A Framework for Identifying Building Energy Models of Localized Utility Service Areas Using Smart Meter Data

Andrew Speake and Andrew Parker

National Renewable Energy Laboratory

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-5500-90443
December 2024

A Framework for Identifying Building Energy Models of Localized Utility Service Areas Using Smart Meter Data

Andrew Speake and Andrew Parker

National Renewable Energy Laboratory

Suggested Citation

Speake, Andrew, and Andrew Parker. 2024. *A Framework for Identifying Building Energy Models of Localized Utility Service Areas Using Smart Meter Data*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5500-90443. https://www.nrel.gov/docs/fy25osti/90443.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Technical Report
NREL/TP-5500-90443
December 2024

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies Office.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

Executive Summary

Bottom-up load modeling of buildings offers a versatile approach to simulating baseline demand and scenarios of future technology evolution and adoption at the individual building level. This capability is essential to understanding how future load shapes may change with the adoption of electric equipment and vehicles, particularly as it relates to grid planning and infrastructure investments. Traditionally, grid planning techniques have used historical load data to predict future load and infrastructure needs. However, with the anticipated rise in adoption of electrification technologies such as heat pumps and electric vehicles, historical data become less reliable predictors of the future. By employing ResStock, a high-fidelity building stock modeling tool, we can fine-tune electrification scenarios and aggregate models to represent varying geographic resolutions of the grid system, while considering the underlying features of homes. This may enable a more accurate and responsive approach to anticipate and plan for the evolving landscape of energy demands. We present a new framework that leverages building stock energy modeling to identify building models that align with the load shapes and housing attributes of buildings with AMI data. This approach applies two model layers: (1) a classification step that identifies the presence of air conditioning, electric heating, and electric water heating, and (2) an optimization routine that identifies building energy models aligning with load profile data from advanced metering infrastructure meters. This report demonstrates one approach to deploying this framework, and presents results for three test cases that use both modeled and AMI data to assess performance. For a test case using AMI data in Fort Collins, Colorado, we observed a median monthly electricity load CV-RMSE of 16.6%, and a top ten daily heating and cooling median absolute percent error of 7.7% and 8.3%, respectively. For each AMI meter, we identify a set of potential energy models so that downstream use-cases can account for uncertainty driven by variability of baseline technologies and occupant behavior, which impact the response to electrification and energy efficiency scenarios. Our results indicate that ResStock has potential as a scalable solution for modeling residential energy demand at local grid resolutions. Its performance depends on location-specific factors, underlying building characteristics, and the level of aggregation, offering a path towards more precise and adaptive distribution grid planning for the evolving energy landscape.

List of Acronyms

A/C	air conditioning
AMI	advanced metering infrastructure
DOE	U.S. Department of Energy
HVAC	heating, ventilating, and air conditioning
IECC	International Energy Conservation Code
NREL	National Renewable Energy Laboratory
RMSE	root-mean-square error
ROC	receiver operating characteristic

Table of Contents

List of Figures

List of Tables

1 Introduction

The landscape of long-term load forecasting for grid solutions is evolving, driven by a transition to clean energy generation, electrification of buildings and vehicles, and changing weather due to climate change (Energy Systems Integration Group 2022; Zhou et al. 2023). A significant area of uncertainty is estimating building load responses under increased adoption of electric vehicles and other electric technologies such as heat pumps (Keen et al. 2022; Blonsky et al. 2019). This challenge is heightened by the lack of historical data for buildings with such technologies and is further complicated by the effect of future weather patterns on heating and air-conditioning loads (Lindberg et al. 2019; Energy Systems Integration Group 2024). Furthermore, the lack of insight into homes for energy-burdened populations poses a significant challenge for equitable grid planning, as it hampers the ability to ensure that all populations benefit from clean energy transitions. One method to simulate future scenarios with respect to housing segments is bottom-up physics-based energy modeling, which relies on the detailed physical parameters of a building, incorporating descriptions of a building's geometry, envelope construction, equipment, and occupant behavior. However, physics-based models are generally not practical to deploy for real buildings at large scales, as there are major limitations in collecting the building inputs, and the process can be computationally expensive and labor-intensive. By using existing, representative energy models, we are able to avoid extensive computation time and uncertainty regarding the housing features. This report outlines a framework that aligns pre-simulated building energy models with load profile data from utility meters, providing a bottom-up modeling approach for any number of homes, which can be used to derive insights at local grid resolutions.

1.1 Overview

The objective of this work is to develop and demonstrate a methodology and framework for aligning utility smart meter data with load profile data from the ResStock™ Analysis Tool (Wilson, Parker, Fontanini, Present, Reyna, Adhikari, Bianchi, CaraDonna, Dahlhausen, Kim, LeBar, Liu, Praprost, White, et al. 2022). This approach uses hourly advanced metering infrastructure (AMI) load data and can incorporate static building inputs to identify similar models in the public ResStock datasets. Our report demonstrates one implementation of this framework across three datasets: two simulated and one AMI dataset. The study is designed to evaluate the versatility of our methodology, but the effectiveness of any custom implementation of this framework will depend on data availability, data quality, geographic resolution, and use cases. ResStock-modeled load profile data are published at 15-minute time steps and include fuel and end-use energy values for modeled residential buildings in the United States. In addition, this dataset includes metadata describing the underlying inputs to the models, giving us flexibility to leverage the data across geographic regions, housing characteristics, and technology adoption. By incorporating a national-scale, public dataset, this framework can be used in many locations across the country.

We focus on two components of bottom-up building energy modeling useful to long-term load forecasting: (1) inferring baseline building characteristics and (2) aligning loads with AMI data. Existing approaches to identify home characteristics include classification and regression models, neural networks, clustering, change-point models, and other non-intrusive load monitoring techniques (El Kontar et al. 2024; Westermann et al. 2020; McLoughlin, Duffy, and Conlon 2015; Perez et al. 2017; Milić, Rohdin, and Moshfegh 2021; Beckel et al. 2014; Schirmer and Mporas 2023). Inferring baseline building characteristics is helpful to identify candidates for technology adoption, however, impacts from adoption, population growth, and climate change scenarios cannot be directly modeled with this information. Other studies have introduced methods of physics-based model identification by mapping smart meter data using clustering methods (Bass, Ezell, and New 2022), with neural network predictions that feed energy model inputs (Deng et al. 2022), with change-point models to assess retrofit impacts (Kissock, Haberl, and Claridge 2002), or through identification of reduced order models to estimate impacts to a specific enduse (Perez et al. 2014).

Our approach predicts housing attributes using a classification model trained on synthetic energy model data, which informs an optimization routine to identify the final building energy models from average load and peak AMI data. The optimization routine minimizes the hourly net and peak electricity root-mean-square error (RMSE) between real and modeled building loads. Additionally, the optimization formulation can be extended to minimize other housing characteristics or features for groups of homes depending on use case and data availability. The building energy model outputs can be used to model scenarios of technology adoption or future weather, and do so at a granular level to forecast distribution system impacts. We present accuracy results for both model layers and demonstrate various levels of aggregation to represent geographic resolutions on the grid.

This report details the methodology to prepare data, infer attributes, and align load profiles in Section 2, presents results for three case studies in Section 3, and provides a discussion of the findings in Section 4.

2 Methods

This section details the methodology of our framework, which imports raw AMI and ResStock load data and assigns the most similar ResStock building model to the AMI data based on various criteria. The three modules that accomplish this are:

1. Preprocessing and Feature Engineering (Section 2.1)
2. Model Layer 1: Input Inference (Section 2.2)
3. Model Layer 2: Load Profile Alignment (Section 2.3).

Figure 1 shows the high-level steps of this methodology, where the enclosed portion includes the three modules within the scope of this report, and the focus of this section. The boxes downstream of the current framework scope ("Process Scenario Data" and "Distributions of Demand Scenarios") show the expected uses of the framework outputs. These outputs represent how identified energy models may be used for scenario analysis, which would use the ResStock electrification and energy efficiency measures to design custom scenarios for the region of interest. This technique aims to protect customer privacy while improving load forecasting predictions. It leverages time-aggregated load profiles to inherently obscure individual energy usage details. This approach is a key focus of the research to ensure individual privacy is maintained. The following sections outline methodological details for preprocessing and feature engineering, the input inference model, and the load profile alignment model.

We detail three test cases: two that train and test the framework using only ResStock data in Los Angeles (LA) County and in International Energy Conservation Code (IECC) climate zone 5B within the state of Colorado, and one that trains the model using ResStock data and tests using AMI data from the Fort Collins, Colorado, area. Testing on modeled data can help assess the upper end of performance as models are trained and tested on like data and all underlying features are known, while AMI data help demonstrate the model's ability to generalize to real-world data.

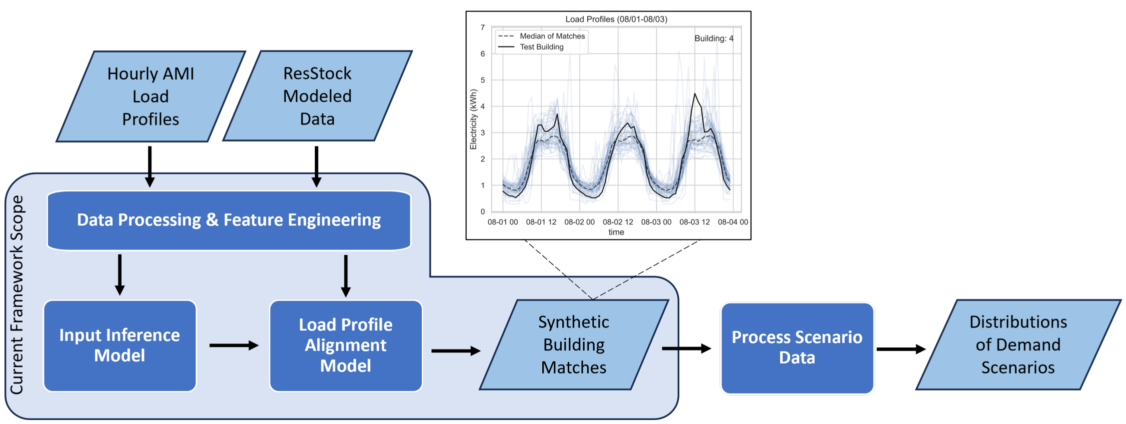


Figure 1. Flowchart illustrating the steps to assign ResStock model profiles to AMI data. The boxes downstream of the current framework scope ("Process Scenario Data" and "Distributions of Demand Scenarios") are included as the expected uses of the framework outputs for scenario analysis.

We present each step of the methodology in the context of a case study using AMI data from Fort Collins, Colorado. This case study demonstrates the potential for matching AMI data with modeled data and validating the identification of housing attributes. Table 1 provides a summary of the AMI data used in the case study. Section 2.1 details the steps to ensure AMI compatibility with our framework, while Sections 2.2 and 2.3 demonstrate the application of this data to each module.

In addition to the time series energy consumption for each premise, the AMI dataset includes some information about the presence of air conditioning (A/C) and electric heating, which assists in validating results from both modules. However, labels for these two parameters were inferred from incomplete data, which may skew the results. Mislabeling could arise from insufficient metadata, unpermitted home modifications, or a mismatch between the AMI and metadata collection dates. Premises with heating, ventilating, and air conditioning (HVAC) types labeled

Table 1. Summary of AMI dataset used in the case study

Location	Fort Collins, Colorado, area
Data points	10,000
Collection year	2018
Time step	Hourly
End use	Electricity usage
Building types	Residential: single-family detached, townhomes, duplexes, and multi-family
Other data	A/C type; presence of electric heating

as central air or heat pumps in the AMI data are identified as having A/C in our framework. This may exclude homes with window A/C units and does not give insight into equipment efficiencies. Information about the presence of electric heating is even less certain; HVAC types labeled as electric baseboard, electric radiant, or heat pump were assigned as having electric heating. Homes with electric furnaces, however, were not labeled as such and could not be classified as electrically-heated, excluding the most common form of electric heating in the region. Homes on electric heating rates are also included, though they represent a small fraction of data points and do not include all electrically heated homes. In practice, it may be important to further verify the accuracy of these labels if additional model validation is needed. Although there is potential for mislabeled data points, this does not affect the overall performance of the framework but rather constrains the accuracy of validation efforts in this specific area.

2.1 Preprocessing and Feature Engineering

To prepare electricity load inputs for downstream models, it is necessary to ensure AMI data compatibility and generate required features, also known as feature engineering. These steps are needed to (1) align features of modeled and AMI data, (2) better isolate the energy impacts of housing attributes, and (3) reduce computational complexity.

2.1.1 Ensure AMI Data Compatibility

This methodology requires multiple steps to refine the AMI data and select the appropriate subset of ResStock models. We cleaned the raw data to reduce outliers and missing data, assessed the AMI data to ensure compatibility with the load mapping model, and selected the geographic resolution of the modeled data. A final check of compatibility was performed after new features were generated, as described in Section 2.1.2.

Data Cleaning: Various methods exist (Wang and Wang 2020; Wilson, Parker, Fontanini, Present, Reyna, Adhikari, Bianchi, CaraDonna, Dahlhausen, Kim, LeBar, Liu, Praprost, Zhang, et al. 2022) to ensure AMI time series data are correct, consistent, and free of outliers, and their selection is contingent upon the quality of available data. After initial data exploration we removed buildings with missing data and filtered outliers. We also filtered all nonresidential meters because these test cases are limited to residential loads. We followed this process to ensure that sufficient data points remained for our analysis; however, more specialized methods could help limit the number of excluded buildings. Given the customized nature of this process, which is adapted to fit the unique characteristics of the dataset, further details on this step are not provided.

Temporal Resolution Constraints: At a minimum, AMI data need to be available at hourly time steps but can be as frequent as 15-minute time steps to align with ResStock and meet the requirement of the load mapping models. Additionally, a single, full year of AMI data is needed, and the year of collection must align with an available simulated year of published ResStock data. In our AMI test case, we leveraged Fort Collins, Colorado, data, which is available at hourly time steps for the year 2018.

Geographic Resolution Selection: The geographic scope of the synthetic loads may need to be fine-tuned depending on the sample size of the region of interest.[1] The goal when selecting the geographic bounds is to include as much diversity in inputs and load profiles as possible while avoiding the influence of dissimilar weather or stock characteristics that may be introduced by a wider geography. To match the region of our AMI data, we originally

[1]This report does not detail how to access the ResStock data; for more information visit https://www.nrel.gov/buildings/end-use-load-profiles.html.

queried the ResStock load profile data to limit data to Larimer County, Colorado. Given the relatively low population in this region, there are a limited number of energy models in the data ($n = 527$); general guidance for ResStock is to use at least 1,000 models for a representative sample. Further, the standard error of the mean annual electricity in Larimer County is 381.0 kWh/yr, with a 95% confidence interval of 10,959 ± 746 kWh. We expanded our region to include all data points in the IECC 5B climate zone in Colorado ($n = 7,871$), which resulted in an annual electricity standard error of 86.3 kWh/yr and a 95% confidence interval of 9,833 ± 169 kWh. If possible, it is best to constrain to a common climate zone and state so that weather and building practices are generally similar across all models. Another method of selecting buildings involves filtering by known stock characteristics that align with the AMI data. If the community of interest is known to be homogeneous across one or more attributes, any of the building inputs to ResStock could be filtered.[2] Figure 2 shows the chosen geography for our modeled data relative to the AMI data.

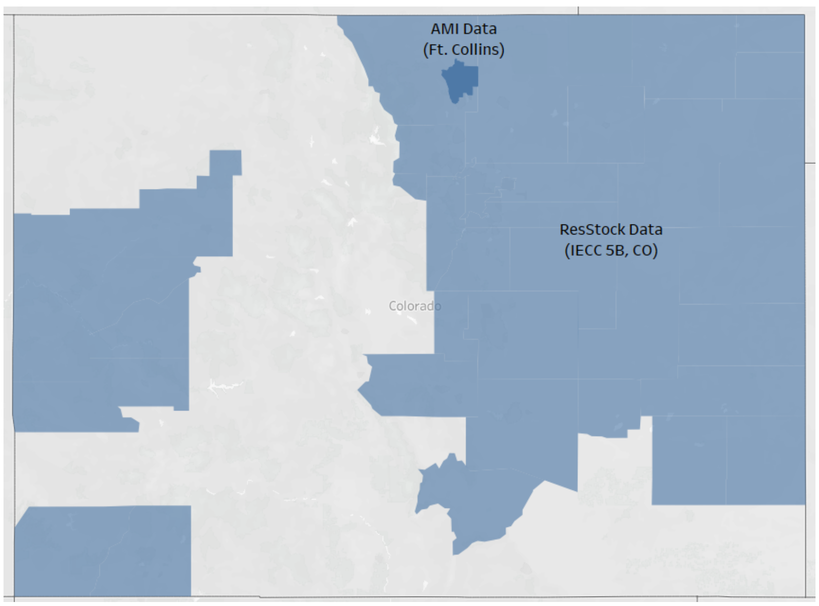


Figure 2. Map of the AMI data location and selected ResStock geographic bounds.

2.1.2 Feature Engineering

Once the raw AMI and ResStock loads are aligned by geography and time, we can engineer features for both datasets to use in the downstream models. The central input features used in our load mapping model are the daily average electricity profiles by month, which produce 12 by 24 arrays for each model and AMI dataset. Figure 3 shows two examples of this data for a ResStock model and an AMI data point. This aggregation was chosen so that seasonal impacts are incorporated as well as the daily shape, which is driven by specific technologies, building characteristics, and use patterns. Although hourly profiles are available to use, that level of granularity may introduce noise from stochastic occupant behavior and equipment cycling, making it difficult to align across datasets. Identifying a ResStock model in which the day-to-day changes in occupant behavior exactly match a real home's day-to-day changes in occupant behavior is unlikely and unexpected.

In addition to average daily profiles, we introduce a penalty for deviating from the peak load. Peak load can be an important consideration, particularly when analyzing grid system impacts from building electrification scenarios. A new feature, "average of top-10 daily peaks," is generated using both datasets for cooling and heating seasons. This value is calculated by pulling the 10 highest daily peaks during each season and averaging, as shown in Equations 2.1–2.3.

[2] Further guidance regarding uncertainty in the ResStock datasets can be found in Section 5.1.3 of (Wilson, Parker, Fontanini, Present, Reyna, Adhikari, Bianchi, CaraDonna, Dahlhausen, Kim, LeBar, Liu, Praprost, Zhang, et al. 2022)

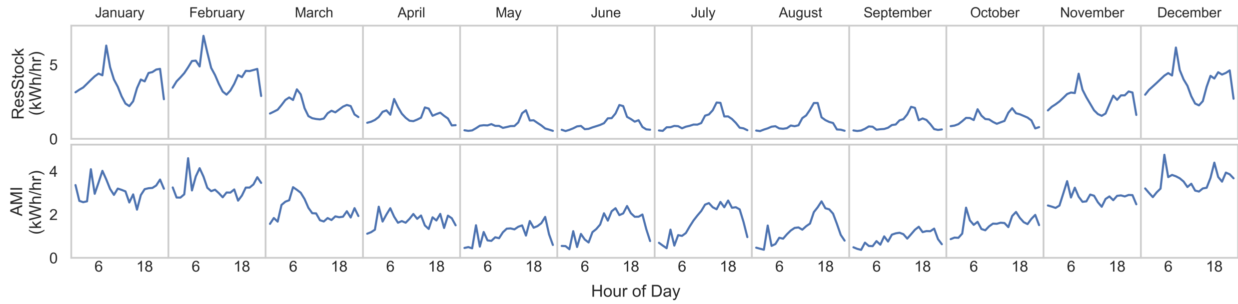


Figure 3. Daily average electricity profiles for each month for two sample data points, one from modeled data (top) and the other from AMI data (bottom).

$$M_{i,s} = \max_j x_{i,j,s} \quad \text{daily max for all hours } j \text{ in day } i \text{ for season } s \tag{2.1}$$

$$M_{\text{sorted},s} = \text{sort}\left(M_{1,s}, M_{2,s}, \ldots, M_{N_s,s}\right) \quad \text{sort the maximum values in descending order for season } s \tag{2.2}$$

$$\bar{M}_{\text{top10},s} = \frac{1}{10}\sum_{k=1}^{10} M_{\text{sorted},s,k} \quad \text{average the top 10 maximum values for season } s \tag{2.3}$$

Other features may be valuable given the specific application. For example, if some housing characteristics are known in the AMI, it is likely that they can be aligned with metadata from ResStock. Common features may include square footage, building type, or build year, which could help align load profiles by their underlying housing attributes in addition to the load shape and peak values.

We compared datasets across three outputs to ensure that the modeled data generally encompass the AMI data. Without this confirmation, there is an increased risk of poorly matched models and greater errors in identifying building attributes. Figure 4 shows distributions of the annual energy, the peak heating energy feature, and the peak cooling energy feature. Although the peak energy distributions show a mismatch, which may impact the accuracy of the matches, the modeled data have a wider range and are likely to contain viable matches for the AMI data. If the range of ResStock features does not fully encompass the AMI data features, as indicated by feature exploration or the final output metrics, an expanded set of models may be necessary. This can be determined using a similar process to the approach for selecting geographic resolution (Section 2.1.1).

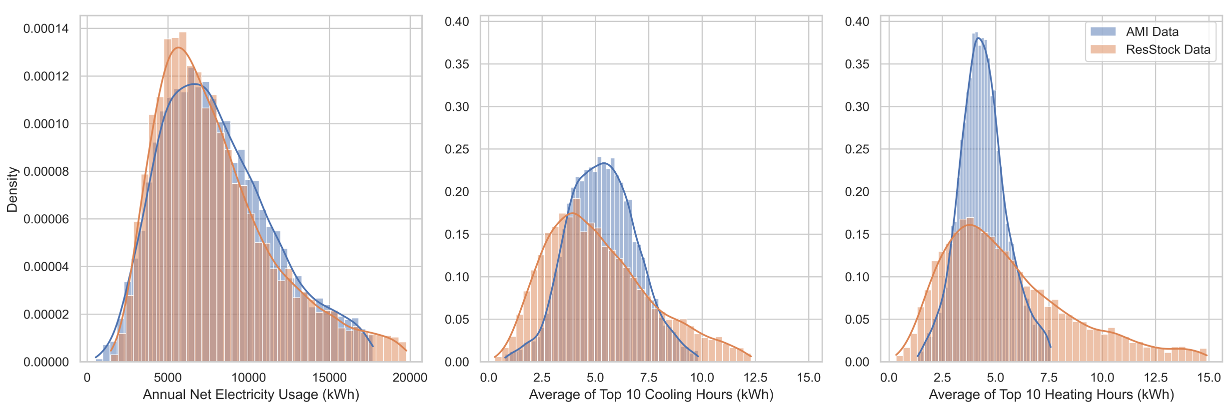


Figure 4. Comparisons of AMI and ResStock data for data within $1.5 \times IQR$ (Inner Quartile Range). The peak cooling and heating outputs are calculated using Equations 2.1–2.3.

2.2 Layer 1: Input Inference Model

Two model layers are employed to identify similar building energy models. The first is the input inference model, which uses a binary classification model to predict high-level building details of the AMI data. Although aligning models using load data will naturally align some of the underlying attributes, load features can be driven by a variety of factors. This model helps embed confidence in identifying a more accurate baseline of existing technologies. Without this step, there is a higher chance of mischaracterizing building inputs, which can influence the accuracy of downstream scenario analyses. We identified three parameters that are reasonably inferred given the data and are important to the load shape and peaks:

- **Has A/C**: The presence of air conditioning
- **Has Electric Heating**: The presence of electric heating systems
- **Has Electric Water Heating**: The presence of electric water heating systems.

Although we developed a straightforward model for this stage, its effectiveness for specific use cases could be enhanced by refining model selection, parameter tuning, and the extension of target variables. We use the XGBoost (Chen and Guestrin 2016) classifier to generate three models that predict probabilities for the target variables using the average daily profile feature described in Section 2.1.2. Training data are generated from ResStock load profiles filtered down to the geographic area of interest. Each of the models are tested under three data scenarios: modeled data in LA County, modeled data in IECC climate zone 5B, Colorado, and AMI data in Fort Collins, Colorado, as described in the following sections.

The modeled data scenarios are trained and tested using a k-fold cross-validation package (Pedregosa et al. 2023), and the AMI data scenario is trained with the full set of ResStock data and tested on 10,000 AMI data points. Since there is a risk that inferences may not generalize well between datasets, we apply the model in a conservative manner. A logistic regression objective function outputs the probabilities of predictions, and only inferences that meet a prediction probability of 0.95 or greater are used. If the prediction probability threshold is not met, then the set of potential models is not filtered for that housing attribute. While binary classification routines classically use a threshold of 0.50 and make predictions for all test data, we use only the predictions with high prediction probabilities, and otherwise do not make any inferences about housing inputs. Section 2.3 describes how these predictions are used in greater detail.

2.2.1 Application to Modeled Data

Testing the models using synthetic data is useful to understand the best-case scenario of model performance, as the training and test data are generated using the same modeling approach. By contrast, testing against AMI data introduces possible differences arising from the physics-based modeling rather than solely from deficiencies in the classification models. Further, all target variables for the modeled data are known exactly, whereas AMI metadata will have more unknowns and greater uncertainty surrounding building attributes. Two locations were simulated: LA County, which is mostly urban and has high summer loads, and IECC climate zone 5B in Colorado, which tends to be more suburban or rural and has colder winters. The results characterize the performance of these models, which are trained and tested using k-fold cross-validation with five folds.

Table 2 and Table 3 present cross-validation precision, recall, and F_1 scores for IECC climate zone 5B, Colorado, and LA County, respectively. Precision measures the ratio of true positive predictions to the total number of positive predictions made, recall measures the ratio of the true positives predictions to the total number of actual positives, and the F_1 score is the harmonic mean of precision and recall. Inferring building inputs with load shapes depends on the geographic region. Colorado has significant A/C and electric heating loads, which drive summer and winter load profile shapes and can therefore be predicted with high confidence, as shown by the F_1 scores of both parameters. Electric water heating can also be predicted well; although it is not as seasonally dependent, it is one of the highest electric loads in a home and often follows a common schedule when averaging across the month, making it a reliable target variable to predict.

LA County, conversely, is warmer and has higher summer peaks due to A/C loads. As a consequence, the highest-confidence predictions are associated with inferring A/C system presence. Although a large fraction of homes have electric heating ($\approx$25%), the F_1 score for predicting the negative class is lower when compared to the Colorado model. This is because the contribution of heating to the overall load shape is relatively small in most cases, making it more difficult to identify when analyzing the total electricity load.

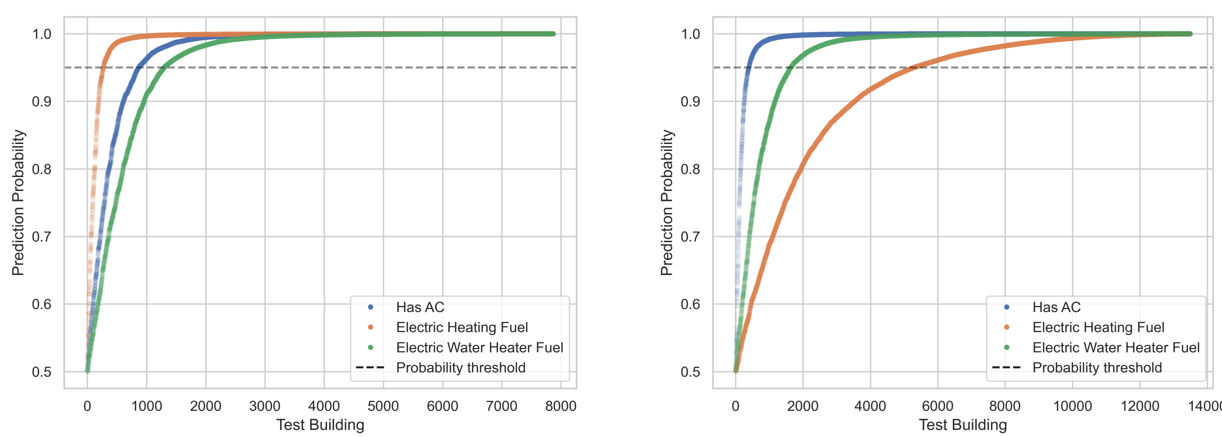
Table 2. Input Inference Model Classification Metrics Using Modeled Data in IECC Climate Zone 5B, Colorado

Parameter	Value	Precision	Recall	F_1 Score	Support
Has A/C	0	0.93	0.98	0.95	2,933
	1	0.99	0.95	0.97	4,938
Has Electric Heating	0	0.97	1.00	0.99	6,066
	1	0.99	0.91	0.95	1,805
Has Electric Water Heating	0	0.94	0.97	0.95	5,715
	1	0.91	0.83	0.86	2,156

Table 3. Input Inference Model Classification Metrics Using Modeled Data in LA County

Parameter	Value	Precision	Recall	F_1 Score	Support
Has A/C	0	0.97	1.00	0.98	4,298
	1	1.00	0.99	0.99	9,197
Has Electric Heating	0	0.87	0.97	0.92	10,027
	1	0.88	0.59	0.71	3,468
Has Electric Water Heating	0	0.95	0.99	0.97	11,783
	1	0.89	0.68	0.77	1,712

Classification metrics are calculated for every test point using a threshold of 0.50 to determine class membership. In practice, this model is applied when prediction probabilities meet or exceed 0.95 for either the negative class or positive class. Figure 5 shows all cross-validation predictions sorted by the maximum prediction probability for each location. The probabilities in these plots can indicate either the positive or negative class, and the value shown is the maximum prediction probability between the two, i.e., greater than 0.50. The chosen probability cutoff of 0.95 is also shown, which is the point at which inferences are included or excluded from the downstream model, where those greater than or equal 0.95 exclude non-matching models, and those less than 0.95 include all possible models. Table 4 shows the percentage of predictions that meet this threshold. These values can be interpreted as confidence levels for each inference model, where the highest confidence values are most likely to meet the threshold. Inference of the “Has A/C” parameter for LA County and the “Has Electric Heating” parameter for IECC climate zone 5B, Colorado, are associated with the highest prediction confidences. Alternatively, the “Has Electric Heating” model for LA County has the lowest average prediction probability and therefore the smallest fraction of predictions that are used.



(a) Colorado, IECC climate zone 5B

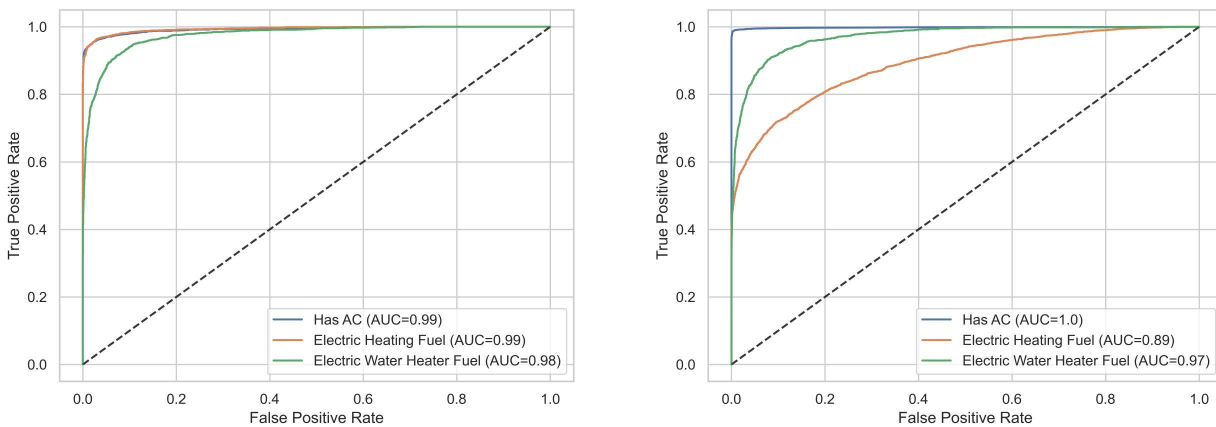
(b) LA County

Figure 5. Prediction probabilities for every prediction from (a) the IECC climate zone 5B, Colorado, model and (b) the LA County model. Values represent the larger probability of each prediction, which may indicate either the positive class or the negative class. The probability threshold of 0.95 represents the point at which predictions are included in the downstream model.

Table 4. Percentage of Predictions That Exceed the Threshold of 0.95 Prediction Probability

Parameter	IECC Climate Zone 5B, Colorado	LA County
Has A/C	89.0%	97.1%
Has Electric Heating	96.6%	60.9%
Has Electric Water Heating	83.5%	87.9%

Receiver operating characteristic (ROC) curves show true positive vs. false positive rates of binary classification models at varying thresholds. These values help assess the overall ability to predict both classes and enable comparison against other models. Figure 6 shows ROC curves for the two locations in modeled data scenarios and reiterates the high accuracy for most of the models. Note that the data shown include all predictions, not just those that meet the probability prediction threshold.



(a) Colorado, IECC climate zone 5B **(b) LA County**

Figure 6. ROC curves of each inference model and location using cross-validation results from the modeled data scenarios. The area under curve (AUC) values are shown in the legends, which quantifies the model's ability to identify classes, where 1.0 is perfect and 0.5 would be equal to a random assignment.

2.2.2 Application to Smart Meter Data

We trained input inference models using ResStock data from IECC climate zone 5B in Colorado and applied them to 10,000 AMI data points from the Fort Collins, Colorado, area. Because the underlying housing attributes of these buildings are either unknown or uncertain, there is less opportunity to validate the models compared to testing with modeled data. The test data in our results have some information regarding the presence of A/C and electric heating, but there is uncertainty surrounding the negative classes for each of these features. Those classified as positive are most likely labeled correctly, but there may be homes with a negative label that lack information, particularly for electric heating. Because precision is impacted by the number of false positives, the values shown for this metric may be lower than actual. No information is known regarding the water heater fuel; therefore, the accuracy results are omitted, even though a model is still trained and deployed for that parameter.

Table 5 shows classification results from the inference models applied to the AMI test dataset. The lowest-performing metric is the precision of the positive class for the "Has Electric Heating" model. There are likely multiple reasons for this: First, the electric heating fuel labels are imbalanced—9,641 vs. 359—meaning just a small number of false positives can have a large effect on precision. Second, the true fraction of homes with electric heating is expected to be much higher than what the labels show, and therefore many of the false positives may in fact be correct if more accurate information was available. For the "Has A/C" model, the recall of the negative class is the worst performing. The recall is driven by the number of false negatives, which in the case of the negative class means that this model is predicting that A/C is present when it is labeled as having none. One explanation may be that the AMI data do not indicate whether homes have room A/C units. This affects approximately 14% of homes that may be mislabeled as not having A/C, whereas in the training data, we categorize all homes with central air, room, or heat pump systems as having A/C. More generally, there are areas in which the model can be refined to improve performance. These models are trained using only their load shapes, and they lack information about other parameters such as building type, square footage, or build year, all of which may influence the response that A/C or electric heating have on the building load.

Table 5. Input Inference Model Classification Metrics Using AMI Data From Fort Collins, Colorado

Parameter	Value	Precision	Recall	F_1 Score	Support
Has A/C	0	0.75	0.27	0.40	4,087
	1	0.65	0.94	0.77	5,913
Has Electric Heating	0	0.99	0.91	0.95	9,641
	1	0.24	0.82	0.38	359

The confidence levels associated with the three inference models are visualized in Figure 7. The prediction probabilities from inferring on AMI data are generally lower than the test case for IECC climate zone 5B, Colorado, but most data points still maintain a probability that exceeds the threshold of 0.95. Inference of A/C and electric heating fuel have the highest prediction probabilities, with 86% or more data points exceeding the threshold, whereas 71% of the electric water heating inferences exceed the threshold, as shown in Table 6.

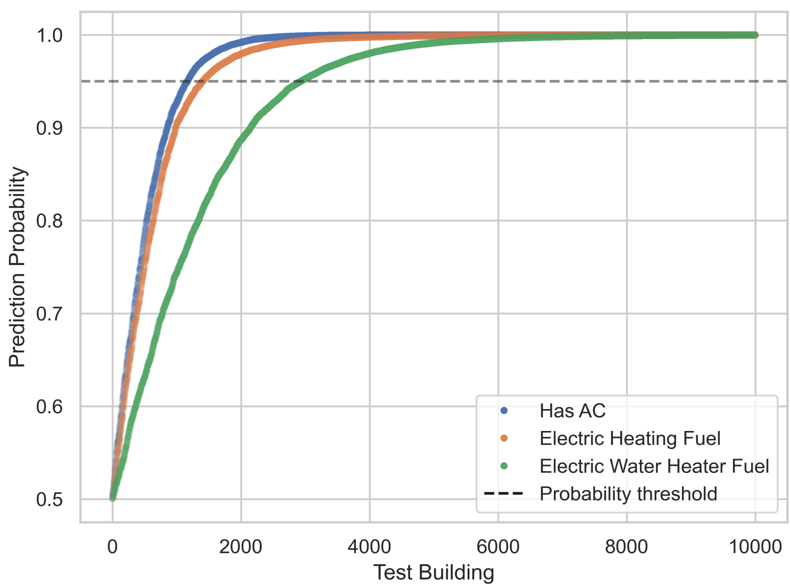


Figure 7. Prediction probabilities for every input inference model prediction of the Fort Collins, Colorado, AMI data. Values represent the larger probability of each prediction, which may indicate either the positive class or the negative class. The probability threshold of 0.95 represents the point at which predictions are included in the downstream model.

Table 6. Percentage of Predictions That Exceed the Threshold of 0.95 Prediction Probability for Fort Collins, Colorado, AMI Data

Parameter	Percentage Above Threshold
Has A/C	88.3%
Electric Heating Fuel	86.0%
Electric Water Heater Fuel	70.7%

The ROC curves generated by the inference model applied to AMI data are shown in Figure 8. Because these curves depend on known labels, only the A/C and heating fuel parameters are shown, as information related to water heater fuel are absent in our AMI test dataset. Despite a higher potential for incorrect labels, the model for predicting electric heating achieves a better balance between true and false positives, as indicated by the ROC curve. Due to the large number of negative samples, the false positive rate remains relatively low despite the high number of false positives and low precision (Table 5), resulting in a consistently high ROC curve.

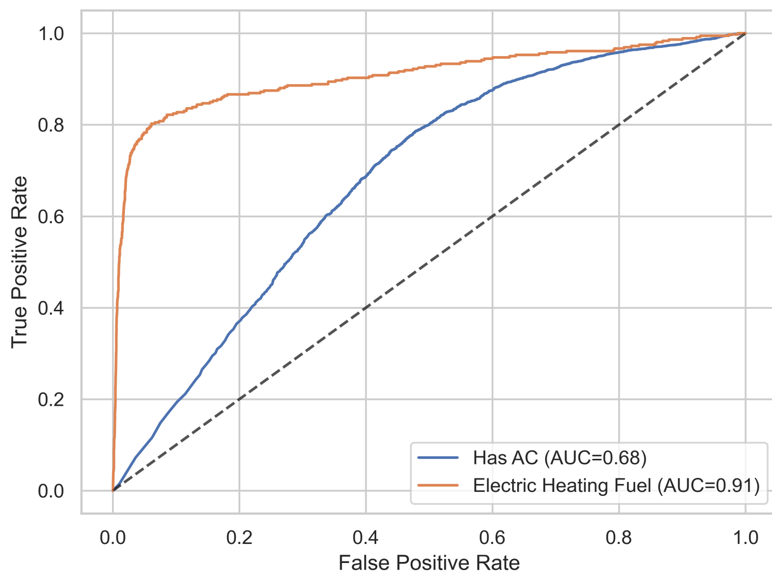


Figure 8. ROC curves of each inference model applied to Fort Collins, Colorado AMI data. The area under curve (AUC) values are shown in the legends, which quantifies the model's ability to identify classes, where 1.0 is perfect and 0.5 would be equal to a random assignment.

2.3 Layer 2: Load Profile Alignment Model

We formulated a mixed-integer linear program to map the set of modeled load profiles to real buildings. Our approach uses a default solver in the PuLP (Roy, Mitchell, and Peschiera 2023) Python package (COIN Branch and Cut Solver); however, further investigation into the best-performing algorithm may be needed depending on the use case. Because our optimization formula does not require complex trade-offs and instead simply minimizes the error across features, this algorithm is sufficient. This layer is formulated as a minimization problem to allow for flexibility of more complex tradeoffs. Given the constraints presented, an optimization is not necessary, but additional constraints to minimize error across building aggregations can easily be incorporated. The model formulation is shown in Equation 2.4.

$$\text{minimize} \quad \sum_{b\in B}\sum_{y\in Y} w_{b,y}(r_{b,y}\mathbf{X}_{b,y} + \alpha^c p^c_{b,y}\mathbf{X}_{b,y} + \alpha^h p^h_{b,y}\mathbf{X}_{b,y}) \tag{2.4}$$

subject to

$$r_{b,y} = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(x_{y,h} - x_{b,h})^2} \quad \text{for } b \in B,\, y \in Y$$

$$p^c_{b,y} = |\bar{p}^c_b - \bar{p}^c_y| \quad \text{for } b \in B,\, y \in Y$$

$$p^h_{b,y} = |\bar{p}^h_b - \bar{p}^h_y| \quad \text{for } b \in B,\, y \in Y$$

$$\mathbf{X}_{b,y} \text{ binary, for } b \in B,\, y \in Y$$

$$\sum_{b\in B}\mathbf{X}_{b,y} \leq n \quad \text{for } y \in Y$$

$$\sum_{b\in B}\mathbf{X}_{b,y} \geq n \quad \text{for } y \in Y$$

where,

$\mathbf{X}$	Indication of a model being selected
$b \in B$	Set of ResStock building models
$y \in Y$	Set of AMI buildings
r	RMSE for AMI data compared to modeled data
x	Average hourly electricity usage (kWh)
H	Total number of hours in the daily average profiles
p^c, p^h	Cooling and heating peak errors, respectively (kWh)
$\bar{p}^c, \bar{p}^h$	Cooling and heating average of top 10 peaks, respectively (kWh)
w	Penalty to exclude ResStock models with non-matching attributes identified by the input inference model
n	Number of matches for each AMI building
α^c, α^h	Cooling and heating peak weights

The three terms in the objective function use the generated features described in Section 2.1.2: daily average profiles by month, the average peak cooling days, and the average peak heating days. The solution to this problem results in a binary 2D matrix indicating the set of ResStock matches that result in the lowest objective value for all AMI data points. We framed this as an optimization problem to allow for flexibility in introducing other terms that may be important and that might require trade-offs. In practice, the RMSE and peak error terms can be precalculated, and the minimum n could be selected to achieve quicker compute times.

Number of matches, n: We elected to match 10 building models to each test data point in our analyses. By identifying more than a single model for each data point, distributions of expected housing characteristics and energy impacts of future scenarios can be outputted. Ideally, there will be many candidates that align with each AMI data point, resulting in higher-confidence distributions. However, increasing samples will result in a trade-off with the average error metrics across matches.

Penalty coefficient, w: As discussed in Section 2.2, the set of potential matches for a given AMI dataset is constrained when a housing attribute is predicted with high confidence, that is, exceeding a prediction probability of 0.95. The modeled buildings are filtered based on the presence of A/C, electric heating fuel, and electric water heating. To incorporate the results of the input inference model, a matrix of penalties, w, is generated that effectively eliminates ResStock building models with attributes that do not match those predicted. This process is mapped out in Figure 9.

Objective function weights, α: To appropriately balance the impacts of individual terms in the objective function, we applied weights to scale the cooling and heating peak error penalties. A parametric analysis informed the values of these weights, during which the cooling and heating peak term weights were incremented, and the impacts to the monthly CV-RMSE (Equation 2.5) and the peak errors were observed, as show in Figure 10. This analysis revealed a gradual increase in the monthly CV-RMSE and a sharper reduction in peak errors as peak weights were increased.

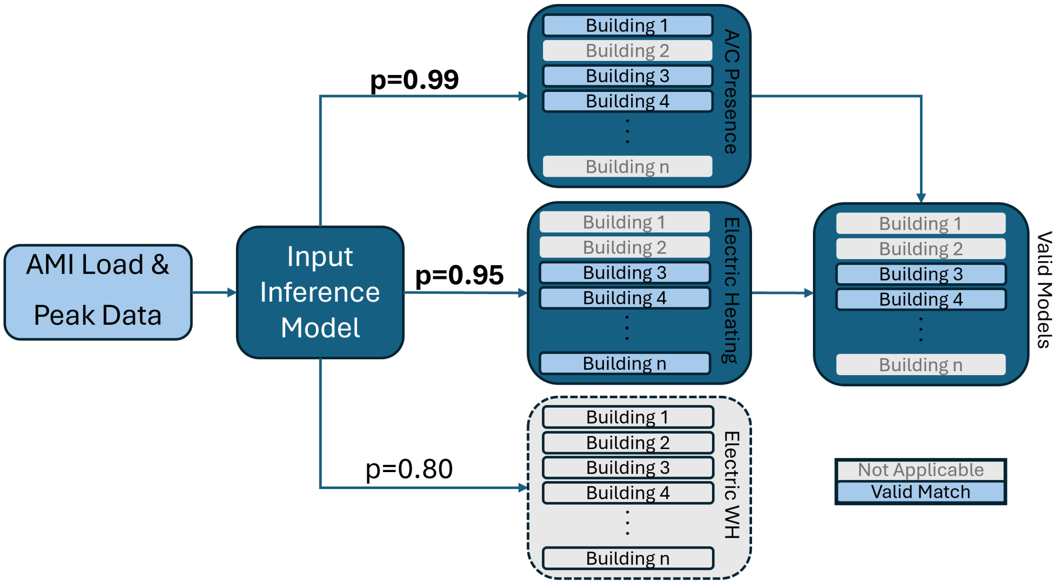


Figure 9. Demonstration of the process by which ResStock models are filtered to target specific housing characteristics with the input inference model. The example probabilities shown represent the prediction probability for each housing attribute inference; those that exceed 0.95 will impact the available ResStock models while the other does not filter any models.

Based on this analysis, we selected equal weights of 0.5 to apply to both heating and cooling in the objective function. Future applications could fine-tune these weights further.

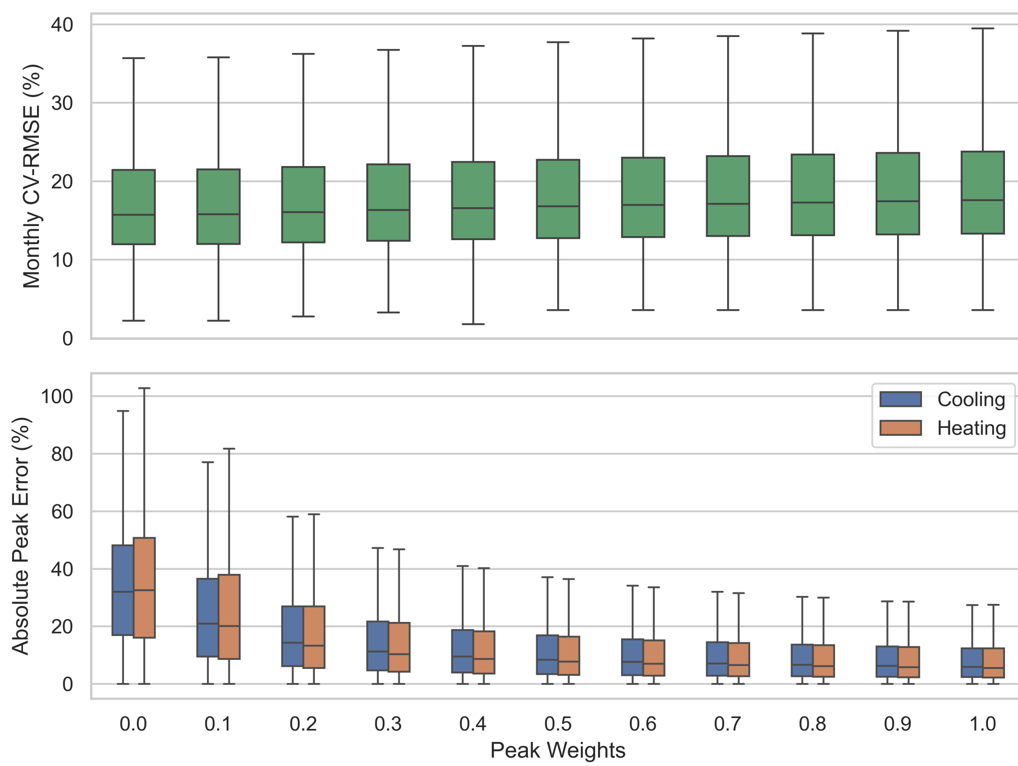


Figure 10. Parametric analysis of weights for the peak heating and cooling features, and resulting trade-off between variance and peak error.

We set the input inference model as the first layer of the framework to identify housing attributes and reduce the parameter space that feeds into the optimization layer. This second optimization layer outputs the *n* ResStock models that minimize the objective function for each test building. We applied this workflow to modeled and AMI test datasets, as shown in Section 3.

2.3.1 Outputs

The optimization layer outputs a set of *n* building IDs that correspond to synthetic ResStock models. Figure 11 shows examples of the AMI and building model loads, which are used to assess one aspect of model performance. The underlying housing characteristics of the matched building models can also be outputted, enabling a better understanding of the baseline housing stock, which homes may be eligible for certain electrification scenarios, and how those scenarios impact loads.

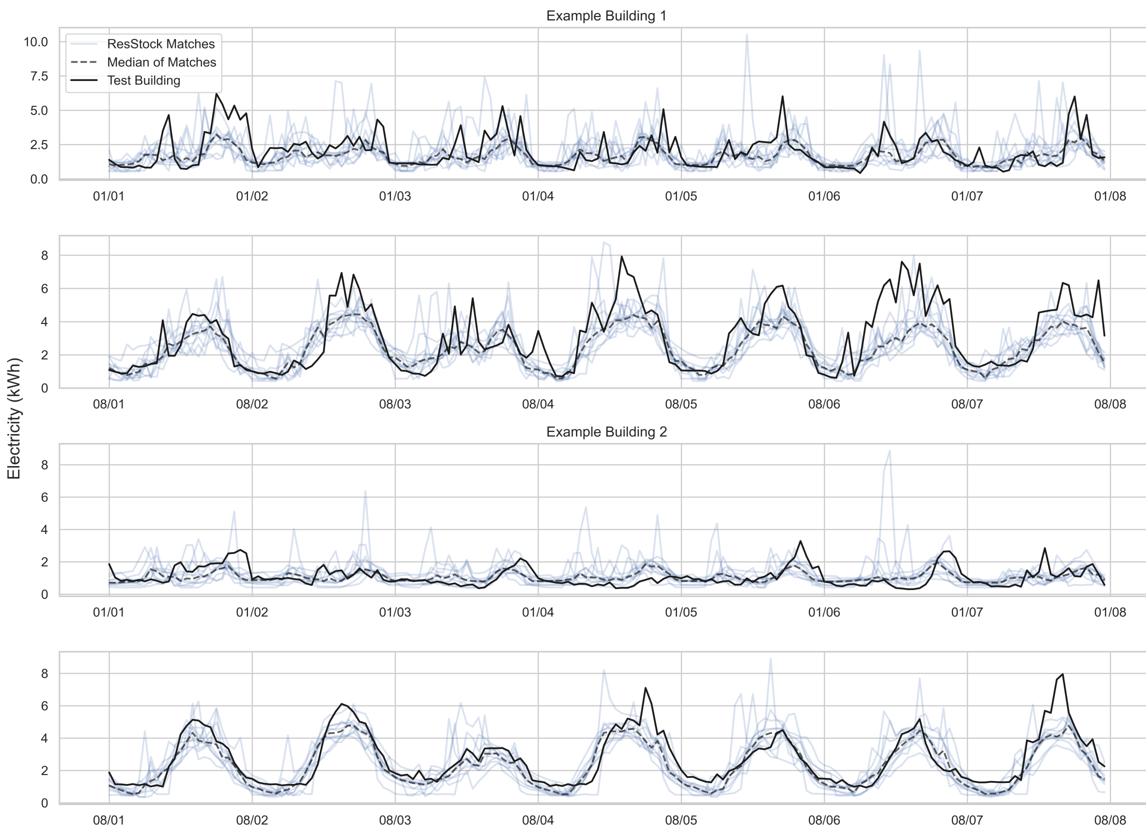


Figure 11. Two AMI test buildings from Fort Collins, Colorado, and their matches for a week in the winter (January) and a week in the summer (August).

To assess the input inference model performance, we output percentages of the correct attributes for each set of matches. Metrics applied to final outputs include the hourly and monthly CV-RMSEs (Equation 2.5) and absolute percent errors for the average of the top ten peaks. Section 3 displays figures and metrics evaluating these outputs in the context of the case studies.

$$\text{CV-RMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\overline{y}} \times 100\% \tag{2.5}$$

where,

n	Number of observations (months or hours)
y_i	Observed value for the i-th observation
$\hat{y}_i$	Modeled value for the i-th observation
$\overline{y}$	Mean of the observed values

3 Results

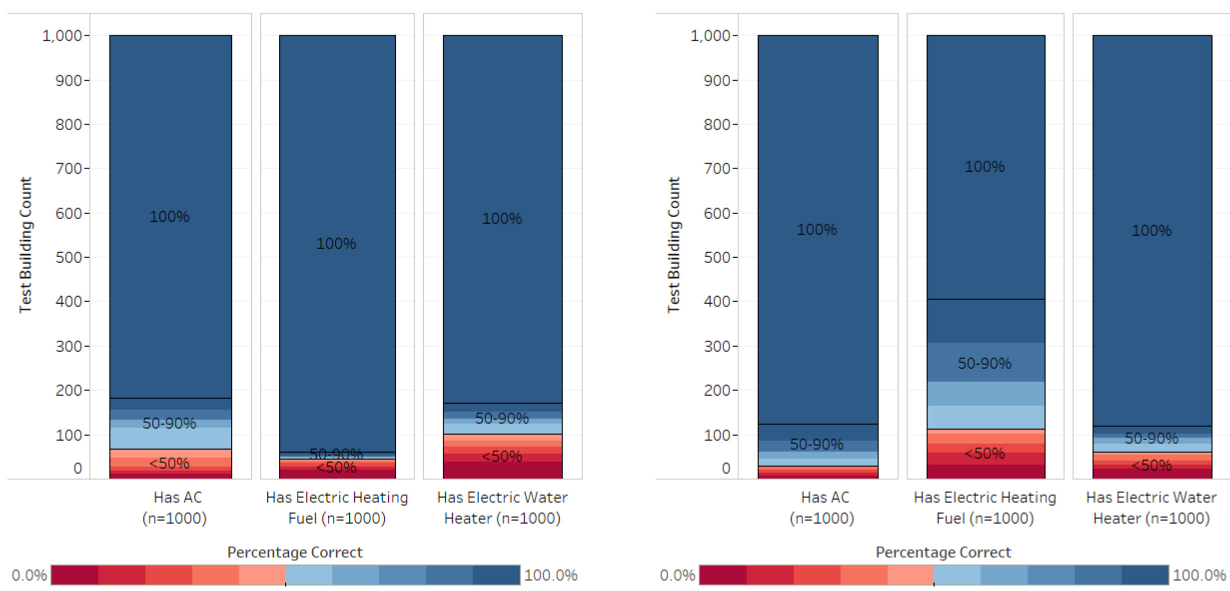
Three aspects of the model are analyzed to assess its performance: (1) accuracy of predicting baseline housing attributes, (2) hourly and monthly electricity usage errors, and (3) summer and heating peak errors. Although these are interrelated, input parameters may still need to be adjusted to enhance specific outputs as needed. Adjustments may include changes to the input inference prediction probability threshold, objective function weights, or feature engineering. This section presents results from applying the framework to modeled datasets in LA County and IECC climate zone 5B in Colorado, as well as the AMI test data in Fort Collins, Colorado. We reserved 1,000 building models as test data in each modeled test case to represent AMI data, and the remaining models were left as potential matches. This resulted in test data that accounted for approximately 13% of the IECC climate zone 5B, Colorado building models and 7% of the LA County building models. The AMI test included the full set of IECC climate zone 5B, Colorado ResStock models as potential matches.

3.1 Prediction of Building Attributes

We assessed the rates of alignment between test buildings and their matches for certain attributes. Although the first layer of the framework helps to identify some of the test building attributes, the final output of building attributes checks the accuracy of all buildings after running the full workflow. For each set of model matches, we collected the values of A/C presence, electric heating fuel presence, and electric water heating presence and compared to known values in the test data. The modeled test runs allow us to check the accuracy of any predicted housing attributes relative to the synthetic modeled dataset, whereas the AMI dataset has only limited information about underlying attributes.

3.1.1 Modeled Data

Figure 12 shows distributions of the two modeled test runs. In general, there is a very high rate of alignment for the three housing attributes shown. Given similar modeled data are used as the test data, these results are expected to be near the upper limit in the framework's ability to infer characteristics. These results are similar to those of the inference models, but because the second layer minimizes across energy metrics without considering housing attributes, results are expected to differ slightly.



(a) Colorado IECC climate zone 5B **(b) LA County**

Figure 12. Percentage of accurate matches across three housing characteristics for each test building, indicating the proportion of matches sharing the same housing characteristic as the test building.

3.1.2 Smart Meter Data

Figure 13 shows the accuracy of A/C presence and electric heating fuel attributes when applying the AMI dataset. We anticipate some reduction in accuracy compared to the modeled data due to (1) differences between the training data (ResStock) and the test data (AMI) and (2) uncertainty surrounding the housing characteristic labels in the AMI data.

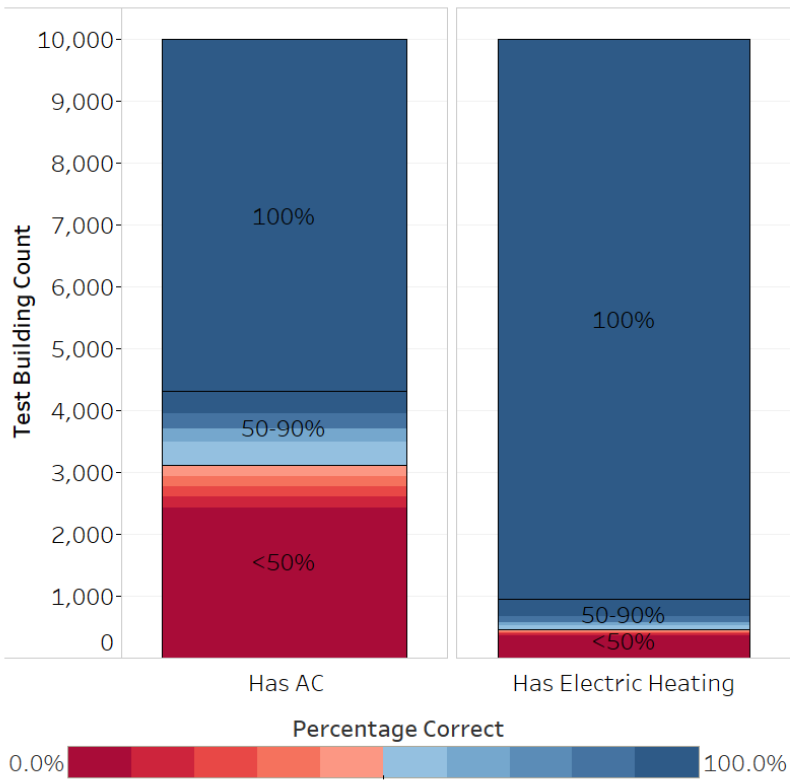


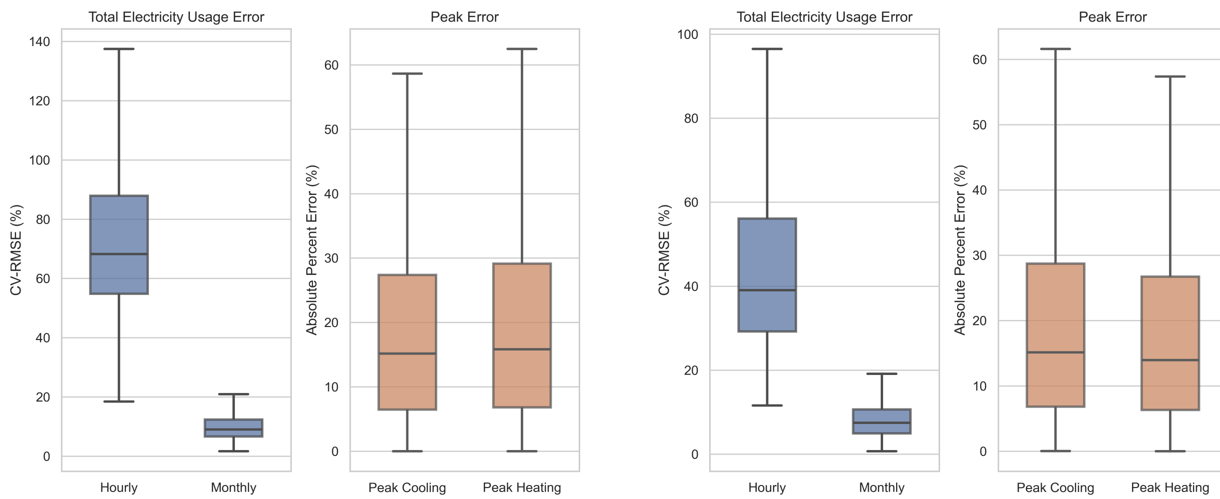
Figure 13. Percentage of accurate matches across two housing characteristics for each Fort Collins, Colorado, AMI test building, indicating the proportion of matches sharing the same housing characteristic as the test building.

3.2 Load Profile Matching Results

This section examines model performance using electricity load data of test and simulation buildings. The CV-RMSE is primarily used to measure the load profile error and the absolute error measures peak error. Beyond these error metrics, we also analyze the effectiveness of the model across aggregations of housing segments, simulating groups of homes that could share distribution grid resources such as transformers or feeder lines.

3.2.1 Modeled Data

Figure 14 displays load and peak error metrics for the two locations in the modeled test case, which uses synthetic data for both the test and simulation datasets. Results from the case studies indicate a regional dependence on the performance of our methodology. The LA County case outperforms the IECC climate zone 5B, Colorado, case for hourly CV-RMSE, but has similar performance for monthly CV-RMSE and peak error. Hourly profiles are more sensitive to occupant behavior and therefore are more difficult to align with similar profiles. The geographic influence on the hourly matches could be due to both the underlying building technologies and the nature of the datasets. First, LA County has a larger number of building models in ResStock compared to IECC climate zone 5B, Colorado (13,495 vs. 7,871), meaning there are more representations of possible occupant behavior in LA, and therefore a greater likelihood of identifying similar profiles. Further, the profiles in LA County are generally more homogeneous than in Colorado, so there is also a greater fraction of models with similar shapes. The majority of LA County homes are summer peaking, even if electric heating is present, while an electrically heated home in IECC climate zone 5B, Colorado, dramatically changes the shape of the load. Monthly errors still have similar distributions across locations, indicating that the hourly error differences are probably driven by occupant behavior and do not lead to broader errors at coarser timescales. Note that the weights applied to peak error were determined through a parametric analysis using AMI data, and refining these values may shift the balance between the peak and CV-RMSEs for modeled data results.



(a) IECC climate zone 5B, Colorado **(b) LA County**

Figure 14. Energy error metrics calculated for each test building and their 10 matches. Each point represented in these distributions is a unique calculation comparing the test building to the simulated match. The whiskers represent values within $1.5 \times IQR$ (Inner Quartile Range), while the outliers beyond this range are not plotted, as homes with low electricity usage can skew percentages to very large values, even if the absolute error is small.

3.2.2 Smart Meter Data

Peak error and CV-RMSE were outputted for the smart meter case study in Fort Collins, Colorado, as shown in Figure 15. The modeled case study (Figure 14) acts as an upper limit to model performance and therefore performs better compared to the application to AMI data. That said, the peak errors in Figure 15 do outperform the modeled data, which may indicate that further parameter tuning of the peak weight values could be needed in the modeled case study, balancing results in favor of the CV-RMSE. These plots summarize metrics across the full AMI dataset, but this framework is expected to be deployed for aggregations across similar housing segments, which has a large influence on the final matching performance, as discussed in the text below.

To estimate how the framework performs when applied to different segments of the housing stock, we sampled AMI data points and their matches and compared the aggregated profiles. This is meant to represent examples of local grid resolutions, such as a group of homes served by a single secondary transformer, all the homes in a subdivision, or all the homes on a distribution feeder. We find that error is reduced with increased aggregation of homes due to smoothing of stochastic events (e.g., occupant behavior and equipment cycling), but is dependent on the features of the housing segments. Figure 16 shows sets of five buildings aggregated across homogeneous housing segments, representing homes near each other that might be served by a single secondary transformer. We present five different samples (one for each row in the figure) per housing segment, as there can be significant diversity for a group of homes even with common attributes. In general, aggregated modeled data align well with the aggregated AMI data, but the housing segment influences to what extent this is true. Single-family detached homes align better than the attached and multifamily homes, most likely because these homes have a higher fraction of HVAC contributing to the overall electricity load, which is highly dependent on timing of weather, and is naturally built into the time series features of the model layers. Alternatively, the attached and multifamily homes have fewer exposed walls and are smaller but still have similar appliance loads as the detached homes, meaning their loads are driven to a greater extent by occupant events. Given the variability of occupant loads, the individual building load is less predictable, and that variability is still observed when aggregating to five homes. The build year influences the aggregate matches in similar ways. The older attached and multifamily homes are better aligned than the newer homes, most likely because HVAC loads are higher in the older, less-efficient buildings.

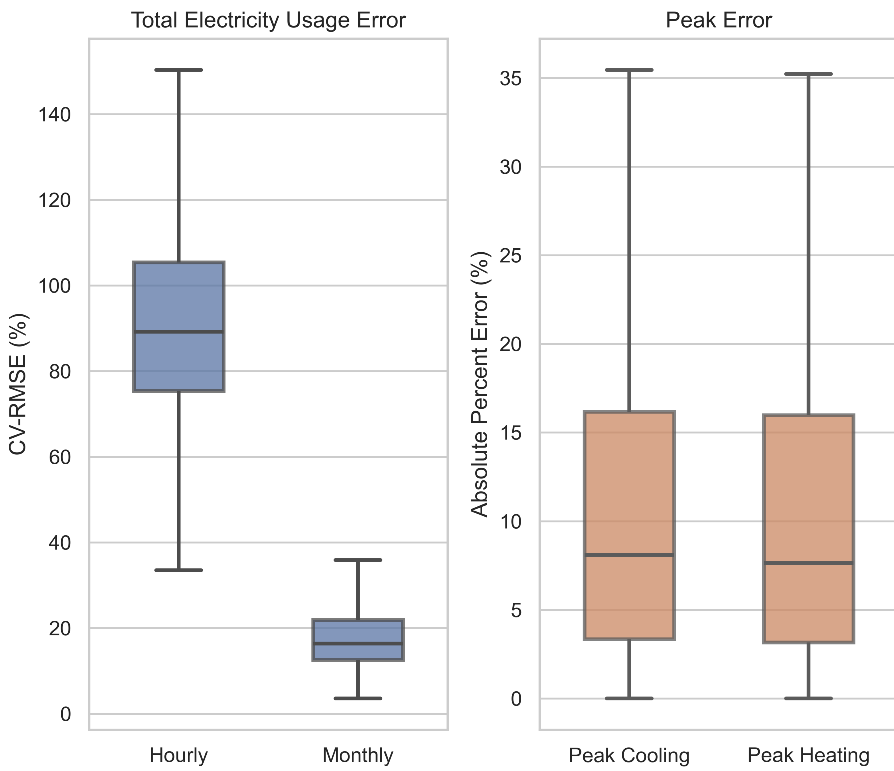


Figure 15. Energy error metrics calculated for each test building and their 10 matches for all 10,000 Fort Collins, Colorado AMI data points. Each point represented in these distributions is a unique calculation comparing the AMI test building to the simulated match. The whiskers represent values within $1.5 * IQR$, while the outliers beyond this range are not plotted, as homes with low electricity usage can skew percentages to very large values, even if the absolute error is small.

Figure 17 shows the aggregation of the monthly daily profiles for segments of 500 homes. Random assignment of homes at this level does not result in much variability within a given segment, and therefore only one sample is taken for each segment. Load alignment and error metrics are generally improved with the increased sample counts; however, there is still a dependency on building characteristics. Similar to the five-building aggregations, the models identified for attached and multifamily buildings are associated with a worse fit. This is likely due to the smaller relative impact of HVAC loads and increased influence from occupant-driven events in smaller and more efficient homes.

Finally, an aggregation of every AMI data point and its matches are shown in Figure 18. This provides insight into the ability of synthetic loads to represent a large and diverse population of residential loads. These results are still specific to the region of interest in Fort Collins, and may not generalize to other locations. Remaining discrepancies between matched modeled data and AMI data could be due to deficiencies in the ResStock model, insufficient tuning and model selection for the input inference model, or simplicity in the objective function definition of the optimization layer.

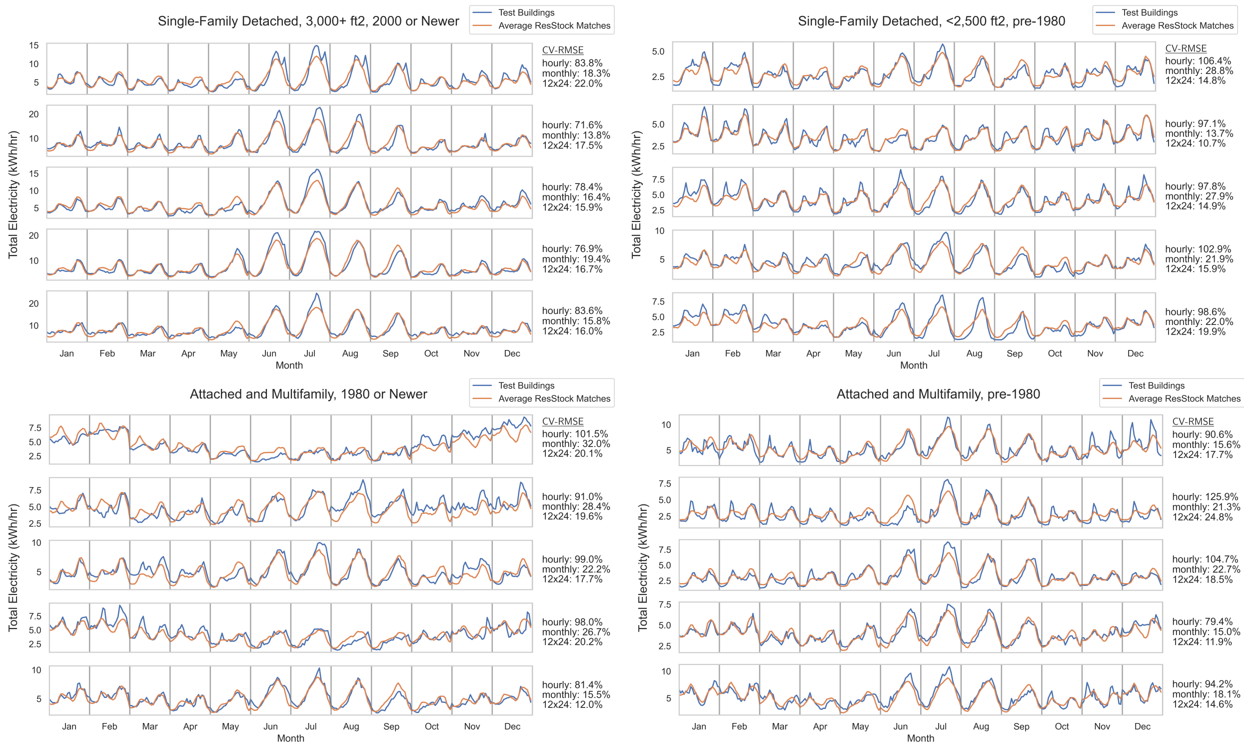


Figure 16. Random aggregations of five Fort Collins, Colorado AMI buildings and their matches across four housing segments. Each row in the plot represents a different random sample of five buildings to illustrate the diversity that can occur between small groups of homes. ResStock profiles are calculated by averaging the ten matches for a given test building and then summing those averages for the five test buildings. The CV-RMSEs on the right of each plot compare the (1) hourly average ResStock matches and summed AMI profiles for every hour of the year, (2) monthly totals, and (3), 12x24 profiles representing the monthly-daily loads shown in the plots and found in the optimization routine.

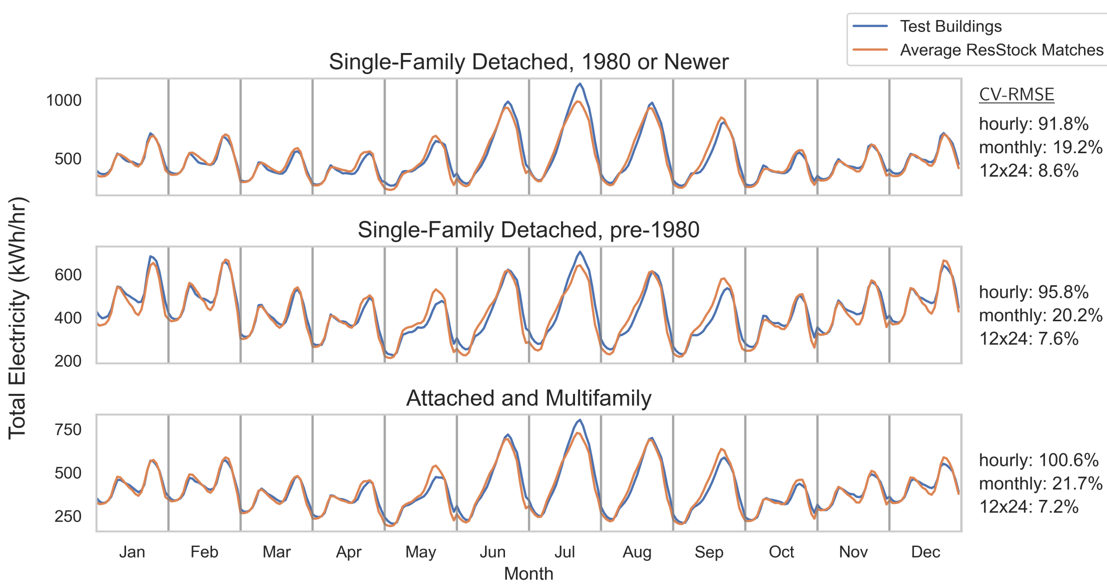


Figure 17. Random aggregations of 500 Fort Collins, Colorado AMI buildings and their matches across three housing segments. ResStock profiles are calculated by averaging the ten matches for a given test building and then summing those averages for the 500 test buildings. The CV-RMSEs on the right of each plot compare the (1) hourly average ResStock matches and summed AMI profiles for every hour of the year, (2) monthly totals, and (3), 12x24 profiles representing the monthly-daily loads shown in the plots and found in the optimization routine.

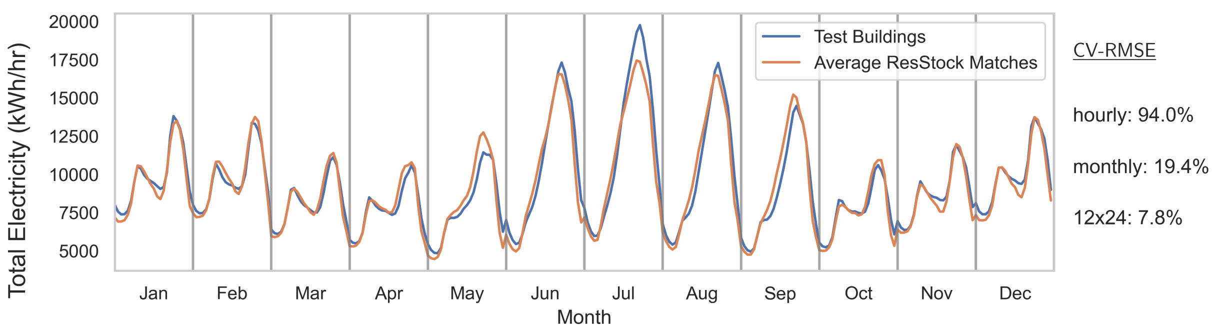


Figure 18. Aggregation of all 10,000 AMI data points and the sum of the average matches for each data point, representing alignment of models for an entire city. The CV-RMSEs on the right of each plot compare the (1) hourly average ResStock matches and summed AMI profiles for every hour of the year, (2) monthly totals, and (3), 12x24 profiles representing the monthly-daily loads shown in the plots and found in the optimization routine.

4 Discussion

This report describes a methodology to identify building energy models for utility service areas using AMI load data. We leverage public ResStock data to generate a model for inferring baseline housing characteristics and align models with AMI data using hourly and peak electricity outputs. The primary application of the workflow is generating localized bottom-up models for building scenario analysis. Although the expansion of smart meter infrastructure has provided valuable information for grid operation and forecasting, very little is typically known about the underlying building details. This impedes the ability to forecast the impacts of electrification or energy efficiency scenarios, which are highly dependent on existing technologies. We detailed one implementation of this framework and simulated case studies with modeled and AMI data.

The first layer of the workflow is an input inference model. This step was isolated to ensure better accuracy of identifying high-level inputs that are important when determining how a home responds to technology changes. Although the inference model predicts only A/C presence, electric heating presence, and electric water heater presence, it could be expanded to potentially capture other inputs, such as electric vehicle (EV) charging or rooftop photovoltaics (PV). The final output of the workflow assigns several ResStock building matches to each AMI data point, meaning distributions of the underlying ResStock inputs can be used to convey probabilities of attributes for a building or group of buildings. We designed the second layer as an optimization routine that minimizes average hourly load error, peak cooling, and peak heating metrics to assign a set of 10 models to a household with AMI data. This approach generally reduces error from the assigned models, but aspects of each model may be improved with updates tailored for specific analyses. Our approach focused on the overall fit of housing attributes and time series data with consideration for peak heating and cooling loads. Other applications may benefit more from ensuring that peak periods are most closely aligned, for example. In this case, a more robust input inference model could be implemented, and an optimization could be used to better minimize peak error instead of relying on it for time series matching and final housing attribute identification.

The following is general guidance on how aspects of the framework may be updated:

- **ResStock data selection**: More targeted selection of buildings that align with the true data beyond just geography, such as building type or build year. If this information is available, it should be used to filter potential ResStock matches or apply penalties in the optimization to improve baseline technology predictions and scenario analysis.
- **Feature selection**: Additional or alternative features can improve usefulness of either layer. Updates to the input inference model should include those that directly inform the housing characteristics; this could include additional load information such as natural gas usage or electric vehicle charging load, or other attributes that are interdependent with housing characteristics. The optimization layer parameters may not need to directly correlate with housing attributes but may focus on specific periods of interest or grid planning priorities. Specifically, introducing the timing of peaks to the optimization formulation can help the model align with peak demand patterns.
- **Model selection**: Alternatives to the input inference model could be explored to improve housing attribute predictions, including different classification models, clustering, neural networks, change-point models, and other methods used in non-intrusive load monitoring.
- **Optimization objective function**: If the information is available for individual buildings, greater penalty for deviating from certain attributes could ensure a more accurate baseline set of synthetic loads. This could help inform the impacts of future scenarios. Additionally, terms that minimize error across aggregations of buildings may be useful to ensure accuracy at distribution network levels.
- **Parameter tuning**: Selection of the input inference threshold and weighting of peak terms in the optimization may be fine-tuned to prioritize specific elements of the model.
- **Introduce weather**: Weather influences the modeled and AMI load data, however, the response of weather is not explicitly introduced as an input to the modules. Extracting the relationships between the load and weather variables could help better identify certain features such as HVAC and envelope efficiency.

Results are presented to understand the error of energy-related outputs and the accuracy of baseline housing characteristics, both of which are important to consider when modeling scenario impacts. On the individual building level, calculating energy error is important to confirm suitability of the matches and their housing attributes. Analyzing energy results at different aggregations helps assess the ability to represent segments of buildings at varying grid resolutions. Uncertainty is highest when representing a single building at hourly timesteps due to the stochastic nature of loads, driven by occupancy behavior. As we increase the number of buildings, variability is averaged and loads become smoother, resulting in a better match between AMI and modeled data. Aggregating housing segments also demonstrates the dependency of variables like building type and build year on model performance.

Each ResStock model corresponds to a set of unique housing characteristics and end-use load data, which allows for results to drive insights in a number of ways. First, the synthetic matches of each AMI data point can be used to produce distributions of housing characteristics and extract probabilities that a characteristic belongs to a building or a group of buildings. With this information, we can (1) provide a better understanding of the building stock characteristics, which inform the applicability of electrification and energy efficiency scenarios and (2) simulate those scenarios and analyze the resulting energy impacts. This can unlock new information for distribution system planners to anticipate how building scenarios influence loads in the future. Second, modeled end-use load data can be used to disaggregate load profiles, informing what drives annual and time series energy usage, which could enable smarter design of energy efficiency and demand-response programs. By generating outputs at the individual building level, these insights may help inform decisions related to demand-side management and distribution networks.

This report documents and demonstrates an approach to align residential building models using AMI load data. One area for future work would be to extend the framework's scope to include commercial building loads. Although the overall concepts should be similar, key aspects would be different, such as the choice of characteristics in the inference models. Based on past experience with commercial building AMI data, another key difference is that commercial building stock model datasets represent whole buildings, whereas commercial buildings are often served by multiple AMI meters. Another area for further work would be evaluating the framework using AMI data from other locations around the country that have different building stock characteristics, different weather, and potentially different occupant behavior. Assessing the sensitivity to varying data characteristics such as temporal resolution or the inclusion of sub-metered data for PV, batteries, or EVs could further clarify the applicability of the methodology. Lastly, extending the framework to natively support scenario analysis would help users derive insights more quickly.

References

Bass, B., E. Ezell, and J. New. 2022. "Using Measured Building Energy Data to Infer Building Characteristics for Urban Building Energy Modeling." In *2022 Building Performance Analysis Conference and SimBuild,* 10:173–180. ASHRAE/IBPSA-USA Building Simulation Conference. Chicago, Illinois: ASHRAE/IBPSA-USA, September. ISBN: 978-1-955516-21-1. https://doi.org/https://doi.org/10.26868/25746308.2022.C021.

Beckel, C., L. Sadamori, T. Staake, and S. Santini. 2014. "Revealing household characteristics from smart meter data." *Energy* 78:397–410. ISSN: 0360-5442. https://doi.org/https://doi.org/10.1016/j.energy.2014.10.025. https://www.sciencedirect.com/science/article/pii/S0360544214011748.

Blonsky, M., A. Nagarajan, S. Ghosh, K. McKenna, S. Veda, and B. Kroposki. 2019. "Potential Impacts of Transportation and Building Electrification on the Grid: A Review of Electrification Projections and their Effects on Grid Infrastructure, Operation, and Planning." Publisher Copyright: © 2019, Springer Nature Switzerland AG. *Current Sustainable/Renewable Energy Reports* 6, no. 4 (December): 169–176. ISSN: 2196-3010. https://doi.org/10.1007/s40518-019-00140-5.

Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785–794. KDD '16. San Francisco, California, USA: Association for Computing Machinery. ISBN: 9781450342322. https://doi.org/10.1145/2939672.2939785. https://doi.org/10.1145/2939672.2939785.

Deng, Z., Y. Chen, J. Yang, and Z. Chen. 2022. "Archetype identification and urban building energy modeling for city-scale buildings based on GIS datasets." *Building Simulation* 15 (January). https://doi.org/10.1007/s12273-021-0878-4.

El Kontar, R., J. Robertson, K. N. Cu, A. Grayson, J. Ling, H. Sotiropoulos, and T. Rakha. 2024. "An Open-Source Framework for Characterizing Urban Energy Models: Integrating Top-Down and Bottom-Up Methods to Predict Residential Buildings Characteristics." In *Proceedings of the 2024 ACEEE Summer Study on Energy Efficiency in Buildings.* American Council for an Energy-Efficient Economy (ACEEE), August.

Energy Systems Integration Group. 2022. *Long-Term Load Forecasting: Workshop Summary.* Workshop held by the Energy Systems Integration Group, June 13-15, 2023, Reston, VA. https://www.esig.energy/event/2023-long-term-load-forecasting-workshop/.

———. 2024. *Grid Planning for Building Electrification.* Technical report. Reston, VA: Energy Systems Integration Group. https://www.esig.energy/grid-planning-for-building-electrification.

Keen, J., J. Giraldez, E. Cook, A. Eiden, S. Placide, A. Hirayama, B. Monson, D. Mino, and F. Eldali. 2022. *Distribution Capacity Expansion Planning: Current Practice, Opportunities, and Decision Support,* November. https://doi.org/10.2172/1898008.

Kissock, K., J. Haberl, and D. Claridge. 2002. *Development of a Toolkit for Calculating Linear, Change-Point Linear and Multiple-Linear Inverse Building Energy Analysis Models.* ASHRAE Research Project RP-1050, Atlanta.

Lindberg, K., P. Seljom, H. Madsen, D. Fischer, and M. Korpås. 2019. "Long-term electricity load forecasting: Current and future trends." *Utilities Policy* 58:102–119. ISSN: 0957-1787. https://doi.org/https://doi.org/10.1016/j.jup.2019.04.001.

McLoughlin, F., A. Duffy, and M. Conlon. 2015. "A clustering approach to domestic electricity load profile characterisation using smart metering data." *Applied Energy* 141:190–199. ISSN: 0306-2619. https://doi.org/https://doi.org/10.1016/j.apenergy.2014.12.039.

Milić, V., P. Rohdin, and B. Moshfegh. 2021. "Further development of the change-point model – Differentiating thermal power characteristics for a residential district in a cold climate." *Energy and Buildings* 231:110639. ISSN: 0378-7788. https://doi.org/https://doi.org/10.1016/j.enbuild.2020.110639.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2023. *Cross-validation: evaluating estimator performance.* Accessed: 2024-08-22. https://scikit-learn.org/stable/modules/cross_validation.html.

Perez, K. X., K. Cetin, M. Baldea, and T. F. Edgar. 2017. "Development and analysis of residential change-point models from smart meter data." *Energy and Buildings* 139:351–359. ISSN: 0378-7788. https://doi.org/https://doi.org/10.1016/j.enbuild.2016.12.084.

Perez, K. X., W. J. Cole, M. Baldea, and T. F. Edgar. 2014. "Meters to Models: Using Smart Meter Data to Predict Home Energy Use." In *Proceedings of the 2014 ACEEE Summer Study on Energy Efficiency in Buildings.* American Council for an Energy-Efficient Economy (ACEEE).

Roy, J., S. Mitchell, and F. Peschiera. 2023. *PuLP (Version 2.4.1) [Software].* https://github.com/coin-or/pulp.

Schirmer, P. A., and I. Mporas. 2023. "Non-Intrusive Load Monitoring: A Review." *IEEE Transactions on Smart Grid* 14 (1): 769–784. https://doi.org/10.1109/TSG.2022.3189598.

Wang, X., and C. Wang. 2020. "Time Series Data Cleaning: A Survey." *IEEE Access* 8:1866–1881. https://doi.org/10.1109/ACCESS.2019.2962152.

Westermann, P., C. Deb, A. Schlueter, and R. Evins. 2020. "Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data." *Applied Energy* 264:114715. ISSN: 0306-2619. https://doi.org/https://doi.org/10.1016/j.apenergy.2020.114715.

Wilson, E., A. Parker, A. Fontanini, E. Present, J. Reyna, R. Adhikari, C. Bianchi, C. CaraDonna, M. Dahlhausen, J. Kim, A. LeBar, L. Liu, M. Praprost, P. White, et al. 2022. *End-Use Load Profiles for the U.S. Building Stock.* V. 2022.1. United States, September. https://doi.org/10.25984/1876417.

Wilson, E., A. Parker, A. Fontanini, E. Present, J. L. Reyna, R. Adhikari, C. Bianchi, C. CaraDonna, M. Dahlhausen, J. Kim, A. LeBar, L. Liu, M. Praprost, L. Zhang, et al. 2022. *End-Use Load Profiles for the U.S. Building Stock: Methodology and Results of Model Calibration, Validation, and Uncertainty Quantification,* March. https://doi.org/10.2172/1854582.

Zhou, E., S. Gadzanku, C. Hodge, M. Campton, S. de la Rue du Can, and J. Zhang. 2023. *Best Practices in Electricity Load Modeling and Forecasting for Long-Term Power System Planning.* Other. https://doi.org/10.2172/1972011.