A scalable and transferable approach for spatio-temporal indoor air temperature forecasting in multi-zone buildings exploiting knowledge graph contextual information

Rocco Giudice [*1]

Abstract

This paper presents a scalable and transferable approach for spatio-temporal indoor air temperature forecasting in multi-zone buildings, utilizing contextual information from a knowledge graph. The method extracts static and dynamic embeddings for devices and zones from the knowledge graph and timeseries data, enabling flexibility across diverse building configurations. By leveraging these embeddings, the model can effectively handle varying HVAC setups and predict temperature evolution while exploiting only half of the training dataset. The approach achieves a mean absolute error (MAE) of 0.4 for a 24-hour prediction horizon over a 6-month test period, demonstrating comparable performance to state-of-the-art methods. Its main advantage lies in its scalability and transferability, as the use of knowledge graph embeddings allows the model to capture contextual information while learning shared thermal patterns.

1. Introduction

Accurate forecasting of indoor air temperature in multi-zone buildings is crucial for optimizing energy efficiency, enhancing occupant comfort, and enabling effective building management, especially when advanced control systems, such as deep reinforcement learning (DRL)-based control, need to be deployed (Silvestri et al., 2024). Traditional methods often rely on physics-based models, which, while typically precise even without calibration, are computationally intensive and challenging to scale across diverse building configurations. In contrast, data-driven approaches, particularly deep learning models, have emerged as promising alternatives, offering adaptability and scalability in dynamic environments. A well-trained, transferable deep learning model can, in fact, theoretically be used to pre-train DRL agents that can be directly deployed in the field (Coraci et al., 2023). This approach eliminates the need to wait for the collection of sufficient data (typically a full year of data with different temperature setpoint configurations is required for thermal dynamics) and avoids the challenges associated with training from scratch, such as data collection, integration, standardization, etc.

1.1. Related works

Spatio-temporal indoor air temperature forecasting for multi-zone buildings has been primarily addressed in the literature using graph neural networks (GNNs) (Wang et al., 2024; Wen et al., 2025), with some studies employing Physics-Informed Neural Networks (PINNs) to enhance scalability. For instance, (Zhang et al., 2023) used a GCN-RNN framework to predict indoor air temperature in an AHU-VAV system within a multi-zone building. A similar approach was employed by (Piscitelli et al., 2025), who analyzed various methods for graph development, comparing physical-oriented and data-driven approaches. Additionally, (Lee & Cho, 2025) incorporated a physics-informed loss function into a Spatio-Temporal Graph Neural Network (STGNN) to improve the scalability of the learning process. However, a limitation of traditional GNNs is their static nature: once the graph structure is defined, it cannot be easily modified without retraining the model. Furthermore, GNNs fit a unique model for the entire multi-zone building. meanings that, often, GNN are not transferable models. This poses challenges in real-world applications where building configurations may change over time. The dynamic nature of building systems and sensor configurations presents a key hurdle that has yet to be fully addressed by existing GNN-based methods. On the other hand, physic-informed deep learning may represent the turning point for scaling and transferring thermal dynamics models. One of the most notable examples of a physics-informed neural network comes from (Jiang & Dong, 2024), who developed ModNN. In their approach, different neural network modules were designed to reflect various heat exchange phenomena, with

[*]Equal contribution [1]Department of Energy, Politecnico di Torino, Torino, Italy. Correspondence to: Rocco Giudice <rocco.giudice@polito.it>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

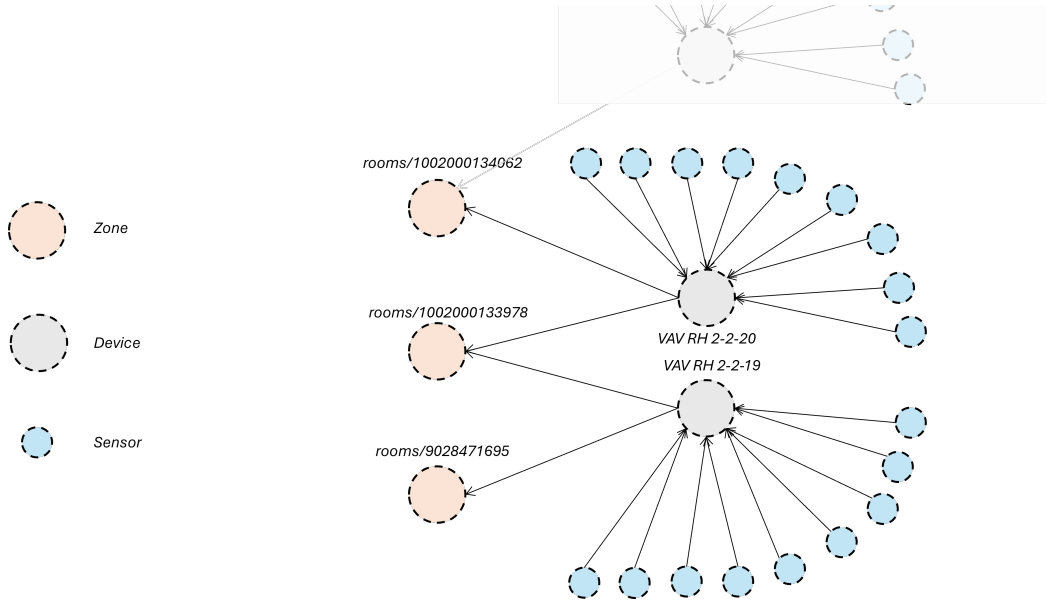


Figure 1: Example of a piece of the knowledge graph developed in this work.

physics integrated both in the problem structure and in the heat and mass balance equations used during training.

1.2. Motivation and novelty of the proposed approach

This paper addresses the prediction of thermal dynamics in multi-zone buildings by tackling two fundamental challenges: transferability and scalability.

Transferability is a common limitation in current forecasting models, particularly those based on Graph Neural Networks (GNNs). These models are typically developed for specific buildings using a fixed graph structure. As a result, transferring them to different zones or buildings requires extensive retraining or architectural redesign. Even when similar HVAC systems or thermal behaviors exist, the lack of standardized and portable representations prevents knowledge reuse. While modular physics-informed models like ModNN have attempted to address this through structured learning, data-driven approaches have not yet achieved similar generalization capabilities.

Scalability concerns arise from the variability in sensor availability and configuration. Real buildings often feature a heterogeneous mix of sensors with different types, quantities, and placements. Conventional deep learning methods assume fixed-size inputs, making them difficult to scale across zones and buildings with differing sensor infrastructures. In practice, this leads to models that are brittle or require heavy data preprocessing.

To overcome these limitations, this work introduces a novel data-driven framework with two key innovations:

- **Transferability through Knowledge Graph Embeddings**: A knowledge graph encodes the building's contextual structure, including zones, HVAC components, and sensors. By applying node2vec, we learn high-dimensional embeddings for zones, devices, and sensor types. These embeddings act as standardized feature vectors that capture the role and identity of each element within the system. The model learns to associate specific thermal dynamics patterns with these embeddings, allowing it to generalize across new zones or buildings without retraining.

- **Scalability through Permutation-Invariant Modeling**: The architecture incorporates DeepSet networks to aggregate sensor and device data, making it inherently flexible to input permutations and varying sensor counts. Masking and padding techniques are used to handle inactive or missing sensors, ensuring that the model remains robust and data-efficient even in sparse sensing environments.

Together, these contributions allow the proposed approach to learn a unified spatio-temporal model of indoor temperature dynamics that is both scalable across different sensor setups and transferable to new buildings or zones with minimal adaptation. In this way, this approach is oriented in the direction of developing a 'foundation model' for building thermal dynamics.

2. Case study

The dataset used is the *Google Open Sources Smart Buildings Dataset* presented in (Goldfeder et al., 2025). The training dataset includes data from the first half of 2022 and the second half of 2023. This allows the model to observe data from different seasons, and consequently, different thermal dynamics and internal setpoints, despite the data coming from different years. The training dataset is split into 80% for training and 20% for validation. The test is conducted on the second half of 2022.

The dataset is aggregated to 15-minute intervals, rather than the default 5 minutes. This choice is made because, for advanced controllers operating at the supervisory level (e.g., adjusting temperature setpoints rather than individual control signals for valves and dampers), a 15-minute granularity is sufficient. This results in a lower noise level. For more specific controllers, a smaller timestep could be used, but it would significantly increase complexity. Finally, data have been pre-processed to avoid physical inconsistencies, filtering out the measurements that were incompatible with physics.

3. Methodology

The model developed in this work integrates contextual information embedded in a knowledge graph. This information, extracted from the KG, is defined in this study as static embeddings, i.e., numerical vectors representing three kinds of information: the zone, the device type (e.g., VAV type), and the sensor type. These three kinds of information are used in two deep learning models, which serve, respectively, to extract dynamic embeddings for the devices and to predict the temperature of each individual zone. It is important to clarify that these deep learning models, together with the static embeddings, are the only trained models. In other words, to predict the temperature of 500+ zones, we use a single architecture composed of two deep learning models and a node2vec model.

In the following sections, the static knowledge graph embeddings extraction and the two deep learning models are elaborated in more details, as well as some details of the training configuration.

3.1. Static KG embeddings extraction

The first stage involves the creation of a Knowledge Graph, representing entities such as building zones, HVAC devices (e.g. VAV boxes), and sensor types, along with their interrelations. Using node2vec, embeddings are generated for each node within this graph, producing standardized, high-dimensional vectors that encapsulate the contextual characteristics of each entity. Specifically, three categories of embeddings are extracted:

- Zone embeddings representing the unique characteristics and metadata of building zones.
- Device embeddings encapsulating the static properties of HVAC components, such as device types and sensor configurations.
- Sensor type embeddings.

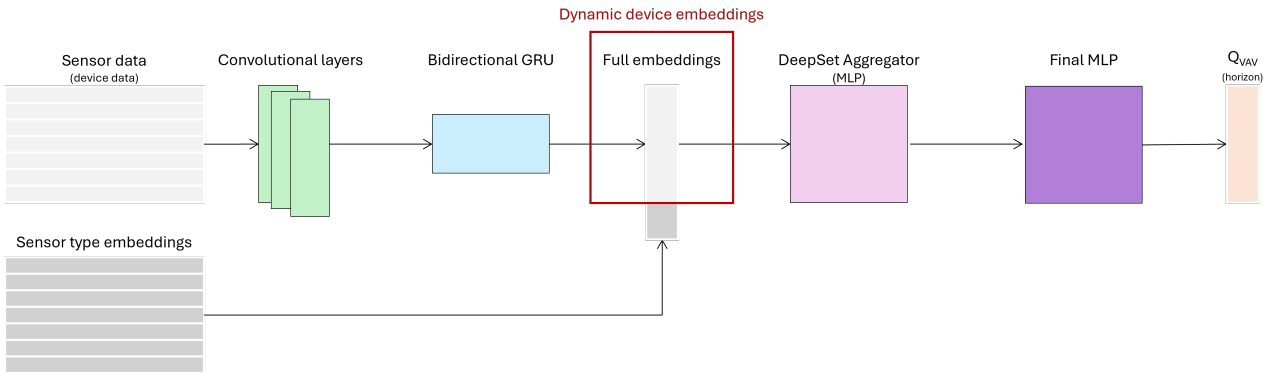
An example of a part of the knowledge graph created is shown in Figure 1. Naturally, if two zones are served by the same VAV and are on the same floor, they will have similar embedding vectors; otherwise, their embeddings will be quite different.

3.2. Dynamic device embeddings learning

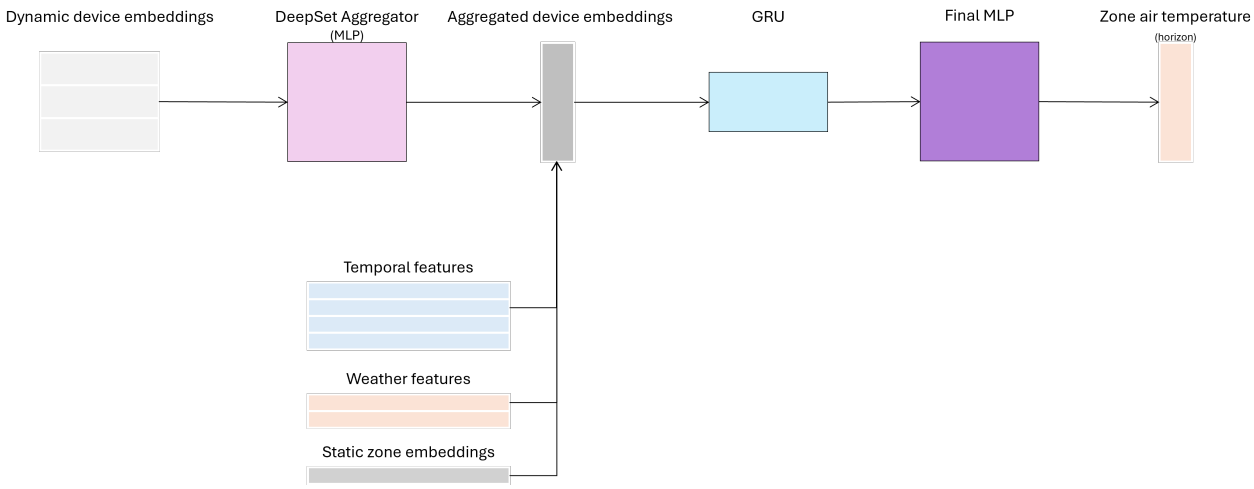
The second stage focuses on dynamically capturing the temporal and spatial behaviors of devices. The model architecture for learning dynamic device embeddings integrates both spatial and temporal features to represent devices in a multi-zone building and capture a signature of each device in a dynamic manner. The architecture is shown in Figure 2a.

The first step consists of an encoder architecture (i.e. Sensor Encoder), composed of convolutional layers and bidirectional GRUs, with the aim to extract both spatial and temporal relationships among all the measurements of a device (i.e., a VAV box). The convolutional layers operate over the sensor features, allowing the network to learn spatial patterns by applying convolution across features (sensor measurements). After extracting spatial features, the data is passed through a recurrent network, which models the temporal dependencies across sensor readings, thus capturing the evolution of sensor behavior over time. The final output from this process is a dynamic embedding for each sensor, which combines the learned spatial and temporal features. Additionally, static sensor type embeddings are integrated with the dynamic sensor embeddings to form a complete representation of each sensor's characteristics, incorporating both time-varying data and fixed sensor properties.

To handle devices with varying numbers of sensors, a permutation-invariant aggregation mechanism is employed, based on the DeepSet architecture. DeepSet networks are specifically designed to handle input sets where the order of elements does not affect the output, making them ideal for aggregating sensor data from devices with different numbers of sensors (Kimura et al., 2024). This mechanism aggregates the individual sensor embeddings into a single device-level embedding, ensuring that the model can process devices with different sensor configurations. The aggregation is performed by summing the sensor embeddings for each device, applying a mask to ignore inactive or padded sensors, and then normalizing the sum by the number of active



(a) Architecture for dynamic device embeddings task.



(b) Architecture of zone air temperature forecasting task.

Figure 2: Model architectures for dynamic device embeddings learning and zone air temperature forecasting.

sensors. This process produces a fixed-size embedding for each device, regardless of the number of sensors associated with it. The aggregated sensor embeddings are then passed through a fully connected network, which further refines the device-level embedding. This network maps the aggregated sensor information into a representation that captures the overall behavior and characteristics of the device for the given horizon.

The device-level embeddings, which encapsulate both dynamic sensor information and static device attributes, are concatenated with static device embeddings to form a comprehensive representation of the device. This combined embedding is then passed through a series of fully connected layers to predict the discharge air temperature injected into the environment by the device.

3.3. Zone air temperature forecasting model

The model for zone temperature prediction aims to forecast the temperature evolution of a certain zone by integrating the dynamic device embeddings related to that zone, static zone embeddings (see Section 3.1), historical zone temperature data, weather features, and temporal features. The architecture of the model consists of two primary components that work together to predict the temperature: a mechanism for aggregating device-level information, and a temporal forecasting network for predicting future temperatures (Figure 2b.

The first component of the model is a DeepSet network responsible for aggregating the embeddings of individual devices within a zone into a unified embedding (Zone Aggregator). Each zone typically consists of multiple devices, and the aggregation mechanism is designed to summarize the information from these devices, provided by their embedding learned as specified in Section 3.2, into a single,

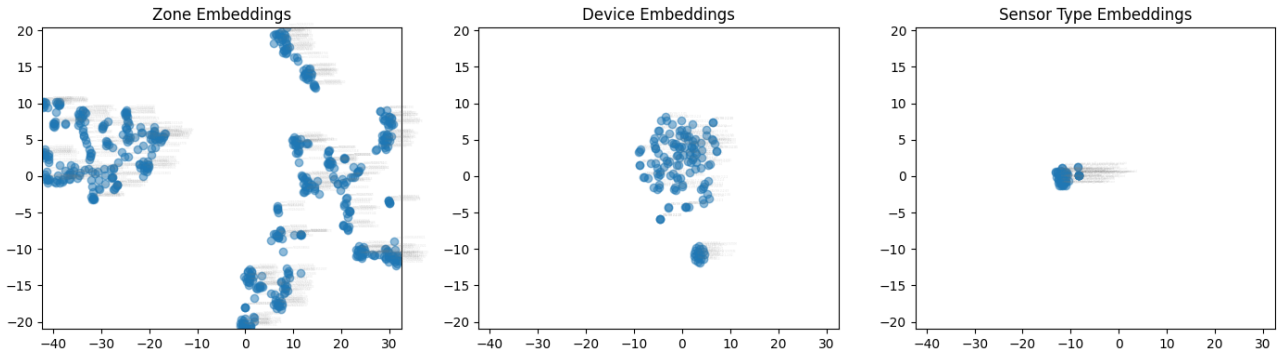


Figure 3: t-SNE plot of node embeddings extracted from the knowledge graph of the case study investigated using *node2vec*.

coherent representation. To ensure flexibility and scalability, the aggregation process is permutation-invariant, meaning it can accommodate varying numbers of devices across different zones. The aggregated device embeddings are passed through a series of layers and a mask is applied to account for inactive or padded devices, ensuring that the aggregation reflects only the number of devices provided. The result of this process is a fixed-size embedding that is used as the base for further forecasting.

The second component of the model focuses on predicting the zone temperature over time, using a combination of the embedding vector just described, the zone embedding extracted from the knowledge graph, temperature history, weather data, and temporal features, such as the day of the week and the hour of the day (we use them as a proxy for occupancy information). The inputs to the forecasting network are concatenated and fed into a GRU network, which is capable of modeling the temporal dependencies inherent in the evolution of zone temperatures. The final hidden state from the GRU, corresponding to the last time step, is used to predict the temperature for multiple future time steps, covering the specified forecast horizon.

3.4. Models architecture and training configuration

This section provides insights into the architectural choices, training setup, and optimization strategy adopted for developing the proposed forecasting model.

First, the static embeddings size is 8 for each node extracted. The low complexity and variability of information contained in the dataset justify such small size.

Regarding networks architecture, the following choices have been made:

- **Sensor Encoder:** Consists of 3 convolutional layers (kernel size 5, padding 2) followed by 2 bidirectional GRU layers.
- **Dynamic Embedding Dimensions:** The GRU output dimension is 128 per direction, resulting in a 256-dimensional dynamic sensor embedding. This is concatenated with static sensor-type embeddings before aggregation.
- **Device Embedding Aggregator:** A DeepSet module processes the set of sensor embeddings per device, applying masking to ignore missing or inactive sensors. The aggregated embedding is passed through a two-layer MLP to produce the final 128-dimensional device representation.
- **Zone Forecasting Module:** The zone model includes the Zone Aggregator to combine multiple device embeddings using a masked and normalized sum. The result is concatenated with zone embeddings and fed, along with temperature history, weather, and temporal features, into a GRU with a hidden size of 256. A two-layer MLP produces the final multi-step temperature prediction.

No extensive hyperparameter tuning was conducted in this study. The architectural choices and embedding sizes were selected based on prior domain knowledge and empirical results from early prototyping.

The model is trained on data from 250 zones out of a total of 500, representing a wide range of configurations and thermal behaviors. The remaining zones are only used for validation and testing. Time series inputs are resampled at 15-minute intervals and segmented using sliding windows with a fixed lookback horizon of `96` time steps (i.e. 1 day).

Training is performed using the Adam optimizer with the following settings:

- Learning rate: `1e-3`
- Batch size: `512`
- Maximum epochs: `100`

To prevent overfitting, early stopping is applied based on validation loss. Training is terminated if no improvement is observed for 5 consecutive epochs. The model state with the lowest validation error is retained for final evaluation.

4. Results

Figure 3 presents a two-dimensional t-SNE plot of the static embeddings extracted from the knowledge graph. As shown in the figure, the zone embeddings are organized into distinct clusters. This organization is useful for identifying zones that share similar device types, and thus exhibit similar thermal dynamics. Furthermore, it can be inferred that zones served by the same device are likely spatially adjacent, and therefore subject to similar thermal loads and gains. In this way the model can map a certain behavior to a specific zone embedding cluster. For device embeddings and sensor types embeddings the difference is smaller, but still sufficient to differ between certain devices or sensor types.

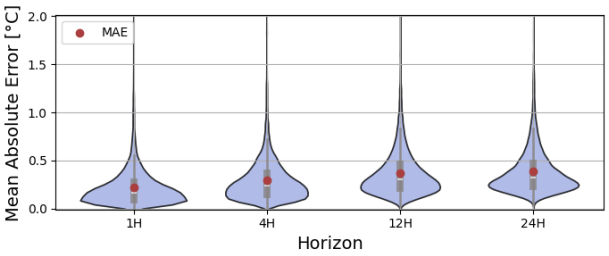


Figure 4: Distribution of MAE for different prediction horizon with evidence of the mean value (red dots).

To evaluate the model's predictive accuracy across various temporal scales, we report the Mean Absolute Error (MAE) for different forecast horizons, ranging from short-term (1 hour) to long-term (24 hours) on the testing set, which comprises six months of data unseen by the model (second semester of 2022). Figure 4 shows the distribution of MAE values as a function of the forecast horizon. As expected, prediction error tends to increase with the length of the prediction window. Nevertheless, the model maintains robust performance, with an average MAE of approximately 0.4 °C at the 24-hour horizon and 0.3 °C at the 1-hour horizon, which is competitive with state-of-the-art results, taking into account that we performed training only on half of the zones.

This result confirms the model's ability to effectively learn both short- and long-term dynamics of indoor air temperature, even under the challenges posed by diverse HVAC configurations and sensor variability across devices. The stability of the error growth trend also indicates the consistency of the learned representations across time.

To better assess the variability in model performance across different zones, we analyze three representative cases on the 24-hour prediciting horizon: the zone with the best predictive accuracy (corresponding to the 10th percentile of the average zone MAE distribution), a zone with average performance (MAE close to the overall mean), and the worst-performing zone (90th percentile of the MAE distribution). Figure 5 illustrates the predicted and actual temperature profiles over the course of two weeks, using a 24-hour forecast horizon with a step size of 24 hours.

As shown in the figure, the predicted patterns generally align well with the ground truth, particularly in capturing the overall seasonal and daily trends. Minor discrepancies are observed, especially around midday peaks, which can be attributed to the model's lack of physical knowledge. Specifically, the model does not incorporate physical variables such as solar irradiance or thermal boundary conditions other than outside air temperature and humidity. Since this information is neither encoded in the knowledge graph nor included in the input features, zones at the building envelope—where irradiance significantly impacts the thermal balance—may exhibit larger prediction errors.

5. Conclusion and future perspectives

The work conducted confirm the effectiveness of the proposed scalable and transferable approach for indoor temperature forecasting by employing knowledge graph embeddings: despite being trained on only half of the available zones, the model achieves robust predictive performance across a variety of configurations, with a Mean Absolute Error of approximately 0.4°C over a 24-hour horizon.

In terms of future work, three main directions will be further explored:

- **Model architecture:** The architecture presented in this work should be regarded as a first step toward a generalizable solution. Future research will involve exploring alternative architectures and conducting more rigorous hyperparameter optimization to identify the most effective configuration. For instance, replacing simple DeepSet aggregators with more expressive mechanisms such as Set Transformers may improve the model's ability to capture complex relationships across sensors and devices.

- **Standardization:** To enhance the model's generalization and transferability, the development of the knowledge graph should follow standardized semantic ontologies, such as Brick (Balaji et al., 2016) or ASHRAE 223P. Embedding contextual information using shared semantic structures would allow the knowledge graph embeddings to be more easily reused and transferred across buildings and applications, facilitating broader

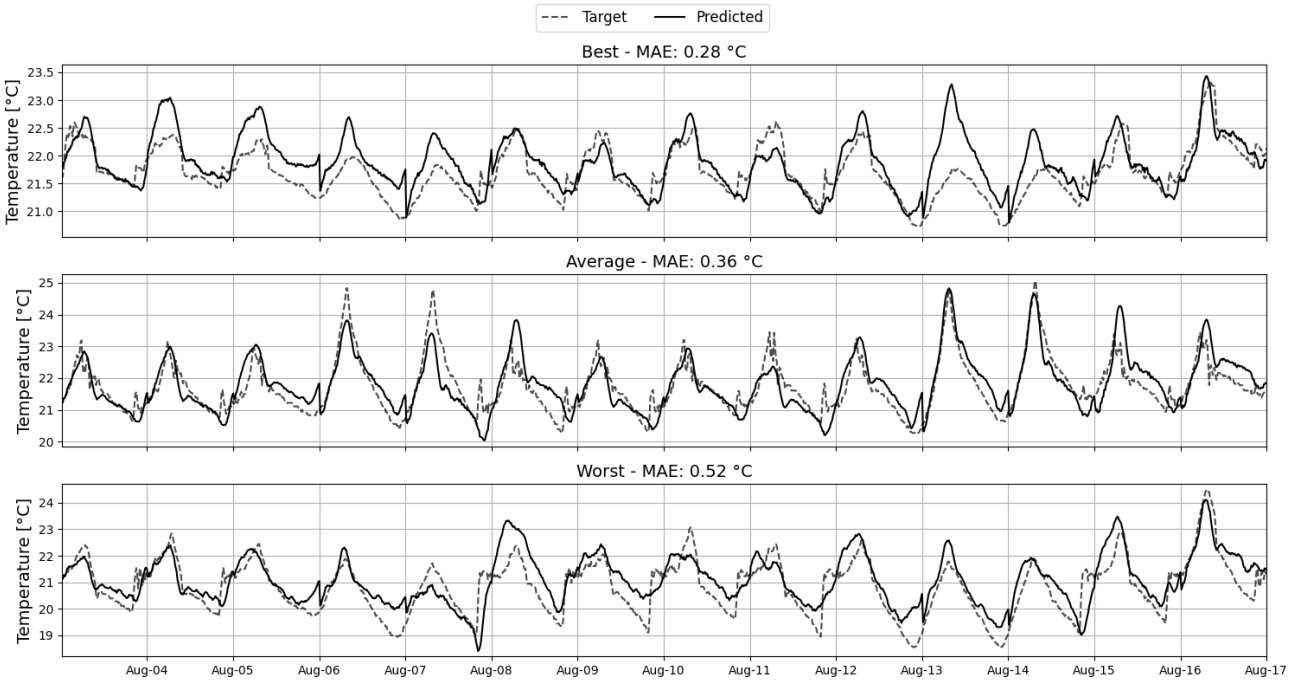


Figure 5: Temperature profiles over a week for zones with three different performance of the model, i.e. Best, Average, Worst. The prediction horizon and the step is 24-hours.

deployment of the model.

This work highlights three key aspects that will undoubtedly be extended in the future to increase the scalability and transferability of thermal dynamics models in buildings, with the goal of defining a ”foundation model” that can efficiently pre-train DRL agents.

The first aspect concerns the training of static embeddings from knowledge graphs that reflect essential building attributes, such as the type of HVAC system used, the properties of the thermal zone under consideration (e.g., surface-to-volume ratio, opaque and window component areas, stratigraphy types, etc.). These attributes are crucial for encoding the building's thermal dynamics, which can vary in speed depending on the building type (high or low mass) and HVAC system. By linking the deep learning model to these embeddings and teaching the model the relationship between them and the thermal dynamics, it will be possible to develop a transferable model. In this work, however, the KG embeddings were trained using node2vec, but in the future these embeddings can be pre-trained on a vast amount of graphs that represent different zone-energy systems configuration and made available.

The second aspect relates to the training dataset for this model. Unfortunately, there are limited datasets available that contain specific building information, particularly in terms of metadata. Therefore, an approach could involve pre-training the model on simulated data and then fine-tuning it for each individual building with a small amount of data. This approach would allow for the pre-training of DRL agents offline using the pre-trained thermal dynamics model, and subsequently, both the thermal model and DRL agent could be updated online via continuous learning.

The third and final aspect concerns the incorporation of physical constraints during the training phase. However, this feature should be adapted based on the availability of data. Specifically, if sufficient sensor data is available to validate the physical principles, the model's output can be grounded in physics, while still leveraging deep learning methods. In this case, the model would use both data and physical constraints to make more accurate predictions. However, in the absence of sufficient data, the model will still provide outputs, but they may not be as strongly grounded in physics. In other words, the precision of the model's output is directly influenced by the amount of available data: more observations lead to higher precision, while fewer observations may result in outputs that are less or not at all constrained by physics.

References

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Berges, M., Culler, D., Gupta, R., Kjærgaard, M. B.,

Srivastava, M., and Whitehouse, K. Brick: Towards a unified metadata schema for buildings. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, BuildSys '16, pp. 41–50, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342643. URL https://doi.org/10.1145/2993422.2993577.

Coraci, D., Brandi, S., and Capozzoli, A. Effective pre-training of a deep reinforcement learning agent by means of long short-term memory models for thermal energy management in buildings. *Energy Conversion and Management*, 291:117303, 2023. ISSN 0196-8904. doi: https://doi.org/10.1016/j.enconman.2023.117303.

Goldfeder, J., Dean, V., Jiang, Z., Wang, X., dong, B., Lipson, H., and Sipple, J. The smart buildings control suite: A diverse open source benchmark to evaluate and scale hvac control policies for sustainability. 2025.

Jiang, Z. and Dong, B. Modularized neural network incorporating physical priors for future building energy modeling. *Patterns*, 5(8):101029, 2024. ISSN 2666-3899. doi: 10.1016/j.patter.2024.101029.

Kimura, M., Shimizu, R., Hirakawa, Y., Goto, R., and Saito, Y. On permutation-invariant neural networks, 2024. URL https://arxiv.org/abs/2403.17410.

Lee, J. and Cho, S. Forecasting building operation dynamics using a physics-informed spatio-temporal graph neural network (pistgnn) ensemble. *Energy and Buildings*, 328: 115085, 2025. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2024.115085.

Piscitelli, M. S., Ye, Q., Chiosa, R., and Capozzoli, A. Spatio-temporal characterization and short-term prediction of indoor temperature in multi-zone buildings. In *Multiphysics and Multiscale Building Physics*, pp. 154–160, Singapore, 2025. Springer Nature Singapore.

Silvestri, A., Coraci, D., Brandi, S., Capozzoli, A., Borkowski, E., Köhler, J., Wu, D., Zeilinger, M. N., and Schlueter, A. Real building implementation of a deep reinforcement learning controller to enhance energy efficiency and indoor temperature control. *Applied Energy*, 368:123447, 2024. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2024.123447.

Wang, X., Wang, X., Yin, X., Li, K., Wang, L., Wang, R., and Song, R. Distributed lstm-gcn-based spatial–temporal indoor temperature prediction in multizone buildings. *IEEE Transactions on Industrial Informatics*, 20(1):482–491, 2024. doi: 10.1109/TII.2023.3268467.

Wen, S., Zhang, W., Zhou, N., and Yuan, L. Adaptive spatio-temporal graph convolutional network for rapid indoor temperature field prediction with limited sensors. *Building and Environment*, 283:113346, 2025. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2025.113346.

Zhang, J., Xiao, F., Li, A., Ma, T., Xu, K., Zhang, H., Yan, R., Fang, X., Li, Y., and Wang, D. Graph neural network-based spatio-temporal indoor environment prediction and optimal control for central air-conditioning systems. *Building and Environment*, 242:110600, 2023. ISSN 0360-1323. doi: https://doi.org/10.1016/j.buildenv.2023.110600.