Energy Policy 168 (2022) 113096

Contents lists available at ScienceDirect

Energy Policy

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

City-level impacts of building tune-ups: Findings from Seattle's building tune-ups program

Travis Walter *, Paul Mathew

Lawrence Berkeley National Laboratory, USA

1. Introduction

Many U.S. cities are implementing policies to reduce greenhouse gas (GHG) emissions of their buildings. These range from building energy benchmarking and disclosure to building performance standards (BPS) that require buildings to meet specific targets of energy use or emissions. The City of Seattle adopted a climate action plan in 2013 that set a goal of zero net GHG emissions in the road transportation, buildings, and waste sectors by 2050 (Seattle Office of Sustainability and Environment, 2013), with a number of near and long term actions. Seattle implemented mandatory building tune-ups in 2016, applying to commercial buildings larger than 50,000 sqft ("Seattle Building Tune-Ups," n.d.). Building tune-ups[1] involve assessment and implementation of operational and maintenance (O + M) improvements to achieve energy and water efficiency, such as changes to thermostat set points or adjusting lighting or irrigation schedules (Sullivan et al., 2020). Seattle's tune-ups program covered 27 such improvements in HVAC, lighting, domestic hot water, and envelope systems (further described in Section 2.2).

Building tune-ups are low cost, typically do not require capital investment, and are therefore an important first step toward reducing energy use and emissions prior to implementing higher-cost measures such as replacing lighting and HVAC equipment and envelope upgrades. Additionally, tune-ups can help ensure savings persistence from previously implemented energy efficiency measures. Several utilities have developed incentive programs for building tune-ups (CEE, n.d.). Tune-ups have also been incorporated into industry standards and certifications, including ASHRAE Standard 202 (ASHRAE, 2013) and LEED (USGBC, n.d.).

Many prior studies have documented the energy savings from building tune-ups (Crowe et al., 2020; Mills, 2011; Mills et al., 2004). conducted large scale meta-analysis studies across hundreds of commissioning projects. The most recent of these studies showed median source energy savings for existing building commissioning (EBCx) ranged from 5% for those conducted under utility programs and 14% for projects outside of utility programs (Katipamula, 2016). conducted a meta-analysis of 24 projects and found 15% median energy savings. The City of Seattle conducted an initial analysis of 10 buildings participating in its tune-ups accelerator program and estimated average energy savings of 8.3% (Ballinger, 2020). (Fernandez et al., 2015) found that many of the individual EBCx measures could save up to 20% of HVAC energy use in large office buildings (Fernandez et al., 2017). conducted a simulation-based analysis of building controls measures and found that savings varied significantly by building type, ranging between −12% and 21%. Some individual measures had negative savings because correcting an underlying operational problem (e.g., inadequate ventilation) resulted in an increase in energy consumption. Savings can vary significantly even within similar buildings depending on existing building conditions. For example (Mills and Mathew, 2014), analyzed 24 projects in university buildings in California and found a wide range of 2–25% source energy savings, with a median of 11%.

The savings data in these studies generally involved building-level measurement and verification (M&V) of savings. However, there is limited information on the impacts of tune-ups across a building stock when implemented as a policy measure, since this is a relatively new development. An analysis of New York City's audit and retro-commissioning policy showed reductions of approximately 2.5% for multi-family residential buildings and 4.9% for office buildings (Kontakosta et al., 2020). However, this analysis did not parse out the effect of just the retro-commissioning measures. Analyzing stock-level impacts of building policies itself is a fairly well-established field and the literature shows a variety of methodological approaches depending on the objectives of the analysis, data availability, and desired level of accuracy (Brøgger and Wittchen, 2018; Langevin et al., 2020). For example, one approach is to use sector level aggregate data and then determine sector level impacts of an intervention by assuming a technical savings rate and adoption rate, e.g., (Langevin et al., 2019). At the other end of the

* Corresponding author.
E-mail address: twalter@lbl.gov (T. Walter).

[1] Building tune-ups is similar to and generally synonymous with existing building commissioning (EBCx) and Retro-commissioning (RCx).

https://doi.org/10.1016/j.enpol.2022.113096
Received 27 December 2021; Received in revised form 25 April 2022; Accepted 6 June 2022
Available online 11 June 2022

spectrum are approaches that use data on individual buildings within a geographical area such as a city or a state. These include empirical approaches using energy benchmarking and audit data, e.g., (Bergfeld et al., 2020; Kontokosta et al., 2020; Walter and Mathew, 2021), as well as simulation model-based approaches, e.g., (Chen et al., 2020; Ferrando et al., 2020; Hong et al., 2016; Wang et al., 2021).

Seattle's building tune-ups program has now been in effect for several years and affords an opportunity for a stock-level empirical analysis to address questions such as: What are the expected stock-level energy savings due to implementing a tune-ups program? Are particular types of buildings expected to save more than others? Do particular issues result in higher savings when corrected? Which issues are most important to check for when conducting a tune-up? Does the tune-up specialist impact the savings and issues found? In this paper, we explore the answers to these questions by analyzing Seattle's building tune-up program data to date, and discuss the implications of our findings for other cities considering a building tune-ups program as part of their suite of policies to meet building sector energy and carbon reduction goals. The paper contributes to the literature about the impacts of tune-ups in two ways: First, it assesses the degree to which impacts of a tune-ups program can be discerned at the stock level just based on measured data, without normalizing for changes in building operations before and after the tune-ups. Stated differently, we look at the extent to which the "signal" from the tune-ups is discernible from the "noise" of other drivers of year-to-year variations in building energy use. Second, it seeks to identify any stock-level empirical evidence of relationships between tune-ups issues and building characteristics, which in turn can help streamline policy implementation through better targeting and tailoring the scope of tune-ups. While there has been limited evidence to date of such relationships (City of Portland, 2018), the tune-ups program affords an opportunity to add to the body of evidence with a new dataset.

2. Data

Seattle's tune-up program requires commercial buildings with floor area greater than 50k sqft to be assessed by trained specialists. The specialists recorded building characteristics (e.g., floor area, use type, year built, occupancy) and systems characteristics (e.g., type, age, and condition of HVAC and lighting systems). They also checked for several potential issues with building equipment and operations, and when practical, rectified the issues during the inspection. Other issues were repaired after the inspection. For this analysis, we also utilized energy use data that was measured and collected using ENERGY STAR's Portfolio Manager (ENERGY STAR, n.d.) as part of a benchmarking program that was independent from the tune-up program ("Seattle Energy Benchmarking," n.d.).

2.1. Data preparation

Data preparation included merging data from several different files, removing erroneous data, aggregating space-level or system-level data into building-level data, computing energy use for the pre- and post-tune-up periods, and computing energy savings.

Since energy use data was only available for the whole building, we aggregated space-level and system-level data to building-level. For example, some buildings are partly retail and partly office space, and we assigned a single use type for the whole building. Likewise for system types, ages, conditions, etc. Space-level data was mostly reported along with the floor area of that space, and system-level data was mostly reported along with the floor area served by that system. To assign a value for the whole building, we area-weighted the values for each space or system, and assigned the value that composed at least 80% of the building's total floor area (i.e., we considered a building that was 85% office and 15% retail to be an office). When no single value comprised at least 80% of the total floor area, or when floor area data was not available, we assigned the value "Mixed" or "Other", indicating there was no single dominant value. Ideally, we would weight the values by energy consumption of the space or system, but this information was not part of the dataset. We used floor area because it was the most reasonable proxy for energy consumption that was included in the data.

Energy data was provided as monthly-updated annual totals (e.g., totals for Jan. 2018 to Jan. 2019, Feb. 2018 to Feb. 2019, etc.) for both individual fuels (electric, gas, steam, etc.) and for site and source energy. Weather-normalized site and source energy were also available, with weather-normalization done by ENERGY STAR's Portfolio Manager (ENERGY STAR, 2021).

For the pre-tune-up period, we used the annual energy total from the last year concluding before the tune-up took place.

The full year of post-tune-up period data proved more difficult, due to many buildings drastically altering their operations and/or occupancy due to the COVID-19 pandemic. Since we did not consider pandemic operations to be a fair comparison to a pre-tune-up time period, we required the post-tune-up time period to end before the pandemic started in March of 2020. So, for the post-tune-up period, we primarily used the annual energy total from the first year starting after the tune-up repairs had been finished, but we made some exceptions to be able to include more buildings in the study: If the end of the first post-tune-up year overlapped with the start of the pandemic, we allowed the post-tune-up year to start up to 3 months earlier (i.e., during the time period between when the tune-up was completed and when the building owner had finished addressing the issues found during the tune-up). While this means that the post-tune-up energy data may include some time before the issues had been addressed (likely causing savings to be underestimated), we suspect that many building owners had actually addressed the issues sooner than their post-tune-up paperwork was submitted (meaning the post-period would correctly include only times after the issues were fixed). Despite this leniency with the post-tune-up time period, we had to exclude the majority (roughly 77%) of the buildings from the energy-related portion of the analysis because they did not have enough post-tune-up and pre-pandemic energy data.

We also removed the energy data for a particular building if either of the following two criteria were met: 1) a monthly-updated annual total changed by more than 20% from one month to the next, or if 2) the post-tune-up energy use changed by more than 20%, relative to the pre-tune-up energy use. We considered energy data meeting these criteria to be untrustworthy, either because the energy data was recorded incorrectly, or the building's operations or occupancy levels changed drastically, or some other unexplained reason. The 20% value was chosen based on a combination of engineering judgement and on the shape of the energy use distributions. Roughly 14% of buildings with enough post-tune-up and pre-pandemic energy data had their energy data removed due to these two criteria (roughly half due to being too high and half due to being too low).

2.2. Data summary

The resulting dataset includes 420 buildings with information on building characteristics, systems information, and tune-up inspection results. As described in Section 2.1, only about 80 of those buildings have energy data.

The dataset includes the following building characteristics:

- building use type
- gross floor area
- year built
- percent occupied

Table 1 shows the number of buildings in the dataset with each use type. Roughly 32% of buildings are offices, while each of the other types have no more than 16% of buildings. When considering only the buildings with energy data, the proportions are roughly the same as in

Table 1
Number of buildings, and number of buildings with energy data, of each type.

Building Type	Number of Buildings	Number of Buildings with Energy Data
Office	134	24
Other	69	15
K-12 School	55	29
University	45	0
Hotel	43	6
Warehouse	32	3
Medical Office	20	3
Retail	16	2
Hospital	6	0

Table 1, except that roughly 35% of buildings are schools (instead of 13%), and there are no university buildings (instead of 11%). We do not believe that the differences in school and university proportions are due to systematic reasons.

All of the 420 buildings have gross floor area >50k sqft, 80% have 50k-250k sqft, 15% have 250k-500k sqft, and the remaining 5% have 500k-1.75m sqft. Of the buildings with energy data, the proportions are similar: 64% are 50k-250k sqft, 27% are 250k-500k sqft, and the remaining 9% are 500k-1.63m sqft.

The dataset includes information for each of the following system types:

- lighting
- heating
- cooling
- ventilation
- distribution
- domestic hot water

For lighting, the dataset includes only the lighting type. For each of the other system types, the dataset includes:

- type (e.g., Heat Pump, Packaged VAV, Chiller, District Steam)
- condition (e.g., New, Fair, Poor)
- age (e.g., 0–5 years, 5–10 years)

In the systems type data (lighting type, heating type, domestic hot water type, etc.), roughly 40% of buildings have the "Other" type, mostly because the building had multiple system types and no single type served over 80% of the building's floor area. For system condition and system age, roughly 15% of buildings have the "Mixed" value.

During the tune-up inspection, a tune-up specialist checked for 18 HVAC issues (10 related to operations, and 8 related to maintenance), 4 lighting issues, 2 domestic hot water issues, and 3 envelope issues. The assessment also included 12 issues related to water use, but we do not consider those here. For each of these 27 issues, the dataset includes a flag for whether or not the issue was identified during the inspection, and another flag for whether the issue was fixed (either during the inspection, or afterwards). Table 2 lists each of the 27 issues checked during the tune-up, and the percentage of all 420 buildings in which the issue was found and fixed. The dataset also includes the name and company of the tune-up specialist that conducted the inspection.

The dataset includes one year of pre-tune-up and one year of post-tune-up energy data for the following:

- electricity
- natural gas
- chilled water
- hot water
- steam
- site
- source
- weather-normalized site

Table 2
Brief descriptions of the issues checked for during the tune-ups inspection, along with whether fixing the issue was required or voluntary, the percentage of the 420 buildings in which the issue was found, and the percentage of the 420 buildings in which the issue was fixed.

Issue	Description	Required or Voluntary	Found (%)	Fixed (%)
HVAC 1	Review HVAC equipment schedules.	Required	45	45
HVAC 2	Review HVAC set points.	Required	48	48
HVAC 3	Review reset schedules.	Required	23	23
HVAC 4	Review optimal stop/start capabilities.	Required	17	17
HVAC 5	Verify HVAC sensors are functioning, calibrated, and in appropriate locations.	Required	44	44
HVAC 6	Verify HVAC controls are functioning as intended.	Required	44	43
HVAC 7	Review HVAC controls for unintended or inappropriate instances of simultaneous heating and cooling.	Required	17	17
HVAC 8	Note any indications of significant air-balancing issues.	Voluntary	22	9
HVAC 9	Identify areas with indications that ventilation rates may vary significantly from standards and be inappropriate for current facility requirements.	Voluntary	22	10
HVAC 10	Identify zones that are dominating multi-zone system operations.	Voluntary	10	4
HVAC 11	Verify HVAC equipment (grilles, coils, and ducts) is clean and adequately maintained.	Required	29	29
HVAC 12	Check filters and strainers. Clean or replace filters and strainers where appropriate.	Required	31	30
HVAC 13	Check filters and strainers. Recommend maintenance as appropriate.	Voluntary	29	16
HVAC 14	Verify equipment observed (motors, fans, pumps, belts, pulleys, bearings, and steam traps) is in good working condition. Repair as appropriate.	Required	36	34
HVAC 15	Verify equipment observed (motors, fans, pumps, belts, pulleys, bearings, and steam traps) is in good working condition. Recommend repairs or replacement.	Voluntary	36	18
HVAC 16	If ducts and pipes are visible and accessible, verify HVAC duct and pipe insulation is in place.	Voluntary	31	10
HVAC 17	Check valves and dampers.	Required	40	40
HVAC 18	Identify equipment approaching the end of its service life.	Voluntary	45	8
Light 1	Identify any areas where lighting levels appear to be significantly higher than appropriate for the space use and occupant needs.	Voluntary	21	5
Light 2	Verify lighting sensors are working and located appropriately for the current functioning of the building.	Voluntary	25	10
Light 3	Review lighting controls schedule and sequences.	Required	12	12
Light 4	Identify inefficient lighting equipment (such as incandescent, T12, or metal halide lighting).	Voluntary	49	14
DHW 1	Review domestic hot water temperature set points.	Required	11	11
DHW 2		Required	7	7

(continued on next page)

Table 2 (*continued*)

Issue	Description	Required or Voluntary	Found (%)	Fixed (%)
	Review circulation pump controls.			
Env 1	Assess for roof penetrations and damage to siding.	Voluntary	30	11
Env 2	Identify duct leaks (such as disconnects and/or holes).	Voluntary	12	5
Env 3	Identify any uninsulated attic areas or where attic insulation has been disturbed.	Voluntary	3	0

- weather-normalized source

As described in Section 2.1, due to too little time between the end of the tune-ups and the start of the pandemic, only about 80 of the 420 buildings have a full year of post-tune-up energy data and were included in the energy savings analysis.

Within the energy data, we observed a significant amount of variability in energy use, even before the tune-ups happened. Not only does energy use vary significantly from building to building, but it also varies significantly over time for a particular building. The variation over time for a given building is due to a combination of changes in occupancy and operation (e.g., operating hours, computers) as well as potentially changes in equipment operation (e.g., HVAC or lighting controls). Overall, we observed a gradual downward trend in energy use in the few years leading up to the tune-ups, but for many individual buildings, the trend is either flat or upward. Likewise with the change in energy use for the year just before the tune-up and the year just after.

3. Methods and results

We analyzed the Seattle tune-ups data in order to better understand how effective building tune-ups are as a low-cost option to reduce energy use and, ultimately, to reduce city-wide emissions. In particular, we were interested in the energy savings due to tune-ups, the relationships between energy savings and building characteristics, and the relationships between building characteristics and which issues were found during the tune-ups. For clarity and readability, we have combined the methods and results for each of these research questions into separate subsections.

3.1. How much energy do tune-ups save?

In this section, we consider the energy savings from the pre-tune-up time period to the post-tune-up time period, under the assumption that changes in energy use can be attributed to the effects of the building tune-up. While this might not be strictly true (e.g., occupancy or level of service may have changed for reasons unrelated to the tune-up), we believe it to be a reasonable assumption.

We primarily focused on weather-normalized site energy use because the tune-ups for different buildings were implemented at different times (sometimes in different years), and we wanted to separate the effects of the tune-ups from potential differences due to weather variations from year to year. Recall that all energy savings results are limited to the roughly 80 buildings with sufficient post-tune-up energy data, not the whole dataset of the 420 buildings that implemented tune-ups. Since we have no reason to believe otherwise, we assumed that the buildings with energy data can be considered representative of the whole dataset with respect to energy savings.

Fig. 1 shows a histogram of weather-normalized site energy savings. Median savings are 4.1%, but 34% of buildings had negative savings (i.e., they used more energy after the tune-up than before). We performed a paired *t*-test on the savings with p=0.05 and accepted the hypothesis that savings are 2.1% or greater.

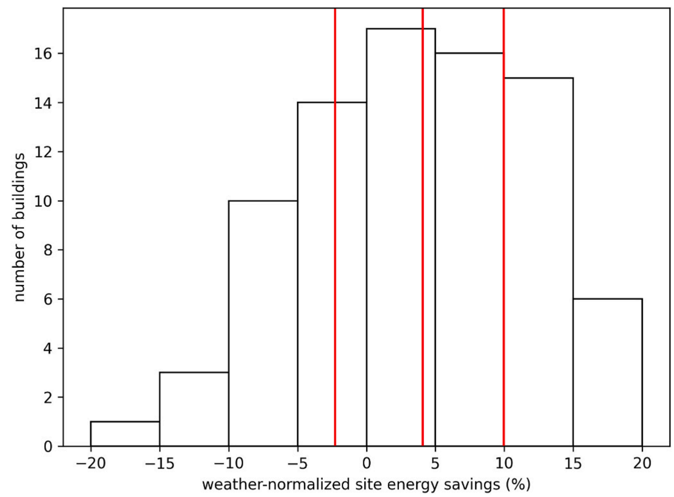

Fig. 1. Histogram of weather-normalized site energy savings percentage. Red vertical lines indicate 25th, 50th, and 75th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Fig. 2 shows similar results for electric energy savings. Median savings are 4.8%, but 20% of buildings had negative savings. With p=0.05, we conclude savings are at least 3.2%. Savings for natural gas were essentially zero (median −0.3%, 52% of buildings with negative savings), and very few buildings used a significant amount of any other fuel. Thus, it appears site energy savings are primarily driven by electricity savings.

Overall, we observed significant variation in savings from building to building. Many buildings have negative savings. Negative savings may be due to buildings increasing occupancy or changing their service levels for reasons unrelated to tune-ups. Negative savings could also be due to tune-ups: systems may have been operating incorrectly and correcting their operation during the tune-up increased energy use. For example, ventilation could be increased to improve indoor air quality, setpoints might have been adjusted, or old lights might have been replaced with LEDs that emit less heat. Despite significant building to building variation, we did find evidence that tune-ups resulted in energy savings. Namely, we would expect at least 2.1% site energy savings with p=0.05, of which nearly all are electricity savings.

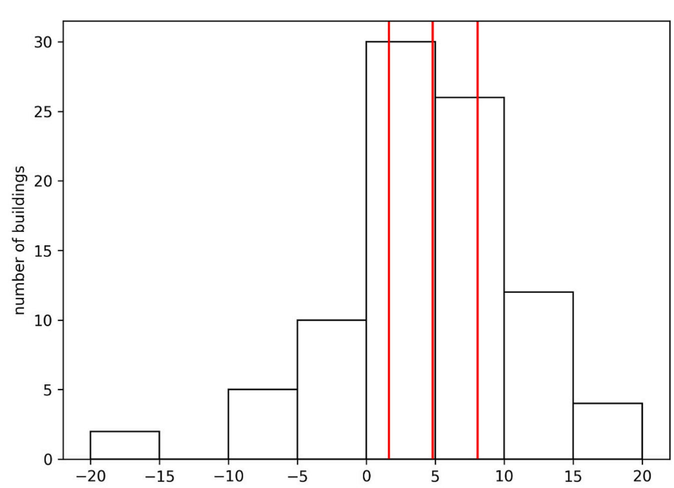

Fig. 2. Histogram of electric energy savings percentage. Red vertical lines indicate the 25th, 50th, and 70th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.2. How do savings relate to building and system characteristics?

In this section, we explore potential relationships between building and systems characteristics and energy savings due to tune-ups. We focused on weather-normalized site energy. We considered all of the building characteristics and system characteristics listed in Section 2.2, and also considered the number of issues fixed due to the tune-up (including total number of issues, number of HVAC issues, number of lighting issues, etc.).

We started by visualizing potential relationships using scatterplots (for numerical variables) and boxplots (for categorical variables). For example, Fig. 3 shows a scatterplot of energy savings against the year in which the building was constructed. One might believe that an older building would have more problems with older equipment, and thus more opportunities for energy savings, but we found no clear relationship. There is significant scatter and the correlation coefficient is low.

Instead, one might believe that energy savings potential is related to the age of the systems in the building, and not the age of the building itself. Fig. 4 shows boxplots of energy savings for buildings with each age of cooling system. Median savings tend to increase as age increases (except for the >51 years category, which has only two buildings). However, there is substantial scatter and significant overlap between subsequent boxplots. Similarly, Fig. 5 shows boxplots of energy savings for buildings with each age of ventilation system. The relationship between median savings and age is less clear, and again there is significant overlap.

In addition to visual inspection of relationships, we fit roughly 20 linear regression models to energy savings as a function of each of the building and system characteristics in Section 2.2. For each model, we tested for statistical significance of the regression coefficients (with p=0.05 confidence). We found very few significant coefficients, and those that were significant were for coefficients for very few buildings. For example, in the model of energy savings as a function of building type, the coefficient for warehouses is statistically significant, but the dataset only includes 3 warehouses with energy data.

We also tested models using combinations of multiple predictors (e.g., energy savings as a function of both building type and year built). The only significant relationships we found were already found in the corresponding single variable models (e.g., energy savings as a function of building type), and all of those relationships involve very few buildings.

In addition, we tested all of the same regression models that we fit to the whole dataset (roughly 80 buildings) against only the portion of the dataset corresponding to office buildings (roughly 24 buildings). Since office buildings likely have less variation in operating characteristics than other building types, we thought the relationships might be clearer within this subset of buildings. However, out of the roughly 20 models, we found no statistically significant relationships between energy savings and building or system characteristics.

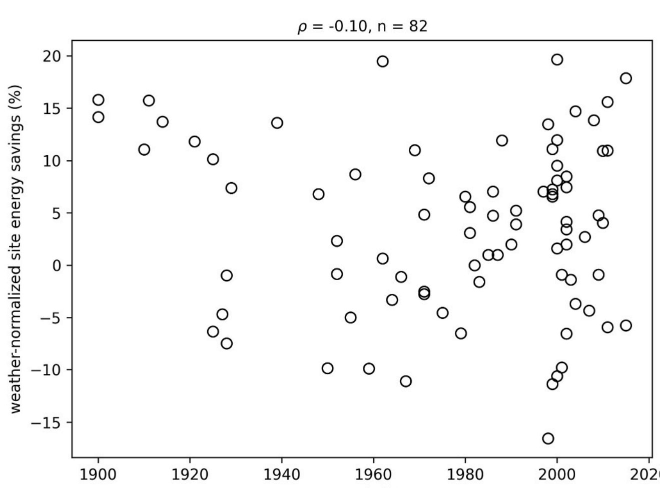

Fig. 3. Scatterplot of weather-normalized site energy savings percentage vs. year built, along with correlation coefficient and number of data points.

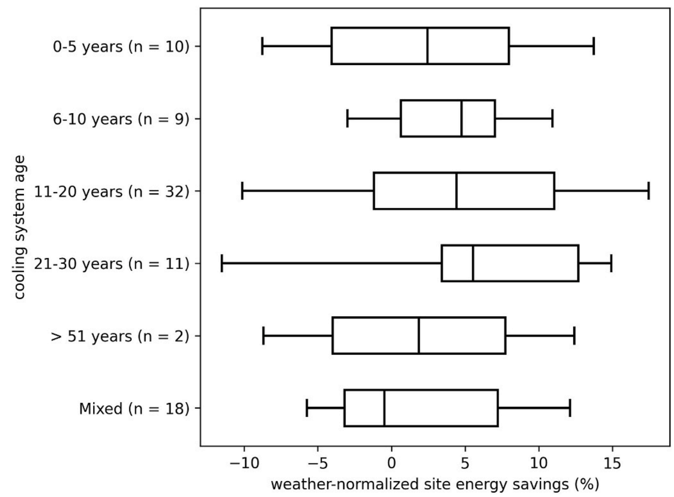

Fig. 4. Boxplots of weather-normalized site energy savings percentage for buildings with each age of cooling system. Values in parentheses are the number of buildings with that age. Boxplots show 5th, 25th, 50th, 75th, and 95th percentiles.

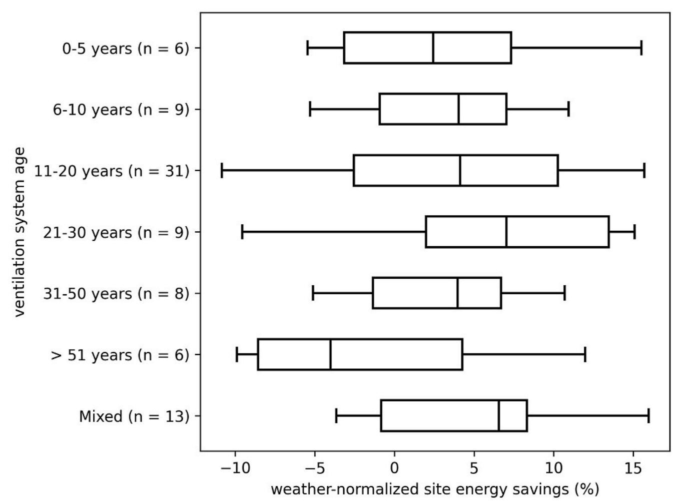

Fig. 5. Boxplots of weather-normalized site energy savings percentage for buildings with each age of ventilation system. Values in parentheses are the number of buildings with that age. Boxplots show 5th, 25th, 50th, 75th, and 95th percentiles.

In summary, we found no evidence that the building or system characteristics included in this dataset could be used to reliably predict energy savings due to tune-ups. We suspect this is due to a combination of factors: the dataset is relatively small (roughly 80 buildings), and savings are both relatively small (median site energy savings were 4.1%) and highly variable (as seen in Section 3.1). In short, there is likely too little “signal” and too much “noise".

3.3. How do tune-ups issues relate to building and system characteristics and tune-up specialists?

In this section, we seek to understand how building and system characteristics relate to the issues found during the tune-ups. From a policy standpoint, the motivation for analyzing this is to determine if the scope of tune-ups can be tailored based on building and system characteristics, toward making program implementation more efficient and reducing the burden for tune-up specialists and building owners.

Here, we consider the full dataset of 420 buildings, not only the roughly 80 buildings with energy data. We started by considering the number of issues found during the tune-up. Fig. 6 shows boxplots of the number of issues found for buildings with different conditions of cooling system. Intuitively, the median number of issues tends to increase as the condition of the cooling system decreases. However, there is significant scatter and overlap between the boxplots. Similarly, Fig. 7 shows boxplots of the number of issues found for buildings with different conditions of heating system. Again, the median number of issues tends to increase as the condition decreases, and there is significant scatter.

We fit several linear regressions models to the number of issues found as a function of each of the building and system characteristics listed in Section 2.2. We also fit all the same models to the number of HVAC issues found, number of lighting issues found, etc. (see Table 2). Out of the roughly 20 models (corresponding to over 100 coefficients), we found roughly 10 regression coefficients that were statistically significant (with p=0.05). For example, we found that having a 100% outside air ventilation system is associated with 1.83 fewer issues being found, while having natural ventilation instead is associated with 3.60 fewer issues found. The large majority of the effects are intuitive (e.g., systems being in worse condition and being older are associated with more issues found). However, we found that the magnitude of the coefficients is relatively small. Other than the 3.60 coefficient for natural ventilation, all coefficients have magnitude less than 3. So, while building and system characteristics may affect how many problems a building may have, the effects are fairly small.

Next, we explored the relationship between particular building or system characteristics and whether or not individual issues were found during the tune-up. For each of the 27 issues in Table 2, and for each of the building and system characteristics listed in Section 2.2 (plus the name and company of the tune-up inspector), we fit a logistic regression model to a binary indicator (i.e., 1 = issue found, 0 = not found) as a function of the building or system characteristic. In total, we fit over 500 models (corresponding to over 2500 model coefficients) and found 286 statistically significant (p=0.05) model coefficients.

Roughly a third of the cases involve coefficients corresponding to one system type being related to the probability of finding issues with a different system type. For example, in the model that predicts the probability of finding HVAC issue 8 (related to air balancing issues)

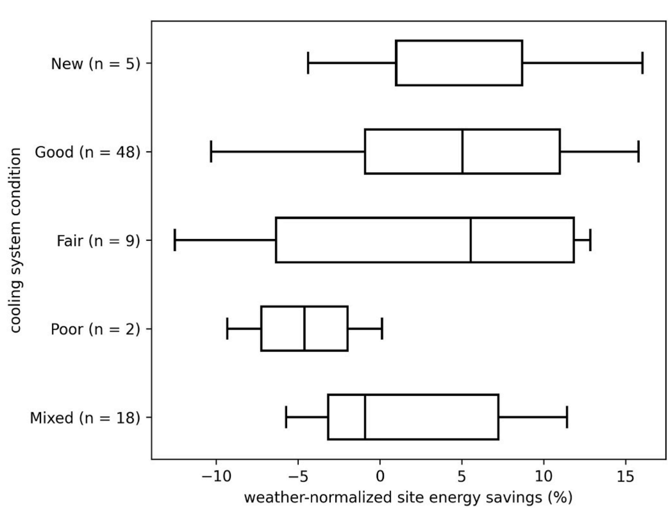

Fig. 6. Boxplots of the number of issues found during the tune-up for buildings with each condition of the cooling system. Values in parentheses are the number of buildings with that condition. Boxplots show 5th, 25th, 50th, 75th, and 95th percentiles.

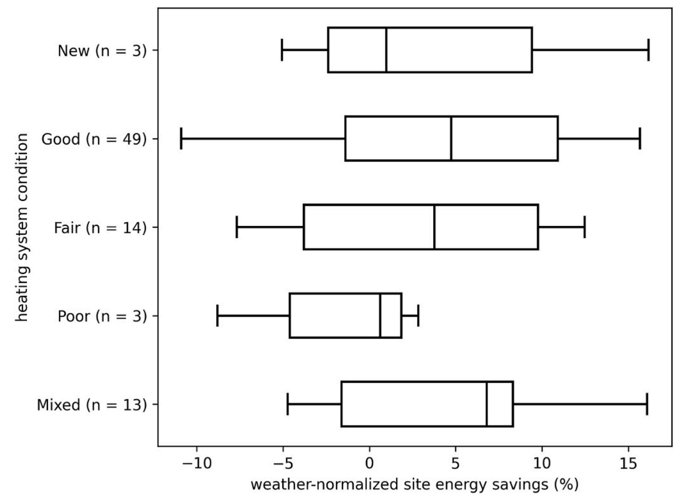

Fig. 7. Boxplots of the number of issues found during the tune-up for buildings with each condition of the heating system. Values in parentheses are the number of buildings with that condition. Boxplots show 5th, 25th, 50th, 75th, and 95th percentiles.

using the type of lighting system, the coefficients for three different light types are significant with $p < 0.05$. However, there is no intuitive explanation for lighting type affecting air balancing issues. Most likely, there is some other factor affecting the probability of finding the issue, and the factor is correlated with lighting type. We consider these cases to be statistical anomalies and not worth consideration.

We also excluded from consideration the many cases in which the model prediction for the value of the coefficient is not sufficiently large (probability>=0.8) or sufficiently small (probability<=0.2). In these cases, the coefficient can be interpreted as the building or system characteristic having a relatively weak effect on whether or not the issue is found. While these results may be interesting in some cases, we think that in general, they have limited utility.

We found 8 cases where the model predicts a probability>=0.8 of the issue being found for buildings with a particular characteristic. As seen in Table 3, all 8 of these cases involve HVAC-related issues. One of

Table 3
The 8 cases where a model was found that predicts a probability>=0.8 that an issue is found for buildings with a particular characteristic. The tune-up specialist names and firms are anonymized for privacy.

Issue	Description	Probability	Characteristic
HVAC 11	Verify HVAC equipment (grilles, coils, and ducts) is clean and adequately maintained.	0.91	specialist firm = Q
HVAC 11	Verify HVAC equipment (grilles, coils, and ducts) is clean and adequately maintained.	0.90	specialist name = F
HVAC 6	Verify HVAC controls are functioning as intended.	0.87	specialist name = W
HVAC 17	Check valves and dampers.	0.83	specialist firm = S
HVAC 18	Identify equipment approaching the end of its service life.	0.82	distribution system age = 21–30 years
HVAC 5	Verify HVAC sensors are functioning, calibrated, and in appropriate locations.	0.80	specialist firm = C
HVAC 14	Verify equipment observed (motors, fans, pumps, belts, pulleys, bearings, and steam traps) is in good working condition. Repair as appropriate. Implementation is required.	0.80	specialist name = Y
HVAC 17	Check valves and dampers.	0.80	specialist name = Y

the cases involves the model relating the age of the distribution system to the probability of finding HVAC issue 18 (identifying equipment near end of service life). The model predicts a probability of 0.82 that HVAC issue 18 will be found during a tune-up if the building has a 21–30 year old distribution system. In the other 7 cases, the predictor is either the name of the tune-up specialist or the specialist's company (i.e., a couple of inspectors and companies are very likely to find some particular HVAC issues during a tune-up).

We found 77 cases where the model predicts a probability<=0.2 of the issue being found for building with a particular characteristic. One fifth of these cases are related to a particular tune-up specialist or company (i.e., specialists or companies that are very unlikely to find particular issues). The majority of the remaining cases involve non-intuitive results, but some made sense. For example:

- buildings with new heating, ventilation, or distribution systems are very unlikely to have HVAC issue 18 (equipment near end of service life), and
- buildings with no cooling system are very unlikely to have HVAC issue 13 (filters and strainers) or HVAC issue 16 (duct and pipe insulation).

4. Conclusions and policy implications

4.1. Summary of findings

In this work, we summarized our findings from an analysis of the results of a tune-ups requirement implemented by the City of Seattle. Our goal was to better understand the energy savings resulting from the tune-ups program, which issues were found and fixed during the tune-ups, and the relationship of savings and issues to building and system characteristics. We analyzed building level changes in energy use before and after the tune-ups were implemented, using weather-normalized whole building energy use. This allowed us to look at the extent to which the "signal" from the tune-ups is discernible from the "noise" of other drivers of year-to-year variations in building energy use at the stock level.

First, we explored the energy savings due to implementing the tune-ups program. There is significant building to building variation in energy savings, and lockdowns due to the pandemic limited the dataset to only about 80 buildings with energy data with normal pre-pandemic occupancy. Despite a fairly small and noisy dataset, we found weather-normalized median site energy savings of 4%, and that essentially all of those savings are due to electricity savings (as opposed to natural gas or other fuels).

Next, we explored relationships between energy savings and building and system characteristics. We found no evidence that the building or system characteristics could reliably predict energy savings.

Finally, we explored relationships between tune-ups findings and building and system characteristics. Namely, we looked at the quantity of issues found during the tune-up, and we looked at whether or not each particular issue was found. We found a handful of building and system characteristics that are associated with more issues being found, but the effects were fairly small (i.e., the number of issues found would not change by more than 3). We found several relationships between building or system characteristics and whether or not particular issues were found. The large majority of these are either suspiciously non-intuitive or are of limited utility (i.e., probability of finding issue between 0.2 and 0.8). Most of the remaining results are either trivial (e.g., buildings without cooling are less likely to have HVAC issues), or are related to the individual or company performing the tune-up inspection.

4.2. Policy implications

4.2.1. Role of tune-ups programs for city-level climate goals

In order to meet climate goals, cities will likely need more energy savings than tune-ups can provide, but they might still implement tune-ups as one low-cost component of a multi-faceted approach to reducing energy use and emissions. In particular, tune-ups can be one component of a prescriptive pathway to meet building performance standards (BPS) that are being adopted across many US cities (IMT, n.d.). Individual building savings can vary widely. Although tune-ups can save significant energy in some buildings, stock-level savings of a broadly applied tune-ups program will generally be lower because other buildings may have low or negative savings (e.g., if the tune-up results in fixes that increase energy use). It is possible that a tune-up policy that is coupled with an energy performance requirement (such as a BPS) might result in more buildings with increased savings, since such a requirement might encourage building owners and tune-up specialists to get more out of the tune-up. It should also be noted that tune-ups can help ensure savings persistence and maintain building performance for energy efficiency measures implemented in other programs. Additionally, some of the tune-up measures can improve indoor air quality and thermal comfort, thereby providing a co-benefit for owners and occupants.

4.2.2. Scope and targeting of tune-up programs

Based on the results, specifically the lack of statistically significant relationships between tune-ups savings and building characteristics, we see no obvious reasons why the scope of a tune-ups program should be limited to particular types of buildings, ages of buildings and systems, or buildings with particular systems installed. However, since energy and emissions are usually closely correlated, cities with a cleaner electric grid may choose to focus on tune-up measures that target on-site fossil fuels.

In a similar vein, these results do not indicate there is any reason to tailor the scope of tune-ups to only checking for particular issues or types of issues based on building characteristics. They do suggest that the selection of the companies and individuals performing the tune-ups can impact the identification of issues.

4.3. Limitations and future work

As with all empirical analyses, the value and validity of the results are constrained by the quantity, quality and characteristics of the dataset. Fortunately, Seattle has a very well-structured data collection system which significantly limited any data quality issues. In particular their system of coding and classifying building system characteristics, as well as tune-ups issues identified and fixed is especially notable and replicable for other cities considering such a program. While we had a reasonably large dataset of 420 buildings, only 80 of them had a whole year of post-tune-up and pre-pandemic energy data. This limitation is further exacerbated by the heterogeneity of occupancy and system characteristics found in buildings. A larger dataset would allow more fine-grained analysis to control for various building characteristics. We would note that tune-up savings may sometimes be masked by changes in building occupancy (e.g., more tenants or computer loads), operation (e.g., longer operating hours), and service levels (e.g., more ventilation). In the future, it would be helpful to collect such information to be able to control for these factors in the savings analysis. Likewise, data on lighting and HVAC energy end-use may help provide additional insights on savings. Indeed, we suspect that more definitive relationships could be found (particularly those related to energy savings) with a larger dataset that also allows controlling for changes in building occupancy, operation and service levels.

Declaration of competing interest

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

Acknowledgments

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

We thank Sandra Mallory, Terry Sullivan, and Nicole Ballinger with the City of Seattle for collaborating on this effort and providing the tune-ups dataset.

References

Ashrae, 2013. Standard 202-2013-Commissioning Process for Buildings and Systems.
Ballinger, N., 2020. Building Tune-Up Accelerator Program. City of Seattle.
Bergfeld, K., Mathew, P., Duer-Balkind, M., Perakis, J., Noori khah, P., Walter, T., Held, A., 2020. Making data-driven policy decisions for the nation's first building energy performance standards. In: Presented at the Summer Study on Energy Efficiency in Buildings, American Council for an Energy Efficient Economy. https://doi.org/10.20357/B7831V.
Brøgger, M., Wittchen, K.B., 2018. Estimating the energy-saving potential in national building stocks – a methodology review. Renew. Sustain. Energy Rev. 82, 1489–1496. https://doi.org/10.1016/j.rser.2017.05.239.
Cee, n.d. Commercial whole building performance committee program summary [WWW document]. https://library.cee1.org/content/commercial-whole-building-performance-committee-program-summary/ accessed 11.5.21.
Chen, Y., Deng, Z., Hong, T., 2020. Automatic and rapid calibration of urban building energy models by learning from energy performance database. Appl. Energy 277, 115584. https://doi.org/10.1016/j.apenergy.2020.115584.
City of Portland, 2018. Policy, permit, perform. Using city benchmarking data and building construction permit history to identify energy performance improvements. URL. https://www.energy.gov/sites/default/files/2018/11/f57/Cities-LEAP%20Portland%20Final%20Report.pdf. accessed 4.15.22.
Crowe, E., Mills, E., Poeling, T., Curtin, C., Bjørnskov, D., Fischer, L., Granderson, J., 2020. Building commissioning costs and savings across three decades and 1500 North American buildings. Energy Build. 227, 110408. https://doi.org/10.1016/j.enbuild.2020.110408.
Energy Star, 2021. Portfolio manager technical reference [WWW document]. https://www.energystar.gov/buildings/tools-and-resources/portfolio-manager-technical-reference-climate-and-weather accessed 4.6.22.
Energy Star. n.d. ENERGY STAR Portfolio manager [WWW document]. https://www.energystar.gov/buildings/benchmark accessed 11.5.21.
Fernandez, N., Katipamula, S., Wang, W., Huang, Y., Liu, G., 2015. Energy savings modelling of re-tuning energy conservation measures in large office buildings. J. Build. Perform. Simulat. 8, 391–407. https://doi.org/10.1080/19401493.2014.961032.
Fernandez, N., Katipamula, S., Wang, W., Xie, Y., 2017. Energy Savings and Peak Load Reduction Benefits from Building Controls Measures in Seattle. Pacific Northwest National Laboratory, Washington.
Ferrando, M., Causone, F., Hong, T., Chen, Y., 2020. Urban building energy modeling (UBEM) tools: a state-of-the-art review of bottom-up physics-based approaches. Sustain. Cities Soc. 62, 102408. https://doi.org/10.1016/j.scs.2020.102408.
Hong, T., Chen, Y., Lee, S.H., Piette, M., 2016. CityBES: a web-based platform to support city-scale building energy efficiency. In: Presented at the 5th International Urban Computing Workshop.
Imt, n.d.. Building performance standards. Institute for market transformation. URL. https://www.imt.org/public-policy/building-performance-standards/. accessed 4.15.22.
Katipamula, S., 2016. Improving Commercial Building Operations Thru Building Re-tuning: Meta-Analysis. Pacific Northwest National Laboratory.
Kontokosta, C.E., Spiegel-Feld, D., Papadopoulos, S., 2020. The impact of mandatory energy audits on building energy use. Nat. Energy 5, 309–316. https://doi.org/10.1038/s41560-020-0589-6.
Langevin, J., Harris, C.B., Reyna, J.L., 2019. Assessing the potential to reduce U.S. Building CO2 emissions 80% by 2050. Joule 3, 2403–2424. https://doi.org/10.1016/j.joule.2019.07.013.
Langevin, J., Reyna, J.L., Ebrahimigharehbaghi, S., Sandberg, N., Fennell, P., Nägeli, C., Laverge, J., Delghust, M., Mata, É., Van Hove, M., Webster, J., Federico, F., Jakob, M., Camarasa, C., 2020. Developing a common approach for classifying building stock energy models. Renew. Sustain. Energy Rev. 133, 110276. https://doi.org/10.1016/j.rser.2020.110276.
Mills, E., 2011. Building commissioning: a golden opportunity for reducing energy costs and greenhouse gas emissions in the United States. Energy Effic. 4, 145–173. https://doi.org/10.1007/s12053-011-9116-8.
Mills, E., Friedman, H., Powell, T., Bourassa, N., Claridge, D., Haasl, T., Piette, M.A., 2004. The Cost-Effectiveness of Commercial Buildings Commissioning: A Meta-Analysis of Existing Buildings and New Construction in the United States. Lawrence Berkeley National Laboratory.
Mills, E., Mathew, P.A., 2014. Monitoring-based commissioning: benchmarking analysis of 24 university buildings in California. Energy Eng. J. Assoc. Energy Eng. 111, 7–24. https://doi.org/10.1080/01998595.2014.10844605.
Seattle Building Tune-Ups [WWW Document]. n.d. http://www.seattle.gov/environment/climate-change/buildings-and-energy/building-tune-ups accessed 9.16.20.
Seattle Office of Sustainability and Environment, 2013. Seattle Climate Action Plan. https://www.seattle.gov/documents/Departments/Environment/ClimateChange/2013_CAP_20130612. (Accessed 16 September 2020).
Seattle Energy Benchmarking [WWW Document]. n.d. https://www.seattle.gov/environment/climate-change/buildings-and-energy/energy-benchmarking accessed 9.16.20.
Sullivan, T., Baker, R., Ringeisen, B., 2020. Scaling commercial building O+M – initial results from mandatory building tune-ups in Seattle. In: Presented at the Summer Study on Energy Efficiency in Buildings, American Council for an Energy Efficient Economy.
n.d Usgbc. LEED rating system | U.S. Green building council [WWW document]. https://www.usgbc.org/leed. accessed 11.5.21.
Walter, T., Mathew, P., 2021. GHG policy impacts for Seattle's buildings: targets, timing, and scope. Build. Cities. 2, 283–301. https://doi.org/10.5334/bc.81.
Wang, Z., Hong, T., Li, H., Ann Piette, M., 2021. Predicting city-scale daily electricity consumption using data-driven models. Adv. Appl. Energy 2 (100025). https://doi.org/10.1016/j.adapen.2021.100025.