STEM Fellowship Big Data Challenge 2023

Using Multi-Objective Optimization to Minimize GHG Emissions and Maximize Energy Efficiency of Multifamily Residential Buildings in Ontario

Shaya Farahmand[1] and Navid Aliahmadi[1]

[1]Upper Canada College

January 15, 2023

Abstract

In order to make human settlements sustainable, it is fundamental for households to be mindful of their energy consumption to reduce greenhouse gas emissions (GHG) and improve energy efficiency. This process begins with data collection at the household level to determine the factors which influence GHG emissions. The government of Ontario has been involved in this process by collecting data about the energy intensity of multifamily residential buildings. Using artificial neural networks, our team predicted GHG emissions of these buildings to a ± 1.35 kgCO2e/m2 margin of error, and predicted Energy Star efficiency scores to a ± 9.60 margin of error. This was based on data about a building's electricity, natural gas, and water consumption as well as the energy needed to produce and distribute energy to a building from the "Energy and water usage of large buildings in Ontario" dataset. We also performed a multi-objective optimization with NSGA-II to determine the energy inputs that maximize efficiency and minimize GHG emissions. With this data, it will be easier for consumers to identify target values of electricity, natural gas, and water that can be consumed sustainability. Furthermore, given the clear correlations shown by the model, we encourage governments to engage in similar data collection for low density single family household units in order to determine a sustainable method of energy consumption.

Keywords
Sustainable Housing, GHG Emissions, Multi-objective Optimization, Energy Star Score

1 Introduction

With urban areas growing by 34% since 2001 [1], collecting sufficient data on the needs of citizens and the environment is critical for successful urban planning. In the face of climate change and Canada's commitment to reduce greenhouse gas (GHG) emissions and reach carbon neutrality by 2050 [2], urban planners are moving towards increased urbanization and population density to combat excessive urban sprawl. A recent example of the harmful impacts of urban sprawl is in Ontario, where the provincial government is legislating a 2,995-hectare decrease in land assigned to the GreenBelt [3], leading to the destruction of farmlands, forests and swamps. The environmental damage is exacerbated by further increases in GHG emissions from new car-dependent suburbs, with transportation already causing 32% of GHG emissions [4]. Thus, it is necessary to increase the population density in urban centres with existing transit infrastructure, reducing urban sprawl's environmental damage. To ensure this is done successfully, new multifamily residential developments must be held at higher energy efficiency standards to further reduce the GHG emissions produced.

In line with goal 11 of the UN Sustainable Development Goals to build more sustainable human settlements, the government of Ontario started collecting data regarding the energy intensity of large residential and commercial buildings. Along with the Energy Star certification run by Natural Resources Canada to certify buildings with, on average, 35% lower energy intensity [5], the government's collected data helps civil engineers and policymakers by providing examples for renovations and future developments that enable sustainable human development in urbanization.

Our study aimed to investigate the relationship between the energy intensity data points and the overall efficiency rating using multi-objective optimization (MOO). The dataset was filtered to remove commercial and industrial buildings, which could serve many functions that lead to wide variations in energy intensity and different approaches to reducing GHG emissions. In addition, building permits issued for multifamily residential properties rose 14.7% last year. In comparison, commercial properties only saw a 3.4% increase in issued permits furthering the importance of studying GHG emissions of multifamily residential buildings [6]. Furthermore, while 18% of national GHG emissions are emitted from buildings, roughly 59% of building GHG emissions originate from residential buildings, with other residential and commercial properties emitting the rest [4], [7]. As a result, increasing energy efficiency in multifamily residential buildings is a crucial step toward lowering GHG emissions.

Using the NSGA-II algorithm for MOO, we found optimal solutions for the Energy Star score and GHG emissions based on the energy intensity explanatory variables that were the model's input. These optimal solutions will provide insights for civil engineers and policymakers to implement improved regulations and designs to reduce GHG emissions and increase the efficiency of multifamily residential buildings as part of the fight against climate change.

2 Materials & Methods

2.1 Data used

The data collected was from the Government of Ontario Data Catalogue, specifically, the annual "Energy and water usage of large buildings in Ontario" collection from 2018-2021 [8]. Building managers self-reported values of energy intensity and water usage for buildings larger than 100,000 square feet including commercial, warehousing, light industrial, and multifamily residential. However, as the purpose of the study is to analyse multifamily residential buildings, those were the only type of buildings selected.

The data was self-reported on the Energy Star Portfolio Manager, which allowed key parameters to be weather normalised. This accounts for the annual variation of weather in a single location over time. The portfolio manager took a 30-year average to determine how much energy the building would have used under average conditions. However, the data was not climate normalised, and does not account for regional variations in average weather conditions. For example, a building in a colder climate will use more energy than that of a warmer climate, and will therefore have a different amount of GHG emissions and energy efficiency.

2.2 Variables Selected

From the 31 data points present in the dataset, 5 were selected as explanatory variables, and 2 were selected as objective variables. Other data points either represented the same value in a different unit or represented a descriptive feature of the data (eg: city, postal code, etc.). The explanatory variables are:

- **Weather-normalized Site Electricity Intensity (kWh/ft2) → (WN_Sit_Elc):** The annual average electricity used per square foot of a building
- **Weather-normalized Site Natural Gas Intensity (m3/m2) → (WN_Sit_Gas):** The annual natural gas used per square metre of a building
- **All Water Use Intensity (m3/ft2) → (All_Water):** The annual water usage per square foot of a building, both indoors and outdoors
- **Weather-normalized Site Energy Use Intensity (ekWh/ft2) → (Site_EUI):** The total annual energy use per square foot at a building
- **Weather-normalized Source Energy Use Intensity (ekWh/ft2) → (Source_EUI):** The total annual energy use per square foot at a building in addition to all the energy used in producing and delivering energy.

The objective variables are:

- **GHG Emissions Intensity (kgCO2e/m2):** The GHG emissions that were emitted due to consumption, production, and transmission of energy
- **Energy Star Score:** This is a value from 1-100 which represents the energy efficiency of a building compared to a Canadian survey of buildings in the same type. A score of 50 indicates the building performed at the median value of nationally surveyed buildings. A score of x indicates that the building performed better than x% of nationally surveyed buildings. To determine energy efficiency, the energy star score algorithm

took into account variables such as size, location, number of occupants, source energy, weather data, etc.

2.3 Data Analysis

Two sub-datasets were created in this analysis. The first has the GHG emissions as the response variable. The second has the same explanatory variables, but energy star score as the response variable. The first action taken to preprocess the data was to remove any outliers from the objective variables. This was done by calculating the interquartile range of the data (IQR) and removing any variables greater than the upper quartile + 1.5*IQR or less than the lower quartile - 1.5*IQR. In addition, it is required to remove missing values from the dataset. The next step was to investigate the pairwise correlations in the dataset by constructing a Seaborn pairplot and heatmap to better understand the parameters which have greater influence on the objective variables.

The next step in preprocessing was to split the data into training, testing, and validation groups at a ratio of 64% train, 20% test, 16% validation. This value was settled on after tuning hyperparameters. From there we performed a standard normal distribution on each of the train, test, and validation sets, fitted on the mean and standard deviation of the training set.

The genetic algorithm we used to perform multi-objective optimization to minimize GHG emissions and maximize energy efficiency was the non-dominated sorting genetic algorithm (NSGA-II). However, to use it, one must develop two functions that can predict each of the objective variables. The first is predicts GHG emissions and the second is predicts energy star scores. For each objective variable, an artificial neural network was developed, and tested with mean absolute error (MAE) from the actual value, becoming the function to minimize in the NSGA-II algorithm.

As seen in figure 1, the NSGA-II algorithm begins by generating an initial population of solutions using a genetic algorithm. Its selection operator then randomly selects the best performing populations (in our case, the ones that have low GHG emissions and high energy star score) and ensures the best solutions from the current population are carried forward. The next generation is then created using genetic operators such as crossover and mutation, with crossover combining the genetic information of two solutions to create a new solution. Mutations introduce random changes to explore new regions of the solution space and maintain diversity in the population. This process is repeated until a certain number of generations is completed [10].

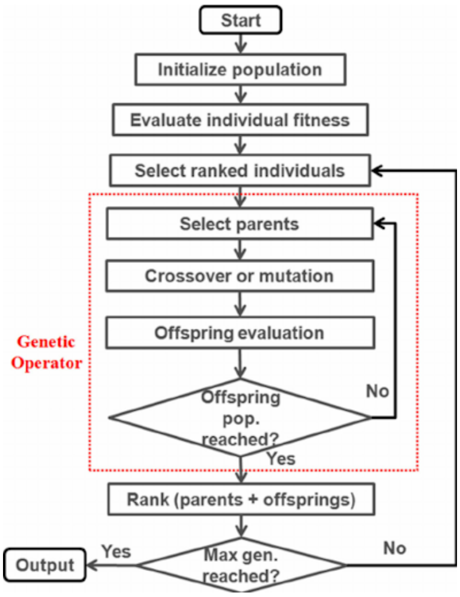

Figure 1: Flowchart outlining the process of the NSGA-II algorithm [9].

3 Results

3.1 Pairwise correlation of data points

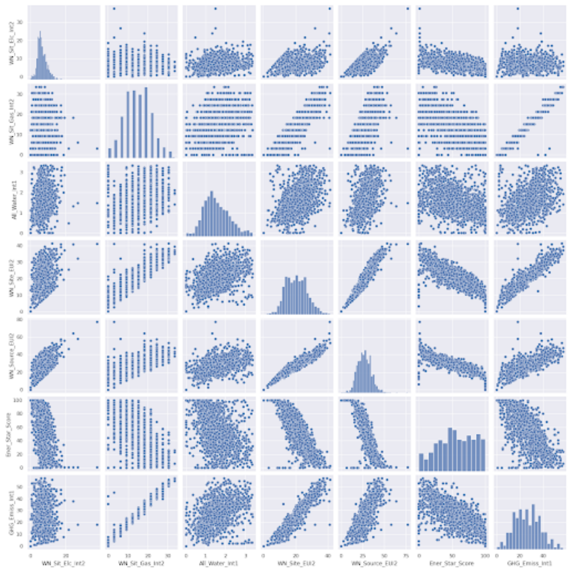

Figure 2: Pairwise correlations of all parameters and objective variables

According to figures 2 and 3, there is a positive correlation between GHG emissions and WN_Sit_Gas_Int (r=0.98), All_Water

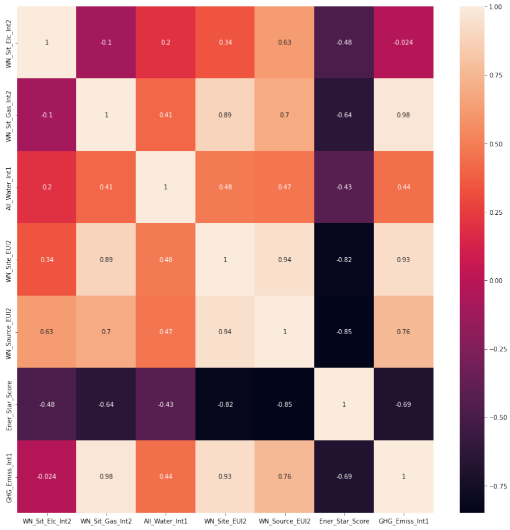

Figure 3: Correlation coefficients of parameters

(r=0.44), WN_Site (r=0.93), and WN_Source (r=0.76). There is a negative correlation between GHG emissions and WN_Sit_Elc (r=-0.024). There is also an negative correlation between energy star score and WN_Sit_Elc (r=-0.48), WN_Sit_Gas_Int (r=-0.64), All_Water (r=-0.43), WN_Site (r=-0.82), WN_Source (r=-0.85). When making a decision to maximize energy efficiency and minimize GHG emissions, the variables that will have a great effect on the data are WN_Sit_Gas, WN_Site, and WN_Source.

3.2 Results of the Artificial Neural Network

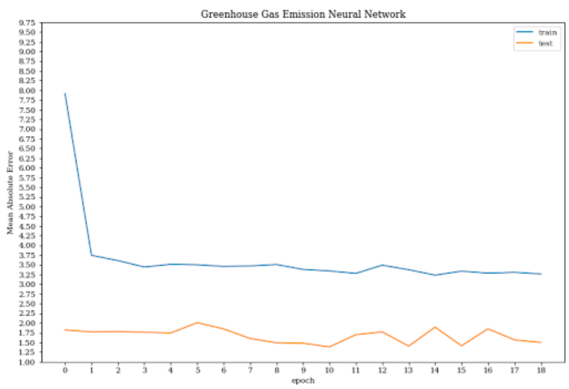

Figure 4: MAE of GHG neural network at every epoch

The first neural network related the explanatory variables with the GHG emissions. It was trained on 200 epochs with early stopping 5 epochs without any improvement. It was trained on 2956 data points, tested on 924 data points, and validated with 739 data points on an adam optimizer with a learning rate of 0.01.

According to figure 4, the model stopped after 19 epochs and the training loss decreased to a range between 3.25 and 4.00. The testing loss decreased to a range between 1.25 and 2.25. The MAE on validation data is approximately 1.35—the model had a margin of error of ± 1.35 on unseen data.

The second neural network related the explanatory variables with the energy star score. It was trained on 200 epochs with early stopping 5 epochs without any improvement. It was trained on 560 samples, tested with 176 samples, and validated with 141 samples on an adam optimizer with a learning rate of 0.007.

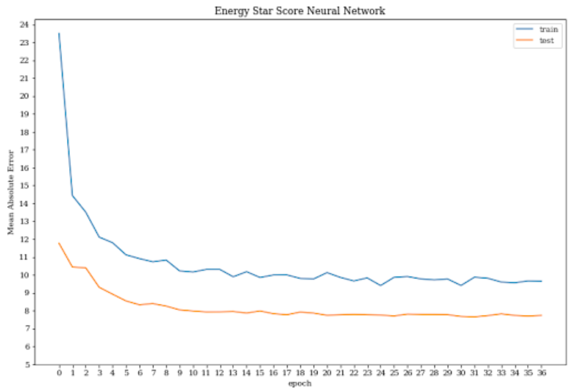

Figure 5: MAE of energy star neural network at every epoch

According to figure 5, the model stopped after 36 epochs and the training loss decreased to a range between 10 and 12. The testing loss decreased to a range between 7 and 9. The MAE on validation data is approximately 9.60 meaning that the model's predicted value on unseen data was approximately 9.60 away from the true value.

Even though the values of energy star score are discrete, the predictions are continuous as the goal of the model is to return the approximate value of energy efficiency. The model will be used for optimization, and the energy star score is relative to other buildings in Canada. Furthermore, the final layer of the model was a ReLU with a maximum value of 100, as that is the maximum energy star score.

3.3 Results of Genetic Algorithm

As mentioned in 2.3, for multi-objective optimization, the NSGA-II algorithm was used, allowing one to optimise two functions simultaneously. The values of each explanatory vari-

	WN_Sit_Elc_Int2	WN_Sit_Gas_Int2	All_Water_Int1	WN_Site_EUI2	WN_Source_EUI2
count	1585.000000	1585.000000	1585.000000	1585.000000	1585.000000
mean	6.804858	14.201699	1.555107	20.520739	27.198097
std	2.988770	6.662538	0.661742	6.657834	8.142581
min	0.100000	0.000000	0.000000	0.210000	0.230000
25%	4.900000	9.131721	1.080181	15.330000	21.400000
50%	6.200000	15.219536	1.461972	20.370001	26.879999
75%	8.300000	18.263443	1.987190	25.090000	32.650002
max	37.500000	33.482979	3.474017	41.270000	77.139999

Figure 6: Descriptive analytics of the features

able must be between the dataset's minimum and maximum, rounded up, as they represent the range of values that the models were trained on. These can be found in figure 6. For the algorithm, the population size at every generation was 100 over 20 generations. However, a constraint was placed on the model that the values of WN_Site must be equal to or greater than the values of WN_Sit_Elec + WN_Sit_Gas + All_Water_Int. This is because the WN_Site represents the total energy values on site and must be made up of at least the sum of the latter values. After 20 generations, there are only 7 points which meet these constraints. These points are visible in figure 7.

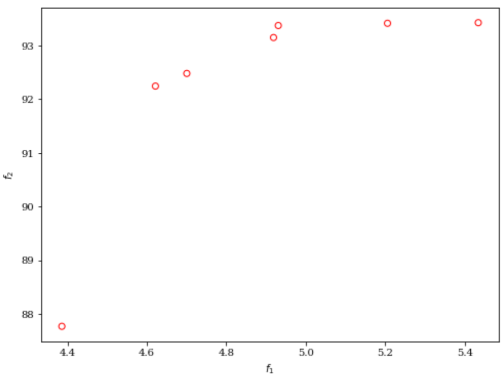

Figure 7: Values of Objective variables at the 20th generation where F1 represents the GHG emissions and F2 represents the EnergyStar Score

	WN_Sit_Elc_Int2	WN_Sit_Gas_Int2	All_Water_Int1	WN_Site_EUI2	WN_Source_EUI2	GHG_Emiss_Int1	Ener_Star_Scre
0	0.973406	0.057143	0.100575	1.222356	0.232073	4.700710	92.476105
1	0.135316	0.073374	0.100533	0.803629	0.236342	5.205852	93.409889
2	3.619793	0.074007	0.330306	4.266633	1.004610	4.386224	87.767540
3	0.974009	0.056289	0.167369	1.236224	0.300705	4.621771	92.240761
4	0.404965	0.061938	0.101780	0.777903	0.306111	4.919014	93.144791
5	0.213133	0.321660	0.021421	0.975571	0.231612	5.434303	93.421402
6	0.255229	0.112980	0.100532	0.584209	0.344273	4.930946	93.368515

Figure 8: Optimised Objective Variables and the explanatory Variables that result in them

In two rightmost columns of figure 8, one can see the optimised values of GHG emissions and energy star scores. The five columns to the left of them represent the parameters which predict the two objective variables.

4 Discussion

4.1 Interpretation of Variables

In Canada, each province's residential GHG emissions are marked by an emissions factor, given that different provinces use a different primary energy source. For example, Quebec and British Columbia derive most of their energy from hydroelectric plants (90% and 88% respectively), and less than 2% from natural gas. Thus, the two provinces have low emission factors of 1.2 and 12.9 gCO2eq/kWh respectively. However, Ontario uses 6% natural gas, thereby rendering an emission score of 40 gCO2eq/KWh [11].

According to figures 2 and 3, between electricity, natural gas, and water usage, the one with the strongest positive correlation with GHG emissions is natural gas usage (r=0.98). This is because in Canada, 63% [12] of household energy consumption is from heating homes, and in single family detached residential zones, most of the energy used to heat homes is from natural gas and heating oil. This results in a GHG intensity of 34.8 kg CO2eq/yr [11]. As seen in figure 6, similar trends are visible when examining how multifamily residential households use energy. The mean of all site electricity use is approximately 6.80 kWh/ft^2 with a standard deviation of 2.99 kWh/ft^2, while that of natural gas usage is approximately 14.20 m^3/m^2 (converted to 13.92 kWh/ft^2) with a standard deviation of 6.67 m^3/m^2 (converted to 6.54 kWh/ft^2). This means that in multifamily residential households, on average more natural gas is used than electricity. Note that WN_Source_EUI is higher than WN_Source_EUI with a higher mean and standard deviation, as seen in figure 6. This is because it includes the total energy needed to produce and deliver energy to a location, in addition to the energy used on the site. According to figure 2 and 3, there is a positive correlation between the WN_Source_EUI and GHG emissions (r=0.89). This is because there are significant GHG emissions in Ontario's electricity generation system, especially in thermal plants with coal and natural gas inputs [13]. This is also why site electricity is negatively correlated with GHG emissions. The electricity that contributes to GHG emissions is that in the source, and not the site.

4.2 Quality of the neural networks

The neural network measuring GHG emissions has a lower MAE on unseen data (1.35) compared to that of energy star score (9.60). The reason that the energy star score had a higher loss than the GHG emissions was that more variables are considered in developing the energy star score, while GHG emissions depend mostly on the energy consumed. For example, it takes into consideration the exact metered energy consumption and it is relative to other buildings [5] . Furthermore, less data was available in training the model (560 data points compared to 2956 data points).

4.3 Results from the Multi-Objective Optimization

Because this is a multi-objective optimization, it will be difficult to converge on a point which maximizes both efficiency while minimising GHG emissions. Hence, the model generated multiple data points. According to figures 7 and 8, all data points had a GHG emission level of 4.3-5.5 GJ/m2 and 6 of the 7 data points had an energy star score of 92-93.5. One data point had an energy star score of approximately 87.77. To determine which point to choose, one must prioritise whether reducing GHG emissions or increasing energy star score is more important. Then, the user can choose an optimal point that matches their criteria. Furthermore, the range of GHG emissions is much less than that of energy star score. Hence, a possible option is to neglect the outlier point. Even though it has the lowest GHG emissions, the increase in GHG emissions resulting from choosing another point will be proportionally less than the increase in energy star score.

The parameters that will result in these predictions are all within the first quartile of data, as seen in figure 6. This was expected, as decreasing consumption is a method to optimize the objective functions.

4.4 Limitations

The model has two limitations. First, there is no consideration of the monetary cost of energy, which can fluctuate over time. The cost of energy including taxes plays a role in determining how much a household will consume. Second, there is no differentiation between locations. Households in colder areas will require more use of energy (due to increased needs for heating) than in warmer areas—and this constraint is not reflected in this model.

4.5 Future Research

An area for further investigation is the correlation between the median income of the residents in multifamily residential buildings and the energy consumption/efficiency. As the model did not account for the energy cost, median income could influence the energy intensity of buildings. The model could improve its optimal solutions to predict the energy consumption of buildings. Furthermore, it could help governments allocate resources and funding to improving energy efficiency if there is a direct correlation with median incomes.

Another potential area to apply insights of this study is in intelligent Home Energy Management Systems (HEMS). Recent studies have found that HEMS, based on deep reinforcement learning techniques developed policies that allowed for an 8% reduction in energy consumption without lowering human comfort in the building [14]. By providing the neural network with the optimal solutions found in our study, the reduction in GHG emissions could be further improved. A slight reduction in energy consumption of even 1% would result in a reduction of 80 kilograms of CO2 per household, which, if implemented in a multifamily residential building, results in over ten metric tonnes of saved emissions per year [15].

Conclusions

Our study found 7 data points for optimal energy efficiency solutions through multi-objective optimization with the NSGA-II algorithm, decreasing GHG emissions and increasing Energy Star scores. This demonstrate a strong correlation between natural gas usage and higher emissions, demonstrating the need for governments to increase the importance placed on programs to promote the usage of new technologies such as heat pumps instead of natural gas furnaces in multifamily residential buildings [13]. Our model will also help developers of existing and new multifamily residential buildings predict how renovations and new technologies affecting the explanatory variables will affect GHG emissions and Energy Star scores.

We encourage governments to start collecting data points regarding energy consumption and intensity from single-household houses so optimal solutions can be found and implemented to lower GHG emissions from all residential buildings. It will be instrumental in helping consumers understand how to increase energy efficiency and help policymakers develop better regulations and guidelines for new developments/renovations. By collecting more data to find efficient methods to reduce our carbon footprint, governments and consumers can help meet Canada's goal to be carbon neutral by 2050 and lead the way in the fight against climate change.

References

[1] Naël Shiab and Isabelle Bouchard. We used ai to measure canada's urban sprawl, March 2022.

[2] Government of Canada. Net-zero emissions by 2050, November 2020.

[3] Ryan Jones. Ford government forges ahead with greenbelt development plan despite 'broad opposition' in public consultation, December 2022.

[4] Deborah Schulte. Better buildings for a low-carbon future, June 2018.

[5] Environmental Protection Agency. Portfolio manager technical reference: Energy star score, 2018.

[6] Leah Schnurr. Canadian building permits jump 10.4% on condos, commercial buildings in August, October 2016.

[7] Senate of Canada. Reducing greenhouse gas emissions in canada's buildings necessary to meeting paris agreement targets, November 2018.

[8] Government of Ontario. Energy and water usage of large buildings in Ontario, November 2022.

[9] Yuqing Chang, Zyed Bouzarkouna, and Deepak Devegowda. Multi-objective optimization for rapid and robust optimal oilfield development under geological uncertainty. *Computational Geosciences*, 19:933–950, 06 2015.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[11] Umberto Berardi and Pouriya Jafarpur. Assessing the impact of climate change on building heating and cooling energy demand in canada. *Renewable and Sustainable Energy Reviews*, 121:109681, 2020.

[12] Ravihari Kotagodahetti, Kasun Hewage, Hirushie Karunathilake, and Rehan Sadiq. Renewable natural gas as a greener energy source: A life cycle cost-benefit analysis. May 2021.

[13] Alex Szekeres and Jack Jeswiet. Effects of technological development and electricity price reductions on adoption of residential heat pumps in Ontario, Canada. *International Journal of Energy and Environmental Engineering*, 9:201–213, March 2018.

[14] Paulo Lissa, Conor Deane, Michael Schukat, Federico Seri, Marcus Keane, and Enda Barrett. Deep reinforcement learning for home energy management system control. *Energy and AI*, 3:100043, 2021.

[15] Environmental Protection Agency. Greenhouse gases equivalencies calculator - calculations and references, December 2018.