Dealing with Corner Cases in Occupant-Centric Control: Do Physics-Informed Models Help?

Sicheng Zhan[1], Audun Botterud[2], Jeremy Gregory[3], Leslie K. Norford[1]
[1]Department of Architecture, MIT, Cambridge, MA, USA
[2]Laboratory for Decisions and Information Systems, MIT, Cambridge, MA, USA
[3]Climate and Sustainability Consortium, MIT, Cambridge, MA, USA

Abstract

Occupant-centric control (OCC) enhances building energy performance and thermal comfort by adapting to occupant behavior, and predictive models are essential for informed control decisions. However, existing studies mainly focused on overall performance metrics, overlooking control robustness in certain unseen scenarios. Given the sparse nature of building data, data-driven models often face extrapolation problems. Physics-informed models recently emerged as a promising solution to integrate domain knowledge and real-world data, but their effectiveness in actual buildings remains unclear. In this study, we investigate four predictive models with an increasing level of physics integration, evaluating their predictive performance in OCC using real-world data collected from a group of classrooms. The results reveal the scenarios and conditions for higher physics integration to benefit the decision-making process. Furthermore, we show that a proper balance between physical constraints and model expressiveness is required to handle the corner cases.

Key innovations

- Inspected four levels of physics-informed models in occupant-centric control scenarios.
- Developed a series of extrapolation tests based on data collected from real-world experiments.
- Conducted investigation of corner cases in optimal building controls.
- Revealed the importance of balancing physical constraints and model expressiveness.

Practical implications

The evaluation of control-oriented models should be based on not only predictive errors but also the robustness in rare unseen cases. Physics-informed models are helpful only when model expressiveness and physics constraints are carefully balanced.

Introduction

The design and operations of building systems are traditionally based on standard and conservative assumptions of occupants, presenting great potential for energy conservation (O'Brien et al., 2020). Occupant-centric control (OCC) provides a new approach to operating building systems by sensing and predicting occupants' presence, preference, and activity. Being able to improve building energy performance and thermal comfort, OCC has drawn substantial research interest over the past two decades (Nagy et al., 2023). However, large-scale applications in actual buildings are still lacking for OCC to move beyond the proof-of-concept stage (Park et al., 2019).

Existing studies have focused on evaluating overall performance during a relatively long period. Hobson et al. (2021) assessed the effectiveness of OCC using one year of actual data and demonstrated up to 37% and 65% reduction respectively in energy use and thermal discomfort; Pang et al. (2020) estimated that the annual energy saving ratio of controlling a medium office building based on occupant count sensors varies between 20% and 45% across the 16 climate zones in the US. Nonetheless, a promising energy-saving percentage is not the silver bullet to promote the adoption of OCC, and robustness is an important yet overlooked issue (Henze et al., 2024).

Corner cases, commonly used in autonomous driving, refer to bugs or problems that occur under unseen, unusual, and perhaps dangerous circumstances (Bogdoll et al., 2021). For OCC in buildings, corner cases can be triggered by extreme values or rare patterns of disturbances that are not covered by historical data used to train the controllers. The backbone of many OCC applications is the predictive model, and the root of potential corner cases is that data-driven models sometimes cannot extrapolate well to support decision-making (Zhan et al., 2022). As the predictive model cannot accurately comprehend the situation or predict what will happen, corner cases could lead to energy wastage or thermal discomfort with suboptimal control actions.

Unlike other data-rich domains, building data typically only covers limited operating conditions and therefore is sparse in the feature space, challenging the training of purely data-driven models (Hiyama

https://doi.org/10.26868/25222708.2025.1808

and Omodaka, 2021). For example, a model trained with data associated with a normal heating setpoint of 21°C may be unreliable to maintain a lower setpoint of 18°C in control (Zhang, 2021). Similar to other complex and open systems that are under-measured, domain knowledge is needed to learn the physical dynamics (Willcox et al., 2021).

Recently, physics-informed neural networks have emerged to enhance the predictive capability of data-driven models by incorporating physics-based prior knowledge (Karniadakis et al., 2021). A common approach is to integrate physical constraints to penalize parameters or predictions that violate physical principles when training the model, such as eigenvalue constraints for physically meaningful parameters (Drgoňa et al., 2021) and an additional loss function term for wall temperatures (Gokhale et al., 2022). Liang et al. (2023) included the residual of heat and mass balance in the loss function to model a chiller and showed better predictive accuracy in new testing conditions. Further, Di Natale et al. (2022) modeled basic heat transfer and uncertain disturbances separately, using a linear model and a parallel long short-term memory (LSTM), to guarantee the physics consistency of open-loop predictions.

While these physics-informed methods seem to offer a good opportunity to improve the optimal control of building systems, their practical value is yet to be shown. Despite such claimed merits as more data-efficient learning and physics-consistent prediction over long horizons, whether and how these theoretical advantages can be translated into control performance remains unknown. Particularly, the efficacy of physics-informed models, compared with traditional data-driven or physics-based models, in improving control robustness is unclear.

To address these knowledge gaps, we designed a series of prediction tests using a real-world dataset collected from multi-zone classrooms, which included diversified operational patterns and representative corner cases resulting from OCC experiments. Four data-driven predictive models with an increasing level of physics integration were developed and inspected. Special attention was paid to the potential decision-making process given the predictions. To the best of our knowledge, this is the first study that bridges occupant-centric control and physics-informed machine learning in buildings.

Experimental design

Development of predictive models

For ease of further discussion, we contextualized the experiments in a typical OCC task of minimizing the HVAC energy consumption while maintaining thermal comfort. Accordingly, the predictive models were developed to project the temperature response over a certain period given the initial temperature, heating and cooling sequences, and disturbances throughout the prediction horizon. In this study, we developed four predictive models to account for the variations in physics integration, including a basic LSTM, a Physics-Informed Neural Network (PINN) with an augmented loss function, a Physics-Consistent Neural Network (PCNN) with a modified model structure, and a basic Resistor-Capacitor (RC) model. `PyTorch` and `FilterPy` were respectively used to implement the neural networks and model identification.

- **LSTM** is a special type of recurrent neural network, using input, forget, and output gates to control the flow of information. It is designed to retain relevant past data and overcome the vanishing gradient problem, making it suitable for grasping long-term patterns in time series prediction. LSTM has been widely applied to predict building temperature or thermal loads and therefore is a fair baseline model to represent the state-of-the-art data-driven modeling methods (Xiao and You, 2023).

  Although the model itself is inherently black-box, domain knowledge is involved in model development by acquiring the related features (Ma et al., 2025). To exemplify a typical situation of data availability, the model inputs included the initial room temperature and the trajectory of HVAC power, CO_2 concentration, outdoor temperature, and solar irradiance. The inputs for all rooms were aggregated and fed to the model without explicit classification. Before each open-loop prediction, the model was warmed up for an hour (twelve steps), when the actual room temperature was used to overwrite the previous prediction when predicting the next step. In terms of key hyperparameters, the model had an encoder(32)-LSTM(64)-decoder(32) structure, a loss function of mean squared error (MSE), a batch size of 256, and a starting learning rate of 5e-3. The training was performed by the Adam optimizer for 500 epochs, with a decreasing learning rate and an early stopping mechanism.

- **PINN** incorporates physical principles through soft or hard constraints, some of which require a satisfactory physics-based model and suffer from scalability issues. Hence, we reproduced the soft constraint proposed by Di Natale et al. (2023) that penalized the negative gradients of the predictions over the power inputs and outdoor temperature as shown in Equation 1:

$$\begin{aligned} \mathcal{L}_{PINN} &= \mathcal{L}_{MSE} + \lambda \mathcal{L}_{grad}, \\ \mathcal{L}_{grad} &= \frac{1}{l} \sum_{k=0}^{l-1} \left[\frac{1}{m} \sum_{z=1}^{m} g_k^z \right], \\ g_k^z &= \sum_{y=1}^{m} ReLU(-\frac{\partial \hat{T}_l^z}{\partial u_k^y}) + ReLU(-\frac{\partial \hat{T}_l^z}{\partial T_k^{out}}) \end{aligned} \tag{1}$$

where l is the prediction horizon, m is the number of zones, k is the time step, z and y are the zone code, and $ReLU$ is the activation function that clips off the positive gradients. Thereby, $\hat{T}_l^z$ denotes the final predicted temperature of the z_{th} room, and u_k^y denotes the heating or cooling power sent to the y_{th} room at the k_{th} step. $\mathcal{L}_{grad}$, with a weighting factor λ, is added to MSE to form the PINN loss function. λ was set as 100 based on preliminary experiments, and the other prediction and training configurations were kept the same as the previous LSTM.

- **PCNN** is not a broadly recognized terminology. Compared with PINN, researchers modified the model structure to ensure that model outputs follow the physical principles. A popular technique is to process different inputs with parallel modules (Di Natale et al., 2022; Xiao and You, 2023; Jiang and Dong, 2024). However, such complex model architectures often challenge the training process and are not suitable for control. Consequently, we adopted a PCNN that relies on the modification of a single multi-layer perception (MLP) (Wang and Dong, 2023).

  Specifically, the MLP was designed to ensure two predictive properties by grouping the neurons and masking some connections: 1) thermal power and other disturbances at a certain time step only had an impact on the temperature responses in the future (not the past); 2) the impact on the next k_{th} time step was stably characterized by the corresponding neurons. Complementarily, the weights were constrained to be non-negative, and the loss function was augmented to promote the impact on nearer time steps and penalize large bias values. The model again took the same inputs to predict the temperature response. Without the receding mechanism of LSTM-based models, the inputs over the prediction horizon had to be concatenated to suit PCNN, and the hidden dimensions were accordingly enlarged.

- **RC** model is a well-established method for building thermal modeling. Yet, there is no standardized modeling framework that applies to all buildings. A lot of customization is needed in the procedure of data collection, model construction, and parameter estimation. To serve as a proper physics-based baseline model for comparison, we used a connected 1R1C model that has been shown effective in control experiments for the case study building (Green et al., 2024).

  While the input and output variables were the same as the above models, small pieces of truncated time series were designed for model identification. First, an unscented Kalman Filter was applied to estimate the RC parameters using seventeen nights of free-floating temperature. Subsequently, based on the fixed thermal properties, the estimators for HVAC power, internal disturbances, and solar heat gain were separately trained using three days of data when the other two were negligible. The obtained model was soldered into a `LTSpice` circuit simulator for further analysis.

Case study and testing scenarios

The dataset used in this study was collected from six side-by-side classrooms in a university building, spanning two consecutive winters. The classrooms are south-facing with large exterior windows. Each room is equipped with a fin tube radiator (FTR), an active chilled beam, and a VAV terminal with a reheat coil. The classrooms were operated by a fixed schedule from 7 am to 9 pm (11 pm for some rooms) in the first winter, with a pair of customizable heating and cooling setpoints during operating hours, and setbacks of 15.6°C (60°F) and 26.7°C (80°F) during unoccupied hours.

In the second winter, a schedule-based OCC was implemented with an expanded deadband. As illustrated in Figure 1, the room was switched to standby mode at 7 am, with the setpoints of 18.3°C (65°F) and 25.6°C (78°F); it then transitioned into occupied mode half an hour before the first class, with the setpoints of 20.6°C (69°F) and 23.9°C (75°F); after the classes ended, it stayed in standby mode till 9 pm and then switched back to unoccupied. As the room temperatures did not exceed the cooling setpoints, the chilled beam was rarely activated during the heating seasons, and the FTR and reheat coil valves were actuated by the same PI control loop to maintain the heating setpoints.

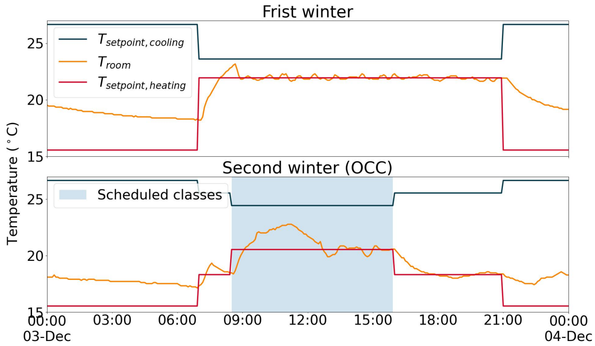

Figure 1: Example of daily room temperature and setpoint profiles in the two winters.

Data points directly used for the predictive models include room temperature, CO_2 concentration, outdoor temperature, solar irradiance, and FTR (or reheat coil) valve positions that indicated heating power. Additionally, room temperature setpoints, supply airflow rate, and supply air temperature were also monitored. The class schedules were downloaded from the university registrar for the OCC experiment, and

https://doi.org/10.26868/25222708.2025.1808

the number of occupants in each room was measured by door counters for reference. Historical data was recorded at 5-minute intervals.

Model training and the first test case followed the standard setup in machine learning studies. The same training data from December 2023 to January 2024 was collated for the neural networks, covering a heavily occupied period in early December, an unoccupied holiday period, and a relatively lightly occupied period in January. The prediction horizon was set as 24 hours (288 steps) to embed more information about the thermal dynamics. After training, data from February 2024 was used to examine the models' open-loop predictive performance. Since the focus was the predictive capability of thermal models, perfect information about the disturbances was assumed in all tests.

The second testing scenario took data from one month of OCC experiments in the second winter, serving as a more demanding extrapolation test. With the class schedule and expanded deadbands, the room temperatures were significantly lower not only in the daytime but also in the nighttime with less heat stored by the thermal mass. Figure 2 compares the room temperatures of the two winters, where both the mean and the variance changed considerably for each room. While the data-generating experiments took a rule-based approach, the control actions could be further refined based on predictive models to optimize the performance or serve more complex objectives. In such cases, the action and state spaces will be expanded to exploit the potential, and the models are expected to be capable of extrapolating the learned dynamics to the new region.

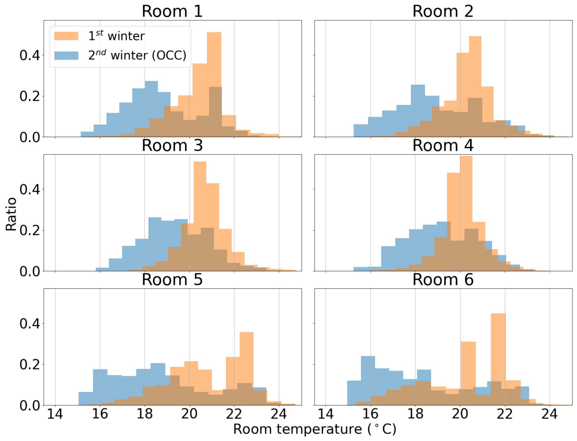

Figure 2: Distributions of room temperature of the six classrooms over the two winters.

As an example of corner cases in classrooms, the internal load may spike with a large cohort of students, which could result in overheating if proper counteractions are not applied. Figure 3 instances one situation in classroom 6 that normally has less than 50 people. This day during the first testing period, more than 90 people came in, boosting the room temperature even though the cooling was activated. Consequently, room temperature exceeded the cooling setpoint and caused discomfort, let alone the waste of cooling energy when the outdoor temperature was around 0°C.

As corner cases are defined by rare and unseen scenarios, it is difficult to gather a repository of test cases for robust quantitative comparison. Instead, we investigated the predictive behavior of alternative models on this day in the OCC context. Assuming the upcoming crowd is known to the supervisory controller, the heating setpoint could be lowered before people come in to allow more capacity to absorb the heat, and the airflow rate of cooled air could be increased earlier, if necessary, to avoid overheating. Thereby, room temperature could remain within the comfort zone, and both heating and cooling energy could be reduced. The prerequisite of such control sequences is the precise prediction of the temperature response across the prediction horizon.

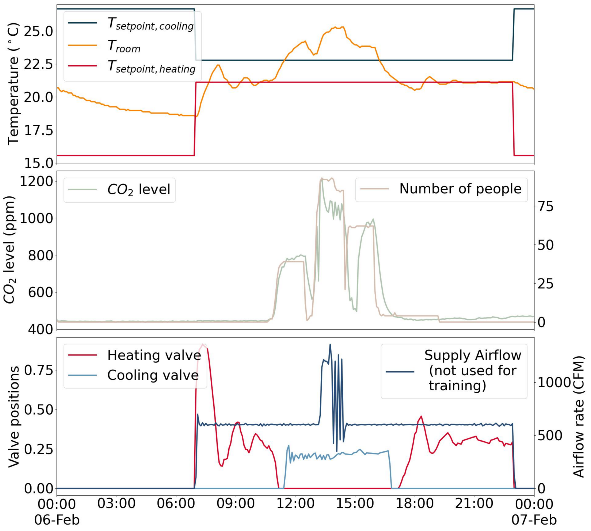

Figure 3: Temperature, internal disturbances, and actuator profiles of room 6 on February 6th 2024.

Results and discussion

Long-term open-loop predictions

The models were first evaluated by the predictive errors over a 24-hour horizon. For each dataset, the models were initialized at midnight each day (after the warm start period when applicable) and then moved forward for 24 hours (288 steps) without correcting the temperatures. Given the predicted trajectory, mean absolute errors (MAE) were calculated for each day, the mean of which during the testing period gave the MAE of each room. The MAEs of each room were then averaged to obtain the final MAE for each model. Table 1 summarizes the average MAEs for the four models during the training and the two testing periods, as well as the standard deviation (*std*) among the six rooms.

Table 1: Predictive errors (MAE) of the training and first two testing cases.

Model	Mean absolute errors (°C)								
	Training[a]			**1st case (standard)**			**2nd case (extrapolation)**		
	by room	*std*	**mean**	by room	*std*	**mean**	by room	*std*	**mean**
LSTM	0.6, 0.65, 0.77, 0.88, 1.29, 0.79	*0.23*	**0.83**	0.69, 0.7, 0.82, 1.11, 1.25, 0.81	*0.21*	**0.89**	1.04, 0.99, 1.17, 0.88, 2.18, 1	*0.44*	**1.21**
PINN	0.54, 0.82, 0.62, 0.51, 1.44, 1.02	*0.33*	**0.82**	0.66, 0.93, 0.7, 0.76, 1.36, 0.95	*0.23*	**0.89**	0.99, 0.88, 0.95, 0.58, 2.19, 1.13	*0.51*	**1.12**
PCNN	0.7, 0.57, 0.66, 0.56, 1.08, 1.59	*0.37*	**0.86**	0.76, 0.65, 0.74, 0.74, 1.16, 1.22	*0.22*	**0.88**	0.69, 0.92, 0.6, 0.64, 1.38, 1.01	*0.27*	**0.87**
RC	1.46, 1.41, 1.19, 1.16, 1.59, 1.73	*0.2*	**1.42**[b]	1.34, 1.53, 1.31, 1.14, 1.48, 1.59	*0.15*	**1.39**	1.78, 1.97, 1.9, 1.62, 1.6, 1.82	*0.13*	**1.78**

[a] This test is different from the training losses, as the 24-hour sequences only started at midnight in testing but were allowed to start from any hour throughout a day in training.

[b] The RC model was trained only with snippets of data during the training period. For fairer comparison, this value was obtained in the same way as other models, by testing the model with the entire training dataset.

The MAEs of the first testing period are comparable to those of the training period for all models, indicating that the models were properly trained and not overfitted to training data. Despite the fluctuation in some rooms, the three neural network models have a close mean of MAEs of the six rooms, indicating their similar capability of fitting the data. In contrast, the RC model has a larger MAE for almost every room and significantly larger means. This is expected considering the simplified physics structure in the RC model and the much fewer parameters to fit the data. Meanwhile, it is worth noting that this RC model has been shown effective for these classrooms in actual control experiments. Given the highly uncertain disturbances in classrooms and the complex multi-zone thermal dynamics, MAEs of all the models are acceptable.

Figure 4 displays the typical predictive behavior of the four models on a day during the first testing period. It can be seen that models based on neural networks were generally better at fitting the profiles, especially in the daytime when heating was on. Noticeably, LSTM and PINN, without the relationships between inputs and outputs explicitly specified, better captured the sudden change of room temperatures at the start of operating hours. However, the flexibility came with a price. The black-box models sometimes gave physically wrong predictions, such as for room 4 in the lower half of Figure 4. The outdoor temperature was lower than 0°C before sunrise, and the room temperature should slowly drop until being heated up by the radiator. On the contrary, LSTM overestimated the impact of the slightly increased CO_2 level (one person was there until 3 am), presuming the temperature to increase and then be "cooled" down by the radiator. PINN alleviated this issue with the augmented loss function, but the prediction was still off. PCNN further constrained the training and yielded physically meaningful predictions. Yet, the reduced flexibility also caused the prediction to be less accurate in the daytime. Synthesizing these factors, better physics consistency usually led to slightly lower 24-hour predictive errors. For example, the MAEs of room 4 on this day for LSTM, PINN, and PCNN were respectively 0.49°C, 0.35°C, and 0.3°C.

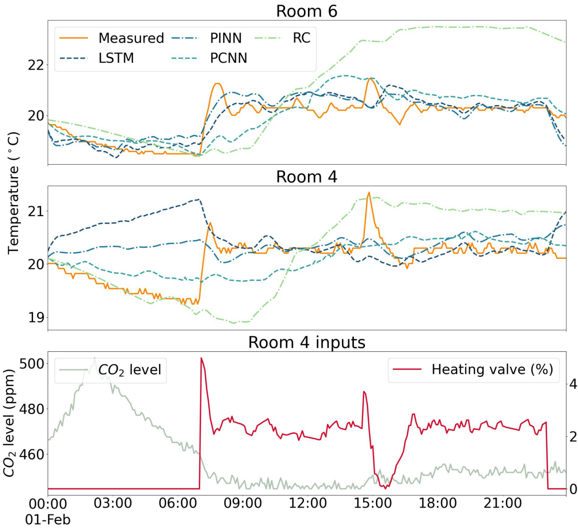

Figure 4: Measured room temperature and predictions of the models for two rooms on February 1st 2024, and the two inputs of room 4 for reference.

The plot also alludes to the lower expressiveness of the RC model. Without the damping capacitors to distinguish the rooms' heat exchange with the radiator, outdoor, and neighboring zones, the training process optimized the long-term predictive error by reducing the impact of disturbances and producing much smoother profiles. Consequently, the RC model was the most robust among the four models when the temperatures were free-floating (the first six hours in Figure 4), but was the least responsive to disturbances and HVAC inputs. Although the model could potentially be further improved by a more complex RC structure (Zhan et al., 2022), it is out of the scope of this study.

Extrapolating to the new operations

The second testing scenario evaluated the models with a significantly different operational pattern and shifted temperature range, quantitatively comparing the models' extrapolation capability. As shown in the last column of Table 1, the average MAE of LSTM was higher than the first standard test case by 36%. PINN, as expected, improved from LSTM but still got a 26% larger MAE. Surprisingly, PCNN maintained the same average MAE as in the first test, while the RC model also observed a 28% increase.

Figure 5 illustrates two dominant causes for accuracy deterioration using two examples on November 27^{th} 2024. The upper half of the figure provides another example of physically wrong predictions of LSTM and PINN resulting from overestimated CO_2 impact. Compared with the previous example in Figure 4, CO_2 caused a larger discrepancy in temperature. Also, the predicted temperature wrongly rose back right after the radiator went off at 9 pm, which manifests more evidently that the radiator was taken by the black-box models as a cooling source.

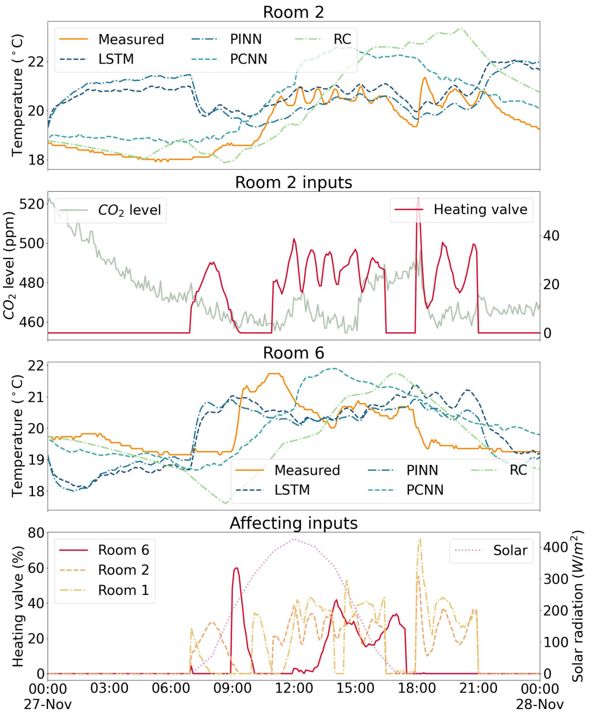

Figure 5: Measured room temperature and predictions of the models for two rooms on November 27^{th} 2024, and the related inputs for reference.

The data of room 6 in the lower half underscores another physical error of overestimated inter-zone impact. On this day, the radiator of room 6 was only activated at 9 am when the room went into occupied mode. However, LSTM and PINN predicted the room temperature to rise much earlier at 7 am, when room 1 and room 2 were being heated to the standby setpoint (room 6 was already above that setpoint). By the end of the day, the actual temperature dropped to the standby after the last class in room 6, but the prediction was still affected by radiators in the other rooms until 9 pm. We verified this finding by testing the models with modified inputs. Disabling the radiators in those two rooms eliminated this issue. In contrast, although also starting to rise at around 7 am, solar radiation was found to have minimal impact on the prediction. Additionally, the black-box models also occasionally produced wrong predictions that cannot be explained by the above analysis, such as the start-of-day temperature drop in room 6.

Overall, it turns out that the change in temperature range was NOT the main reason for the decreased predictive performance. Since the models were designed to learn the sequential nature of time-series data, the predicted trajectories were in the right range given the initial temperature. On the other hand, it is the new operating patterns that challenged the models. Due to the schedule-based OCC, each classroom was operated differently on each weekday, making the diurnal patterns more diversified and less correlated, negating the relationships learned by the neural networks. Recall that the RC model tried to minimize the predictive error by producing smooth profiles, with the shape of predicted profiles staying about the same as plotted in Figure 5, it became less effective for the more irregular actual data.

In addition to the mean of MAEs, the standard deviation *std* also changed in the extrapolation test. For example, PCNN had almost the same mean of MAEs in the two testing scenarios, but the *std* of the second period was much larger. Similarly, the *std* of LSTM and PINN also significantly increased, with the model performance worse in most rooms but slightly better in some. There was no sufficient evidence to consolidate the causal analysis, but some clues implied the contribution of internal disturbances. During the second testing period, in room 5, where both LSTM and PINN obtained the largest MAE among all tests, the *std* of CO_2 level was also the highest. Vice versa, the *std* of CO_2 decreased in room 4, and so did the MAE. Conversely, the performance of PCNN was not as sensitive to the variability of inputs, which may be another advantage of physics consistency.

Handling the corner case

Regarding the corner case identified in Figure 3, the four models were first compared using a 24-hour prediction horizon. As the first plot in Figure 6 displays, none of the models were near the ground truth when the temperature spiked. After all, capturing the corner case in a 24-hour ahead temperature prediction could be an almost impossible task with the available measured data. Building operations are subject to many uncertain disturbances, most unmonitored. For public spaces like the classrooms, occupants could participate in the room's heat balance in various ways, including but not limited to entering or leaving the room, using electrical devices such as laptops and projectors, and adjusting blinds. Besides, the heating and cooling effect coming from the ventilation could also be significant but unquantified. Missing some of these factors created discrepancies at each time step, accumulating to the big error when the corner case happened.

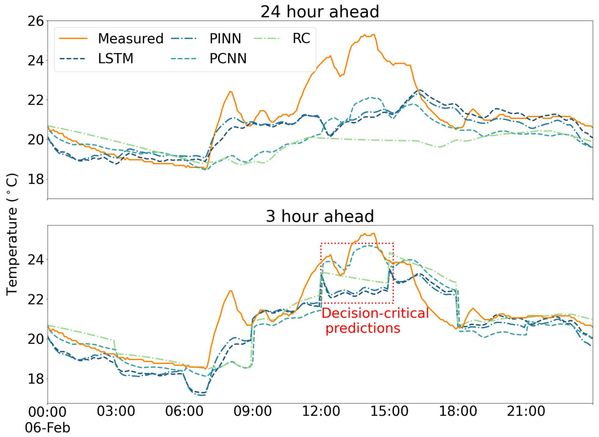

Figure 6: 24-hour and 3-hour ahead predictions on February 6th 2024.

Noticeably, despite the much lower absolute value, PCNN was the only model that responded to the input of CO_2 level (middle plot of Figure 3) with a temperature rise in the trend. Ideally, the models are expected to replicate the actual building, but being able to partially interpret the key inputs could be just enough to support the control decisions. Depending on the control objectives, accurate prediction is required at different horizons. Meanwhile, the room temperature can be measured at each time step to initialize the next optimization. For the OCC task of minimizing the HVAC energy while maintaining thermal comfort, a much shorter prediction horizon could be useful to obtain near-optimal solutions.

The second plot in Figure 6 compares the 3-hour ahead predictions, where the room temperature was overwritten by the measured value every three hours. Empirically, this is the time window needed to take precautions against overheating in this case. At a glance, all four models captured the trend more accurately than the 24-hour-ahead predictions. However, zooming into the mid-day period of the corner case, physics-inconsistent models predicted the response in the wrong direction that could mislead the control decisions. As highlighted in the red dotted rectangle, before the temperature was about to drastically increase, LSTM and PINN, even the RC model, presumed the temperature to decrease. Correspondingly, the controller would have stayed idle until it became too warm hours later. Only PCNN expected the temperature to rise in view of the upcoming crowd, which would activate the cooling in time.

Implications and future directions

To summarize, the above experiments showed that the capability of the predictive models to extrapolate into unseen input scenarios is critical for the robustness of model-based optimal control. Moreover, the traditional approach to evaluating machine learning models could disguise the potential risk. Compared with a low predictive error, either short-term or long-term open-loop, correctly reacting to the disturbances could be more beneficial for the decision-making in controls. Besides, it is useful to develop models for an appropriate horizon. While a longer-term predictive error may be a stricter criterion for physics representation, it could fail to distinguish the fit-for-purpose model due to the lack of input information.

Integrating physics altered the learning process. Nevertheless, penalizing the unphysical predictions cannot guarantee physics consistency, as it could be overwhelmed by the goal of fitting the data. In the experiments, although PINN achieved a lower MAE in the extrapolation tests, it essentially processed the inputs in the same way as LSTM. As critical as it is to explicitly embed the physical relationships in the model structure, too strong constraints could restrict the models' ability to fit training data, often referred to as model expressiveness for data-driven models. For example, without sufficient degrees of freedom, the RC model could not characterize the dynamics and overlooked the disturbances. A systematic approach is needed to balance the expressiveness and physical constraints in model development.

The experimental results clearly showed the merits of proper physics integration in PCNN, whereas the corresponding advantages in OCC have not been quantitatively shown. Hence, measuring the models' effectiveness in a control framework, particularly in improving robustness, is one key direction for further development. More broadly, the definition and solutions for corner cases in optimal building control are critical but largely untapped research topics. Future research in the field should go beyond reporting the overall performance metrics and pay attention to the associated risks. Realistic control robustness can only be investigated in long-term real-world experiments.

Conclusion

This paper investigated the predictive performance of four data-driven thermal models with an increasing level of physics integration. Dedicated extrapolation tests were defined using measured data from real-world occupant-centric control experiments in six connected classrooms. The results reveal the benefits of integrating physics knowledge when developing control-oriented models, highlighting the importance of a physics-consistent model structure that still preserves sufficient expressiveness. Such an extrapolatable modeling framework provides the foundation for scalable occupant-centric control in buildings.

Acknowledgment

The authors acknowledge the technical and financial support of the MIT Department of Facilities and Schneider Electric. We also thank Dr. You Lin for assistance with data collection.

References

Bogdoll, D., J. Breitenstein, F. Heidecker, M. Bieshaar, B. Sick, T. Fingscheidt, and M. Zöllner (2021). Description of corner cases in automated driving: Goals and challenges. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1023–1028.

Di Natale, L., B. Svetozarevic, P. Heer, and C. N. Jones (2022). Physically consistent neural networks for building thermal modeling: theory and analysis. *Applied Energy 325*, 119806.

Di Natale, L., B. Svetozarevic, P. Heer, and C. N. Jones (2023). Towards scalable physically consistent neural networks: An application to data-driven multi-zone thermal building models. *Applied Energy 340*, 121071.

Drgoňa, J., A. R. Tuor, V. Chandan, and D. L. Vrabie (2021). Physics-constrained deep learning of multi-zone building thermal dynamics. *Energy and Buildings 243*, 110992.

Gokhale, G., B. Claessens, and C. Develder (2022). Physics informed neural networks for control oriented thermal modeling of buildings. *Applied Energy 314*, 118852.

Green, D. H., Y. Lin, A. Botterud, J. Gregory, S. B. Leeb, and L. K. Norford (2024). Pareto-optimized thermal control of multi-zone buildings using limited sensor measurements. *IEEE Transactions on Smart Grid 15*(5), 4674–4689.

Henze, G. P., K. J. Kircher, and J. E. Braun (2024). Why has advanced commercial hvac control not yet achieved its promise? *Journal of Building Performance Simulation 18*(2), 217–228.

Hiyama, K. and Y. Omodaka (2021). Operation of climate-adaptive building shells utilizing machine learning under sparse data conditions. *Journal of Building Engineering 43*, 103027.

Hobson, B. W., B. Huchuk, H. B. Gunay, and W. O'Brien (2021). A workflow for evaluating occupant-centric controls using building simulation. *Journal of Building Performance Simulation 14*(6), 730–748.

Jiang, Z. and B. Dong (2024). Modularized neural network incorporating physical priors for future building energy modeling. *Patterns 5*(8), 1–15.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang (2021). Physics-informed machine learning. *Nature Reviews Physics 3*(6), 422–440.

Liang, X., X. Zhu, S. Chen, X. Jin, F. Xiao, and Z. Du (2023). Physics-constrained cooperative learning-based reference models for smart management of chillers considering extrapolation scenarios. *Applied Energy 349*, 121642.

Ma, Z., G. Jiang, Y. Hu, and J. Chen (2025). A review of physics-informed machine learning for building energy modeling. *Applied Energy 381*, 125169.

Nagy, Z., B. Gunay, C. Miller, J. Hahn, M. M. Ouf, S. Lee, B. W. Hobson, T. Abuimara, K. Bandurski, M. André, et al. (2023). Ten questions concerning occupant-centric control and operations. *Building and Environment 242*, 110518.

O'Brien, W., A. Wagner, M. Schweiker, A. Mahdavi, J. Day, M. B. Kjærgaard, S. Carlucci, B. Dong, F. Tahmasebi, D. Yan, et al. (2020). Introducing iea ebc annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation. *Building and Environment 178*, 106738.

Pang, Z., Y. Chen, J. Zhang, Z. O'Neill, H. Cheng, and B. Dong (2020). Nationwide hvac energy-saving potential quantification for office buildings with occupant-centric controls in various climates. *Applied Energy 279*, 115727.

Park, J. Y., M. M. Ouf, B. Gunay, Y. Peng, W. O'Brien, M. B. Kjærgaard, and Z. Nagy (2019). A critical review of field implementations of occupant-centric building controls. *Building and Environment 165*, 106351.

Wang, X. and B. Dong (2023). Physics-informed hierarchical data-driven predictive control for building hvac systems to achieve energy and health nexus. *Energy and Buildings 291*, 113088.

Willcox, K. E., O. Ghattas, and P. Heimbach (2021). The imperative of physics-based modeling and inverse theory in computational science. *Nature Computational Science 1*(3), 166–168.

Xiao, T. and F. You (2023). Building thermal modeling and model predictive control with physically consistent deep learning for decarbonization and energy optimization. *Applied Energy 342*, 121165.

Zhan, S., Y. Lei, Y. Jin, D. Yan, and A. Chong (2022). Impact of occupant related data on identification and model predictive control for buildings. *Applied Energy 323*, 119580.

Zhang, L. (2021). Data-driven building energy modeling with feature selection and active learning for data predictive control. *Energy and Buildings 252*, 111436.