-act (accessed on 12 July 2020).

61. FedCenter. EO 13693 (Archive)—Revoked by EO 13834 on 17 May 2018, Sec 8. 2019. Available online: https://www.fedcenter.gov/programs/eo13693/ (accessed on 12 July 2020).

62. Wenninger, S. Advanced Metering Ensures Cost Effective and Efficient Federal Facilities. 2013. Available online: https://www.gsa.gov/blog/2013/01/22/Advanced-Metering-Ensures-Cost-Effective-and-Efficient-Federal-Facilities (accessed on 2 January 2020).

63. GSA Public Buildings Service. The Building Commissioning Guide. 2015. Available online: https://www.gsa.gov/cdnstatic/BCG_3_30_Final_R2-x221_0Z5RDZ-i34K-pR.pdf (accessed on 16 April 2020).

64. GSA. Lighting. 2018. Available online: https://www.gsa.gov/governmentwide-initiatives/sustainability/emerging-building-technologies/gsa-technology-deployment-maps/lighting (accessed on 12 September 2018).

65. Kontokosta, C.E. Modeling the energy retrofit decision in commercial office buildings. *Energy Build.* **2016**, *131*, 1–20. [CrossRef]

66. Deschênes, O.; Greenstone, M. Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US. *Am. Econ. J. Appl. Econ.* **2007**, *3*, 152–185. [CrossRef]

67. Dell, M.; Jones, B.; Olken, B. What Do We Learn from the Weather? The New Climate-Economy Literature. *J. Econ Lit.* **2013**, *52*, 740–798. [CrossRef]

68. Auffhammer, M.; Mansur, E.T. Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Econ.* **2014**, *46*, 522–530. [CrossRef]

69. Wager, S.; Athey, S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *J. Am. Stat. Assoc.* **2018**, *113*, 1228–1242. [CrossRef]

70. Tibshirani, J.; Athey, S.; Wager, S. Generalized Random Forests. 2020. Available online: https://CRAN.R-project.org/package=grf (accessed on 29 June 2021).

71. Mills, E.; Friedman, H.; Powell, T.; Bourassa, N.; Claridge, D.; Piette, M.A. *The Cost-Effectiveness of Commercial-Buildings Commissioning*; Berkeley Lab: Berkeley, CA, USA, 2004; p. 98.

72. Chidiac, S.; Catania, E.; Morofsky, E.; Foo, S. Effectiveness of single and multiple energy retrofit measures on the energy consumption of office buildings. *Energy* **2011**, *36*, 5037–5052. [CrossRef]

73. Faruqui, A.; Arritt, K.; Sergici, S. The impact of advanced metering infrastructure on energy conservation: A case study of two utilities. *Electr. J.* **2017**, *30*, 56–63. [CrossRef]

74. Paschmann, M.H.; Paulus, S. The Impact of Advanced Metering Infrastructure on Residential Electricity Consumption: Evidence from California. EWI Working Paper No. 17/08. 2017. Available online: https://www.econstor.eu/handle/10419/172838 (accessed on 29 June 2021).

75. Lin, G.; Kramer, H.; Granderson, J. Building fault detection and diagnostics: Achieved savings, and methods to evaluate algorithm performance. *Build. Environ.* **2020**, *168*, 106505. [CrossRef]

76. Carli, R.; Dotoli, M.; Pellegrino, R.; Ranieri, L. Using multi-objective optimization for the integrated energy efficiency improvement of a smart city public buildings' portfolio. In Proceedings of the 2015 IEEE International Conference on Automation Science and Engineering (CASE), Gothenburg, Sweden, 24–28 August 2015; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2015; pp. 21–26.

77. Diakaki, C.; Grigoroudis, E.; Kabelis, N.; Kolokotsa, D.; Kalaitzakis, K.; Stavrakakis, G. A multi-objective decision model for the improvement of energy efficiency in buildings. *Energy* **2010**, *35*, 5483–5496. [CrossRef]

78. Lok, J.; Gill, R.; Van Der Vaart, A.; Robins, J. Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Stat. Neerlandica* **2004**, *58*, 271–295. [CrossRef]

79. Hernán, M.A.; Brumback, B.; Robins, J.M. Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments. *J. Am. Stat. Assoc.* **2001**, *96*, 440–448. [CrossRef]
80. Robins, J.M. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*; National Center for Health Service Research: Washington, DC, USA, 1989; pp. 113–159.