2022 Building Performance Modeling Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA
Chicago, IL
September 14-16, 2022

Using Measured Building Energy Data to Infer Building Characteristics for Urban Building Energy Modeling
Co-organized by ASHRAE and IBPSA-USA

Brett Bass[1], Evan Ezell[2], Joshua New[1]
[1]Oak Ridge National Laboratory, Oak Ridge, TN
[2]University of Tennessee, Knoxville, TN

Abstract

Buildings in the United States used 40% of total energy use in the United States in 2020, providing a significant opportunity to reduce energy use and the carbon footprint of the US. Modeling large numbers of buildings in a particular region maximizes this impact, allowing cities, utilities, or other stakeholders to determine the optimal solutions to reduce energy consumption in buildings based on modeling simulations. Building type is a critical input variable from which many significant building characteristics such as occupancy, equipment, lighting, etc., can be inferred if this data is not directly available. Aggregating the building type in a non-intrusive manner for large scale analyses can be difficult as there is no public database with the function of each building. For this reason, another method of assigning building type using measured building energy use was developed. Measured building energy use was compared to Department of Energy (DOE) prototype building energy models for about 46 thousand buildings from the Electric Power Board (EPB) of Chattanooga service area to determine which prototype building was most representative of each individual building. Two methods of comparing the energy usage to the prototype building energy models at two different temporal resolutions were compared, with the building type assignments used to generate building energy models. Simulation results were compared to measured data to determine which method had the highest accuracy. It was determined that Euclidean distance was the optimal comparison metric and reduction of temporal resolution from hourly to monthly did not result in a major reduction in simulation quality.

Introduction

Buildings in the United States used 40% of total energy use in the United States in 2020. While individual building energy modeling is a critical piece of carbon reduction in the US, modeling large numbers of buildings allows stakeholders such as utilities to optimize their service area and grid by understanding how their buildings will perform under various conditions and with different technologies. This large scale building analysis, often called urban building energy modeling (UBEM) requires aggregation of a significant amount of metadata describing each building to develop models. Building type and building vintage are crucial input variables in most UBEM analyses as they may be used to infer heating ventilation and air conditioning (HVAC) type, occupancy, lighting, equipment, and other energy related building characteristics. While physical properties describing the building such as building footprint may be publicly available, properties describing the performance of the building such as building type or age are less readily available and there is no publicly available national database that can be used to determine building type. Obtaining the building type is typically done using public databases such as tax-assessor's data or other relevant databases. This can be difficult at a scale beyond an individual municipality as databases for different cities or counties can be difficult to aggregate due to formatting differences. Even if the building type data is available, the function of the building may not match the energy performance of the building. For this reason, a method was developed to assign building types without these databases and will capture the energy performance of the building.

Measured sub-hourly meter data used in this analysis was provided for more than 80 thousand meters for the year 2019. These meters correspond to about 46 thousand buildings. Measured building energy use can be compared to DOE prototype building energy models to determine which prototype is closest to the measured data with this measured data often being available when working with utilities. DOE prototype building energy models are a set of energy models developed which represent common buildings in the US (US Department of Energy 2019). Each of these prototype buildings has a set of vintages that represent various construction years and contains different levels of technology and efficiency. The commercial prototype buildings cover 75% of commercial building floor area in the US across all climate zones. Various methods of comparison can be used to compare the measured energy time series data and the DOE prototype simulation energy time series data at different temporal resolutions. The building energy models can be generated with the assigned building type as determined by which DOE prototype building is closest to the measured energy use. AutoBEM (Automatic Building Energy Modeling), a software UBEM software developed at Oak Ridge National Laboratory, then uses the assigned prototype building and vintage as well as other physical characteristics about the building to generate building energy models (New et al. 2020). These buildings can then be simulated and compared to the measured data to determine simulation accuracy and understand which of the comparison methods is most effective in assigning building type.

Assigning correct input data to each building is a critical piece of generating a representative building energy model and building type is one of the most impactful variables. This is illustrated in Figure 1 below where a single building geometry was generated and simulated using each building type and vintage combination. The difference in simulated annual energy use is significant across building types for the same area.

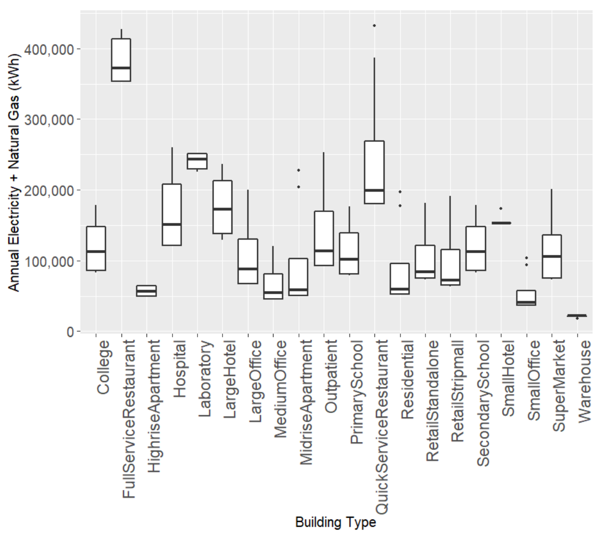

Figure 1: The impact of building type can be seen as the building geometry remained constant for each simulation. Some building types such as restaurants used more than 10x the amount of energy as buildings such as warehouses for the same building geometry.

Building vintage is also an important variable as the age of the building affects several building properties such as efficiencies, infiltration, insulation, etc. As expected, older buildings tend to use significantly more energy than newer buildings which utilize the latest technology and are up to code. The energy use differences for a building with the same geometry and different vintages is shown in Figure 2

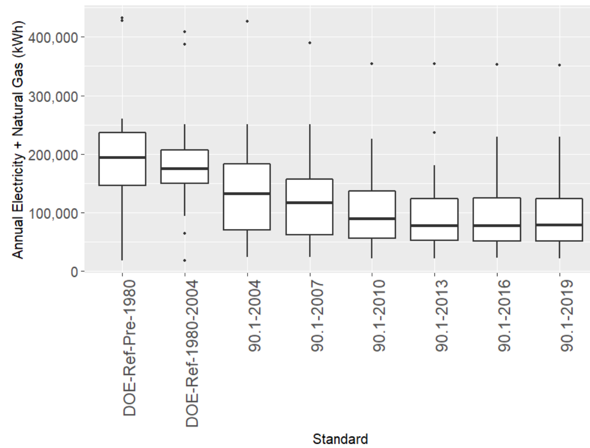

Figure 2: The impact of building vintage can be seen as the building geometry remained constant for each simulation. Older vintages tend to use more energy.

It is important to develop a representative baseline building energy model so that any modification made to the building simulation environment will be representative of the actual outcome. There are multiple examples of simulation perturbations that can be explored in UBEM. One example of this is the energy, emissions, and cost savings estimates of various building technologies or retrofits. A representative building energy model should provide representative savings estimates of the technology and allow a stakeholder to estimate if implementation is worthwhile. Another example is simulation of the building energy models under different weather conditions. This could include future weather taken from climate models or estimation of the impact of extreme weather events such as heat waves or draughts allowing stakeholders to understand the projected energy use and demand under varying conditions.

This 46 thousand building analysis for the year 2019 builds upon previous exploratory work of 100 buildings from the EPB service area for the year 2015 (Bass et al.). In addition to the scale of the previous analysis, another limitation was the use of sub-hourly (15-minute) data which is not widely available. This analysis investigates that shortcoming by using more common resolutions of hourly and monthly. The building type of the 100 buildings from the previous study was classified by hand in the in an effort to both a quantitatively (correct building energy usage) and qualitatively (correct building function) evaluate the assignment methods. It found that different building type assignment methods were better for different methods of evaluation. For this study, it was not possible to qualitatively label each building type, but the quantitative results are typically more valuable in large scale analyses.

Methodology

Measured electric meter data for the year 2019 was provided for 80,017 meters that were a part of the EPB service area. Each electric meter was aggregated with any other meters that share the same building and compared to each DOE building type and vintage combination to determine which was most similar. To ensure the closest comparison of measured meter data and simulated prototype data, weather data from the same year was used. For this reason, each prototype building and vintage combination (110 models) was simulated using a Chattanooga weather file for the year 2019. For the comparison, meters were aggregated per building and compared to prototype building energy models using two different comparisons at two different temporal resolutions. The comparison process was a multi-step process: (1) the electrical meters for each building were aggregated into a single time series; (2) the energy use was divided by the area to obtain the building energy use intensity (EUI); (3) the building was removed if it did not contain the proper number of hours in a year (8760 hours); (4) the measured aggregated time series EUI was compared to each prototype building and vintage combination simulated time series EUI for the same year (110 combinations) for each method; (5) the building type and vintage combination with the minimum distance (or highest correlation) was assigned to the building. The comparison metrics and temporal resolutions are shown in Table 1.

Table 1: Two different methods of comparison were evaluated in comparing the measured data to the prototype simulation data. These methods were evaluated at two different temporal resolutions.

Comparison Method	Temporal Resolutions
Euclidean Distance	Hourly
Dynamic Time Warping	Monthly

It is important to note that only electrical utility data was shared and compared (no natural gas or other energy use data was available). While this is a limitation, only an estimated 33% of residential buildings use natural gas in the south (US Energy Information Administration), minimizing the number of buildings this applies. Still, an attempt was made to filter these buildings out of the dataset by removing buildings that used less than 20% of their annual energy use in the winter months (January, February, December). Buildings that used less than 2% of their annual energy in a single month were also not included in the final dataset. This filtered out 12,432 from the dataset for a remaining total of 34,303 left for evaluation.

Euclidean distance is a straightforward metric from the the distance between each point in the measured data is compared to each point in the prototype simulation data. Euclidean Distance assumes each energy measurement corresponds directly to the simulated energy use at the same time, not allowing for offsets. As the same weather data from the year 2019 was used, this may be a positive as the simulated energy use should be similar during the equal weather conditions. While this is true, individual buildings do not all perform consistently as a prototype simulation would. For this reason, dynamic time warping (DTW) was also considered (Sakoe and Chiba 1978). DTW allows for non-linearity in the time dimension, meaning each measured point is compared to points at different times in the simulation data, allowing for time shifts. Specifically, we use a FastDTW implementatin of DTW (Salvador and Chan 2004). The building type and vintage combination with the smallest distance was assigned.

Once the most similar building type and vintage was assigned for each method, the building energy models were generated using AutoBEM (Automatic Building Energy Modeling) (New et al. 2020). AutoBEM is a UBEM software which takes a set of input data about a building and develops a building energy model. The most coarse set of input data necessary to develop a model include 2D footprint, height, building type, standard, and climate zone. More specific data about the building (such as number of floors, lighting, HVAC, etc.) may be used if available but this data may be inferred using these general characteristics. These inferences are made using the estimates from each of the DOE prototype building values, further increasing the impact building type and vintage have on the simulation output.

The generated buildings are then simulated using a weather file from Chattanooga in 2019 and compared to the measured data to evaluate quality of each simulation with the assigned building type and vintage. Results are adjusted so the sum of the simulation data is equal to the sum of the measured data. While more complex ways of adjusting or calibrating the building energy simulations are possible, this simple and consistent method for eliminating scalar bias allows for equal comparison of the classification methods. *(It could be useful to con-*

sider unadjusted values and that could be considered for future work)

Results

The computing time of the two methods is a major factor when considering a large number of buildings. The average time to classify a single building by comparing aggregating measured EUI to simulated prototype EUI of the same year is shown in 2. The massive difference between the time to classify a building using DTW is a major drawback of the method. DTW takes more than 100x longer to classify a building than the other methods mentioned. This weakness is amplified when considered thousands of buildings for UBEM analyses.

Table 2: The computing time to classify a single building is orders of magnitude larger for the dynamic time warping method and should be considered when choosing a comparison method.

Comparison Method	Time To Classify Building
Euclidean Distance	0.117s
Dynamic Time Warping	133.5s

The building type and vintage classifications of each method provide a preliminary glimpse into each method. The building type classifications for each method are shown in Figure 3. In the visualization, building subtypes are grouped together for easier visualization (i.e. "small", "medium" and "large" offices are all grouped together as "office"). One would expect a typical building type distribution in the United States to contain mostly residential buildings, followed by offices, retail, and warehouses though it is important to remember the assigned building type may not be the same as the functional building type due to differences in individual building energy use patterns. Though classification by building energy performance rather than function may provide a better baseline model, it is important to consider that the properties (equipment, materials, etc.) are more likely to be different to the actual building than if classified using the function of the building. If a high energy intensity house is classified as a restaurant, the equipment of the building is more likely to be different and could lead to discrepancies when considering savings and other analyses. However, these buildings that perform differently than their function are inherently outliers and are only a small number of the total buildings.

An interesting observation about the Eucldiean and DTW is the number of buildings classified as warehouses. The amount of warehouses in these classification methodologies warrants further investigation. It is possible that lack of information derived from the measured and building data may be the cause of these classifications. Lack of measured natural gas data (estimated to be in 33% of residential buildings in the south (US Energy Information Administration)), over-estimation of conditioned floor area, or missing data could lead to lower measured EUI values and warehouse classifications, as the warehouses have the lowest prototype EUI. While this is a limitation of the available data, it could be remedied with additional energy data or more descriptive building data.

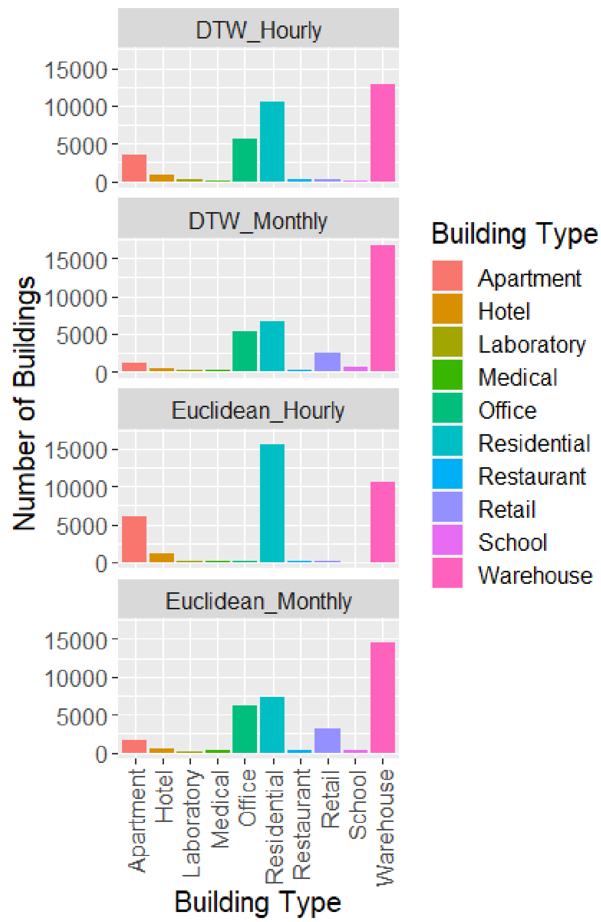

Figure 3: A majority of residential buildings would be expected for this sample. The distance based metrics (DTW, Euclidean) has large numbers of warehouse likely caused by low measured EUI values.

When considering vintage classifications, it should be noted that every building type does not have a model for every vintage. Residential buildings, laboratories, and high-rise apartments, for example, do not have pre-2004 vintages. This skews the vintage distributions for some classifications methods that classify many buildings as one of these types. According to the Commercial Buildings Energy Consumption Survey (CBECS) 2018, 55% of commercial buildings were constructed before 1980 while 82% were constructed before 2000 (US Energy Information Administration 2018). According to the Residential Energy Consumption Survey (RECS) 2015, 46% of residential buildings were constructed before 1980 while 75% were constructed before 2000 (US Energy Information Administration 2015). Considering all US buildings, 54% were constructed before 1980 while 82% were constructed before 2000. These US building age distributions are interesting to note as the assigned distributions are tend to be newer. This is not caused by the lack of pre-2004 residential prototype models as for the Euclidean hourly classification schema, only 2% of the 18,096 buildings were classified as 2004 while only 4% were classified in the 2007 vintage while 62% were classified in the newest vintage (2019). It is again possible that the inherent issues with the measured energy data such as lack of measured natural gas or over-estimation of conditioned floor area could lead to lower measured EUI values and newer vintage classifications. Another possibility is that building age distribution may not necessarily represent the standard performance of a building of that age if renovations or updates have occurred, which is possible based on these vintage assignments.

The vintage with the most buildings for these methods is 2013. For the Euclidean hourly classification, 85% of the buildings classified as 2013 were warehouses. 2013 was by far the most common of the vintages for warehouses for the Euclidean hourly classification with 94% of the warehouses classified as 2013. The same trend was true for the Euclidean monthly assignment. This is the building type and vintage combination with the lowest overall EUI. This further attributes the low measured EUI values to this classification. Even though these buildings may not actually be warehouses or the newest vintage, the energy performance of these buildings may be better represented by the warehouse model.

Measuring the quality of the simulation compared to the measured data is the important factor in the analysis. In building energy modeling, coefficient of variation of root-mean squared error (CVRMSE) is a common metric used to determine the quality of a baseline simulation compared to measured data. The CVRMSE values shown in this document are base AutoBEM models and have no further modifications. This is because the evaluation of different methods is most important for work. If

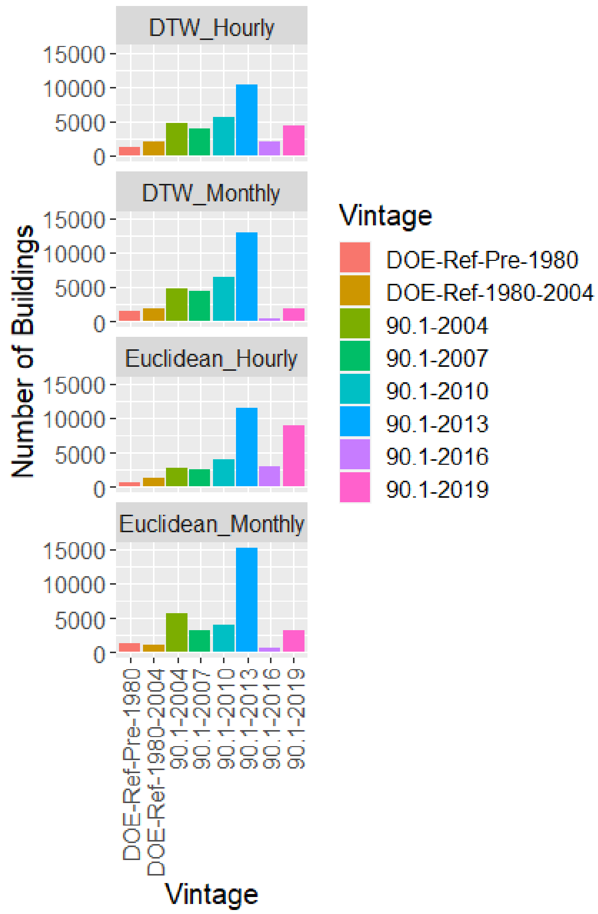

Figure 4: The large amount of 2013 vintages are likely caused by the high number of warehouses. The newest vintage of warehouses used for this analysis was 2013. This vintage had the lowest EUI of any prototype and was likely classified for low energy intensity buildings.

other UBEM analysis was being done with these models, additional adjustments (electrification, insulation, COP, etc.) would be taken based on the area of interest to further improve the models. This leads to lower than typical hourly and monthly CVRMSE values.

The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14 states that <15% monthly or <30% hourly CVRMSE are often considered "investment grade" (ASHRAE 2014). The equation for CVRMSE is in Equation 1. Though ASHRAE guideline 14 is used as a benchmark, it was created as a benchmark for evaluating individual building energy models that were developed by hand. It is much more difficult to attain these benchmarks for individual buildings when modeling thousands and less descriptive building data is available. For typical large-scale workflows, the models can be further improved with building or location specific data, but no further improvement was explored for equal comparison of each method.

$$CVRMSE = \frac{1}{\bar{Y}}\sqrt{\frac{\Sigma_{i=1}^{n}(Y_i - \hat{Y})^2}{N}} \tag{1}$$

The CVRMSE values for each method are ultimately most important metric as it measures the quality of the baseline model. The hourly and monthly CVRMSE values for Euclidean and DTW at hourly and monthly comparison resolution are shown in Figure 5. Monthly CVRMSE values are naturally lower than hourly as the comparison is between 12 points rather than the 8760 point hourly comparison and thus has less variance. The monthly classification resolution has lower monthly CVRMSE rates while the hourly classification resolution has lower hourly CVRMSE rates. This makes sense as some variance is lost during monthly aggregation. It is interesting to note that similar monthly and hourly CVRMSE rates can be achieved by comparing monthly data as hourly data. This implies that building type and vintage may be assigned when only low resolution data is available. It may even be possible that an annual EUI value may be sufficient for a rough estimation for certain use cases.

The DTW and Euclidean distributions are very close for monthly CVRMSE but the Hourly Euclidean classification has the lowest hourly CVRMSE. This is important as the Hourly Euclidean classification captures the higher resolution signal of the measured data best out of these methods.

Viewing CVRMSE values separated by building type and vintage provide insight into the quality of the classifications for each method and resolution. The hourly and monthly CVRMSE values for each method and resolution separated by building type and vintage are shown in Figure 6.

There are not many significant differences between Euclidean and DTW at hourly and monthly resolution in terms of CVRMSE. The building type with the largest CVRMSE is the warehouse, indicating that the issues mentioned previously with the warehouse could be grounded. It also could be that this is the building type with the highest number of classifications and is most susceptible to individual building variance. Hotels, medical, and restaurants have the best overall hourly CVRMSE and apartments and restaurants have the best

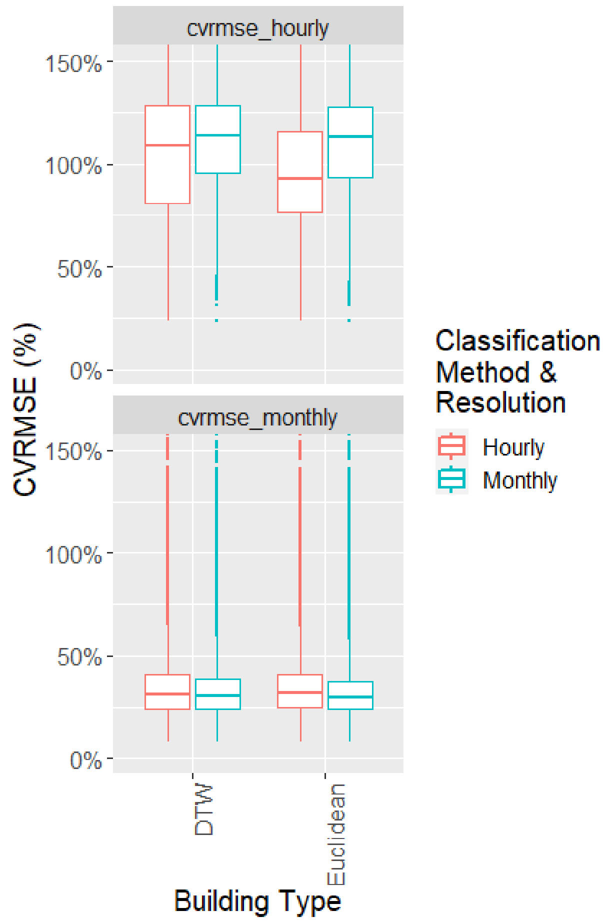

Figure 5: The hourly classification resolutions have lower hourly CVRMSE values while the monthly classification resolutions have lower monthly CVRMSE values. The Hourly Euclidean method has the best overall hourly CVRMSE by a significant margin.

overall monthly CVRMSE. It is interesting to consider differences in monthly and hourly CVRMSE. Though building types such as offices and retail have some of the better monthly CVRMSE metrics, the higher fidelity signal of the measured data is not being captured as they have some of the higher aggregate hourly CVMRSE values.

The other major difference is low hourly CVRMSE values for the offices and retail classified by the Hourly Euclidean method compared to the other methods. Offices and retail only make up a combined 0.6% of the classifications for the Hourly Euclidean method and therefore do not tell the full story as to why this method has the best hourly CVRMSE in Figure 5. The Hourly

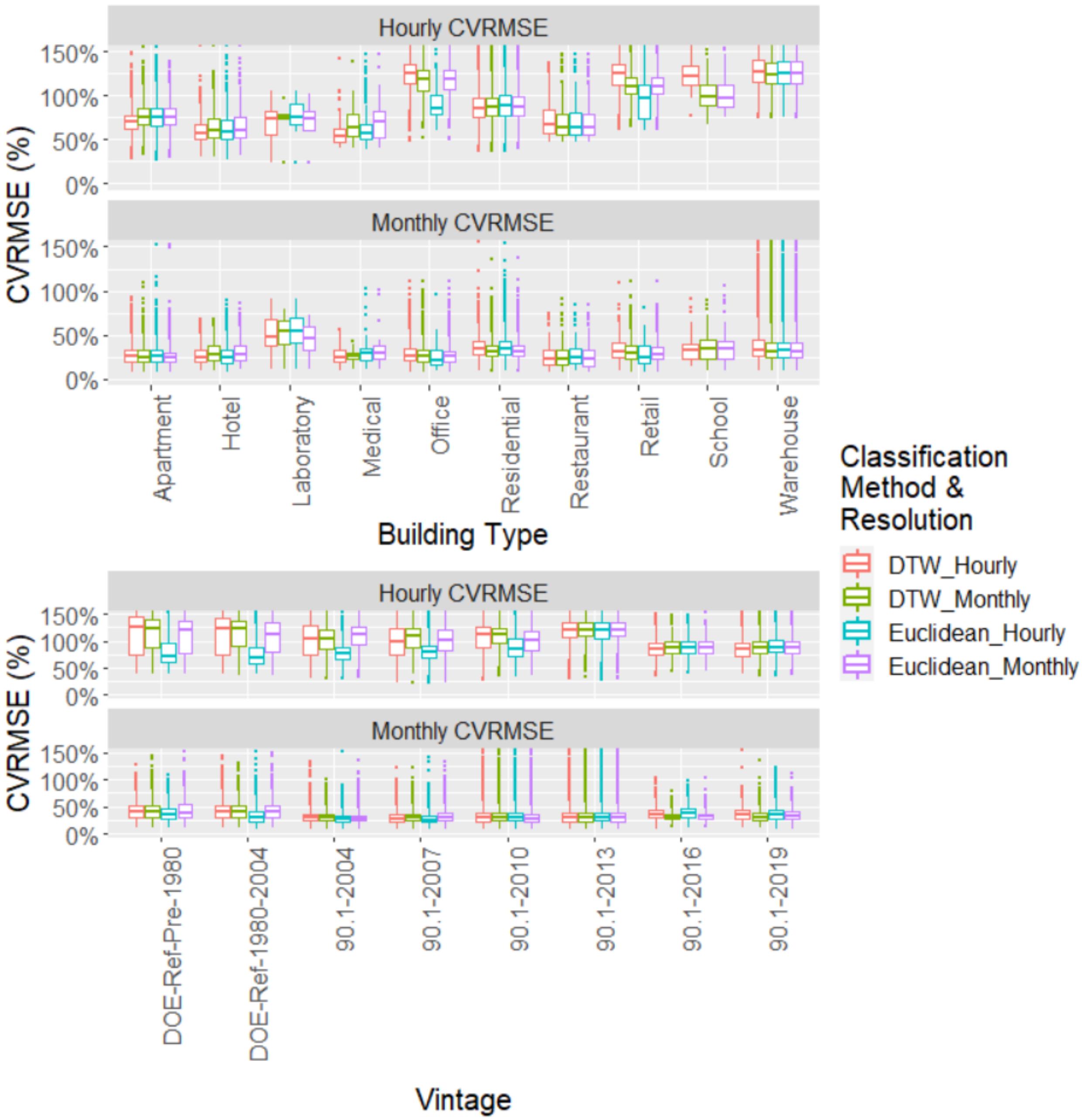

Figure 6: Warehouses have the highest CVRMSE but this may be caused by the large amount of buildings classified as warehouses. Medical and offices have the lowest. The Hourly Euclidean has the lowest hourly CVRMSE for most vintages.

Euclidean classification has the best hourly CVRMSE for all vintages 2010 and before by a significant margin. This shows why the hourly CVRMSE metric quality from Figure 5. This method seems to be the best at properly identifying the vintage of the building.

Conclusion

This work highlights the importance of assigning proper building type and vintage in UBEM. Measured building energy data can be used to infer building type and vintage by comparing the measured energy data to prototype building energy model simulations. Two comparison methods are evaluated including Euclidean distance and DTW. These methods are evaluated at hourly and monthly temporal resolution as high resolution data may not be available for all cases. This analysis utilized a sample of 46 thousand buildings that were filtered to 34 thousand buildings that were determined to have a full year of electricity data. These buildings were from the EPB service area in Chattanooga, TN. These buildings were generated as building energy models using the assigned building type and vintage from each method and temporal resolution and simulated to be compared to measured data to better understand their performance.

The building type and vintage assignments were considered for each method. The Euclidean and DTW methods classified a large number of buildings as warehouses. This is likely due to a combination of three primary factors: (1) Lack of natural gas data; (2) over-estimation of conditioned floor area; (3) missing or faulty data. This issue could be remedied using scaling or with additional data and could be addressed in future work.

CVRMSE was used as the evaluation metric for evaluating the quality of the simulation compared to the measured data. CVRMSE can be evaluated at an hourly or monthly level. The hourly classification resolution had better hourly CVRMSE values while the monthly classification resolution had better monthly CVRMSE values for each method. Euclidean and DTW had similar overall CVRMSE but the Euclidean comparison at hourly resolution was the best overall method. Adding that classification using DTW took about 100x longer than Euclidean, Euclidean distance was the best comparison method used in this analysis. As the temporal resolution of the measured data decreases, the hourly quality of simulation fit decreases. Still, the increase in CVRMSE from hourly classification resolution to monthly is low enough that it could be used in analyses. In fact, the difference is small enough that it is possible the annual measured EUI value may be compared directly to annual simulation data for classification data if only aggregated annual values are available. This is fully dependent on the type and level of granularity of a particular analysis. This could be considered in future work.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Building Technologies Office. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan.

References

ASHRAE. 2014. Guideline 14-2014: Measurement of Energy, Demand, and Water Savings.

Bass, Brett, Joshua Ryan New, Evan Ezell, Piljae Im, Eric Garrison, and William Copeland. "Utility-scale Building Type Assignment Using Smart Meter Data."

New, Joshua, Eric Garrison, Mark Adams, Brett Bass, and Tianjing Guo. 2020. "Scaling Beyond Tax Assessor Data."

Sakoe, H., and S. Chiba. 1978. "Dynamic programming algorithm optimization for spoken word recognition." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26 (1): 43–49.

Salvador, Stan, and Philip Chan. 2004. "FastDTW: toward accurate dynamic time warping in linear time and space." *Prcoeedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshop on Mining Temporal and Sequential Data*. ACM, 70–80.

US Department of Energy. 2019, December. User Quick Start Guide: Create DOE Prototype Building Measure.

US Energy Information Administration. Everywhere but Northeast, fewer homes choose natural gas as heating fuel.

US Energy Information Administration. 2015. RESIDENTIAL ENERGY CONSUMPTION SURVEY (RECS).

US Energy Information Administration. 2018, March. COMMERCIAL BUILDINGS ENERGY CONSUMPTION SURVEY (CBECS).