Utility-scale Building Type Assignment Using Smart Meter Data
Building Simulation 2021 Conference

Brett Bass[1], Joshua New[2], Evan Ezell[1], Piljae Im[2], Eric Garrison [1], and William Copeland [3]
[1]University of Tennessee, Knoxville, TN, United States
[2]Oak Ridge National Laboratory, Oak Ridge, TN, United States
[3]Electric Power Board, Chattanooga, TN, United States

Abstract

United States building energy use accounted for 40% of total energy use, 74% of peak demand, and $412 billion in 2019. Building energy modeling allows researchers to simulate building physics, gain insights into possible energy/demand saving opportunities, and assess cost-effective resilience amidst climate change. Many building features needed to create building energy models are readily available such as 2D footprints and LiDAR (height). A critical feature that is not generally obtainable is the building type. In partnership with a utility, a years worth of real-world, 15-minute electrical use data has been examined. The smart meter data is compared to 97 different prototype building energy models to assign building type. Real-world considerations including data preparation, quality assurance, and handling of missing values for advanced metering infrastructure data are addressed. Euclidean distance for pattern-matching of energy use, dynamic time warping, and time-window statistics with machine learning are compared for determining building type from measured electricity use.

Key Innovations

- Data preparation for real-world, sub-hourly electricity data
- Compare three novel methods for assigning building type using utility data
- Illustrate method for assigning confidence in predictions on an individual building basis

Practical Implications

This framework can be used to assign individual building archetypes at a utility level; scaling beyond where tax-assessors may be easily available.

Introduction

There has been increasing attention paid to energy in recent years as the effects of climate change are felt around the world. There are many ways to decrease energy use, demand amidst increased cooling loads, and building resilience for extreme weather events. In the United States in 2019, approximately 125 million commercial and residential buildings consumed about 39% of the nations primary energy (Energy Information Administration (2020)). Building energy use constitutes one of the greatest opportunities to mitigate or adapt anthropogenic forcings on climate change in the coming years. The United States Department of Energy (DOE) has made it a national priority to reduce energy use intensity (EUI) of the building stock 30% by 2030 compared to 2010 usage. Building energy modeling (BEM) allows researchers to quickly explore thermal and energy flows within buildings to inform investments in promising technologies, assess total energy and demand savings potentials for the US building stock, provide a level playing field of information and energy tools for the private sector, and close market gaps through field deployment to de-risk efficient building technologies and practices.

In recent years, the prevalence and accessibility of software tools, cloud services, and high-performance computing have led to a wide array of uses for machine learning and the ability to process large amounts of data. Among related emerging applications is urban-scale energy modeling which includes generating, simulating, and analyzing energy use in a way that scales from individual buildings to the scale of cities, utilities, or larger. These approaches often benefit from methods for obtaining building-specific descriptors (e.g. 2D footprints) from deep learning computer vision segmentation algorithms on satellite imagery. For this analysis, building footprints were obtained from the Microsoft building footprint dataset which contains over 125 million computer generated footprints across the country (Microsoft (2018)). In this project, the utility has smart meters collecting electricity use of buildings. This advanced metering infrastructure is referred to as premises, which were associated with buildings using Euclidean distance on the latitude and longitude of each premise and each building. Each building footprint was associated with one or more premises and the polygon of the footprint was used to calculate each building's 2D footprint area. Building height was obtained from a LiDAR data set that was developed over the course of five years for the state of Tennessee State of Tennessee (2019). The 2D footprint was extruded based on LiDAR processing for the height of the building and used as an estimate of total conditioned area. In partnership with the Electric Power Board (EPB) of Chattanooga, TN, 15-minute electricity data for calendar year 2015 from smart meters for each of 178,368 buildings was shared in order to analyze energy and demand savings opportunities. With the building 2D area, building height, and whole-building electricity use, the electrical portion of actual energy use intensity (EUI) was calculated.

Computation of EUI can be a guide for utilities in assessing demand or emissions reduction opportunities for energy-intensive buildings, but EUI doesn't provide sufficiently detailed metrics necessary to make utility-scale investment decisions. To make such data actionable, building energy models have the potential to provide energy, demand, emissions, or cost impacts dynamically at the scale of individual buildings, feeders, substations, or the entire service territory. While building footprint data is readily available today and LiDAR data for building height is becoming more available at reasonable costs, the geometry of the building must be populated with thousands of details to create a modern building energy model. There are ASHRAE audit processes (levels 1-3) with prioritized lists for collecting major energy use contributors in a building. However, both due to the intractability of such labor-intensive efforts at an urban scale combined with the privacy concerns associated with extracting some data automatically, accurately identifying the building type is an outstanding issue in the area of urban-scale energy modeling. Several countries have defined reference, prototype, or other canonical building energy models by common building functions (e.g. offices, schools) over different building code vintages based on surveys of energy-relevant characteristics (e.g. HVAC type, water heating, lighting, hours of operation) in these buildings. Such intelligent defaults are often necessary since modern simulation engines such as EnergyPlus, used in this study, has on average approximately 3,000 inputs (depending on size and complexity of the building) and an input/output reference manual with over 3,000 pages of everything that can be inserted or simulated/output from a building energy model. If building descriptors can be used to accurately assign building type, all internal characteristics can be assigned as an intelligent guess, used to generate answers to initial investment-grade questions, and refined as necessary. This paper explores the use of 15-minute measured electrical use of each building as one building descriptor which, if available, could be used to assign building type in a useful manner.

Obtaining the inputs about individual buildings physical characteristics may be easy on a county-sized scale with tax assessor's data, but is more difficult to obtain as the number and geographical regions of buildings increase. Simulations can be performed with relatively few parameters if major internal, energy-use characteristics are assumed from prototype buildings. The more input information available to generate the model, the more accurate the simulation. DOE's prototype buildings are a set of typical residential and commercial building models derived from the DOEs earlier commercial reference building models. The 16 different commercial prototype models represent approximately 80% of commercial buildings in the U.S. (US Department of Energy (2019a)). There are also residential prototype buildings that cover a large number of residential variation. The prototype models provide a consistent baseline of comparison for actual buildings of these types (US Department of Energy (2019b)). As one would expect, the building type assignment to a prototype has a large effect on the type of equipment (sizing and schedules), envelope characteristics, occupancy, and many other related characteristics that determine the anticipated energy use or energy savings opportunities from that BEM. If the simulations are of sufficient quality, a cogent digital twin of buildings at city- or utility-scales can be generated, simulated, analyzed, and intuitively visualized. A digital twin of a utility can facilitate infrastructure-scale investments from the quantified energy, demand, cost, and emissions reduction opportunities for relevant buildings.

Each DOE prototype building, in addition to having a building type, has a vintage based on the building codes relevant at different years. While building codes in the U.S. are updated every three years, each state or local region elects which code-release year they choose to enforce: the ASHRAE (i.e. 90.1, 90.2), International Energy Conservation Code (IECC), or some variant of those standards within its geographical region. Each new code release typically requires higher levels of energy efficiency, with current building codes requiring a building that uses approximately 40% of the energy than they did in 1985 (Liu and Athalye (2018)). For urban-scale energy modeling of buildings, building type and vintage are an integral part of the simulation with much of the accuracy or value of the resulting analysis deriving from the accuracy by which building type and vintage are assigned.

The impact that building type and vintage have on simulation results are shown in Figures 1 and 2 for climate zone 4A (Chattanooga). Each building type and vintage combination was simulated using the same geometry to illustrate the impact these parameters have on simulated annual energy use. The difference in electricity use across building types is significant with warehouses using less than 500 GJ annually whereas hospitals use 4000 GJ annually with the same geometry. These findings change across different climate zones throughout the U.S., but this is representative of building energy models in Chattanooga and the EPB service area.

For EPB's service area of 178,368 buildings, data for footprint and height were used to determine estimated conditioned square footage and calculate electrical use intensity from measured smart-meter data. This was then compared to all possible building type and vintage combinations to determine which annual, 15-minute electricity use intensity pattern most closely matched each building's actual electricity use pattern. An example of a real building compared to two sample prototype buildings is shown in Figure 3. Three methods of assigning building type and vintage labels to each of the buildings were compared. The first method is calculating Euclidean distance between each sample building 15-minute EUI to the 97 prototype combinations 15-minute EUI with the shortest distance being the label. The second method uses the minimum dynamic time warping (DTW) distance to select the appropriate building label. The last approach is to create a machine learning (ML) model from time-based statistics from the 15-minute EUI to predict labels. These methods are described and compared in greater detail within the Methods and Results sections.

A critical piece of this analysis was how to handle missing data from the real EPB 15-minute electricity. Premises had differing amounts and different types of missing data. Missing data was ultimately categorized into three different types. The first type of missing data was small gaps. Small gaps were considered less than one week. Small gaps were dealt with using autoregressive integrated moving average (ARIMA) in which ARIMA was fitted to the signal and the small missing gaps were imputed. The next type of missing data was large gaps. These large gaps were greater than one week but less than three quarters of the year. Large gaps were either left out (where imputation was unnecessary), or they were imputed using univariate DTW imputation. The last type of missing data was premises that had more than 75% of their data missing. These files were removed as classifying buildings on that small amount of data could lead to error. Since results can vary significantly based on the methods of preliminary data preparation, our methods are disclosed within the following Data section.

Data

The real 15-minute electricity data was provided by EPB Chattanooga. The data covered more than 178,000 premises while this study focuses on a small sample to test out several methods. The electrical energy values were reported in average kilowatts (kW) consumed during each 15-minute interval with the sum of four consecutive values resulting in kilowatt-hours (kWh). With this real, raw data, there were several issues with the data and its format which we attempt to succinctly describe and address here.

First, the data was provided in a streaming format at a resolution of 15-minutes from each utility-grade meter as it was available. This meant that the reporting and interleaving of data use information needed to be assembled into time-series for each unique meter's ID. The result of the initial pre-processing phase was a spreadsheet with 178,368 rows and 35,040 columns of 15-min energy use (kW) of each building's electrical use for calendar year 2015. Second, there was missing data for many of the 15-minute intervals for many of the premises. While previous work has reported over 93% of buildings were missing less than 2% of their data, out of the small sample that was analyzed for this comparison, none of the premises had "full" sets of data. Third, some data is incorrectly formatted, being either a wrongly formatted date/time string or a non-numeric value for energy use. Fourth, as this data represents real electricity use, business operations occur in the middle of time series that can raise complications for data processing, such as customers changing rate structures, meter swapouts with different IDs mid-year, or stopping power to a building. Duplicate reporting was also present in the data. Properly dealing with such issues is a critical part of any analysis.

The 97 combinations of prototype building and vintage combinations were simulated using EnergyPlus and climate zone ASHRAE-169-2006-4A building codes and actual meteorological year (AMY) weather data for the year corresponding to the year of the EPB data. Table 1 contains the prototype building types and standards. The 15-minute electricity data output was measured in Joules and converted to kWh for comparison to the EPB data. For this analysis, and to address confusion between electrical and energy use intensity, all building energy models were converted to all-electric HVAC for comparison purposes.

Missing Values

Many researchers underestimate the time required for data preparation. While this may vary dramatically by project, total project time spent on data preparation often ranges from 70 - 92% (KDnuggets (2003)). Likewise, preparing data files was a major part of this analysis. The prototype files were outputs from the simulation and therefore did not need cleaning. For comparison, the real electricity data had to be cleaned. The distributions of the number of missing values and the number of 0 values of the sample of 100 premises is shown in Figure 4.

Many of the premises had a relatively low percent

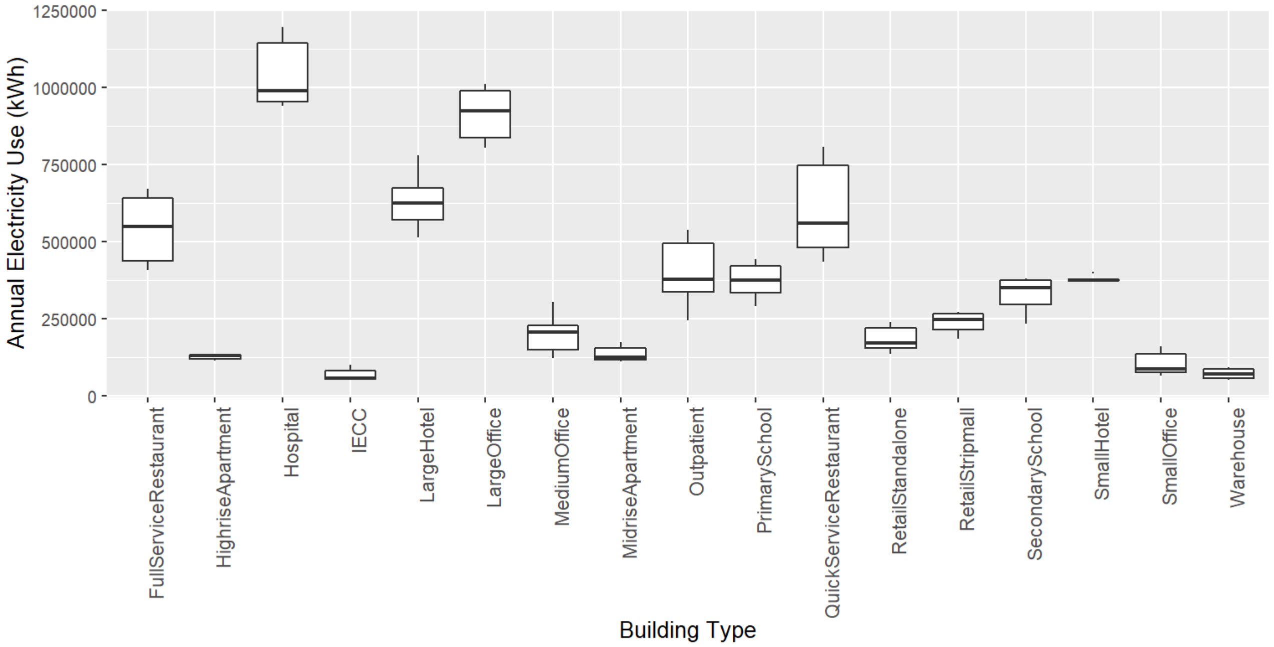

Figure 1: Building Type has a significant impact on annual electricity use with some buildings using orders of magnitude more electricity for the same geometry.

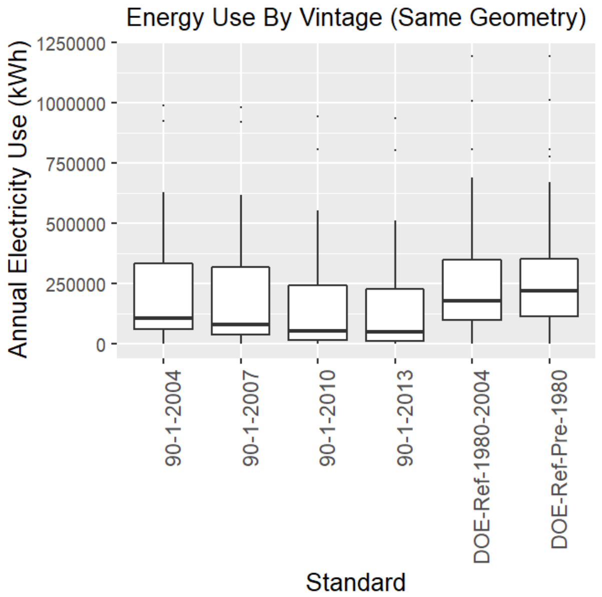

Figure 2: The building vintage affects electricity use with older vintages using more energy than newer ones.

of “NA” or “0” values; however, there are several premises in which much of the data is missing or 0, including several premises that had values of 0 for the majority of the year. First, it is important to note that 0s are treated differently than NA values. Only gaps of 0 values greater than one week were treated as missing values and thereby imputed. Any zero value that contained values greater than 0 within a week before or after was kept in the data while these longer strings of 0s were treated as missing. It is impossible for one to know if these longer than one week 0 strings are legitimate values or some sort of sensor or recording failure during these periods of time. The reason they are being treated as missing is because these long 0 strings will not compare to the prototype buildings which have no missing or 0 data. The first thing that was done was to categorize the missing data into three types.

Table 1: Every valid combination of Building Type and Standard were used to create 97 prototype models.

Building Type	Standard
Small Office	DOE-Ref-Pre-1980
Medium Office	DOE-Ref-1980-2004
Large Office	90.1-2004
Standalone Retail	90.1-2007
Retail Stripmall	90.1-2010
Primary School	90.1-2013
Secondary School	
Outpatient	
Hospital	
Small Hotel	
Large Hotel	
Warehouse	
Quick-service Restaurant	
Full-service Restaurant	
Mid-rise Apartment	
High-rise Apartment	
Residential	

Small Gaps

This study defines small gaps to be less than one week of consecutive missing values. If these small

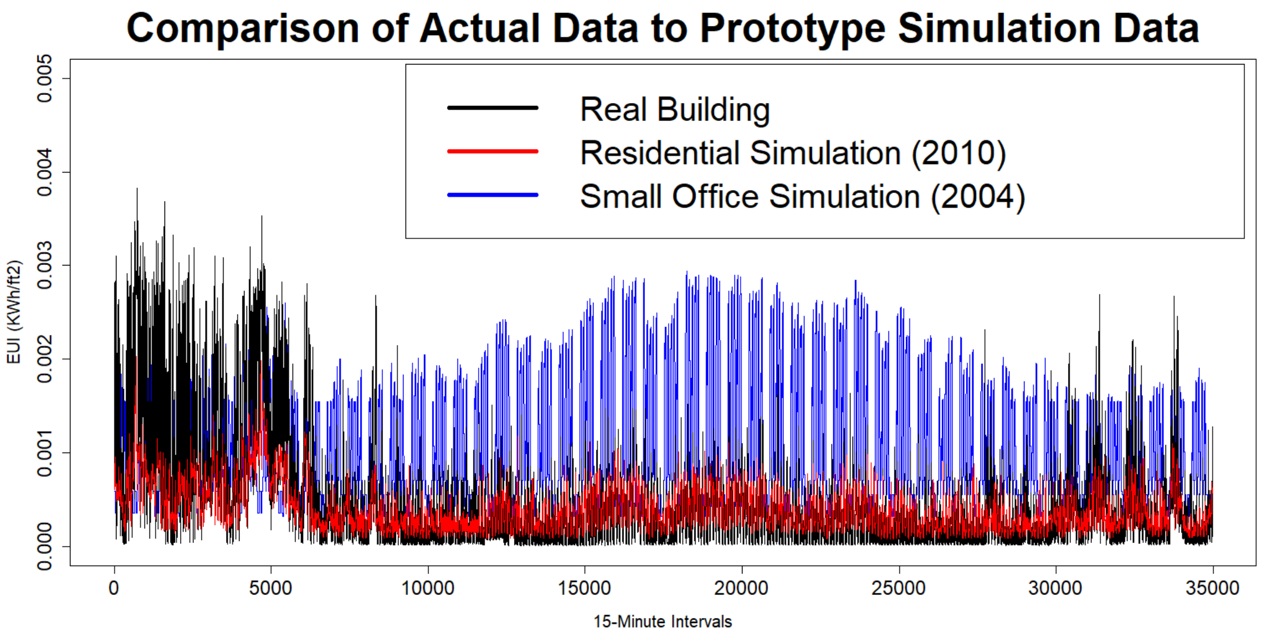

Figure 3: Comparing real building 15-minute EUI to two prototype simulations shows this building more closely matches a residential prototype rather than the office prototype. In this study, every building's real data is compared to all 97 Building Type and Vintage combinations.

gaps were 0 values, they were left in the data. If these small gaps were missing values, they were imputed using Auto Regressive Integrated Moving Average (ARIMA). ARIMA models forecast a time series based on past values. The "Autoregression" (AR) refers to model that regresses on lagged values. "Integrated" (I) represents the differencing of raw observations to allow for the time series to become stationary. And "Moving Average" (MA) incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations. These combined features attempt to fit the data and forecast future points (Hyndman and Athanasopoulos (2020)). An ARIMA model was fit to each of the series of 15-minute data for each premise. The model was then used to forecast the small gaps throughout the 35,040 points of the year. ARIMA was a good choice for this imputation as it can naturally handle missing data and provided the imputation (forecast) with more depth than interpolation various other methods that were explored. An example of the ARIMA imputation is shown in Figure 5.

Large Gaps

This study defines large gaps to be any number of consecutive missing values or 0s greater than one week and less than three months. These gaps were dealt with in two different ways. Both ways started out the same in which small gaps were imputed as previously shown using ARIMA. Once the small gaps were imputed, a copy of the premises was stored with NA values remaining for DTW which will be shown later. Another copy of the premises was imputed using univariate dynamic time warping for large gaps of missing data. ARIMA and several other imputation methods were attempted for the large gaps, but most could not handle this much missing data in a row and the result was imputation that looked nothing like its surrounding points. An example of the univariate dynamic time warping imputation is shown in Figure 6.

The three types of gaps as well as their respective imputation strategy are shown in Table 2. This overall imputation methodology allowed for a full comparison of 15-minute prototype simulation EUI to 15-minute actual building EUI for classification of building type and standard.

Table 2: Different imputation strategies were used for different gap sizes in the meter data.

Missing Data Type	Imputation Strategy
Small Gaps (< 1 Week)	Auto Regressive Integrated Moving Average (ARIMA)
Large Gaps (> 1 Week)	Univariate Dynamic Time Warping (DTW)
> 75% Missing	Omitted

Methods

Euclidean

The first and most straightforward method of comparison of the real 15-minute EUI data to the prototype 15-minute EUI data was measuring Euclidean distance between the EPB sample to each of the 97 prototype simulation combinations. For this analysis,

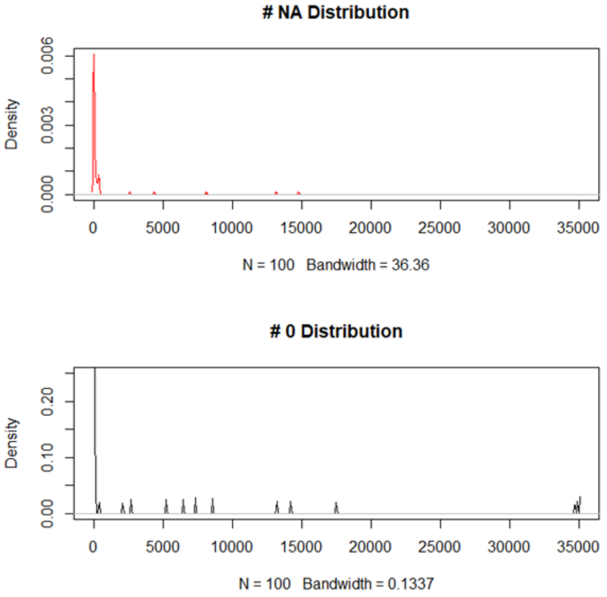

Figure 4: Most meter data had less than 1000 “NA” or “0” values, some had significant gaps, and a few had nearly all “0” values. Different strategies were used to address these.

the previous sections about imputation were ignored as time did not factor into this method. This resulted in a comparison of however many points were in the EPB sample to the same number of points from each of the prototypes. The prototype and standard combination with the smallest distance to each of the observations was chosen as the label for that observation. This method has been previously analyzed (Garrison et al. (2019)).

Dynamic Time Warping

The next method of comparing the EPB data to the prototype combinations was DTW. DTW is a commonly used measure of the similarity between two time series. DTW works by finding the optimal global alignment between two time series accounting for temporal distortions. The algorithm optimally maps one time series onto another and similarly to Euclidean by comparing each point in one time series to every other point and returns the warped distance as a result. By doing this, even if time series are not exactly in phase, their points are compared and the warping distance would be small. This method may be a good fit for electricity data as the same patterns may occur at different points throughout the year. As one would expect, this vast comparison is very computationally expensive (quadratic time and space complexity) and many modifications have been made in an attempt to expedite this process. For this analysis, an approximation called ”FastDTW” was used (Salvador and Chan (2004)). A comparison of Euclidean to DTW is shown in Figure 5. The DTW warps to another section of the time series and maps similar queries together which may result in a better match. DTW cannot be used directly on time-series with missing data. The data either had to be omitted or imputed. For this analysis, the missing data was imputed for comparison using the small and large gap strategy previously described.

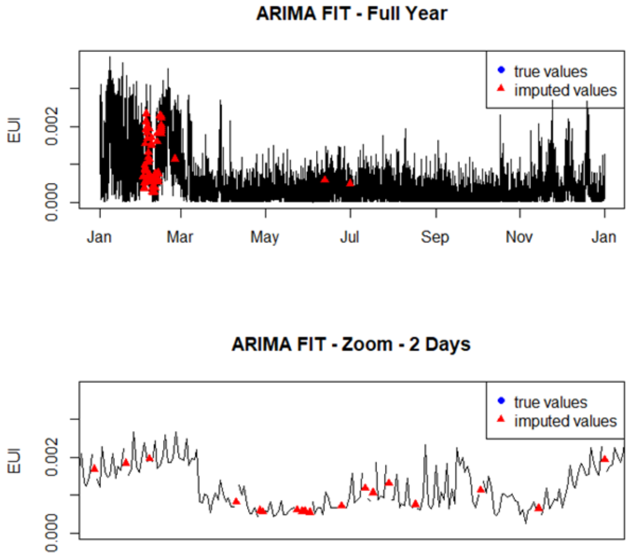

Figure 5: Example of Small Gap ARIMA Imputation, shown for a single building at full year and 2-day scales, qualitatively illustrates filling in logical values within the time series data.

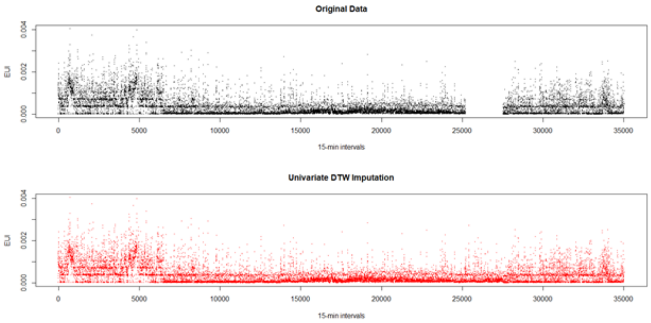

Figure 6: Example of Large Gap Univariate DTW Imputation, shown for a single building at full year scale, demonstrates original data (top) filled with data based on the imputation strategy (bottom) in a way that preserves characteristics of the time series.

Windowed Statistics Machine Learning

The final method of labeling building type and standard was using a Machine Learning (ML) classifier. The ML classifier was done in a different way than the previous two methods. Instead of directly comparing the data to the prototypes, this method extracted time and statistics-based features from the data with the prototypes considered the labels. The Caret package in R was used to build, tune, and compare these models (Kuhn (2019)). The first thing that had to be done was to impute the data. Similar to the second DTW method, time series small gaps were imputed using ARIMA and large gaps were imputed

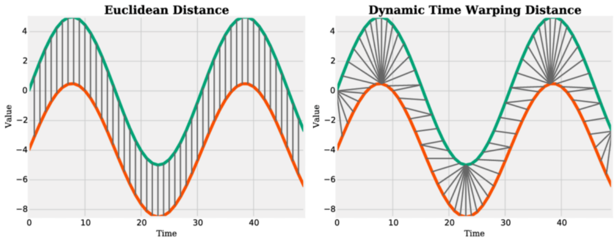

Figure 7: Euclidean Distance vs Dynamic Time Warping (Schfer (2015)). Sine curves demonstrate Euclidean distance may not adequately compare two time series.

using univariate DTW. Several time-based statistics were extracted from the time series. They are shown in Table 3 below.

Table 3: Combinations of Time windows and Statistics were calculated to identify unique statistical variations between buildings at different time scales. For example, weekends can traditionally serve as a significant differentiator between building types.

Time Window	Statistic
Monthly	Maximum
Yearly	Mean
Weekends	Median
	Minimum
	Standard Deviation

These time windows were chosen as they summarize critical structures of the time series. For example, one would expect the EUI of a large office on the weekend to be different than normal and completely different when compared to other building types. Weekly windows were originally used but removed as they resulted in lower cross-validation metrics. Three different models were evaluated with a hyperparameter grid search being used for each to compare optimal models for each. These models were k-nearest neighbor (KNN), random forest (RF), and extreme boosting (xgbTree). KNN is a classifier that works by assessing the distance of a test vector to all training vectors with the label being the vector at which distance is minimized (Ripley (1996)). xgbTree and RF are both decision tree classifiers which work by recursively partitioning data based on feature values for which each of the partitions serves as a test for on a feature of test data (Quinlan (1986)). Boosting (xgbTree) relies on shallow trees for which error is minimized by minimizing bias while RF utilize fully grown decision trees and minimizes error by minimizing variance (Chen and Guestrin (2016); Breiman (2001)).

As there were 97 different classes with one observation per class, cross-validation could not be done with the raw labels. Instead, the labels were changed to building type only (removing vintage), thereby incorporating at least 3 labeled observations (6 for most) into the training data set which allowed for classic cross-validation to get a rough estimate of what the hyperparameters should be to split the building types. The random forest ultimately had the highest classification accuracy and was the final model used to create the building energy models for the EPB samples. The hyperparameter grid values are shown in Table 4 while the cross-validation results are shown in Table 5.

Table 4: Hyperparameter values used for grid search are shown with optimal hyperparameter values from cross-validation in bold. For more on these metrics, see Ripley (1996), Schliep et al. (2016) Breiman (2001), Chen and Guestrin (2016).

Method	Hyperparameter	Value
KNN	Kernel	**Rectangular**
		Gaussian
		Triangular
		Epanechnikov
	Kmax	30, 40, **50**, 60
RF	Trees	**500**
	Mtry	2, **125**, 390
	Min Node Size	**1**
	Split Rule	Gini
		Extra Trees
xgbTree	N rounds	50, **100**, 150
	Max Depth	**1**-3
	Eta	**0.3** -0.5
	Gamma	**0**
	Col Sample By Tree	**0.6** , 0.8
	Min Child Weight	**1**
	Subsample	**0.5** , 0.75, 1

Table 5: Cross-validation metrics for KNN, xgbTree, and RF. RF had superior mean and median classification accuracy as well as κ, which considers the chance of randomly classifying correctly.

Method	Median Acc	Mean Acc	Median κ	Mean κ
KNN	78.4%	80.1%	77.1%	78.8%
RF	84.3%	82.2%	83.3%	81.1%
xgbTree	80.3%	81.0%	79.0%	79.7%

Results

For this analysis, three different datasets were created with building type and standard being classified from one of the methods. The models were then simulated using EnergyPlus and compared to the actual data to obtain accuracy metrics. There are two ways of comparing the methods; error rates of simulated electricity to actual electricity usage and assigned building type compared to actual building type. Actual building type was found manually by searching each of the 100 buildings. Building electricity data was adjusted using a single, annual adjustment factor. Both of these metrics may be valuable depending on the goal of an analysis.

Runtime will become an increasingly important factor as these results are scaled beyond 100 buildings. The Euclidean distance calculation was the fastest with a runtime 25x faster than the very slow dynamic time warping which must compare all sets of points. This runtime gap could be improved with additional numerical methods. The random forest was took only a few seconds to train as it only needed to be trained on 97 samples.

Quantitative Summary

Coefficient of Variation of Root-Mean Squared Error (CVRMSE) is a quantitative metric used for building energy modeling that measures error between simulation output and real data. The equation for CVRMSE is shown in equation 1.

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

It should be noted that CVRMSE is often computed for building energy models on monthly or hourly data for a year, whereas these numbers are computed for 15-minute data over a year. Also, missing data for this calculation was addressed by omitting "NA" values and the aforementioned imputation strategies. The performance metrics based on comparison of the resulting BEM and measured data is shown in Table 6.

Table 6: While all three methods tested have similar error values, random forest demonstrates the lowest maximum error. For assessment of these quality results, <15% monthly or <30% hourly CVRMSE are often considered "investment grade" and errors here are 15-minute resolution. (ASHRAE (2014))

Method	Min	Median	Mean	Max
RF	.7%	38.6%	44.1%	206%
Euc	.5%	38.5%	48.6%	545%
DTW	.5%	38.7%	49.1%	560%

Qualitative Summary

Qualitative results are sorted into three categories; direct accuracy, general accuracy, and commercial accuracy. Direct accuracy was determined by comparing the exact prediction with the actual building type. This was difficult to classify in some instances as a church or a car dealership could not be directly classified into building prototypes. General accuracy corrects this issue slightly by classifying actual buildings into their closest representative prototype building as well as combining categories such as small, medium and large office into a general "office" label. The final category is commercial accuracy which is simply residential (detached) or commercial (other). The accuracy values are shown in Table 7.

Table 7: Euclidean distance classifier demonstrated the highest accuracy, primarily due its number of residential predictions coupled with the amount of residential buildings in the sample.

Method	Direct	General	Commercial
RF	62%	63%	78%
Euc	80%	80%	81%
DTW	71%	71%	77%

While accuracy is a reasonable metric, it can be a bit misleading given the dataset was comprised of about 80% residential detached houses. The ability of these classifiers to differentiate between residential buildings and commercial buildings is important as it has a large impact on building properties and energy use. Though this is a multi-class problem with 17 different buildings types, an estimate of the binary commercial classification quality can be obtained by simplifying the actual building type and the prediction to commercial or residential. As the simplification to commercial vs residential is done in a post-processing step, a single decision threshold is used.

Sensitivity (true positive rate - method predicts commercial and building is commercial) and specificity (true negative rate - method predicts residential and building is residential) are useful metrics for binary classification exercises as they highlight class imbalance issues. The sensitivity and specificity values are shown in Table 8. While the Euclidean classifier had the best direct accuracy of the three methods (Table 7), it's over-prediction of residential in a highly residential dataset is problematic as it shows the classifier does not have the ability to separate these important building distinctions. In contrast, the lower direct accuracy of the Random Forest (Table 7) may be caused by certain real commercial buildings in the dataset that behave more closely to a different commercial prototype, likely making these predictions more representative than a residential classification.

Table 8: Random Forest was best at differentiating commercial vs residential buildings while Euclidean Distance over-predicted residential.

Method	Commercial Sensitivity	Commercial Specificity
RF	78.9%	78.3%
Euc	0.05%	100%
DTW	36.8%	87.8%

There are some other interesting observations in the qualitative building classifications. For the Euclidean distance classification method, the large number of residential building assignments could be explained if the 15-minute electricity data for an apartment building was for a single unit which may closely resemble a small house and would be the closest classification. The number of warehouses classified is interesting as there were no actual warehouses in the dataset. This would lead one to believe that some of the houses electricity signature functioned more closely to a warehouse than a house based on scheduling an electricity use. This could also be due to the assumption that all buildings were 100% electric HVAC; if a building was heating with gas, it may more resemble a warehouse.

It should be noted that no post-processing was done for any of these methods. If 15-minute electricity is available, often billing rates may be available and may be used to change a building classification with awareness that assigning the correct building type may not lead to a more representative energy simulation.

Confidence

All of these methods utilize some sort of distance metric from Euclidean distance, to warping distance (DTW), and class probability (RF). These distances can be viewed as a pseudo-confidence factor with the premises with the smallest EUI distance (i.e. highest similarity) to prototypes being labels of the highest confidence. This confidence level allows the utility to determine to what degree they trust certain labels as well as if some buildings need to be labeled in another way if the distance is out of the normal range for that method. For visualization, the distances for each method were scaled between 0 and 1 and are shown in Figure 8.

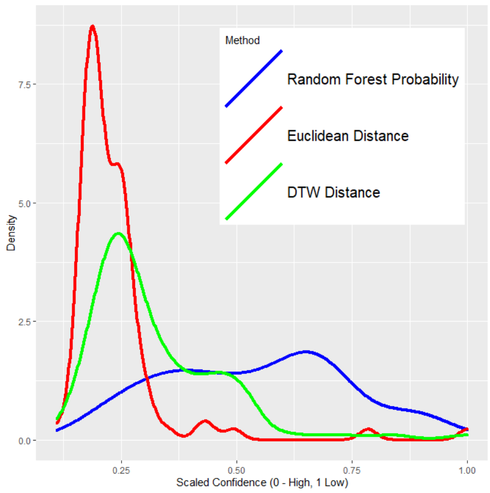

Figure 8: Distance-based classification methods perform well at finding a close EUI signature match for building types whereas the RF classifier is least confident with its building type assignment.

For both the Euclidean distance and DTW distance methods, the majority of prototype-actual building matches has a small distance that increases for several observations. The observations with the smallest distance would be the predictions with the highest confidence. The shapes of these two methods is different than the random forest because of the way they are calculated. For this dataset, the maximum random forest probability was 37.4%. The reason the percent was this low is because of the similarity between prototype building vintages within building type bins which decrease the maximum probability per class. For example, the EUI signature of a 2010 small office may be similar to a 2007 small office; leading to a split in probability voting. A confidence could still be used for this method, but one would have to consider the top probability classes to ensure a high-confidence prediction.

If was one to filter the dataset to only predictions below the mean, and below the first quartile distance for each method, the error rates would be expected to decrease. These scenarios are shown in Tables 9 and 10. The increased filtering does limit the number of buildings in each; taking the average to 45 buildings for the mean filter and 20 buildings for the quartile filter.

Table 9: The mean and median CVMRSE improves for Euclidean and DTW methods but worsens for RF compared to Table 6.

Method	Min %	Median %	Mean %	Max %	Acc %
RF	18.7	45.8	51.8	138	71
Euc	5.5	35.2	37.5	78.4	97
DTW	20.2	35.2	41.3	206	87

Table 10: When filtering data for higher confidence levels, mean and median CVMRSE improves for the Euclidean and DTW methods but worsens for the RF compared to Table 9.

Method	Min %	Median %	Mean %	Max %	Acc %
RF	18.7	48.1	57.2	138	59
Euc	5.5	29.9	31.6	66.8	94
DTW	20.2	29.9	32.4	55.4	100

These results are generally as expected with a decrease in CVRMSE for the Euclidean distance and DTW methods but the CVRMSE increases for the random forest method. This leads one to believe that the random forest probability may not be as effective as measuring confidence as the distance methods or building type probabilities across vintages need to be included.

Conclusion

Assignment of building type and vintage is currently an outstanding challenge in the emerging area of urban-scale energy modeling. This study leveraged 15-minute whole-building electricity use and building energy models of 97 prototypes to assess data preparation and algorithmic accuracy for accurately assigning building type and vintage. Omission of large gaps (>75% of data), Auto Regressive Integrated Moving Average (ARIMA) for filling small gaps (<1 Week), and Univariate Dynamic Time Warping (DTW) for filling large gaps (>1 Week) was found to be effective in this study. Three building type classification methods were compared involving Euclidean distance, DTW, and machine learning with random forest and time-based statistics. Euclidean distance was the fastest and had the best overall classification accuracy, whereas the random forest performed better for commercial buildings. The run-time of DTW would be a significant hindrance as the number of buildings to be classified increased. For each method, a pseudo-confidence was obtained via the similarity distance (Euclidean, DTW) or the class probability (RF). The distance metrics proved to yield higher-confidence building type assignments with lower error metrics. The imputation strategies and building type assignment methods demonstrated in this paper result in building energy models with error rates comparable to ASHRAE Guideline 14 requirements.

This study suffers from several limitations. Most of this work derives directly from the use of sub-hourly, whole-building electricity use information at utility scales. Many utilities do not collect sub-hourly data, and any such energy data is rarely shared outside the utility (esp. at the building-specific level). The authors are hopeful that responsible data sharing and the demonstrated value of this data, both for utilities and building owners, will in some small part help make such data more prevalent in the future. While lacking comparison to other quality assurance and building type classification methods, the authors provide quantitative results using industry-standard metrics to facilitate apples-to-apples comparison to other techniques toward the establishment of best practices for urban-scale energy modeling.

Acknowledgment

This work was funded by field work proposal CEBT105 under US Department of Energy Building Technology Office Activity Number BT0305000, as well as Office of Electricity Activity Number TE1103000. The authors would like to thank Amir Roth and Madeline Salzman for their support and review of this project. The authors would also like to thank Mark Adams for his contributions to the AutoBEM software for model generation, and Jibonananda Sanyal for AutoSIM contributions for scalable simulation.

References

ASHRAE (2014). Guideline 14-2014: Measurement of energy, demand, and water savings.

Breiman, L. (2001, October). Random forests. *Mach. Learn. 45*(1), 532.

Chen, T. and C. Guestrin (2016, Aug). Xgboost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Energy Information Administration (2020, December). How much energy is consumed in u.s. residential and commercial buildings?

Garrison, E., J. New, and M. Adams (2019). Accuracy of a crude approach to urban multi-scale building energy models compared to 15-min electricity use.

Hyndman, R. and G. Athanasopoulos (2020). *Forecasting: Principles and Practice.*

KDnuggets (2003). What percent of time in your data mining project(s) is spent on data cleaning and preparation?

Kuhn, M. (2019, March). Caret.

Liu, Bing, R. M. and R. Athalye (2018). National impact of ansi/ashrae/ies standard 90.1-2016. *2018 Building Performance Analysis Conference and SimBuild (BPACS) co-organized by ASHRAE and IBPSA-USA.*

Microsoft (2018, July). Microsoft building footprints.

Quinlan, J. R. (1986, March). Induction of decision trees. *Mach. Learn. 1*(1), 81106.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press.

Salvador, S. and P. Chan (2004). Fastdtw: Toward accurate dynamic time warping in linear time and space.

Schliep, K., K. Hechenbichlier, and A. Lizee (2016, August). Package kknn.

Schfer, P. (2015, October). Scalable time series similarity search for data analytics.

State of Tennessee (2019, December). Elevation lidar - a coordinated effort with the us geological survey.

US Department of Energy (2019a). Commercial prototype building models.

US Department of Energy (2019b). Residential prototype building models.