Estimating Heating Loads in Alaska using Remote Sensing and Machine Learning Methods

Madelyn Gaumer [*], Nick Bolten[2], Vidisha Chowdhury[3], Philippe Schicker[4], Shamsi Soltani[5], and Erin Trochim[6]

[1]Department of Applied Mathematics, University of Washington
[2]Paul G. Allen School of Computer Science and Engineering, University of Washington
[3,4]Heinz College, Carnegie Mellon University
[5]Department of Epidemiology Population Health, Stanford University School of Medicine
[6]University of Alaska Fairbanks

Abstract

Alaska and the larger Arctic region are in much greater need of decarbonization than the rest of the globe as a result of the accelerated consequences of climate change over the past ten years [10]. Heating for homes and businesses accounts for over 75% of the energy used in the Arctic region. However, the lack of thorough and precise heating load estimations in these regions poses a significant obstacle to the transition to renewable energy. In order to accurately measure the massive heating demands in Alaska, this research pioneers a geospatial-first methodology that integrates remote sensing and machine learning techniques. Building characteristics such as height, size, year of construction, thawing degree days, and freezing degree days are extracted using open-source geospatial information in Google Earth Engine (GEE). These variables coupled with heating load forecasts from the AK Warm simulation program are used to train models that forecast heating loads on Alaska's Railbelt utility grid. Our research greatly advances geospatial capability in this area and considerably informs the decarbonization activities currently in progress in Alaska.

1 Introduction

Compared to the global average, global warming is occurring almost four times as fast in the state of Alaska [10]. In colder climates like those of Alaska and other arctic regions, heating homes and commercial buildings is a major contributor to global warming. In 2020, Alaska had the second highest petroleum usage, specifically for electricity needs, in the nation [12]. In addition, Alaska has high per capita energy consumption compared to the rest of the United States [12], which is often attributed to the energy needs due to the harsh climate.

In order to decarbonize Alaska, domain experts require accurate, fine-grained estimates for heating loads at the neighborhood level. Here we provide these estimates at the building level. While providing these estimates at the building level yields privacy concerns (one reason our current estimates are not publicly available), in future work, building level estimates provide an opportunity for optional public participatory refinement of the heating load estimates. Researchers in Alaska need these heating load estimates to inform grid modernization, diversify energy sources, and advise retrofitting and other energy efficiency update programs in the state. All of which will directly assist in decarbonizing Alaska. Not only do researchers need these heating load estimates, but these

[*]megaumer@uw.edu

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

estimates are also in demand at multiple levels of government. This proposal undertakes a proof of concept work to show it is possible to obtain these estimates through a top-down approach that is in direct contrast to the micro-level approaches currently used by researchers.

2 Previous Work

The main limitation of current Alaskan heating load estimation work is its micro-level scale. This work often consists of invasive methods like visiting homes and installing various fuel use monitoring apparatuses, such as the PuMA pump [1]. These methods can have high levels of variation in fuel estimates between identical stove models because of variation in the mechanisms and sensitivity to pump placement. While these approaches can provide some estimations of household fuel use, they only do so at a small scale because of the labor intensive nature of the approach. It is impractical to visit each of the approximately 230,000 buildings in Alaska's railbelt to install fuel use monitoring apparatuses or attempt other micro-level approaches to generate heating load estimates. While these projects do not yet provide us with an adequate sample of heating estimates for our machine learning approach, in the future, they provide an interesting opportunity for cross validation of our methods.

In 2014, the Alaska Housing Finance Corporation published two regression relationships between the decade a home was built and the average energy use per square foot [2]. One of these relationships was for homes built in the Fairbanks Northstar Borough, and the other was for homes built in the Municipality of Anchorage. These regression relationships use input features from a comprehensive heating loads database coming from retrofitting efforts in Alaska that is currently in the process of becoming open source and heating load estimates from AK Warm, the standard software used for modeling heat in Alaska. This synthetic data and the corresponding regressions were not set up to be used for predictive purposes and do no generalize to the entire state of Alaska. Since the comprehensive heating loads database coming from retrofitting efforts in Alaska is not yet publicly available, we chose to assign buildings in Anchorage and Fairbanks BTU estimates coming from these two respective regression relationships, the only public version of this data at this time, to get preliminary results on our overall methods.

3 Methods

3.1 Data and Data Sampling

In this project, we used Google Earth Engine [6], a tool used for open source geospatial big data analysis, to extract features from satellite data that we then passed to our model. All seven of the datasets used in this project are from Google Earth Engine's public archive and consist of satellite images captured across time as well as geometries of building outlines in Alaska. We aggregated the data both temporally and spatially to get building level features that were passed to our model. Alaska and other arctic regions present unique challenges when it comes to data collection. As a result, data in this region can often be sparse and inaccurate. We combat that in our approach by using multiple data sources.

The features we extracted included information on local climate conditions and on building features like height, base area, and age. All of these features were chosen because they are relevant to how much heat is needed to heat a building.

We compared the Open Street Maps Building Footprints dataset [9] to the Microsoft Buildings dataset [11] in Alaska and found that the Open Street Maps Building Footprints dataset was more accurate for our task. Using these building footprints ensures there are no duplicates in our data processing.

To incorporate local climate conditions into our model, we calculated ten and thirty year averages of thawing and freezing degree days at a building level using the ERA5 daily aggregated air temperature dataset [4]. Thawing and freezing degree days are measures of how often the temperature is above or below zero degrees celsius.

To calculate building age in Alaska, we used three different datasets: World Settlements Footprint Evolution (1985-2015) [8], Word Settlements Footprint 2019 (2019) [8], Dynamic World (June 2015-present) [3]. We first reduced the World Settlements Footprint Evolution dataset over the Open Street Maps Building Footprints dataset by taking the mean age value in each building outline. If a building did not yet have an age, we then repeated this with the Word Settlements Footprint 2019 dataset, making any new age values 2019. We then repeated this process with the Dynamic World dataset, assigning any new age values 2020. Finally, for any buildings left without an age at this point in the process, we assigned them an age of 1984.

To calculate building height, we used two different digital elevation model datasets: Copernicus Digital Elevation Model (GLO-30 DEM) [5] and FABDEM (Forest And Buildings removed Copernicus 30m DEM) [7]. GLO-30 DEM contains information on elevation with buildings and forest whereas FABDEM contains information on elevation without buildings and forest. We took the difference in elevation between these two models and reduced it over the Open Street Maps Building Footprints in Alaska dataset, taking the mean elevation value. Both of these models have a 30 meter resolution.

To calculate building base area, we used the Open Street Maps Building Footprints dataset and calculated the area of each building geometry.

In this project, we explored the effects of data sampling on our models. The motivation for doing so is to balance the data across a number of different factors in order to avoid biasing the model towards particular types of data. Particularly in this problem, Anchorage has more total buildings than Fairbanks, but we want our model to be able to generalize to the entire state of Alaska. We explored the effects of upsampling buildings from Fairbanks as well as downsampling buildings from Anchorage. In addition, due to the time ranges covered by the datasets used to calculate building age, some building age categories have many more samples than others. We explored the effects of up and downsampling age categories to certain equilibrium points.

3.2 Training Procedure and Models

After extracting training features including building base area, building height, building age, and climate from Google Earth Engine, we assigned each building in the Fairbanks Northstar Borough and in the Municipality of Anchorage, the two regions included in the 2014 Alaska Housing Finance Corporation regression relationships, a heating load estimate. We then created a train/test split of 70/30 and ran our data through the regression models. We then predict on buildings in the rest of the railbelt, not including those in the Fairbanks Northstar Borough or in the Municipality of Anchorage. It is worth noting here that due to finite number of heating load estimates generated from on-the-ground micro-level approaches, traditional validation methods are not possible. Instead, we validate with the regression relationships from the Alaska Housing Finance Corporation.

Regression was chosen as the general model type in this project due to the constraint of needing a continuous output variable, the heating load of a building. We compared five different regression models including linear regression, ridge regression, ridge regression with polynomial features and cross validation, decision tree regression, and random forest regression.

4 Preliminary Results and Future Work

In comparing the mean squared error of the five models tested, we found that the random forest model performed the best when the data was not sampled to be balanced in any way. The models tested gave mean squared errors ranging from 6.976 x 10^{-3} to 1.221 x 10^{-7}. When the data from Fairbanks was upsampled, the random forest model once again performed the best, even better than when the data was not balanced at all or when the data from Anchorage was downsampled. The models tested gave mean square errors ranging from 8.214 x 10^{-3} to 2.026 x 10^{-8}. Finally, the lowest mean squared error was achieved by the decision tree model when building age was balanced and location was not. The models tested gave mean square errors ranging from 5.801 x 10^{-3} to 3.321 x 10^{-9}.

After calculating feature importances of the random forest model with no data balancing, the most important features were building age, freezing degree days 30 year average (1981-2010), and freezing degree days 10 year average (1991-2000). This is an interesting finding that could have to do with the way that BTU estimations for training were calculated. However, this does show that our model was using relevant local climate information to generalize across the different climates found in Fairbanks and Alaska. In future work after incorporating the comprehensive heating loads database, further experiments need to be run to better understand the role and importance of the features being inputted into the model.

In this proof of concept project, our top down geospatial approach generated necessary data at the building level for estimating heating loads and yielded a model that learned to make predictions from this data. Our method allows for heating load estimates for areas in Alaska with no previously available building data or heating estimates to be predicted.

This project's preliminary success has yielded future directions for research, especially as the previously mentioned comprehensive heating loads database coming from retrofitting efforts in Alaska becomes publicly available. Incorporating this database into our models would provide more relevant neighborhood and building level information and most likely allow more generalization of our model. In addition, since the climate data features were important to the model, adding more climate information at varying scales could provide a benefit to the models. Performing a sensitivity exploration as the project progresses could also allow us to explore some of the assumptions made by our geospatial first approach. Finally, due to the severe nature of climate change in Alaska and other arctic regions, expanding this work to other arctic regions is a top priority to inform arctic decarbonization.

Acknowledgments and Disclosure of Funding

We would like to thank the supporters of this work including Micron, the University of Washington, and the University of Washington eScience Institute. We are very grateful to the University of Washington Data Science for Social Good Program for all their support.

References

[1] Alaska Center for Energy and Power. Puma fuel meters capture fuel use in fairbanks homes. URL `https://acep.uaf.edu/acep-news/2019/4-march-acep-this-week/puma-fuel-meters-capture-fuel-use-in-fairbanks-homes.aspx`.

[2] Alaska Housing Finance Corporation. 2014 alaska housing assessment. URL `https://www.ahfc.us/pros/energy/alaska-housing-assessment/housing-assessment`.

[3] C. Brown, S. Brumby, B. Guzder-Williams, et al. Dynamic world, near real-time global 10m land use land cover mapping. *Nature Scientific Data*, 9:251, 2022. doi: https://doi.org/10.1038/s41597-022-01307-4.

[4] Copernicus Climate Change Service (C3S) (2017). Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate. URL `https://cds.climate.copernicus.eu/cdsapp#!/home`. Copernicus Climate Change Service Climate Data Store (CDS).

[5] Copernicus DEM. URL `https://samapriya.github.io/awesome-gee-community-datasets/projects/glo30/#citation`. Accessed in August 2022.

[6] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017. doi: 10.1016/j.rse.2017.06.031. URL `https://doi.org/10.1016/j.rse.2017.06.031`.

[7] L. Hawker, P. Uhe, L. Paulo, J. Sosa, J. Savage, C. Sampson, and J. Neal. A 30m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 2022.

[8] M. Marconcini, A. Metz-Marconcini, T. Esch, and N. Gorelick. Understanding current trends in global urbanisation – the world settlement footprint suite. *GI_Forum*, 1:33–38, 2021. doi: https://doi.org/10.1553/giscience2021_01_s33.

[9] Open Street Maps. URL `https://osmbuildings.org/`.

[10] M. Rantanen, A. Karpechko, A. Lipponen, et al. The arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth Environment*, 322(168):1–10, 2022. doi: https://doi.org/10.1038/s43247-022-00498-3.

[11] S. Roy. Global ml building footprints. URL `https://samapriya.github.io/awesome-gee-community-datasets/projects/msbuildings/`. Created by Microsoft.

[12] U.S. Energy Information Administration. Alaska state profile and energy estimates. URL https://www.eia.gov/state/?sid=AK.