Towards Equity in Energy Efficiency Analyses

John Wamburu, Emma Grazier, David Irwin, Christine Crago and Prashant Shenoy
University of Massachusetts Amherst
{jwamburu,egrazier,irwin,ccrago,shenoy}@umass.edu

ABSTRACT

The electric grid has begun a profound transition from primarily using carbon-intensive energy to instead using carbon-free renewable energy. In parallel, smart meters and other sensors are now providing us unparalleled visibility into the energy-efficiency of building and grid operations. Researchers are actively using building and grid energy data from these sensors to develop analytics techniques, e.g., using machine learning, that can improve energy-efficiency and facilitate the energy transition. Unfortunately, much of this research ignores the impact of these analytics on *equity*. That is, while current data analytics techniques may accurately identify energy-inefficiencies, they generally do not contextualize the underlying reasons for these inefficiencies. For example, data analytics that identify the most energy-inefficient homes might motivate new programs that target these homes for subsidies to improve energy-efficiency. However, the most energy-inefficient homes might also correlate with those with the highest income that have less need for subsidies, and engage in the most energy wasteful behavior. In contrast, the most energy-efficient homes might be the homes that can least afford to waste (or even use) energy. In this paper, we use an example from recent research to illustrate the inequity of state-of-the-art energy analytics, and argue that energy analytics research should elevate equity to a first-class concern.

CCS CONCEPTS

• General and reference → Empirical studies; **• Information systems → Data analytics**; **• Hardware → Impact on the environment**; **Energy metering**.

KEYWORDS

Equity, Energy Efficiency Analysis, Decarbonization

ACM Reference Format:
John Wamburu, Emma Grazier, David Irwin, Christine Crago and Prashant Shenoy . 2021. Towards Equity in Energy Efficiency Analyses. In *The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '21), November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3486611.3492411

BuildSys '21, November 17–18, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9114-6/21/11…$15.00
https://doi.org/10.1145/3486611.3492411

1 INTRODUCTION

Buildings consume nearly 40% of the total energy consumption and 70% of the total electricity in many countries. They contributed over 1850 million metric tons of greenhouse gases in 2019 [1]. As the grid makes a profound energy transition towards a carbon-free future, improving the energy efficiency and carbon footprint of the buildings sector will play an important role in meeting our society's sustainability goals. Building energy efficiency has been an active area of research in recent years, and approaches that use data-driven and machine learning techniques for energy efficiency are increasingly commonplace [3, 4, 6, 7].

While newly built buildings can be made zero carbon by design, the bigger challenge lies in making existing, and particularly older, buildings energy efficient in order to reduce their carbon footprint. This is typically achieved through retrofits such as installing better insulation in the building envelope, or replacing the building's HVAC system. In many parts of North America and Europe, building heating is achieved through the use of fossil fuels such as natural gas, oil or propane heating. Replacing older HVAC equipment with energy-efficient heat pumps is a promising approach to not only enhance efficiency but also substantially lower the building's carbon footprint — assuming that supplied electricity is less carbon intensive than the carbon-intensive heating fuel it replaces. As the percentage of green energy (e.g. from renewables) in the grid rises, it lowers the carbon footprint of the building's heating systems.

As we move towards decarbonization of the grid and energy system, it is imperative to not only consider how buildings can reduce their energy and carbon footprint using technical approaches, but to also factor social equity into this process. Consider, for instance, data driven methods that have been increasingly used to optimize building energy efficiency and detect inefficiencies in building energy usage [4, 7, 8, 11]. For example, machine learning has been used to identify inefficient residential buildings from a cohort of buildings in a city, and pinpoint the underlying cause of inefficiency [7].

In this paper, we argue that data driven approaches for building energy efficiency may have inherent biases that prevent them from producing equitable results. For example, approaches such as [7] that choose highest energy consuming buildings from a community may have a bias towards choosing larger residential homes due to their higher energy footprint. However, such larger homes may belong to higher income residents, which inadvertently biases the technique and may prevent middle and lower income homes from sharing the benefits of decarbonization schemes (which include generous subsidies for retrofits).

In this paper, we conduct an experimental study to identify the degree to which such biases are present in machine learning based analytic methods. We consider the WattHome approach from [7] as a representative example of ML-based analytic approaches and use

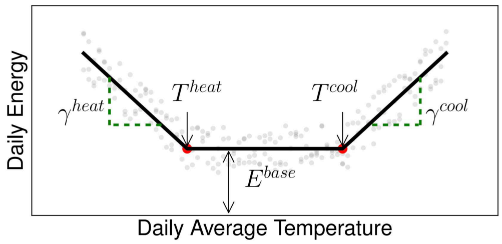


Figure 1: Energy model for a building showing breakpoint temperatures T^{heat}, T^{cool} and average energy usage E^{base}.

it to conduct our equity analysis. Our results show that purely technical analyses can cause skewed towards certain demographics by data bias present in the underlying datasets. Based on these insights, we argue for design of equitable and fair analytic approaches to ensure that benefits of energy improvements and decarbonization schemes are seen equitably across our society.

2 BACKGROUND

Energy analytics for buildings has emerged as an important tool for decarbonization and the energy transition. In particular, the use of AI, machine learning and data-driven techniques to identify energy inefficiencies in buildings and recommend improvement measures has risen in popularity in recent years. For example, [11] propose an outlier based technique that detects abnormal day to day energy consumption in buildings. Other studies using similar techniques have also been conducted [7, 8].

The goal of our work is to understand whether these analytical techniques are equitable, and whether they make improvement recommendations fairly across a population. We study this question through the lens of WattHome, a recently proposed analytical technique for energy efficiency analysis [7].

WattHome is a data-driven approach that uses Bayesian inference to analyze the energy efficiency of a building and detect possible faults. It does so in three key steps; (i) Learning a building's energy model using energy usage data, (ii) Creating a partial order of buildings using their parameter distribution from the learned model, and (iii) Detecting building faults that might be the underlying cause of energy inefficiency.

Figure 1 depicts the energy model learned by WattHome for each building. During winter months, the outside temperature is colder than a building's inside temperature, which results in a net thermal loss where the inside heat flows outside causing the inside temperature to drop. Each building therefore has a specific temperature T^{heat} below which there is need for a heater to heat the building. Conversely, during summer months, the outside temperature is warmer than the inside temperature. The outside heat flows into the building causing the temperature to rise, and the building experiences a net heat gain. Similarly, a specific temperature T^{cool} therefore exists for each building above which there is need for an air conditioner to cool the building. Finally, a specific temperature T_b exists, during which there is neither thermal loss nor thermal gain. During this time, a buildings energy consumption comes from inside appliances.

Following this model, an analysis of the heating slope i.e. γ^{heat} and cooling slope i.e. γ^{cool} can be used to determine whether a building is energy efficient or not. Buildings with a high γ^{heat} lose heat at a higher rate, hence the need for more energy to compensate for the high heat loss. Similarly, buildings with a high γ^{cool} absorb outside heat at a higher rate, hence the need for more energy to cool the inside of the building. WattHome uses Bayesian inference to learn a parameter distribution of the heating slope and uses stochastic dominance to identify inefficient buildings. Buildings whose learned parameters are higher than 75% of the population are singled out as having energy inefficiency of some sort.

3 EXPERIMENTAL METHODOLOGY

The methodology for our work comprises using WattHome [7] to evaluate the energy efficiency of a group of residential buildings to identify the least efficient homes as candidates for energy efficiency improvements. We then analyze whether the choices made by the algorithm are fair and equitable. To do so, we use the following datasets.

Energy usage data. This dataset is collected from smart meters installed at homes in a city in the New England region during the year 2019. The data contains electricity usage data recorded at 5 minute granularity, as well as gas data recorded at hourly granularity. The data is anonymized, and only includes a mapping of electric and gas meters by home but not the specific address. In addition to energy usage data, this dataset also includes building information such as age, size, and type of home e.g. single family vs multi-family. Before comparing energy usage across different homes, energy usage is first normalized with size of the home.

Demographic data. We collected demographic data for this city including race, median household income and house value from the Geocodio API [1]. The data is provided by the Census Bureau and is available by census block i.e. the smallest geographical unit that the U.S. Census Bureau provides statistical data for. We map each home in our dataset to its specific census block and use the demographic information for that block to perform our equity analyses.

Weather data. We use weather data for the year 2019 gathered from the Dark Sky API [2] at hourly granularity.

We apply WattHome to these datasets and use an outlier based technique to select homes as candidates for improvements. Our broader goal is to analyze how equitable such an approach is in making recommendations in such a community.

4 EQUITY ANALYSIS

To analyze whether WattHome's algorithms makes equitable choices, we apply WattHome to our energy usage data and combine the results with demographic data from the same city to study the extent of bias present in the approach.

We begin by selecting only a subset of the whole dataset which have gas heating. This subset is made up of 6,368 homes. Since gas heating is a high emitting source of heating, this subset represents a good target for decarbonization goals. For each home, we compute γ^{heat}, T^{heat}, γ^{cool}, T^{cool} and E^{base}. We then determine the median income for each home by mapping the home to its parent census block and using the median income of the block as the income in the home. We do the same for house value and consider the median

[1]https://www.geocod.io/docs/
[2]https://darksky.net/dev

house value for each block as the value for each home in the block. We then apply an outlier based technique to identify homes with energy inefficiency – for each parameter, we select the top 10% homes as an indicator of energy inefficiency i.e. the top 10% of homes with the highest heating slope are considered as having heating inefficiency. Finally, for the identified homes, we analyze the distributions of median income and real-estate house values to determine whether the results are biased in any way by these factors.

Figures 2a, 2b and 2c depict learned energy models for sample homes in our dataset. Figure 2a depicts a home with a high heating and cooling slopes γ^{heat} and γ^{cool}, indicating a home with high heating energy usage as well as an AC unit that is in use during summer months. Figure 2b depicts a home with a high heating slope but lower cooling slope during summer months indicating an AC unit which is conservatively kept running hence the lower cooling energy requirement. Figure 2c depicts a home without an AC unit as evidenced by the low energy requirement on high temperature days. Each of these homes present different profiles that must be taken into consideration while designing equity goals. For instance, a home with a low cooling slope might be interpreted as being energy efficient – however, at the same time, it could also mean that the home lacks funds to purchase a cooling unit, and such insights are useful while designing energy efficiency improvement measures.

Figure 3 shows the results of analyzing houses with the lowest cooling slope against household income. For this analysis, we first divide household income into deciles. For each decile, we count the number of homes that fall in the top 10% of homes with the lowest cooling slope. The figure shows that homes with the lowest cooling slope are predominantly lower income homes with the first decile comprising of more that 42% of all homes in the top 10%. An outlier based technique used on this data would therefore flag these homes out as being energy efficient. However, since they fall in the lower income decile, it is also possible that these homes cannot afford an AC unit, and therefore do not consume cooling energy during summer months. At the same time, homes in the lower income deciles tend to have poor insulation and building envelope, and such an outlier based technique would leave these homes out when choosing homes that would benefit from a retrofit program. It is clear that this technique is biased against lower income neighborhoods, and not equitable in its choices.

Figure 4 shows the results of analyzing homes with the highest cooling slope against the value of the house. Here, we begin by computing house value deciles. For each decile, we count the number of homes that fall in the top 10% of homes with the highest cooling slope. A high cooling slope could be indicative of a home with a poor building envelope. A home with a poor building envelope absorbs outside heat at a high rate, hence the high energy requirement to rapidly cool the house. In a purely technical analysis, such buildings could be selected as candidates for insulation retrofits. Figure 4 shows that homes with the highest cooling slope are predominantly in the higher home value deciles. The 7th decile accounts for more than 24% of all homes in the top 10%. Conversely, the lowest three deciles account for 4%, 8% and 4% of homes in the top 10% respectively. An outlier based technique used on this data would therefore disproportionately select higher value homes as candidates for efficiency improvement, leaving out lower value homes from accessing such benefits. However, lower value homes are more likely to have poor building envelope, and should be selected first for such improvements. It is therefore clear that this analysis is biased against lower value homes neighborhoods, and equity should be a key design goal for such a technique.

Figure 5 shows the results of analyzing homes with the highest heating slope against the household income. A high heating slope can be interpreted as a home with a poor building envelope. A home with a poor building envelope loses heat to the outside at a high rate, hence the high energy requirement to continue heating the inside of the house. Such buildings could be selected as candidates for insulation retrofits reduce the thermal loss. Such homes could also be selected for replacing gas heating (which is present in all these homes) with greener sources of heating energy such as electric heat pumps. Figure 5 shows that homes with the highest heating slope are predominantly in the higher income deciles. The 7th decile accounts for more than 25% of all homes in the top 10%, which is ≈5x the number of homes whose income level falls in the 2^{nd} decile. These homes in higher income deciles would therefore be selected as candidates for efficiency improvement (and potential recipients of subsidies), squeezing out financial incentives that can instead be offered to lower income households.

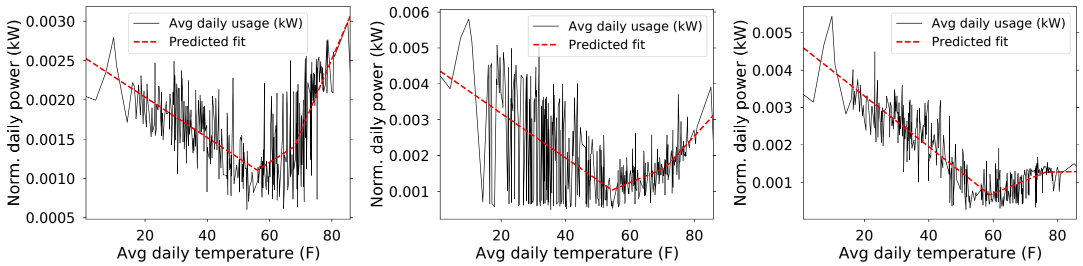
5 INCORPORATING EQUITY

In this section, we discuss techniques that can mitigate the bias that comes from performing purely technical analyses.

In WattHome, homes were collected into peer groups based on building attributes such as age, type of home and the area covered by a home. This was done to make the comparison of energy parameters more meaningful e.g., newer houses are built while adhering to new design specifications, and should therefore be compared to other newer homes. However, building attributes are not the only cause of inequity in energy analysis. The demographics of the household, especially income, are an equally important component in explaining the home's heating and cooling intensities. One technique to mitigate this effect is to create peer groups based on economic factors such as income, value of the home and other demographic factors. Comparisons should then be made within these peer groups to identify inefficient homes and generate recommendations for energy improvements.

Another technique would be to collect homes based on geographical census tracts and blocks. Typically, homes in the same geographical block tend to have similar economic and demographic profiles. Using this approach, a simple heuristic such as selecting at least n homes from each peer group as candidates for energy efficiency improvements would ensure that such improvements are uniformly distributed across the whole geographical region. This would also ensure that selected homes are equitably distributed across the multiple demographic profiles associated with each block.

As future work, we will devise a comprehensive approach with equity as a key design goal. In this approach, household metrics such as a census block group's median household income, racial and ethnic composition, level of energy poverty, etc., will be considered from the onset of energy efficiency analysis. All recommendations



(a) (b) (c)

Figure 2: (2a) Sample energy model for a home with varying heating and cooling intensities, (2b) Sample energy model for a particular home with higher heating energy requirements than (2a), and (2c) Sample energy model for a building with no cooling load which can be interpreted as a house lacking an AC unit.

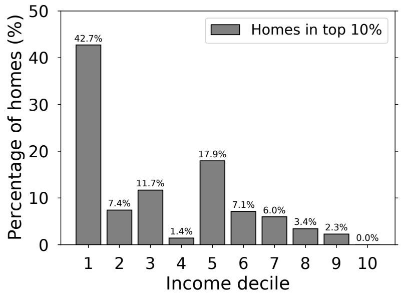


Figure 3: Bias introduced by household income on buildings with lowest cooling slope.

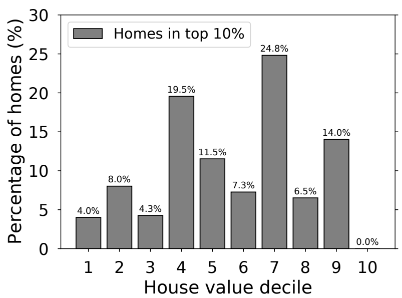


Figure 4: Bias introduced by house value on buildings with highest cooling slope.

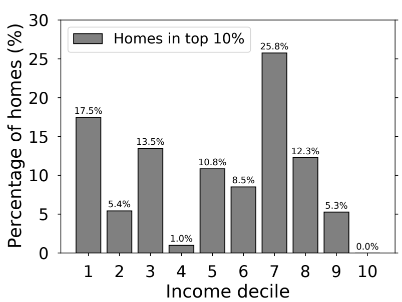


Figure 5: Bias introduced by household income on buildings with highest heating slope.

will then be weighted by these metrics to ensure that recommendations for energy efficiency improvements do not accrue to certain demographics only, but rather span across the whole population.

6 RELATED WORK

In this section, we discuss prior work in energy efficiency analysis and equity.

Data driven approaches for energy efficiency analysis in buildings have been widely studied. Multiple studies have analyzed energy usage data to identify inefficiencies and their underlying causes in buildings [4, 7, 8, 11]. By identifying inefficiency and detecting possible causes, energy improvement measures can be recommended. This work analyzes such data driven approaches and examines whether the recommendations they make are equitably distributed in the society.

The role of equity in the energy transition has also gained attention in the recent past. For example, Andor et al [2] demonstrate that reducing the inequity in cost burden of green energy for the general population significantly increases household willingness to pay for greener electricity. Rezec and Scholtens [9] discuss the role of financial markets and equity indices in the energy transition. Roberts [10] discusses the inequity that exists in access to energy in low income households. Carley et al [5] analyze the injustice that will be perpetuated by the energy transition against specific communities and socio-economic groups. Our work complements these studies by reiterating the importance of equity in the energy transition and showing various biases that affect purely technical analyses.

7 CONCLUSIONS

In this paper, we conducted a data driven study to identify the presence and extent of biases in machine learning and analytical methods in energy efficiency analysis. We analyzed WattHome, a recently proposed energy efficiency and fault detection technique. Our results showed that such technical analyses are not equitable and can be affected by bias introduced by demographic profiles of households such as level of income and house value. We showed that equity and fairness should be considered key design goals for such techniques to ensure that benefits of energy improvements and decarbonization are distributed equitably across the whole society. We hope that this work spurs future work that improves energy equity in the journey towards decarbonization of the entire grid.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful comments. This research was supported by NSF grants 2020888 and 2021693.

REFERENCES

[1] [n. d.]. Total Energy Monthly Data - U.S. Energy Information Administration (EIA). https://www.eia.gov/totalenergy/data/monthly/. (Accessed on 10/07/2021).

[2] Mark A Andor, Manuel Frondel, and Stephan Sommer. 2018. Equity and the willingness to pay for green electricity in Germany. *Nature Energy* 3, 10 (2018), 876–881.

[3] Gaby Baasch, Adam Wicikowski, Gaëlle Faure, and Ralph Evins. 2019. Comparing gray box methods to derive building properties from smart thermostat data. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation*. 223–232.

[4] Gaby M Baasch and Ralph Evins. 2019. Targeting Buildings for Energy Retrofit Using Recurrent Neural Networks with Multivariate Time Series. In *Neural Inf. Process. Syst.* 2019.

[5] Sanya Carley and David M Konisky. 2020. The justice and equity implications of the clean energy transition. *Nature Energy* 5, 8 (2020), 569–577.

[6] François Culière, Laetitia Leduc, and Alexander Belikov. 2020. Bayesian model of electrical heating disaggregation. In *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring*. 25–29.

[7] Srinivasan Iyengar, Stephen Lee, David Irwin, Prashant Shenoy, and Benjamin Weil. 2018. Watthome: A data-driven approach for energy efficiency analytics at city-scale. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 396–405.

[8] Imran Khan, Alfonso Capozzoli, Stefano Paolo Corgnati, and Tania Cerquitelli. 2013. Fault detection analysis of building energy consumption using data mining techniques. *Energy Procedia* 42 (2013), 557–566.

[9] Michael Rezec and Bert Scholtens. 2017. Financing energy transformation: The role of renewable energy equity indices. *International Journal of Green Energy* 14, 4 (2017), 368–378.

[10] Simon Roberts. 2008. Energy, equity and the future of the fuel poor. *Energy Policy* 36, 12 (2008), 4471–4474.

[11] John E Seem. 2007. Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy and buildings* 39, 1 (2007), 52–58.