Beyond Average: Evaluating Indoor Average Temperature in Grey Box Modeling

Ozan Baris MULAYIM[1*], Mario BERGÉS[1]

[1] Carnegie Mellon University
Pittsburgh, PA, USA
omulayim@andrew.cmu.edu, marioberges@cmu.edu

* Corresponding Author

ABSTRACT

Grey Box Modeling (GBM) strategies have proven effective in identifying the thermal behavior of buildings. The widespread adoption of Smart Thermostats (STs) has unveiled large datasets, which are now being harnessed for such modeling strategies. A notable limitation of earlier GBM approaches for STs was their sole focus on a single sensing point. With the proliferation of additional sensors in homes, STs have begun integrating more sophisticated control algorithms such as conditioning based on the average temperature of occupied rooms. When conducting GBM using the Indoor Average Temperature (IAT), the modeling assumptions can falter due to the inconsistent inclusion of sensors in averaging for the control temperature. In this study, we leverage a publicly available dataset from 1,000 homes across the United States, each outfitted with varying numbers of sensors, to identify grey box models. Beyond one-step ahead prediction models, we apply Model Predictive Control Relevant Identification to assess the effectiveness of room-level modeling in one-day ahead predictions. Our findings indicate that varying occupancy profiles typically result in a 0.5 to 3.3°F discrepancy in the accuracy of one-hour-ahead predictions when utilizing IAT for modeling. Also, distinctly modeling individual sensors yielded a 1.8 to 25.8% enhancement in the accuracy of one-hour and one-day ahead predictions, respectively.

1. INTRODUCTION

Model Predictive Control (MPC) has demonstrated promising results in several residential applications. The modeling required for MPC can adopt a white-box, grey-box, or black-box approach. White-box models, although highly interpretable, demand significant engineering efforts to develop. On the other hand, black-box models, while easier to implement, often suffer from a lack of interpretability. As a result, grey-box modeling (GBM) has gained popularity as a hybrid approach since it effectively merges the precision of physics-based modeling with the adaptability of machine-learning techniques, thus achieving a balance between computational efficiency and interpretability.

Many studies have conducted parameter identification for GBM using smart thermostat datasets due to their common availability, supported by their high adoption rate (11% in the US (King, 2018)). However, most of these studies use either Indoor Average Temperature (IAT) (Yu, Abhari, Fung, Raahemifar, & Mohammadi, 2018) or thermostat temperature (Li, Pinto, Capozzoli, & Hong, 2022) as their primary indication of house temperature. Each method has its limitations. The former might represent the thermal dynamics in the house better since it averages the sensor readings. However, it tries to model the combination of different rooms in every step since the averaging would only consider *active sensors* (i.e., sensors that are either assigned by a schedule or activated by motion detection). The working mechanism of an *ecobee* thermostat is visualized in Figure 1. The latter is beneficial in the sense that it always captures the thermal dependencies of the same sensor (i.e. thermostat), but it might fail to capture the thermodynamics of the house sufficiently due to significant temperature variances existing in single-zone multi-node systems (Mulayim & Bergés, 2023).

In this study, we aim to investigate the consequences of such assumptions in the predictive accuracy of the developed models. We consider one-hour and one-day ahead prediction capabilities during the cooling season of around 850 houses with number of additional sensors varying from 1 to 5. While the models trained with thermostat temperature are included in our analysis, our main focus is on the effect of different occupancy patterns on the accuracy of Average-Based Modeling predictions. Given that these are prevalent practices in GBM, pinpointing conditions where it might fall short could prove beneficial for enhancing the reliability of GBM for smart thermostats. Additionally, we offer and analyze the benefits of a new GBM framework, called Distributed Sensor Modeling, where each sensor is modeled separately, and then their predictions are averaged to reach the IAT prediction.

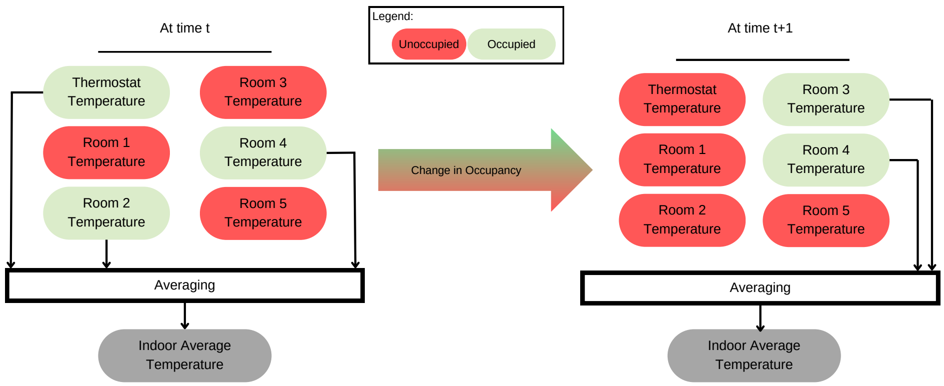

Figure 1: The dynamic process of indoor temperature averaging by an *ecobee* thermostat across different rooms over two consecutive time points. At time t, the thermostat gathers temperature readings from multiple rooms—Rooms 1, 2, 3, 4, and 5—distinguishing between occupied and unoccupied spaces as denoted by the color-coded legend. These readings are then averaged to compute the overall indoor temperature. A subsequent change in room occupancy leads to a revised temperature averaging at time t+1, reflecting the change in the physical space considered in averaging.

The remainder of the paper is structured as follows: Section 2 provides a background on the related work. Section 3 gives an explanation of our methodologies and dataset, followed by the discussion of our results in Section 4. The paper completes with a detailed discussion and conclusion in Section 5.

2. BACKGROUND

The current literature has extensively leveraged large subsets of *ecobee* Donate Your Data (DYD) dataset for various modeling endeavors. For instance, a subset dataset comprising 10,000 houses was used to estimate thermal time constants (John, Vallianos, Candanedo, & Athienitis, 2018). Three different methods to identify grey box modeling parameters have been compared using winter months' data from 4,000 houses (Baasch, Wicikowski, Faure, & Evins, 2019). The predictive capabilities of data-driven models developed by using 1,000 houses within the dataset have been investigated by (Huchuk, Sanner, & O'Brien, 2022). The authors (Wang, Chen, Li, & Hong, 2021) also used a large dataset to train several grey box models but their focus was on developing a simulation environment. Bayesian Neural Networks have been used to identify grey box models with data consisting of 8,884 houses (Hossain, Zhang, & Ardakanian, 2021). The recent work (Vallianos, Candanedo, & Athienitis, 2023) bears the most similarity to our study. They utilized data from 7,800 houses in Canada to explore the data length, data interval, and calibration horizon of building models for MPC trained for one-step and one-day ahead predictions. Huchuk, Sanner, and O'Brien (2019) were the first one to use the occupancy data provided by remote sensors. The authors considered the occupancy status from all sensors to find a ground truth binary occupancy level of the whole house for 100 houses. They have also demonstrated that as the number of sensors in a house increases, the frequency of the occupancy increases, depicting the insufficiency of only using thermostats to evaluate occupancy. Lastly, significant temperature discrepancies of multiple rooms from the thermostat temperature and IAT have been identified in our previous work utilizing the same dataset as this work (Mulayim & Bergés, 2023), marking attention to the need for considering room-level temperature readings when making predictions for MPC.

The existing literature underscores that grey box model development using *ecobee* DYD datasets (Ecobee,

2023) remains a contemporary research area. However, to the best of the authors' knowledge, no studies have considered the prediction capabilities of these models with an emphasis on the role of additional sensors. Given these research gaps, this study provides two contributions: (1) investigating the effect of different occupancy patterns in average-based grey box models for one-step and one-day ahead predictions, and (2) offering a potential improvement to the existing modeling status quo by analyzing the benefit of distinctly modeling each sensor.

3. METHODOLOGY

3.1 Data

ecobee DYD program is estimated to have around 120,000 houses (Mulayim, Severnini, & Bergés, 2024). Researchers so far have utilized different sizes of subsets of this data for their research based on their agreements with *ecobee*. In this study, we are utilizing a publicly available subset of the *ecobee* DYD dataset consisting of 1,000 houses (Luo & Hong, 2022). The dataset has a resolution of 1°F and a sampling rate of 5 minutes, featuring houses from four states (i.e., TX, CA, NY, IL) in the US that represent a range of different climates. Our initial investigation revealed that making one-step ahead predictions using the original 5-minute sampling rate was insufficient to capture temperature changes due to external factors, primarily because of the 1°F resolution. As a result, we resampled the data to a 60-minute interval, which is commonly used by most utilities (Rhodes, Cole, Upshaw, Edgar, & Webber, 2014). For temperature values, resampling was done by averaging the values within the hour, while the cooling equipment runtime was summed. Motion values were resampled by taking the maximum of the values within that hour. Using the metadata provided, we supplemented our dataset with global horizontal irradiance data from the National Solar Radiation Database API (National Renewable Energy Laboratory, 2023) based on the latitude and longitude of the city, state pairs we extracted from the metadata, a methodology similar to the work of Vallianos et al. (2023). We narrowed our focus to the cooling season and extracted 32 continuous days of data for each house. We allocated 14 days for training and 2 days for testing our one-day and one-step ahead prediction models, aligning with recent studies for direct comparison (Vallianos et al., 2023). The remaining 16 days were used to evaluate the long-term performance of our Average-Based Modeling and Distributed Sensor Modeling approaches. Though the initial dataset had 1000 houses, 841 of them had at least one or more additional sensors. Since our focus was on identifying the limitations of using IAT, we scoped our analysis to those with at least one or more additional sensors. The distribution of the houses with additional sensors can be seen in Figure 2. The number of sensors in each house is determined by checking the existence of 32 consecutive days of data for each remote sensor.

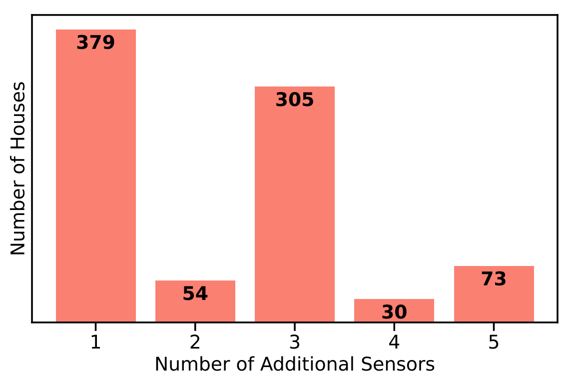

Figure 2: Bar chart depicting the distribution of the number of additional sensors installed in homes, following a data curation process that extracted 32 consecutive days. The bars represent the count of houses that are equipped with 1 to 5 additional sensors.

3.2 Grey Box Modeling

After preprocessing, the data is used to construct a linear approximation to first-order thermal resistance-capacitance models with ridge regression as the optimization method. The structure of the grey box model can be seen in Equation 1.

$$T_{in,t+1} = \beta_1 T_{in,t} + \beta_2 T_{out,t} + \beta_3 q_{cool,t} + \beta_4 q_{solar,t} + \epsilon_t \tag{1}$$

T represents the temperature reading while subscripts t, in, out stands for time, indoor and outdoor, respectively. $q_{cool,t}$ stands for cooling runtime at time t while $q_{solar,t}$ is the term for global horizontal irradiance. Time-dependent error is termed as ϵ_t. This model is calibrated by minimizing the error between predictions in the next step (i.e. 1 hour) and ground truth. Separate models are trained using different temperature data such as IAT, thermostat, and remote sensors.

Though one-step ahead prediction models have been used in multi-step predictions by making regressive predictions, their accuracy seems to suffer considerably in such cases. Models calibrated through the use of MPC Relevant Identification (MRI) have demonstrated better prediction performance when considering the entire prediction horizon, as compared to recursively using one-step ahead prediction models (Žáčeková, Váňa, & Cigler, 2014). The objective function in MRI is given below in Equation 2.

$$J_{\mathrm{MRI}} = \sum_{k=0}^{N-P} \sum_{i=1}^{P} \left(T_{in,k+i} - \hat{T}_{in,k+i|k} \right)^2 \tag{2}$$

N represents the number of samples, P is the prediction horizon, $T_{in,k+i}$ is the actual temperature reading at time $k+i$ while $\hat{T}_{in,k+i|k}$ is the i-step prediction of temperature using data starting from k. Levenberg–Marquardt algorithm has been used for optimization. Due to the nature of the genetic algorithm, an initial guess for the parameters is required. To achieve an educated guess, coefficients identified for the first model ($\beta_{\{1,2,3,4\}}$) have been used as the initial guess. Further details of this approach can be found in the work of Žáčeková et al. (2014).

It is important to note that the one-step ahead prediction models are equivalent to one-hour ahead prediction models, as we are utilizing an hourly sampling rate. Similarly, our one-day ahead prediction models are, in fact, multi-step forecasts that leverage the subsequent 24 time steps. Throughout this paper, we will use 'one-step ahead' and 'one-day ahead' predictions to refer to those concepts.

3.3 Model Predictive Control

After training models for each individual sensor and the IAT with varying occupancy patterns, we introduce two methodologies for predicting the control temperature (i.e., IAT) to facilitate a comparative analysis.

- *Average-Based Modeling*: This model, trained using the historical IAT data, predicts the average indoor temperature one-day or one-hour ahead. This approach represents the existing methodologies in the literature.
- *Distributed Sensor Modeling*: Here, distinct sensor models generate predictions of their own. Subsequently, an average of the predictions from *active sensors* at the given prediction horizon is computed and outputted as the estimated IAT.

Root Mean Squared Error (RMSE) is employed to evaluate the accuracy of the predictions. The ground truth is taken as the true value of the IAT at the times for which predictions are made. In order to deploy Distributed Sensor Modeling, we need to identify the *active sensors* contributing to the IAT at each time step so the predictions of distinct sensor models can be averaged accordingly. Details of our approach regarding that are explained in Section 3.5.

3.4 Occupancy Profile Definitions

Though we have an IAT in our dataset, only using it would limit the analysis since there might be different occupancy patterns that are not present in our dataset. Thus, an additional analysis is required to understand the effect of occupancy on the predictive capabilities of the Average-Based Modeling approach. To evaluate the situations where averaging might either fail or succeed, we have outlined several occupancy profiles. These are described as follows:

- *Actual*: The provided occupancy data is retained. This is done by utilizing the IAT available in the dataset as is.

- *Full*: All rooms are considered occupied.
- *Thermostat*: Only the room with the thermostat is deemed occupied.
- *Random*: Room occupancy is determined randomly.
- *Motion*: Occupancy is determined based on the detected motion. If there is no motion detected, the thermostat temperature is taken as the IAT.
- *Worst-Case*: This represents a scenario where users move between the hottest and the coldest room at each time step, displaying the extreme scenario.

For each occupancy scenario, we compute a new average. These new averages are then used to construct grey box models for both one-step and one-day ahead prediction models.

3.5 Defining *Active Sensors*

The *ecobee* dataset includes residences with additional sensors, ranging from 1 to 5. However, the dataset does not indicate whether sensors are active or inactive at any particular moment. *Active sensors*, which contribute to IAT calculations, are central to our study. Thermostats have two ways of having *active sensors* at a given time. They either rely on motion sensors or use pre-defined schedules based on the preferences of the user. To further explain, if users prefer setting a sensor schedule, *active sensors* would be different from the occupied rooms detected by motion sensors. Thus, the specific identification of the *active sensors* contributing to the IAT at any given point in time is not available in the dataset.

Consequently, we have defined two scenarios for *active sensors* and computed their corresponding IATs. Since the motion data is available, we considered that as the active sensors at each timestep, which would be identical to houses relying on motion sensors for occupancy detection. In the case of houses relying on pre-set schedules for defining *active sensors*, we needed to create a general schedule since it is not available in the dataset. The defined occupancy schedule is visualized in Figure 3, where occupancy is marked with green boxes. We defined separate schedules for weekdays and weekends, dividing each day into five phases: morning, day, evening, late evening, and night, with an assumption of the house being largely unoccupied during the daytime on weekdays, representing a common workday schedule. We have made assumptions regarding the positioning of the sensors to have a generally applicable schedule. The thermostat is assumed to be located in an area that has activity in the morning such as a kitchen or a hallway, whereas remote sensor 1 is placed in a setting akin to a living room. Other sensors are assumed to be in bedrooms, with the exception of remote sensor 5, which corresponds to a multi-purpose room. This investigation is tailored to homes equipped with five additional sensors. For a comprehensive assessment of their performance over an extended period, we analyzed data spanning 16 days that were not part of the training dataset.

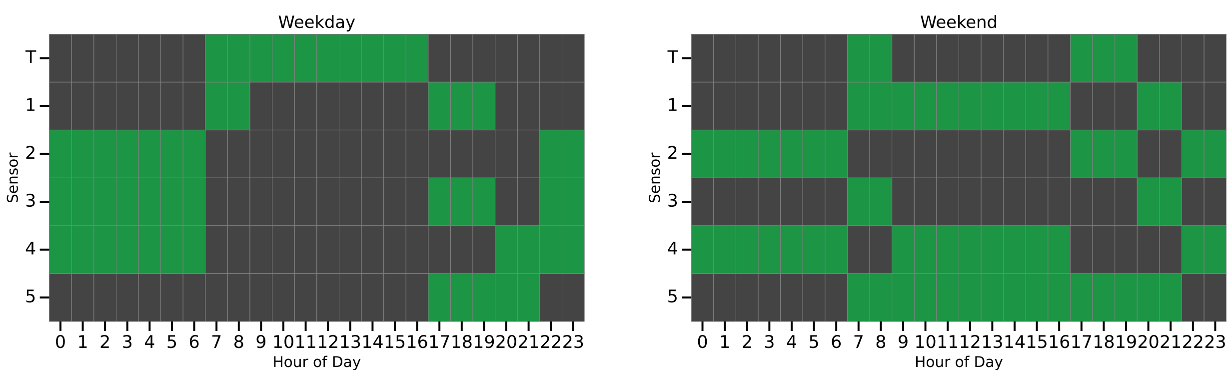

Figure 3: Pre-defined occupancy patterns for each sensor throughout the hours of a typical weekday and a weekend. Each row represents a sensor, with the thermostat indicated as 'T', while each column corresponds to an hour of the day. Green blocks signify periods of occupancy, and black blocks indicate non-occupancy.

Overall, the purpose of this analysis is to explore the benefits of Distributed Sensor Modeling in GBM compared to the traditional Average-Based Modeling for houses that rely on either assigned schedules or motion detection for detecting occupancy.

4. RESULTS

4.1 Effect of Occupancy Profile

The preprocessing of data, as detailed in Section 3.1, results in the categorization of houses based on the number of additional sensors. Each house have IATs computed with six distinct occupancy profiles (Section 3.4). Each IAT data is used to train two models: one-step ahead and one-day ahead predictions (Section 3.2). Considering six occupancy profiles for 841 houses, 5,046 models were trained for each predictive framework. Figure 4 presents the distribution of prediction RMSE for the testing datasets under these two predictive frameworks. In the one-step ahead analysis, scenarios involving *Full*, *Actual*, and *Thermostat* demonstrate commendable performance, with median RMSE values approximately around 0.5-0.6°F. An increase in errors is noticeable for the *Motion* and *Random* occupancy scenarios, with the *Worst Case* scenario depicting the lowest accuracy levels.

Transitioning to the one-day ahead predictions, an increment in errors is observed across all scenarios, except for the *Worst Case*. This observation aligns with the hypothesis suggesting that the MRI enables the model to adjust adequately for temperature variations that typically occur in long-term forecasts, thereby minimizing errors. Although the error range for other scenarios increases, their relative performance rankings remain somewhat stable, with a noticeable reduction in the performance disparity between the *Motion* and *Random* scenarios against the *Full*, *Actual*, and *Thermostat* scenarios. The *Worst Case* scenario tends to perform comparably to other scenarios, particularly at lower sensor counts. Overall, the change in error range across different sensor counts is minimal for most scenarios except for the *Worst Case*, where an increase in error range is noted as sensor numbers rise. Notably, the *Full* occupancy scenario consistently outperforms others for both prediction types, a predictable outcome given the stable physical environment it represents. The *Actual* schedule often mirrors the performance of the Thermostat, potentially due to the houses utilizing motion sensors for occupancy detection, which equates the thermostat temperature to the IAT when no room is occupied.

It is important to note that the *Worst Case* scenario is inherently expected to perform poorly as the number of sensors increases, owing to the potential for significant temperature disparities between rooms. However, fewer sensors do not necessarily imply smaller houses or fewer rooms (Mulayim et al., 2024). Thus, it is crucial to recognize that similar deviations from the predictions might still persist in houses with fewer sensors, but the temperature variations in unmonitored areas remain undetected.

Table 1 provides a comprehensive overview of RMSE statistics across different occupancy scenarios, encompassing all houses to present a broader perspective rather than focusing on specific sensor counts. For one-step ahead predictions, the *Full* scenario secures the lowest mean RMSE at 0.49°F, setting a reliable benchmark for accuracy. In contrast, the *Worst Case* scenario reveals the most significant increase in mean RMSE relative to the *Full*, with a substantial 571% surge, highlighting its volatility and the challenges in maintaining predictive accuracy under such conditions. Surprisingly, the *Actual* occupancy, which leverages actual occupancy data from the dataset, performs surprisingly well, slightly outperforming the *Thermostat* scenario. Other scenarios, such as *Random* and *Motion*, show considerable increases in RMSE, displaying their reduced reliability compared to the *Full*. Notably, the *Motion* scenario's performance is nearly double that of the *Actual* occupancy, highlighting potential inefficiencies in motion sensors for detecting static occupants, thus indicating inconsistent occupancy patterns across many houses. This observation may point to the suboptimal performance of motion sensors with static occupants. The impact on houses using a scheduled approach to detect occupancy remains unexplored here but will be further investigated in the upcoming section.

Conversely, in one-day ahead predictions, the variations in error increases among different occupancy scenarios are less marked, suggesting improved consistency and diminished volatility in long-term forecasts, although mean errors remain elevated. For most scenarios, there is a discernible reduction in error increase from the *Full* scenario when switching to one-day ahead predictions. Occupancy scenarios still maintain the same performance ranking across both prediction intervals, which depicts the consistency in the performance outcomes for each occupancy scenario. This trend may signify the benefits of MRI strategies empirically,

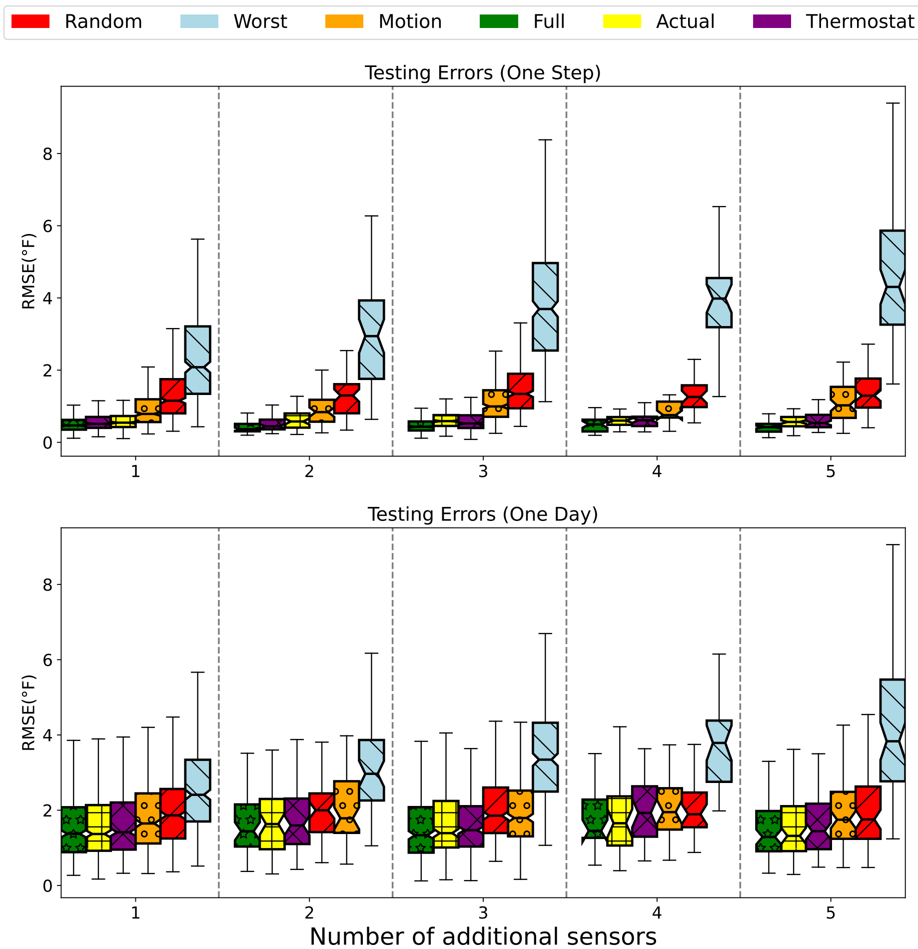

Figure 4: Distribution of RMSE across varying sensor counts for different occupancy scenarios during testing phases of one-step ahead and one-day ahead predictions. This visualization aids in evaluating the prediction accuracy of each scenario against the number of additional sensors. Boxes are ranked by their median error value ranging from small to large.

which calibrates models based on their predictive performance over the prediction horizon. Such findings underscore the substantial variability that can manifest in GBM predictions based on the occupancy patterns utilized to compute the IAT, emphasizing the critical role of adaptive prediction strategies in enhancing long-term reliability and accuracy in residential settings.

4.2 Performance of Distributed Sensor Modeling

In an effort to evaluate the efficacy of Distributed Sensor Modeling in contrast to the Average-Based Modeling approach as detailed in Section 3.3, models were trained for six sensors and two IATs in 73 of the houses as outlined in Section 3.2. To discern the variance between houses employing motion sensors and those utilizing a predetermined schedule, we have defined *active sensors* as described in Section 3.5, subsequently generating two distinct IAT scenarios: (1) leveraging available motion data, and (2) following a pre-defined schedule.

Figure 5 showcases the RMSE distribution for both Average-Based Modeling and Distributed Sensor Modeling methods. These values represent the performance for houses with five additional sensors over an unseen 16-consecutive-day dataset. The mean RMSE for each modeling approach is denoted by black diamonds on the plot. Initial findings indicate that one-step ahead predictions improve by 1.7% and 2.0% for *Motion* and *Schedule* occupancy scenarios, respectively. For one-day ahead predictions, these improvements markedly

Table 1: Summary of Testing RMSE (°F) statistics for each occupancy scenario, including difference and percentage increase from Full

	One-Step Ahead Predictions					One-Day Ahead Predictions				
Occupancy	**Mean**	**Std**	**Min**	**Max**	**Diff (% Inc)[1]**	**Mean**	**Std**	**Min**	**Max**	**Diff (% Inc)[1]**
Full	0.49	0.22	0.11	1.77	0.00 (0%)	1.71	2.23	0.12	57.70	0.00 (0%)
Actual	0.62	0.27	0.10	2.78	+0.13 (27%)	1.78	1.83	0.15	43.06	+0.07 (4%)
Thermostat	0.62	0.34	0.08	2.88	+0.13 (27%)	1.88	3.13	0.13	85.53	+0.17 (10%)
Motion	1.07	0.92	0.23	21.25	+0.58 (118%)	2.17	4.79	0.16	136.27	+0.46 (27%)
Random	1.43	0.80	0.31	7.40	+0.94 (192%)	2.18	2.15	0.36	54.30	+0.47 (28%)
Worst Case	3.29	2.03	0.43	31.31	+2.80 (571%)	3.30	1.98	0.52	22.84	+1.59 (93%)

[1] *Diff (% Inc) indicates the difference and percentage increase in Mean RMSE compared to the Full Average.*

increase to 25.3% and 26.4% for *Motion* and *Schedule* occupancy scenarios, respectively. In general, we observe that utilizing a *Schedule* achieves lower errors than *Motion*, which is in line with our expectations for Average-based Modeling since the sensors contributing to the average are more stationary. While the Distributed Sensor Modeling approach does offer a reduction in error, its advantages are not significantly evident for one-step ahead predictions, likely due to minimal occupancy shifts within such a short timeframe. However, it is important to recognize that averaging temperatures predicted by individual sensor-based models might restrict the breadth of available data to mere averages. Demand response aggregators or utilities could employ these room-level predictions more granularly to better assess the thermal comfort implications of their actions. Moreover, the proposed control algorithm could also enhance the reliability of predictions in extreme cases, such as the *Worst-Case* scenario.

This analysis underscores the advantages of distinctly modeling sensors within GBM for predictive accuracy. Collectively, the findings suggest that Distributed Sensor Modeling surpasses traditional Average-Based Modeling, showing a 1.8% and 25.8% enhancement in one-step and one-day ahead predictions, respectively. Yet, if this method were to be integrated into an MPC system, it raises a question about predicting which sensors will be *active* in the next hour/day. For houses that utilize a schedule to determine *active sensors*, such identification could be achieved using historical data. In cases where motion sensors are utilized for occupancy, room-level occupancy prediction algorithms could be deployed by extending existing house-level occupancy prediction methodologies in the literature.

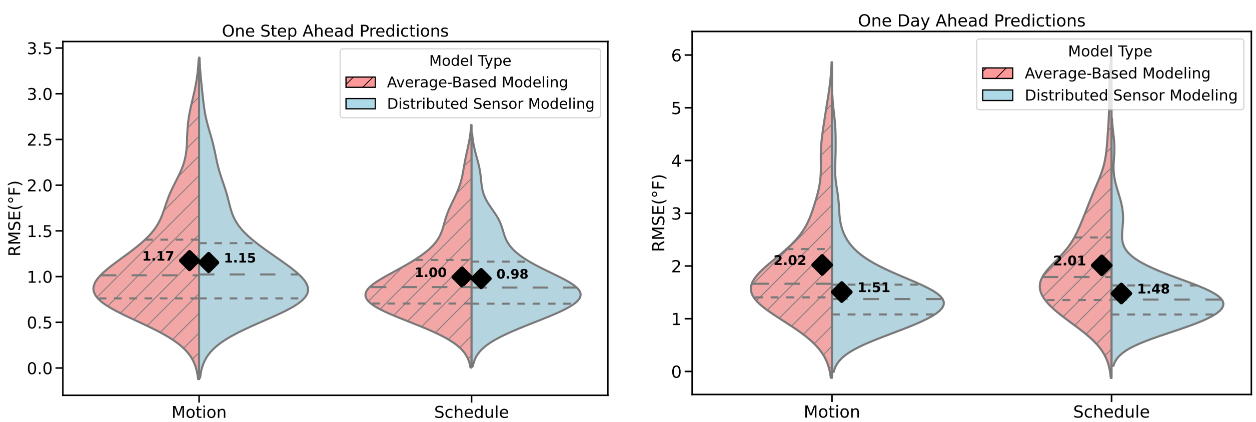

Figure 5: Violin plots representing the RMSE distribution of temperature prediction errors for models utilizing either Motion or Schedule-based occupancy detection strategies. The left plot shows the error distribution for one-step ahead predictions, while the right plot shows the distribution for one-day ahead forecasts. Within each plot, the two types of models—Average-Based Modeling and Distributed Sensor Modeling—are compared. The central diamond markers denote the mean RMSE values, with the dashed lines indicating the interquartile ranges.

5. CONCLUSION

Our research provides compelling empirical evidence challenging the use of a dynamically calculated average of active sensors as the representative temperature in the GBM of houses equipped with STs. First and foremost, we noted that the IAT delivers optimal performance when all sensory measurements are considered in the average, regardless of the sensor count or prediction horizon. This performance can be attributed to the static physical structure considered in the parameter identification process. Moreover, the accuracy of models trained with existing occupancy patterns in the dataset slightly surpassed the accuracy of models trained using thermostat temperature. This suggests that even with potential variations in the physical structure considered during parameter identification, due to averaging temperatures across rooms, high precision can still be achieved in certain cases. However, we observed substantial prediction errors in houses where occupants predominantly use the coldest or hottest rooms or have random occupancy. This underscores the need for a more granular control algorithm to enhance prediction reliability with minimal dependence on occupancy. Therefore, we have utilized a control algorithm using distinct sensor models and then averaging their predictions, which resulted in 1.8% and 25.8% improvement in the accuracy of one-step and one-day ahead predictions, respectively. By discerning user preferences from available data, utilities can identify users who schedule their sensor activation. This knowledge enables utilities to produce more accurate one-day or one-hour ahead IAT forecasts by leveraging individual sensor models for separate predictions. Overall, it is important to clarify that the results of this study do not make conclusions regarding the use of average indoor temperature for existing thermostat controls. Our goal in this paper was to analyze how using such averages as the representative temperature influences the GBM paradigm. Some limitations of this study include the low resolution of the dataset. Future studies could also investigate room-level occupancy prediction to supplement the Distributed Sensor Modeling method. Additionally, our analysis was limited to the cooling season and used a single modeling approach, more extensive modeling approaches and a detailed analysis for the heating season should also be conducted.

NOMENCLATURE

The nomenclature should be located at the end of the text using the following format:

Symbol

$\beta_1, \beta_2, \beta_3, \beta_4$	coefficients	(–)
q_{cool}	cooling runtime	(seconds)
ϵ	time-dependent error	(°F)
q_{solar}	global horizontal irradiance	(W/m^2)
J_{MRI}	objective function in MRI	(–)
N	number of samples	(–)
P	prediction horizon	(–)
$\hat{T}$	predicted temperature	(°F)
T	temperature	(°F)

Subscript

t	time index
in	indoor
out	outdoor
k	time index in MRI
i	step index in prediction

Abbreviation

DYD	Donate Your Data
GBM	Grey Box Modeling
IAT	Indoor Air Temperature
MPC	Model Predictive Control
MRI	Model Predictive Control Relevant Identification
RMSE	Root Mean Squared Error

REFERENCES

Baasch, G., Wicikowski, A., Faure, G., & Evins, R. (2019). Comparing gray box methods to derive building properties from smart thermostat data. In *Proceedings of the 6th acm international conference on systems for energy-efficient buildings, cities, and transportation* (pp. 223–232).

Ecobee. (2023, July). Retrieved from `https://www.ecobee.com/en-us/`

Hossain, M. M., Zhang, T., & Ardakanian, O. (2021). Identifying grey-box thermal models with bayesian neural networks. *Energy and Buildings*, *238*, 110836.

Huchuk, B., Sanner, S., & O'Brien, W. (2019). Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. *Building and Environment*, *160*, 106177.

Huchuk, B., Sanner, S., & O'Brien, W. (2022). Evaluation of data-driven thermal models for multi-hour predictions using residential smart thermostat data. *Journal of Building Performance Simulation*, *15*(4), 445–464.

John, C., Vallianos, C., Candanedo, J., & Athienitis, A. (2018). Estimating time constants for over 10,000 residential buildings in north america: Towards a statistical characterization of thermal dynamics.

King, J. (2018). Energy impacts of smart home technologies. *Report A1801*.

Li, H., Pinto, G., Capozzoli, A., & Hong, T. (2022). Building thermal dynamics modeling with deep learning exploiting large residential smart thermostat dataset. In *Proceedings of the 9th acm international conference on systems for energy-efficient buildings, cities, and transportation* (pp. 242–245).

Luo, N., & Hong, T. (2022). Ecobee donate your data 1,000 homes in 2017. Retrieved from `https://www.osti.gov/biblio/1854924` doi: 10.25584/ecobee/1854924

Mulayim, O. B., & Bergés, M. (2023). Unmasking the thermal behavior of single-zone multi-room houses: An empirical study. In *Proceedings of the 10th acm international conference on systems for energy-efficient buildings, cities, and transportation* (pp. 21–30).

Mulayim, O. B., Severnini, E., & Bergés, M. (2024). *Unmasking the role of remote sensors in comfort, energy and demand response.* ([Online]. Available: arXiv:2404.15368)

National Renewable Energy Laboratory. (2023). *National solar radiation database (nsrdb) api documentation.* `https://developer.nrel.gov/docs/solar/nsrdb/`. (Accessed: 2024-04-10)

Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., & Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, *135*, 461–471.

Vallianos, C., Candanedo, J., & Athienitis, A. (2023). Application of a large smart thermostat dataset for model calibration and model predictive control implementation in the residential sector. *Energy*, *278*, 127839.

Wang, Z., Chen, B., Li, H., & Hong, T. (2021). Alphabuilding rescommunity: A multi-agent virtual testbed for community-level load coordination. *Advances in Applied Energy*, *4*, 100061.

Yu, D., Abhari, A., Fung, A. S., Raahemifar, K., & Mohammadi, F. (2018). Predicting indoor temperature from smart thermostat and weather forecast data. In *Proceedings of the communications and networking symposium* (pp. 1–12).

Žáčeková, E., Váňa, Z., & Cigler, J. (2014). Towards the real-life implementation of mpc for an office building: Identification issues. *Applied Energy*, *135*, 53–62.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the support provided by the Pennsylvania Infrastructure Technology Alliance (PITA). We would like to thank *ecobee* and its customers for their data. Mario Bergés holds concurrent appointments as a Professor of Civil and Environmental Engineering at Carnegie Mellon University and as an Amazon Scholar. This paper describes work at Carnegie Mellon University and is not associated with Amazon.