Predicting Building Envelope Construction from In-Situ Thermal Testing

Tyler James Pilet$^{1,2,*}$ and Tarek Rakha$^2$

$^1$Pacific Northwest National Laboratory, Richland, WA
$^2$Georgia Institute of Technology, Atlanta, GA

Abstract

When embarking on a retrofit of a building envelope, it is extremely important to understand the composition of the assembly. This practice is currently done by destructive and invasive material testing or demolition, which is sometimes not possible when in historic or protected buildings. To address this problem, in-situ thermal testing can be utilized along with machine learning classification algorithms to infer the composition of an assembly. In this paper, a proof-of-concept K-nearest neighbors classification model is developed to classify assembly composition from effective thermal resistance, effective thermal mass, and assembly cladding. This model was trained and tested utilizing a synthetic dataset producing an F1-score of 94.6%. This model was also validated with experimental data from a 100-year old wall assembly, confirming the model’s real-world validity. This study serves as a framework for inferring as-built envelope assembly composition, all without having to damage or disturb the building.

Introduction

In envelope retrofits, it is important to understand the composition of an existing building’s envelope before embarking on the retrofit decision-making process. With 50% of the aging US building stock constructed before 1980 (IEA, 2019), there is a significant number of buildings requiring energy retrofits in the coming future. In practice, it is difficult to work with existing buildings due to missing documentation and uncertainties in the composition and integrity of a building’s envelope. To solve this problem, destructive or non-destructive forensic testing is employed during the building retrofit and repair process. While simple and obvious, destructive testing cannot provide information without damaging the building and disturbing occupants. Additionally, destructive testing may not be possible for in-use or essential buildings, leaving few ways to identify envelope performance. In these cases, thermal testing, namely R-value testing, can be employed to non-intrusively characterize an envelope’s thermal properties and quantify the performance of as-built or deteriorating assemblies.

Currently, R-value testing is the industry-standard method of gathering information on in-service envelope assemblies. Transient thermal characterization, which characterizes assembly thermal resistance and effective thermal mass, is also a method frequent in literature but has not yet grown into a mainstream industry practice (Pilet & Rakha, 2020; Šuklje et al., 2019; van Schijndel, 2009). The advantage of this method is the ability to measure thermal mass, alongside thermal resistance, to better appreciate the time-dependent performance of an assembly.

While R-value testing is typically performed to generate model inputs for thermal modeling, building energy modeling, or retrofit decision making, this testing data can serve another purpose—classification modeling. Machine learning classification algorithms, such as K-nearest neighbors (KNN) (Altman, 1992), random forests (Breiman 2001), and neural networks (Wan, 1990), can be utilized to correlate physical parameters of envelope assemblies to assembly construction types to enable more informed retrofit decision-making. Research utilizing machine learning classification algorithms is becoming more popular in the building industry; however, the bulk of these applications are mostly relegated to computer vision and geometry reconstruction applications (K. Chen, Reichard, Akanmu, et al., 2021; K. Chen, Reichard, Xu, et al., 2021; Deeb & LeWinter, 2018; Park & Guldmann, 2019; Rakha et al., 2018). While geometry reconstruction and feature recognition do provide context to a building’s structure, these advances have little to do with predicting what is inside of an envelope assembly.

Currently, there are a few methods to infer the make-up of an envelope assembly—(1) destructive testing, (2)
borescope drilling, (3) thermography, and (4) ultrasound or penetrating radar techniques. Outside of destructive testing, each of these methods has one major commonality—the requirement of professional judgment. Each of these techniques can provide context to the construction’s performance or structure, but no matter how much context is provided via modern in-situ testing techniques, no technique can provide context on the material layering of an envelope assembly. To remove the requirement of experience and professional judgment in this situation, machine learning can be utilized alongside in-situ thermal testing to infer assembly make-up. Measured thermal resistance values, along with effective assembly thermal mass values measured from transient thermal characterization, can be leveraged alongside machine learning techniques to infer assembly make-up. The details of this procedure will be described in subsequent sections.

Methods

This paper proposes a proof-of-concept methodology to infer the materiality of a building envelope assembly from in-situ testing data. This methodology’s framework is displayed graphically in Figure 1. From Figure 1, it can be noted that the proposed methodology has three major stages: (1) in-situ data collection, (2) classification, and (3) interpretation of results. It is proposed that this methodology utilize in-situ thermal data, namely effective thermal resistance and effective thermal mass, alongside a visual indication of assembly cladding, to infer the materiality of the candidate assembly. This model is built such that the input thermal resistance and thermal mass can be stochastic or deterministic, with stochastic inputs of N size producing N number of classifications and deterministic inputs producing singular, deterministic outputs. From there, the N outputs can be counted and presented as percentages, which can be examined and grouped to infer the assembly’s construction make-up. Further details on each stage of the proposed methodology will be further outlined in subsequent subsections.

In-Situ Assembly Data

As previously discussed, thermal resistance and thermal mass will serve as inputs to this classification model. These metrics can be computed via many different approaches, but the selected approach must be per ASTM C1046 measurement techniques (ASTM, 2013a). For this paper, a Bayesian inference methodology adapted from that which was proposed by Pilet and Rakha will be utilized for the computation of in-situ thermal data (Pilet & Rakha, 2020). The advantage of utilizing Bayesian inference for the computation of in-situ thermal data is that Bayesian inference is a stochastic approach that will produce a distribution of thermal metrics via Markov chains. This differs from deterministic methods, such as that of ASTM C1155 (ASTM, 2013b), which would output a singular thermal resistance and a singular thermal mass value which resulted in the best non-linear fit between measured data and the thermal characterization.

Classification Model and Model Training Data

For the task of classifying assemblies from in-situ data, many multi-dimensional classification algorithms can be utilized. When accounting for the categorical data cladding information alongside the thermal resistance and thermal mass features, this becomes a mixed-feature classification problem, which narrows the number of applicable classification algorithms. Support vector
molecules, neural networks, decision trees, and $K$-Nearest Neighbor (KNN) algorithms are examples of classification algorithms that can be applied to this problem. For this proof-of-concept model, a KNN classifier will be utilized for its ease of computation and easy model interpretation.

For training the model and identifying assemblies that the model will classify data as a robust dataset must be selected. This dataset must be representative of the diverse set of in-service assemblies in the built environment, such that most assemblies which a user may interact with should be present in the training dataset. For this work, the ASHRAE wall database present in Table 16 and Table 18 in the ASHRAE Handbook of Fundamentals: Chapter 18 (ASHRAE, 2021) was utilized as a starting point. The ASHRAE Fundamentals database is a collection of typical new construction wall assemblies with their associated material layering and material thermal properties. In total, 66 wall assemblies are represented in this database. Two examples of wall assemblies from this database are displayed below.

Table 1: Two example assemblies with layer IDs from the ASHRAE Fundamentals Assembly Database.

<table>
<thead>
<tr>
<th>Layer Number (Ext. to Int.)</th>
<th>Wall #13</th>
<th>Wall #63</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25mm Stucco (F07)</td>
<td>200mm Heavyweight Concrete (M15)</td>
</tr>
<tr>
<td>13mm Fiberboard Sheathing (G03)</td>
<td>89mm Batt Insulation (I04)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>89mm Batt Insulation (I04)</td>
<td>89mm Batt Insulation (I04)</td>
</tr>
<tr>
<td>16mm Gypsum Board (G01)</td>
<td>16mm Gypsum Board (G01)</td>
<td></td>
</tr>
</tbody>
</table>

From Table 1, two example wall assemblies, Wall #13 and Wall #63, from the ASHRAE Fundamentals assembly database are displayed. The layer ID located in parentheses should also be noted, as these layer IDs correspond to a list of material thermal properties based upon ASHRAE Fundamentals Chapter 26 material data. In total, this assembly database is comprised of 66 typical wall assemblies which range from stud walls to CMU and tilt-up concrete constructions.

While the ASHRAE fundamentals database provides a diverse set of wall assemblies to train the classification model with, there is a major limitation of this dataset—this assembly database is designed for new construction applications. Because of this dataset’s application towards new construction, uninsulated or historic constructions are not represented in the dataset. This omission can potentially be problematic, as 42% of residential wall assemblies in the United States are uninsulated (National Renewable Energy Laboratory, 2019). Due to this, there is a high likelihood that one may encounter assemblies not represented in the ASHRAE Fundamentals assembly database.

To address this shortcoming, 19 additional wall constructions were generated and appended to the ASHRAE Fundamentals assembly list. These additional assemblies were designed to represent assemblies typically found in in-service buildings including:

- Uninsulated structural masonry walls
- Uninsulated non-structural masonry walls
- Uninsulated stud cavity walls
- Stud walls with continuous insulation

Each of these additional assemblies is comprised of materials inside of the ASHRAE Fundamentals material list so no additional materials were added to the database. In total, the edited construction dataset comprised of 85 wall constructions which are intended to represent a majority of constructions that would be encountered in a retrofit scenario.

With the 85 assembly classes identified, the effective thermal resistance and effective thermal mass of each assembly were required for model training and testing. To address this, each of the 85 assemblies was simulated in EnergyPlus for the TMY3 weather file of Atlanta, Georgia, and characterized utilizing the Bayesian inference adoption of Pilet and Rakha’s thermal characterization methodology. To generate distinct training and testing datasets, walls were characterized during two different periods: (1) during the simulated first week of April, and (2) during the simulated second week of December. While these are comprised of the same construction on the same simulated building, transient thermal performance is expressed differently in cold outdoor conditions versus temperate outdoor conditions.

For the two seasons, each assembly was stochastically characterized, providing 2,000 samples of effective thermal resistance and thermal mass for each of the 85 assemblies. These characterizations serve as a synthetic dataset of sorts since the model will be trained and evaluated based upon physics-based modeling simulating the real world. A scatter plot of each of the characterized effective thermal resistance and effective thermal mass values for each of the 85 assemblies are plotted in Figure 2. It should be noted that each shape and color corresponds to an individual assembly, with 2,000 data points plotted for each of the 85 assemblies. These were then separated into training and validation datasets for model training and validation, which will be described in more detail within the next section.

© 2022 U.S. Government
Model Training and Testing

Dataset Preparation

As previously mentioned, effective thermal resistance, effective thermal mass, and material information from visual inspections will be utilized within the classification model. Thermal resistance and thermal mass will be input into the model as numeric features, both of which differ greatly in magnitude. Due to the difference, both features were scaled via standard score normalization, also known as a z-score normalization. The equation for this scaling is displayed below:

$$ z = \frac{x - \mu(x)}{\sigma(x)} \quad (1) $$

Where $x$ is a numeric feature, $\mu(x)$ is the mean of feature’s sample, $\sigma(x)$ is the standard deviation of feature’s sample, and $z$ is the standard score normalized feature. The normalization can be applied to both of the numeric features, allowing for scaling of these features to be of similar orders of magnitude. Feature scaling is especially important for distance-based algorithms such as KNN, where differing orders of magnitude between features can skew results and bias the model toward higher magnitude features.

Alongside numeric features, a categorical feature corresponding to cladding material was also utilized in the model. The goal of this feature is to encode information of an assembly’s cladding material into the classification algorithm. This information can be easily determined via a photograph or visual inspection of the assembly. This feature is meant to distinguish relevant assemblies from irrelevant assemblies based upon cladding material. For example, if a wall with a brick cladding is instrumented, tested, and classified, it does not make sense for the classification algorithm to suggest that the assembly is constructed with wood siding.

Encoding the exterior cladding information allows for the model to implicitly reduce the number of potential assemblies, which is especially important in the highly populated region of thermal mass values below 150 kJ/m$^2$-K as displayed in Figure 2.

To encode categorical features such as cladding material in a format usable for a machine learning algorithm special care must be taken. The cladding feature is a categorical, non-ordinal feature, so the popular One-Hot Encoding (OHE) method was utilized to encode categorical variables into numeric data (W. Chen, 2016). OHE operates by transforming a list of $N$ unique categories and $M$ samples into a matrix of size $M$-by-$N$ where every column corresponds to a specific category. When a category is selected for a given sample, that specific category will be denoted with a one and all other categories will be zeros. An example of the One-Hot Encoding is shown below.

**Table 2** An example visualization of One-Hot Encoding for usage on categorical features.

<table>
<thead>
<tr>
<th>Baseline Categorical Variable</th>
<th>One-Hot Encoded Dummy Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cladding Category</td>
<td>&quot;Brick&quot; &quot;Wood Siding&quot; &quot;Stucco&quot;</td>
</tr>
<tr>
<td>Sample 1 “Brick”</td>
<td>Sample 1 1 0 0</td>
</tr>
<tr>
<td>Sample 2 “Wood Siding”</td>
<td>Sample 2 0 1 0</td>
</tr>
<tr>
<td>Sample 3 “Stucco”</td>
<td>Sample 3 0 0 2</td>
</tr>
<tr>
<td>Sample 4 “Brick”</td>
<td>Sample 4 1 0 0</td>
</tr>
</tbody>
</table>

Model Selection and Training

As previously mentioned, a $K$-nearest neighbors (KNN) classification model was utilized in this study. This model is considered to be a lazy learner as it does not learn from training data but instead evaluates new model inputs against the training data for classification. In short, KNN calculates the distance between the candidate data and every datapoint within the training set. Once every distance is computed, the distances and their associated class labels are ordered from nearest to farthest, and the mode label of the first $K$ nearest data points is reported. One large advantage of KNN is that the model is explainable. Model parameters such as the distance function and the $K$-number of nearest neighbors can be tuned to increase KNN accuracy. To address modeler bias in model tuning, MATLAB’s built-in hyperparameter optimizer was utilized to identify the
optimum distance metric and best number of neighbors for the spring synthetic training dataset. The results of this model training and parameter tuning are displayed in Figure 3. From the figure, it can be noted that the classification error for standardized Euclidean, Euclidean, and city-block distance metrics produced similar results. It was also noted that any number of neighbors below 15 for these three distance metrics also seemed to produce similar results. This phenomenon is also displayed in the asymptotic present after evaluation number 6 in the classification error versus evaluation number graph. While results were similar for these three metrics, the hyperparameter optimizer’s best-observed parameter pair is that of a standardized Euclidean distance metric with 11 nearest neighbors. Because this parameter set was the optimum value, this set was selected for the proof-of-concept model evaluation and deployment.

Model Testing

With the model trained and optimized on the spring dataset, the model was evaluated on the winter dataset for testing. This testing dataset comprised of 85,000 testing samples, with 1,000 occurrences of each of the 85 walls simulated in EnergyPlus and characterized for winter climatic conditions. To evaluate model performance, the results of this testing set were evaluated via the standard metrics of Precision, Recall, and the F1-Score, which is also known as the Sørensen–Dice coefficient (Dice, 1945). The equations for each of these testing metrics are displayed in Eq. 2–4.

\[
\text{Precision} = \frac{\text{Number of Correct Class Predictions}}{\text{Number of Class Samples}}
\]

\[
\text{Recall} = \frac{\text{Number of Correct Class Predictions}}{\text{Number of Class Samples}}
\]

\[
F1 - \text{Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}
\]

From the above equations, the F1-score can be computed. It should be noted that the class macro-averaging scoring approach is utilized due to the large number of classes present in this classification model. The results of this testing are displayed in Table 3.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Macro-Averaged Score (%)</th>
<th>Standard Deviation of Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>95.8%</td>
<td>18.7%</td>
</tr>
<tr>
<td>Recall</td>
<td>94.1%</td>
<td>20.6%</td>
</tr>
<tr>
<td>F1-Score</td>
<td>94.6%</td>
<td>19.5%</td>
</tr>
</tbody>
</table>

From Table 3, the testing metric scores can be seen. It should be noted that all scores lie near an average value of approximately 95%, indicating strong model performance for the testing set. The similarities in value between precision and recall indicate that this classification model is approximately misclassifying classes approximately as often as it is predicting false-positive occurrences of certain classes. This phenomenon is also visible due to the non-zero standard deviation of scores. For example, the F1-score’s standard deviation of 19.5% indicates that there are discrepancies in scores between classes. When viewing these classification results in detail, it can be seen that 3 classes are never predicted by the model. While these classes are problematic from a modeling perspective, these misclassifications are quite understandable—e.g. two walls contain the same materials yet one wall has an air space and the other doesn’t, and two walls with two inches of continuous insulation are misclassified as similar walls but instead with double stud construction. Regardless of these small discrepancies in three classes, the model fulfills the goal of classifying the general construction of an assembly.
Abridged Case Study

With the model being trained and tested on synthetic datasets, it is important to validate the classification model with real-world data. In another work by the authors, a wall assembly was instrumented with temperature and heat flux sensors and was characterized to compute the effective thermal resistance and thermal mass of the assembly. A photo of the instrumented wall assembly is displayed in Figure 4.

This building was constructed around 1920 in Atlanta, Georgia, and is confirmed to be an uninsulated, brick mass wall by the building’s owner. With this knowledge, this classification of this assembly can serve as a real-world validation case for the assembly classification model.

Following four days of data collection, this assembly was characterized to have a mean thermal resistance of 0.546 W/m²·K with a standard deviation of 0.0294 W/m²·K and a mean thermal mass of 636 kJ/m²·K with a standard deviation of 43.7 kJ/m²·K. This assembly was characterized via 10,000 Markov chain iterations utilizing Bayesian calibration and inverse transient heat transfer modeling. Utilizing each step of the thermal characterization’s Markov chain, 10,000 discrete predictions of assembly composition were made to stochastically classify the composition of this assembly. Utilizing this data alongside the cladding information, which was identified as brick via visual inspection, this assembly was classified via the proof-of-concept KNN classification model. The results of this classification are displayed in Table 4.

From Table 4, the results of the assembly’s classification can be seen. From these results, 100% of the Markov chain iterations were classified to be a brick mass wall, with Wall #70 being two wythes thick, Wall #71 being three wythes thick, and Wall #72 being four wythes thick. In short, these predictions all suggest that this assembly is a mass wall, with the highest number of occurrences suggesting that this is, an uninsulated mass wall constructed of four wythes of brick. This prediction aligns with the information provided by the building owner, which supports the real-world validity of this KNN assembly classification algorithm on masonry walls in hot-humid climates.

Conclusion

In conclusion, the composition of building envelope assemblies can be predicted via in-situ thermal data and basic material identification. This paper developed an assembly dataset and proposed, trained, tested, and validated a proof-of-concept K-nearest neighbors classification model to predict wall assembly material composition via effective thermal resistance, effective thermal mass, and cladding material identification. This classification model was trained and tested to produce an average F1-score of 94.6% via synthetic datasets generated from the physics-based simulation. This model was then verified via characterization of a 1920s mass wall that was tested in Atlanta, Georgia, where it correctly classified a 100-year-old, in-service wall assembly. This classification further suggests that this classification model has real-world validity with the potential to be deployed alongside current state-of-the-art in-situ thermal measurement techniques. This work enables retrofit authorities to non-destructively assess the make-up of in-service assemblies to inform envelope retrofit decisions, all without disturbing the assembly.

For future work, this classification model should be further evaluated to identify additional features or sources of data that may increase its performance and usability for non-technical audiences. Furthermore, work should be conducted to identify alternative classification algorithms, such as neural network or deep learning-based approaches which may provide additional value above the simplistic K-nearest neighbors approach utilized in this work. Finally, this model was trained to classify assemblies per the 85 assemblies utilized for model training. These 85 classes do not represent every assembly that may be in service, so additional experiments on known assemblies should be conducted to further validate this classification model on different climates. Further experimentation also provides the opportunity for iterative improvement of the model’s training dataset to represent assemblies not represented in the current model.

Table 4  Assembly classification predictions for the Atlanta brick wall.

<table>
<thead>
<tr>
<th>Wall ID</th>
<th>Percentage of Predictions (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wall #72</td>
<td>98.0%</td>
</tr>
<tr>
<td>Wall #71</td>
<td>1.70%</td>
</tr>
<tr>
<td>Wall #70</td>
<td>0.300%</td>
</tr>
</tbody>
</table>
Acknowledgment
This work was supported by the U.S. Department of Energy, Building Technologies Office under contract DE-AC05-76RL01830 with the Pacific Northwest National Laboratory.

References