

From Spatial to Temporal: A Geometry-Aware Approach to Indoor Thermal Prediction

Cross-Building Zone-Level Prediction for Rapid Summer Thermal Assessment

CHENGRUI YANG

Chalmers School of Architecture
Department of Architecture and Civil Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026

Supervisors: Alexander Hollberg, Yinan Yu, and Sanjay Somanath
Examiner: Jonas Lundberg



CHALMERS
UNIVERSITY OF TECHNOLOGY

From Spatial to Temporal: A Geometry-Aware Approach to Indoor Thermal Prediction

Cross-Building Zone-Level Prediction for Rapid Summer Thermal Assessment

CHENGRUI YANG

Institution: Chalmers School of Architecture
Department of Architecture and Civil Engineering

Master's Programme: Architecture and Urban Design
Profile: Architectural Experimentation / Matter & Media

Supervisors: Alexander Hollberg, Department of Architecture and Civil Engineering
Yinan Yu, Department of Computer Science and Engineering
Sanjay Somanath, Department of Architecture and Civil Engineering

Examiner: Jonas Lundberg, Department of Architecture and Civil Engineering

Gothenburg, Sweden 2026

From Spatial to Temporal: A Geometry-Aware Approach to Indoor Thermal Prediction
Cross-Building Zone-Level Prediction for Rapid Summer Thermal Assessment

Chengrui Yang

© Chengrui Yang, 2026.

Supervisors: Alexander Hollberg, Yinan Yu, and Sanjay Somanath
Examiner: Jonas Lundberg

Master's Thesis ACEX35
Department of Architecture and Civil Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31 772 1000

Typeset in LaTeX.
Gothenburg, Sweden 2026

Acknowledgements, dedications, and similar personal statements in this thesis reflect the author's own views.

Abstract

Climate change is increasing warm-season indoor heat exposure and building overheating risk, especially in Central and Northern Europe, where many homes have limited active cooling. Overheating varies within a building because floor level, orientation, glazing, shading, envelope properties, and spatial position affect each thermal zone differently. Many overheating indicators also depend on hourly operative-temperature trajectories. Therefore, rapid assessment requires both zone-level spatial resolution and dynamic temporal resolution.

This thesis develops and evaluates a cross-building surrogate framework trained on EnergyPlus simulations for rapid prediction of zone-level thermal conditions. Parametric buildings provide hourly operative-temperature reference data. Their geometry, component locations, and thermophysical properties are converted into sparse three-dimensional representations. The model combines whole-building and target-zone spatial information with weather, solar conditions, air exchange rate, and known dynamics of adjacent zones. It predicts hourly operative temperature under a fully separated building-level training, validation, and test set.

Across 200 independent test buildings evaluated using three fixed 168-hour periods, the proposed model achieves an operative-temperature MAE of 0.92 °C, an hourly temperature-change MAE of 0.087 °C/h, and a temperature-change direction accuracy of 94%. Compared with a baseline based on manually aggregated descriptors, the 3D model provides more consistent performance for overall temperature, strong temperature transitions, and the complete static-parameter set. Controlled experiments show that it recovers the main responses to window-to-wall ratio, exterior shading, orientation, and glazing. Climate-transfer tests retain good performance in Copenhagen and show systematic bias in Munich and Tromsø, with corresponding changes in high-temperature detection. These results define the current climate domain of the framework. Reusing the static geometric representation reduces the marginal cost of repeated evaluation: cached CPU inference are approximately 12.57× faster than repeated EnergyPlus simulations.

The framework supports rapid zone-level prediction for large-scale urban overheating screening, building retrofit decision-making, and early-stage building design.

Keywords: building overheating; indoor thermal environment; operative temperature; zone-level prediction; building-performance simulation; surrogate model; 3D building representation

Contents

Abstract iv

List of Figures vii

List of Tables ix

1 Introduction 1
- 1.1 Summer warming and building overheating challenges in Europe . . . 1
- 1.2 Spatial differences in building thermal conditions . . . 3
- 1.3 Overheating as a continuous temporal process . . . 3
- 1.4 Computational challenges in large-scale thermal assessment . . . 4
- 1.5 Cross-building prediction from building-performance simulation . . . 4
- 1.6 Aim, research questions, and thesis position . . . 5
- 1.7 Scope and limitations . . . 6

2 Related Research 7
- 2.1 Dynamic representation and prediction of indoor thermal conditions . . . 7
- 2.2 Dynamic building-performance simulation and surrogate models . . . 8
- 2.3 Building representation: from summary parameters to spatial structure . . . 8
- 2.4 Temporal modeling of indoor thermal dynamics . . . 9
- 2.5 Generalization in building-performance surrogate models . . . 10
- 2.6 Research gap and study positioning . . . 10

3 Method 12
- 3.1 Research framework and problem definition . . . 12
- 3.2 Parametric building data generation . . . 13
  - 3.2.1 Generation framework and building form . . . 13
  - 3.2.2 Thermal zoning and three-dimensional models . . . 14
  - 3.2.3 Envelope, openings, and operating conditions . . . 15
- 3.3 Conversion from HBJSON to EnergyPlus . . . 16
- 3.4 Sparse 3D building representation and zone indexing . . . 16
  - 3.4.1 HBJSON surface parsing and equivalent thermal properties . . . 16
  - 3.4.2 Panel extrusion, occupancy tests, and conflict resolution . . . 17
  - 3.4.3 Voxel fields and zone indices . . . 18
- 3.5 Geometry-aware spatio-temporal prediction model . . . 19
  - 3.5.1 Overall model structure . . . 19
  - 3.5.2 Whole-building sparse encoding and target-zone cross-attention . . . 20

3.5.3 Neighboring-zone dynamic context . . . 22
3.5.4 Temporal tokens and causal autoregressive decoding . . . 23
3.5.5 Output layer and training objective . . . 24
3.6 Flattened representation baseline . . . 25
3.7 Dynamic boundaries and information availability . . . 25
3.8 Training and inference protocol . . . 26
3.9 Evaluation framework . . . 26

4 Results **28**
4.1 Training and validation performance . . . 28
4.2 Response to static building-parameter perturbations . . . 28
4.2.1 Window-to-wall ratio and glazing . . . 29
4.2.2 Exterior and interior shading . . . 30
4.2.3 Other static factors . . . 31
4.3 Temporal-response fidelity under changing thermal conditions . . . 31
4.3.1 Overall and strong-transition response fidelity . . . 32
4.3.2 Zone-level paired response comparison . . . 32
4.3.3 Recursive error accumulation over the forecast horizon . . . 34
4.4 Generalization across unseen buildings and climates . . . 35
4.4.1 Unseen buildings under the source climate . . . 35
4.4.2 Transfer to different climate sources . . . 35
4.5 Sensitivity to training-building coverage and voxel resolution . . . 36
4.5.1 Number of training buildings . . . 36
4.5.2 Voxel resolution . . . 37
4.6 Geometry reuse and computational cost . . . 37

5 Discussion **39**
5.1 Cross-building prediction and the value of detailed spatial representation . . . 39
5.2 Thermal-response fidelity and temporal stability . . . 40
5.3 Generalization across building and climate domains . . . 41
5.4 Reusable spatial representation and the computational role of the surrogate . . . 42
5.5 Scope, limitations, and research outlook . . . 43

6 Conclusion **45**

References **47**

Student Background and an Evolving Approach to Architecture and Urban Design **51**

A Generative AI Use Statement **53**

List of Figures

1.1 Summer surface-air-temperature anomalies and long-term summer temperature changes in western Europe. . . . 2
1.2 Heat stress during the record-breaking July 2025 heatwave in Fennoscandia. . . . 2
1.3 Key facts on heat vulnerability and cooling challenges in European buildings. . . . 2
1.4 Cross-building zone-level dynamic thermal prediction. Target zones in buildings with different geometries are queried under specified weather and operating boundaries to generate hourly operative-temperature trajectories. . . . 5

2.1 Limitations of aggregated descriptors for zone-level thermal prediction. (a) Similar aggregated building parameters can correspond to different spatial configurations and geometric relationships. (b) Local spatial differences among zones can produce distinct thermal responses within the same building. . . . 9

3.1 Parametric building generation and preprocessing pipeline. Building parameters are converted into zoned three-dimensional HBJSON models, which provide a common source for EnergyPlus simulation and sparse-voxel generation. . . . 13
3.2 Examples of the parametric building-generation process. The two rows illustrate different generated building instances. . . . 14
3.3 Representative building forms generated by the parametric framework. The examples illustrate variation in footprint geometry, story count, envelope configuration, and window distribution across the synthetic building data set. . . . 15
3.4 EnergyPlus simulation curves showing air temperature and operative temperature for a room from July 15 to July 30 of a standard year. . . . 16
3.5 Schematic illustration of the voxelization workflow for two example buildings. . . . 18
3.6 Overall architecture of the proposed geometry-aware autoregressive model for zone-level indoor thermal forecasting. . . . 21

4.1 Static temperature-response profiles for WWR and glazing perturbations. EP denotes EnergyPlus, PM the proposed model, and FM the Flattened baseline. Points show the median temperature response across paired base buildings and horizontal bars show the interquartile range. The black dashed line indicates zero response, while the gray dashed lines indicate the corresponding EnergyPlus median responses for the tested perturbation levels. . . . 30
4.2 Representative zone-level WWR-response example for the proposed model and the Flattened baseline. Each panel shows the predicted and EnergyPlus temperature trajectories under different WWR perturbation levels relative to the same baseline case. . . . 30

4.3 Static temperature-response profiles for exterior and interior shading perturbations. EP denotes EnergyPlus, PM the proposed model, and FM the Flattened baseline. Points show the median temperature response across paired base buildings and horizontal bars show the interquartile range. The black dashed line indicates zero response, while the gray dashed lines indicate the corresponding EnergyPlus median responses for the tested perturbation levels. 31
4.4 Binned comparison between the reference hourly temperature change and the predicted hourly response of the proposed model and the Flattened baseline. The dashed diagonal indicates a one-to-one response. The reported response gain summarizes the retained amplitude of the hourly thermal response. 33
4.5 Representative zone-level dynamic prediction example comparing the proposed model and the Flattened baseline. The dashed black line denotes the EnergyPlus reference, the red line denotes the proposed-model prediction, the blue line denotes the Flattened-baseline prediction, and the gray line shows the outdoor temperature. 33
4.6 Zone-level paired comparison of hourly temperature-change prediction between the proposed model and the Flattened baseline. The left panel includes transitions with $|\Delta T_{\mathrm{EP}}| > 0.1\,^{\circ}\mathrm{C/h}$, while the right panel includes transitions with $|\Delta T_{\mathrm{EP}}| > 0.8\,^{\circ}\mathrm{C/h}$. Positive paired differences indicate lower zone-level $\mathrm{MAE}_{\Delta T}$ for the proposed model. 34
4.7 Change in temperature MAE across the 168-hour recursive forecast relative to the initial 0–24 h period. Values close to zero indicate that prediction error remains comparable to the initial forecast period, while positive values indicate increased error. 34
4.8 Summer climate-transfer performance of models trained on 2,000 and 4,000 Gothenburg buildings. Bars show MAE, triangles show mean prediction bias, and the dashed gray line indicates the latitude of each climate location. 36
4.9 Sensitivity analyses of training-data scale and voxel resolution. (a) Validation MAE and the proportion of predictions with absolute error above 2°C for different numbers of training buildings. (b) Validation MAE, relative number of active voxels, and relative observed training cost for different voxel resolutions; relative values are normalized to the 0.20 m setting. 37
4.10 Computational cost with geometry reuse. The upper panel shows cumulative runtime as the number of repeated dynamic scenarios increases. The lower panel decomposes the one-time setup cost and the marginal scenario cost for EnergyPlus, Proposed model (CPU), and Proposed model (GPU, batch size = 1). 38

5.1 Comparison of flattened and geometry-aware input representations. Flattened inputs provide direct low-dimensional signals for individual building factors, whereas geometry-aware inputs encode these factors less directly through component geometry while retaining richer spatial relationships. 40

List of Tables

2.1 Comparison of representative building-performance surrogate and thermal-prediction models. 11

3.1 Default sparse-voxel input features. 18
3.2 Availability of different information during warm-up and the prediction horizon. 25
3.3 Key training and inference configuration. 26

4.1 Training and validation performance of the proposed model and the Flattened baseline with 2,000 training buildings. 28
4.2 EnergyPlus mean absolute response magnitude and parameter-response errors of the proposed model and the Flattened baseline for major static building factors. 29
4.3 Overall and strong-transition temporal-response fidelity of the proposed model and the Flattened baseline on the independent Test buildings. 32
4.4 Cross-building generalization on the independent Test set under the source-climate condition. 35
4.5 Joint input-distribution shift relative to the Gothenburg training domain. Values are the proportions of target samples with nearest-neighbour distance above the 99th percentile of the training distribution. 36

Chapter 1

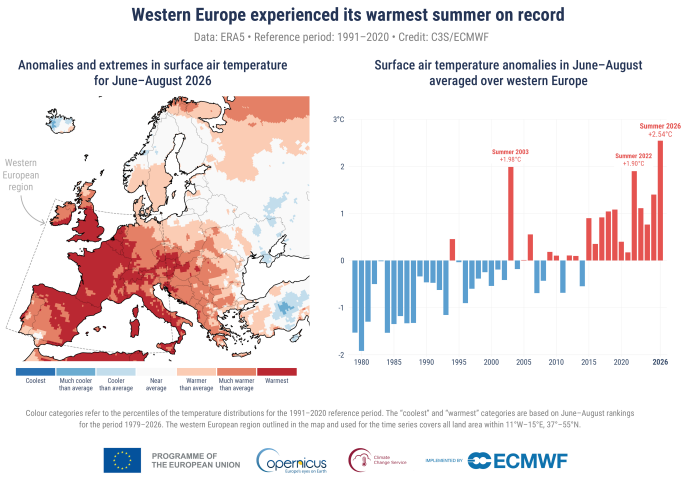
Introduction

1.1 Summer warming and building overheating challenges in Europe

Climate change is increasing warm-season temperatures and the frequency and intensity of heat events across Europe. As a result, building overheating has become an important concern even in regions that have historically experienced relatively moderate summer conditions. This issue is increasingly relevant in Central and Northern Europe, where buildings have traditionally been designed primarily for winter performance and summer overheating has received less attention (Intergovernmental Panel on Climate Change, 2022; Taylor et al., 2023). In 2026, western Europe experienced its warmest summer in the ERA5 record, with the June–August regional temperature anomaly reaching approximately $+2.54\,^{\circ}\mathrm{C}$ relative to the 1991–2020 reference period (Copernicus Climate Change Service (C3S), 2026), as shown in Figure 1.1. Recent extremes have also extended to much higher latitudes: in July 2025, sub-Arctic Fennoscandia experienced its most severe heatwave on record, lasting 21 days, while parts of the region experienced up to 12 days of strong heat stress (C3S/ECMWF & World Meteorological Organization, 2026), as shown in Figure 1.2.

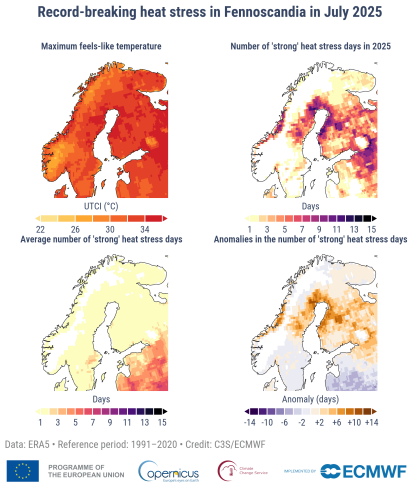
Recent heat events also show the growing public-health consequences of high temperatures. Based on mortality and meteorological data from several European countries, Jano et al. (2025) estimated approximately 62,775 heat-related deaths in 2024. These estimates describe the whole effects of heat exposure. However, Europeans spend approximately 80–90% of their time indoors, making building thermal conditions an important part of heat exposure (European Environment Agency, 2022). In addition, household air-conditioning equipment was only around 20% in 2019 (European Environment Agency, 2022) in Europe. In buildings without active cooling, indoor temperatures are strongly affected by outdoor conditions, solar gains, shading, thermal mass, ventilation, and other building characteristics. Evidence from both building and health studies shows that indoor conditions cannot be represented by outdoor temperature alone. Monitoring of homes occupied by older adults in the Netherlands found stronger associations between indoor temperature and heat discomfort, and related symptoms (van Loenhout et al., 2016). Recent reviews have similarly reported associations between high residential indoor temperatures and adverse effects on health, well-being, and thermal comfort (Edwards et al., 2025). These findings highlight the importance of assessing the thermal conditions that buildings actually provide during hot weather.

The challenge is increasingly visible in Nordic building stocks. Studies of Finnish apartment buildings



Source: Copernicus Climate Change Service (C3S), ECMWF (2026), 10 September 2026. Data: ERA5; reference period: 1991–2020; credit: C3S/ECMWF. Copernicus Climate Change Service.

Figure 1.1: Summer surface-air-temperature anomalies and long-term summer temperature changes in western Europe.

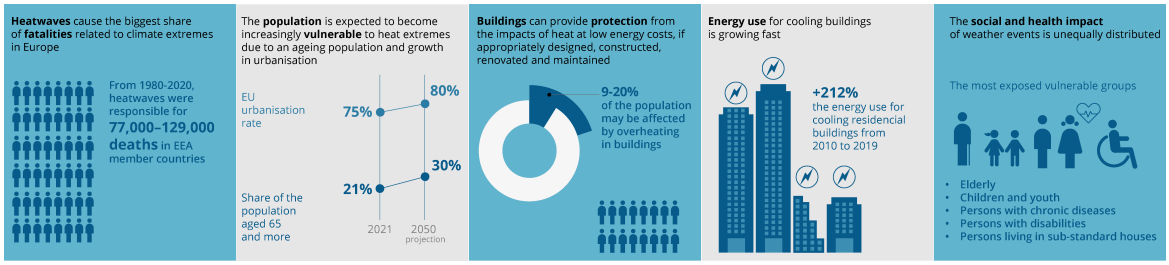


Source: C3S/ECMWF and WMO (2026), *C3S–WMO European State of the Climate 2025*, Data: ERA5-HEAT; reference period: 1991–2020; credit: C3S/ECMWF. DOI: 10.24381/zy93-sb27.

Figure 1.2: Heat stress during the record-breaking July 2025 heatwave in Fennoscandia.

have reported notable overheating under both present and future climates, particularly during extreme warm summers, while shading, window opening, ventilation, and active cooling strongly influence indoor conditions (Farahani et al., 2021). Similar concerns have been identified in Swedish multi-family housing, where buildings developed for cold-climate conditions may provide limited protection against increasingly frequent and intense summer heat (Schade et al., 2025). These studies show that buildings developed primarily for cold-climate conditions can face significant summer overheating under current and future climates.

European policy is also moving in this direction. The recast Energy Performance of Buildings Directive, Directive (EU) 2024/1275, gives greater attention to indoor environmental quality and climate adaptation in new buildings and major renovations. Its preamble specifically addresses summer thermal performance and measures such as shading, thermal capacity, and passive cooling (European Parliament and Council of the European Union, 2024). This policy direction reflects a broader adaptation challenge in the European building stock. As summarized by the European Environment Agency, buildings can provide protection from heat when appropriately designed and renovated, while overheating vulnerability, increasing cooling demand, and unequal health impacts make summer thermal performance an increasingly important building-performance issue (European Environment Agency, 2022), as illustrated in Figure 1.3.



Source: European Environment Agency (EEA) (2022), *Cooling buildings sustainably in Europe: exploring the links between climate change mitigation and adaptation, and their social impacts*, Figure 1, "Key facts on vulnerability." DOI: 10.2800/61355.

Figure 1.3: Key facts on heat vulnerability and cooling challenges in European buildings.

1.2 Spatial differences in building thermal conditions

Residential overheating has a strong spatial dimension. A building contains thermal zones with different external exposure and boundary conditions. Facade orientation and window placement determine the path of solar radiation. Floor and room position change the exposure to the roof, ground, and external walls. Spatial subdivision determines zone volume, component area, and adjacency between zones. Geometry organizes thermal boundaries and physical properties in space and shows significant differences.

Measurements and simulations show that these spatial differences can directly affect overheating. Arriazu-Ramos et al. (2023) compared summer overheating across apartments and found significant effects from floor position, window area, and external shading. Windows larger than $4\,m^2$ get almost three times the probability of overheating of windows smaller than $2\,m^2$, while apartments on the middle-floor were less affected than apartments on the upper floor. In typical Swedish multi-family housing, overheating risk also increased on higher floors, and west- and south-facing rooms with large windows showed higher risk (Schade et al., 2025). Finnish housing studies also show that orientation, exterior shading, surrounding Environment, window performance, and ventilation strategy all together shape indoor temperature (Farahani et al., 2021).

Whole-building average indicators can conceal high exposure in heterogeneous conditions. Spatial differences can influence decisions of window and shading design, facade and plan organization and local retrofit measures. Building-stock assessment also benefits from identifying individual rooms or thermal zones that require attention.

These observations motivate the use of three-dimensional building information in this study. Aggregated variables such as floor area, window-to-wall ratio, orientation, and compactness have clear physical meaning, but their effects on thermal response are coupled and often nonlinear. The resulting indoor thermal state depends on how these properties are combined and distributed within the building. For zone-level temperature prediction, spatial information such as component locations, zone boundaries, adjacency, and the position of the target zone within the whole building can provide additional information that aggregated descriptors do not fully capture. Therefore, the proposed representation need to retain the three-dimensional relationships between building geometry, component location, thermophysical properties, and thermal zoning.

1.3 Overheating as a continuous temporal process

Overheating varies over time. Many residential overheating methods derive indicators from hourly or finer temperature trajectories, including exceedance hours, degree-hours, peak temperature, and other time-integrated measures (Rahif et al., 2021). The duration and timing of high temperature and the capacity for night-time cooling can all affect the final assessment.

Indoor temperature is affected by outdoor weather, solar radiation, envelope properties, thermal mass, ventilation, and heat exchange with adjacent spaces. These influences are highly time-dependent because heat is continuously stored, transferred, and released within the building. As a result, the indoor thermal state at a given time depends on both current conditions and the preceding thermal history. This temporal dependence is particularly important during consecutive warm days, when heat can accumulate over multiple hours or days (Kuczyski et al., 2021).

The fundamental output of this study is a *zone-level hourly operative-temperature trajectory*. The sequence retains thermal differences among zones and the processes of warming, cooling, thermal lag,

and sustained high temperature. It can support direct analysis of thermal response and the calculation of high-temperature hours, threshold exceedance, and related overheating indicators. Model evaluation need to cover average temperature error, rapid transitions, controlled changes in building parameters, and high-temperature states.

1.4 Computational challenges in large-scale thermal assessment

Dynamic building-performance simulation provides a physically grounded account of building thermal processes. Simulators such as EnergyPlus can calculate transient heat transfer, solar gains, envelope heat storage, air exchange, internal gains, and system operation at sub-hourly or hourly resolution (Crawley et al., 2001). The computational cost is usually manageable for one building and a limited number of scenarios.

The computational cost grows rapidly with the scale of assessment. European climate-adaptation guidance emphasizes a combination of building-envelope renovation, passive cooling measures, efficient ventilation, and, where necessary, efficient active cooling to improve building resilience under warmer climates (European Environment Agency, 2022). Evaluating these strategies requires comparing multiple building parameters and intervention options, such as orientation, glazing, shading, envelope construction, ventilation, and cooling measures. Retrofit studies may further consider different measures and their combinations, while building-stock and climate-adaptation assessments can involve hundreds or thousands of buildings under multiple weather and operating scenarios. The resulting number of simulations therefore grows with the number of buildings, intervention scenarios, climate conditions, and simulation periods. Repeated dynamic simulation can consequently become a major computational constraint in large-scale scenario assessment and optimization (Nguyen et al., 2014). Automated scanning, modelling, and model-preparation tools can reduce the manual effort required to construct simulation models, while each new dynamic scenario still requires a separate thermal-balance solution. Even when geometry and envelope properties remain unchanged, a change in weather, air exchange, or another dynamic boundary requires a new simulation. Rapid prediction tools that can efficiently evaluate many buildings and repeated design or retrofit scenarios are therefore necessary for large-scale applications.

1.5 Cross-building prediction from building-performance simulation

Surrogate models reduce repeated calculation by learning an approximate mapping between the inputs and outputs of a high-fidelity simulator. They are widely used in building-design optimization, energy-demand prediction, and scenario analysis (Roman et al., 2020; Westermann & Evins, 2019). Design-stage surrogates commonly use manually selected low-dimensional descriptors, including building area, orientation, window-to-wall ratio, and envelope properties. Component-based and deep-learning approaches have extended their ability to process complex building inputs (Singaravel et al., 2018, 2019). More recent work learns spatial representations directly from building form (Singh & Smith, 2023).

Most building-performance surrogates predict whole-building heating, cooling, or annual energy indicators. These targets support building-level comparison but provide limited information about temperature differences among rooms and the three-dimensional relations that produce local overheating.

A related research stream predicts indoor temperature for HVAC control, operational optimization, and short-term forecasting. These methods can achieve high short-term accuracy by using historical indoor temperatures, sensor records, control signals, or occupancy data from the target building (Drgoa et al., 2021; Fang et al., 2021; Wang, Wang, et al., 2024; Xing et al., 2023; Yang et al., 2024). However, early design, topology-changing retrofits, and poorly monitored building stocks often lack this history data. These models may also have limited generalization capability across different buildings.

1.6 Aim, research questions, and thesis position

Based on the above discussion of spatial thermal response, temporal indoor-temperature prediction, building overheating, and the need for large-scale screening, this thesis aims to develop and evaluate a rapid method for zone-level dynamic thermal prediction across buildings. Given building geometry, thermophysical properties, weather, and operating conditions, the method estimates hourly operative temperature for each zone of a unseen building. The study further examines how such predictions can support overheating-sensitive screening in building design, retrofit assessment, and building-stock analysis, reducing the need to apply detailed building-performance simulation to every building and scenario.

Building geometry, component locations, zoning, and thermophysical properties are encoded as a sparse 3D representation. The model extracts a target-zone-specific spatial condition from whole-building information and combines it with weather, solar conditions, prescribed air exchange, and known dynamics of adjacent zones to generate hourly operative temperature through a causal temporal process. Parametric EnergyPlus simulations provide the training labels, while building-level separation is used to evaluate generalization to unseen buildings. As illustrated in Figure 1.4, the resulting zone-level temperature trajectories can be compared across buildings, zones, parameter settings, and climate scenarios to support rapid identification of high-temperature conditions.

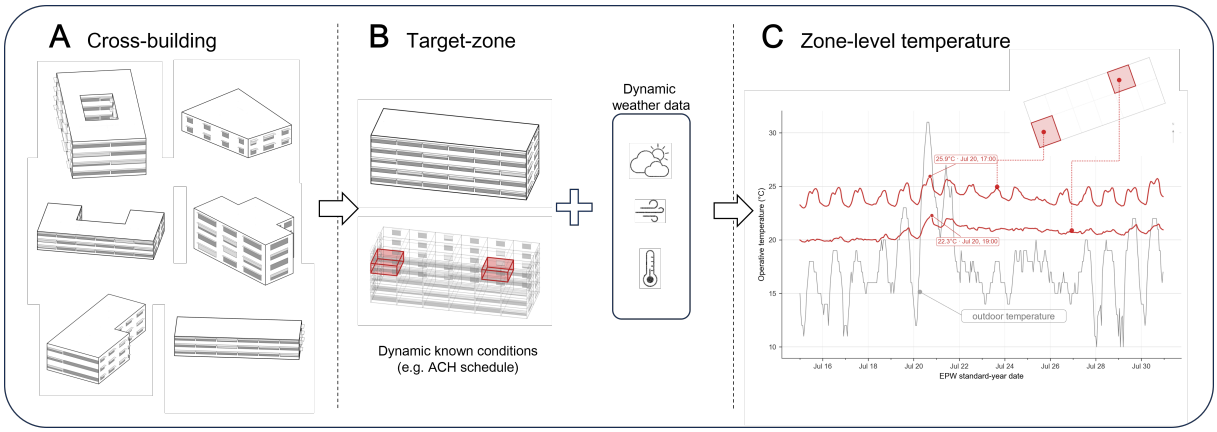


Figure 1.4: Cross-building zone-level dynamic thermal prediction. Target zones in buildings with different geometries are queried under specified weather and operating boundaries to generate hourly operative-temperature trajectories.

The study addresses four research questions:

RQ1: Can a cross-building surrogate model predict zone-level operative-temperature dynamics on unseen buildings with sufficient accuracy? How much value does detailed 3D spatial information add beyond manually aggregated low-dimensional building descriptors?

RQ2: How well does the framework retain thermal responses relevant to overheating screening, including changes in building parameters, hourly warming and cooling at different intensities, and the identification of high-temperature states?

RQ3: How does predictive performance change across unseen building geometries and climate conditions, and what application boundaries follow from these changes?

RQ4: How much can a reusable static 3D building representation reduce the cost of evaluating multiple weather or operating scenarios for the same building compared with repeated building-performance simulation?

The evaluation in this study is organized around building thermal performance. It covers prediction on unseen buildings, responses to building parameters, rapid thermal dynamics, high-temperature states, climate transfer, and repeated-scenario efficiency.

1.7 Scope and limitations

The main experiments use controlled free-running conditions to focus on the thermal response of building space, envelope properties, and external dynamic boundaries. Active HVAC is disabled, and internal gains from occupants, lighting, and equipment are set to zero. Prescribed hourly air changes remain as an explicit dynamic boundary. This setting provides a controlled physical baseline for comparing the thermal response of spaces and building parameters.

Free-running operation is relevant to many Central and Northern European homes without continuous active cooling, although actual buildings also include occupants, internal gains, window operation, shading behavior, and mechanical systems. These processes also affect indoor temperature but are excluded from current research.

The prediction target is zone operative temperature. Complete regulatory assessment additionally require occupant exposure, adaptive comfort, active cooling, urban heat-island effects, construction uncertainty, and building-use information.

The training buildings are parametrically generated, so geometric generalization is bounded by the sampled design space. Gothenburg provides the only reference climate and EnergyPlus outputs serve as the reference data throughout the study. Real-world indoor temperature measurements remain limited in the current study.

Within these boundaries, this thesis presents an initial framework for rapid, high-accuracy, and high-fidelity zone-level indoor temperature prediction. The main contribution is to demonstrate the feasibility of using reusable spatial building representations and temporal modeling for dynamic thermal prediction across variable buildings. The results provide a basis for further development towards large-scale building performance screening, retrofit assessment, and design-stage prediction.

Chapter 2

Related Research

2.1 Dynamic representation and prediction of indoor thermal conditions

Indoor thermal conditions result from the combined effects of building geometry, envelope properties, weather, and operating boundaries. Annual heating and cooling demand aggregate performance over long periods, while indoor temperature describes the thermal environment over time. It is a central output of building-performance simulation and an important variable in building-operation forecasting and thermal-comfort research.

Hourly trajectories are particularly useful for warm-season assessment, as heat storage in the building envelope, solar gains, and air exchange produce cumulative and delayed effects on indoor temperature. Overheating indicators such as exceedance hours, degree-hours, and sustained high temperature are commonly derived from continuous indoor-temperature series (Rahif et al., 2021). The time series therefore contains more information about the thermal response than an isolated peak or average.

Indoor thermal conditions are commonly represented by air temperature or operative temperature. This study uses zone operative temperature. Under indoor conditions with low air speed, operative temperature can be approximated as a weighted combination of air temperature and mean radiant temperature:

$$T_z^{\mathrm{op}} = wT_{a,z} + (1-w)T_{r,z}, \qquad w = 0.5, \tag{2.1}$$

where $T_{a,z}$ is zone air temperature and $T_{r,z}$ is mean radiant temperature. Air temperature represents the state of the zone air, while mean radiant temperature reflects the influence of windows, walls, roofs, solar exposure, and indoor surface temperatures. Operative temperature links envelope response to the thermal environment experienced within a space.

Indoor temperature evolves over time in response to weather, solar gains, air exchange, and heat storage in the building envelope. At the zone level, this temporal response is further shaped by spatial differences within the building, including zone position, exterior exposure, geometric configuration, and interactions with adjacent spaces. In conclusion, Predicting zone-level thermal dynamics requires both the temporal evolution of thermal conditions and the spatial context in which each zone is located. The following sections review how building-performance simulation, surrogate modeling, geometric representation, and multi-zone forecasting represent these aspects.

2.2 Dynamic building-performance simulation and surrogate models

Dynamic building-performance simulation is a principal method for studying building thermal processes. Simulators such as EnergyPlus solve building heat balances at hourly or finer temporal resolutions using information on building geometry, envelope construction, weather, and operation. The simulation accounts for heat transfer through the envelope, solar gains, air exchange, internal heat sources, and building-system operation to produce zone thermal conditions (Crawley et al., 2001). Their explicit account of transient processes supports building design, operation, and climate-adaptation studies.

Surrogate models provide a faster approximation when optimization and large scenario studies require many simulation runs. Repeated simulation can become a major computational cost in simulation-driven building optimization (Nguyen et al., 2014). In building-performance applications, such training data can be generated from a set of high-fidelity simulations. Once trained, the surrogate replaces many repeated simulation runs with rapid predictions.

Building-performance surrogates have used linear regression, Gaussian processes, support-vector machines, tree ensembles, and artificial neural networks (Roman et al., 2020; Westermann & Evins, 2019). Tsanas and Xifara (2012) predicted heating and cooling loads from relative building arguments. Similar parametric-based surrogate models are widely used for early design evaluation and design-space search. Wang, Harrison, et al. (2024) used machine-learning models to approximate heating demand across a parametric residential design space and rapidly compare many design combinations.

2.3 Building representation: from summary parameters to spatial structure

A building surrogate model requires a machine-readable representation of the building. A common approach uses manually defined low-dimensional variables such as floor area, volume, compactness, orientation, window-to-wall ratio, and envelope properties (Tsanas & Xifara, 2012; Wang, Harrison, et al., 2024). These variables are computationally efficient and correspond directly to familiar building-design parameters, making them well suited to regular design spaces described by a limited number of parameters.

However, such flattened representations compress the building into a small set of global descriptors and do not explicitly preserve many spatial relationships among building components or the heterogeneity of zone-level thermal responses, as illustrated in Figure 2.1.

Later studies increased the expressive capacity of these representations. Singaravel et al. (2018) decomposed building-performance prediction across building components and their associated design variables. Singaravel et al. (2019) applied convolutional neural networks to learn feature interactions from predefined building variables. Such approaches retain the efficiency of low-dimensional inputs while reducing the need to specify all interactions between design parameters manually.

More recent studies have introduced building geometry and spatial organization directly into the learning process. Singh and Smith (2023) used convolutional neural networks and autoencoders to derive latent representations of building form. Cao et al. (2022) represented floor-plan organization as a graph, while Wu et al. (2025) used graph neural networks to describe spatial relations among buildings

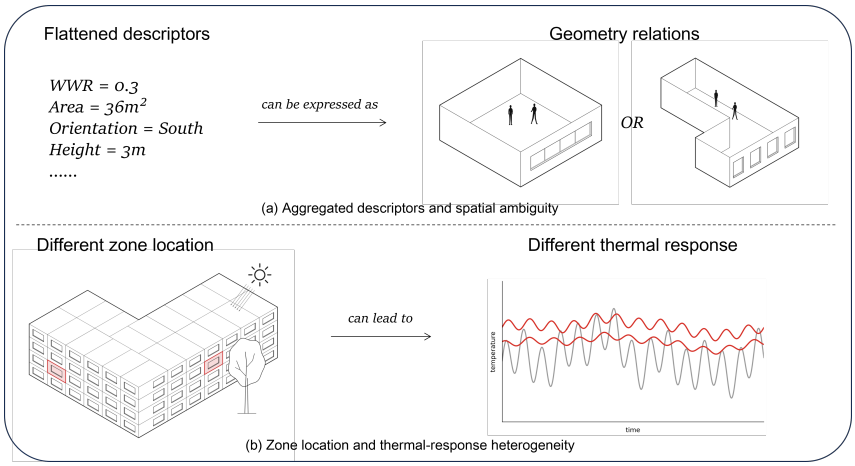


Figure 2.1: Limitations of aggregated descriptors for zone-level thermal prediction. (a) Similar aggregated building parameters can correspond to different spatial configurations and geometric relationships. (b) Local spatial differences among zones can produce distinct thermal responses within the same building.

in urban blocks. These approaches extend building representation beyond aggregated scalar descriptors and allow geometric or topological relations to contribute directly to prediction.
Several representation formats can be used to encode building geometry, including point clouds, surface meshes, graphs, and voxels. Each format provides a different balance between geometric detail, connectivity, and computational efficiency. Voxel representations discretize the three-dimensional space into a regular grid, allowing building components, locations, and physical properties to be represented within a common coordinate system. Because building elements occupy only a fraction of the full bounding volume, sparse convolution can restrict computation to active voxels and avoid processing of empty space (Choy et al., 2019; Graham et al., 2018). This makes sparse voxels suitable for learning from three-dimensional building information while retaining a regular spatial structure for convolution and aggregation.

2.4 Temporal modeling of indoor thermal dynamics

Indoor temperature varies continuously in response to weather, solar gains, air exchange, thermal storage, internal loads, and building operation. Data-driven forecasting studies commonly represent these dynamics using recent observations of indoor temperature together with weather, control signals, equipment states, and occupancy. Depending on the application, prediction horizons range from minutes to several hours or days, particularly for model-predictive control and operational optimization. Recurrent and sequence models are widely used for this purpose. Fang et al. (2021) developed an LSTM sequence-to-sequence model that used historical temperature dynamics from multiple zones to predict future indoor temperatures. Xing et al. (2023) combined multiple sensor and operational inputs for short-term multi-zone temperature forecasting. These studies show that recent thermal and operational histories provide strong information about subsequent indoor-temperature evolution.
Temporal models have also been extended to account for interactions among multiple zones. Wang, Wang, et al. (2024) combined graph convolution with distributed LSTM modules to represent spatial relations between zones together with their temporal sequences. Yang et al. (2024) incorporated inter-zone heat-flow relations and physical constraints into a graph-based thermal model. In these approaches, neighboring-zone states provide additional information for predicting the thermal response of an individual zone.

Another line of research introduces physical structure into data-driven thermal prediction. Drgoa et al. (2021) incorporated thermodynamic structure into a neural predictor for building control, while related studies developed physics-constrained neural networks for thermal-dynamics modeling (Di Natale et al., 2022, 2023). These methods use heat-balance relations or other physical constraints to guide temporal prediction and improve consistency with building thermal behavior.
Most existing operational forecasting methods rely on historical measurements from the target building. Indoor-temperature records, control signals, occupancy information, and other operational data provide direct evidence of building-specific thermal dynamics and are highly effective for short-term forecasting. However, in early design, building-stock assessment, and cross-building screening, the target building may have no historical indoor-temperature measurements. In this case, prediction must rely on the building geometry, physical properties, and external dynamic conditions rather than on historical indoor-temperature records.

2.5 Generalization in building-performance surrogate models

The scope of generalization varies considerably among building-performance surrogate models. Many models are developed for a specific building, building type, or predefined parametric design space. In these cases, the model learns relationships within a relatively controlled set of conditions and is evaluated on unseen parameter combinations generated from the same underlying design rules (Singaravel et al., 2019; Wang, Harrison, et al., 2024). Such evaluation mainly measures interpolation within the sampled design space.
A broader objective is to develop models that remain applicable when changing target buildings. This requires the model to respond to differences in geometry, spatial organization, envelope characteristics, and other building-specific conditions rather than relying on a fixed building configuration. The prediction problem then shifts from learning the response of one predefined system to learning thermal-response patterns that are shared across a collection of buildings.
Recent studies have also examined generalization across geographic and climatic domains. Canli Akyol et al. (2025) investigated model transfer between cities in urban building-energy modeling and showed that differences in building stock, urban context, and climate can affect predictive performance. Dai et al. (2025) combined static building descriptors with weather time series for hourly urban-building load prediction, illustrating how building-specific characteristics and changing external conditions can be represented within the same predictive framework.
These different forms of generalization define different levels of difficulty. Prediction within a fixed building or parametric design space relies on a relatively stable system definition, whereas cross-building prediction must accommodate new geometric and spatial configurations. Cross-climate prediction introduces an additional shift in external thermal conditions. Different generalization requirements also influence the choice of model architecture and input representation.

2.6 Research gap and study positioning

The reviewed research provides complementary foundations for the present study.
Building-performance surrogate models demonstrate how simulation data can be used to train reusable

predictive models for rapid performance assessment. Geometry-learning studies extend building representation beyond low-dimensional summary variables by introducing convolutional, graph-based, and other spatial representations. Temporal forecasting studies show how indoor thermal dynamics can be represented from evolving thermal states and external conditions, while generalization studies distinguish models developed for specific target buildings from those intended for use across different buildings, cities, or climates.

These research directions define the main requirements for this cross-building zone-level thermal prediction. The building representation needs to accommodate different geometries and zone configurations. Zone-specific thermal conditions must be extracted from the spatial context of the whole building. Static geometric and physical information must be combined with changing weather and operating conditions in hourly prediction. The resulting model should also remain applicable to buildings with previously unseen geometry and spatial organization.

The present study addresses these requirements through a sparse three-dimensional building representation, target-zone-conditioned spatial aggregation, and causal temporal prediction. Table 2.1 positions this formulation relative to representative studies according to their building representation, prediction scale, temporal formulation, dependence on target-building history, and generalization setting.

Table 2.1: Comparison of representative building-performance surrogate and thermal-prediction models.

Study	**Representation**	**SC**	**PT**	**TF**	**TH**	**Generalization**
Singaravel et al. (2018)	Component-based design variables	B	E	×	×	Unseen designs and transfer settings
Singaravel et al. (2019)	Predefined design variables with learned features	B	E	×	×	Defined design space
Singh & Smith (2023)	CNN and autoencoder representation of building form	B	E	×	×	Unseen geometries
Cao et al. (2022)	Floor-plan layout graph	B	E	×	×	Cross-building generalization not a primary focus
Fang et al. (2021)	Multi-zone time-series variables	Z	T	✓	✓	Building-specific
Drgoa et al. (2021)	Physics-structured recurrent representation	Z	T	✓	✓	Building-specific
Wang, Wang, et al. (2024)	Zone graph with distributed LSTM	Z	T	✓	✓	Single building
Canli Akyol et al. (2025)	Tabular building descriptors	U	E	×	✓	Cross-city transfer
Dai et al. (2025)	Static building descriptors and weather sequences	U	E	✓	×	Cross-climate evaluation
This study	**Sparse 3D representation with target-zone conditioning**	**Z**	**T**	✓	×	**Held-out buildings and cross-climate tests**

Abbreviations: SC = scale: B = building, Z = zone, U = urban building; PT = prediction target: E = energy-related performance, T = temperature or thermal dynamics; TF = temporal formulation: ✓ = dynamic temporal modelling, × = static prediction; TH = target-building history: ✓ = target-building or target-domain data required, × = not required.

Target-building history refers to indoor-temperature or operational observations from the building being predicted. Limited target-domain data used for model adaptation are also marked as required. Simulation outputs used as supervised training labels are excluded.

Chapter 3

Method

3.1 Research framework and problem definition

The research framework consists of parametric building generation, dynamic building-performance simulation, cross-building surrogate training, and multi-dimensional evaluation. EnergyPlus provides zone-level operative-temperature references for different buildings and dynamic boundaries. The surrogate model learns a mapping from building space, thermophysical properties, and known dynamic conditions to operative-temperature trajectories. Complete separation of training, validation, and test buildings measures the transfer of learned spatial thermal responses to unseen buildings.
The model inputs follow the information structure of a multi-zone heat balance. For zone z, air-temperature dynamics can be expressed conceptually as

$$C_{a,z}\frac{\mathrm{d}T_{a,z}}{\mathrm{d}t} = \dot{Q}_{\mathrm{surf},z} + \dot{Q}_{\mathrm{oa},z} + \dot{Q}_{\mathrm{mix},z} + \dot{Q}_{\mathrm{int},z} + \dot{Q}_{\mathrm{sys},z}, \tag{3.1}$$

where the terms represent convective surface heat transfer, outdoor-air exchange, inter-zone air exchange, internal gains, and active-system effects. Surface heat transfer can be written as

$$\dot{Q}_{\mathrm{surf},z} = \sum_{s\in\mathscr{S}_z} h_{c,s}A_s(T_s - T_{a,z}), \tag{3.2}$$

where the building geometry defines the enclosing surfaces $\mathscr{S}_z$, their areas A_s, orientations, and geometric relationships to the target zone. Material properties determine conduction and heat storage, while boundary type defines each connection to outdoors, the ground, or an adjacent zone. Solar radiation changes surface temperature T_s through exterior surfaces and window systems. Geometry therefore describes the spatial organization of zone heat-transfer boundaries and exposure.
The main data set uses controlled free-running conditions. Active HVAC is disabled, and internal gains from occupants, lighting, and equipment are set to zero. Outdoor-air exchange remains an explicit exogenous boundary. With air changes per hour $\mathrm{ACH}_z(t)$, its sensible-heat effect can be approximated as

$$\dot{Q}_{\mathrm{oa},z}(t) = \rho_a c_p \frac{\mathrm{ACH}_z(t)V_z}{3600}[T_{\mathrm{out}}(t) - T_{a,z}(t)]. \tag{3.3}$$

Randomly generated schedules are known in advance and allow air exchange to vary as an independent dynamic boundary.
The information boundary is defined as follows. Static building geometry and material properties are

known. Weather, solar variables, and prescribed ACH are known over the forecast horizon. Measured temperatures from neighboring zones are excluded. The model recursively generates the target-zone state after the forecast begins. Complete buildings are assigned to a single data subset. The learning task is

$$\widehat{\mathbf{T}}^{\mathrm{op}}_{b,z,t+1:t+H} = f_\theta(\mathscr{G}_b, \mathscr{G}_{b,z}, \mathbf{X}^{\mathrm{met}}, \mathbf{U}_{b,z}, \mathscr{N}_{b,z}, \widehat{T}^{\mathrm{op}}_{b,z,t}), \tag{3.4}$$

where $\mathscr{G}_b$ and $\mathscr{G}_{b,z}$ denote static spatial and physical information for the whole building and target zone, $\mathbf{X}^{\mathrm{met}}$ contains weather and solar variables, $\mathbf{U}_{b,z}$ contains exogenous operating boundaries for the target zone, $\mathscr{N}_{b,z}$ contains neighboring-zone topology and known exogenous dynamics, and $\widehat{T}^{\mathrm{op}}_{b,z,t}$ is the latest available predicted operative temperature of the target zone, which provides the autoregressive initial state for the subsequent forecast.

3.2 Parametric building data generation

This section describes the parametric workflow used to construct the building data set and to maintain consistent geometric, physical, and thermal information across simulation and learning inputs. As summarized in Figure 3.1, building samples are progressively converted into zoned three-dimensional building models, exported to HBJSON, and then processed through two parallel branches for EnergyPlus simulation and sparse-voxel representation.

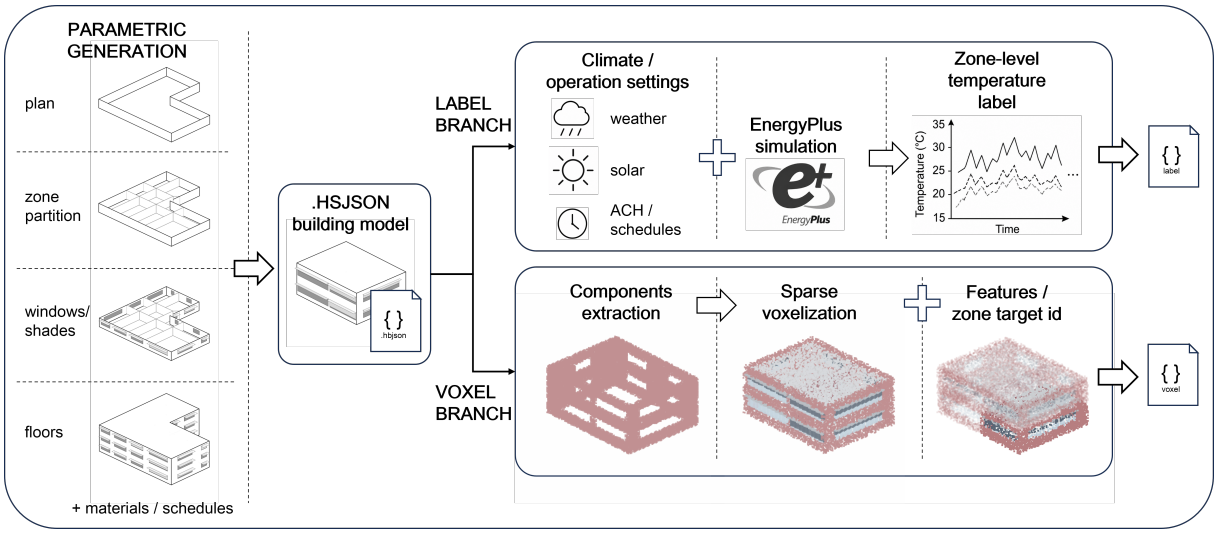


Figure 3.1: Parametric building generation and preprocessing pipeline. Building parameters are converted into zoned three-dimensional HBJSON models, which provide a common source for EnergyPlus simulation and sparse-voxel generation.

3.2.1 Generation framework and building form

The main training data set contains 2,000 independent parametric building instances. A common generation framework combines building form, zone topology, windows, shading, envelope construction, exterior obstruction, and air exchange. Parameters are assigned at different scales. Building-level variables, including floor-plan type, story count, structural system, glazing system, and exterior context, remain consistent inside each building. Zone- and opening-level variables, including

air exchange and local shading, vary under building-level constraints. Two examples of the progressive generation process are shown in Figure 3.2.

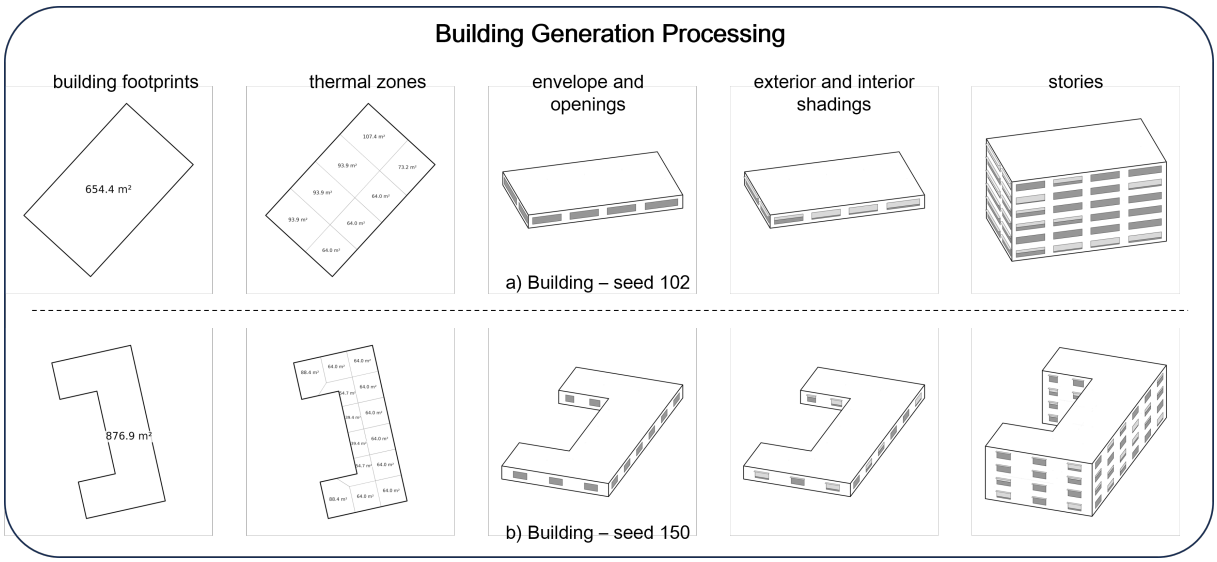


Figure 3.2: Examples of the parametric building-generation process. The two rows illustrate different generated building instances.

Building footprints are sampled from six forms commonly observed in the Gothenburg area with target proportion: rectangle (25%), convex quadrilateral (15%), L-shape (20%), T-shape (15%), U-shape (15%), and courtyard (10%). Effective floor area per story ranges from 100 to 1,000 m^2, width and depth range from 8 to 48 m, with a maximum aspect ratio is 5. All footprints are checked for validity, connectivity, and area.

3.2.2 Thermal zoning and three-dimensional models

Thermal zones are generated through clipped Voronoi subdivision from footprint above. The internal zoning is generated to produce rooms with a typical dimension of approximately 8.0, m. The resulting Voronoi cells are clipped to the building outline. Fragment zones smaller than 12 m^2 are merged into the neighbor with the longest shared edge. Zones contain the exterior perimeter are classified as perimeter zones, and the remaining zones are classified as core zones. Courtyard boundaries are treated as exterior boundaries. This process creates spatial units with controlled size, closed topology, and valid multi-zone heat balances.

The two-dimensional model is rotated in $[0°, 360°)$ with a probability of 90%; the remaining cases retain their initial orientation. The same zoning is extruded across stories with floor-to-floor heights sampled from 2.7 to 4.0 m. When floor area does not exceed 200 m^2, one-, two-, and three-story buildings have target proportions of 30%, 50%, and 20%, respectively. Larger buildings cover one to six stories and include middle-section models for taller buildings. These middle sections represent consecutive typical floors and use adiabatic top and bottom boundaries to approximate neighboring floors outside the model.

Exterior surfaces receive `Outdoors`, `Ground`, or `Adiabatic` boundary conditions according to their location. Ground floors mainly use ground boundaries and top floors mainly use outdoor boundaries. Interior walls and intermediate slabs use paired `Surface` boundaries. The resulting model retains heat exchange with outdoors and the ground as well as conductive connections across shared components.

3.2.3 Envelope, openings, and operating conditions

Window parameters are sampled at the building level. The training set assigns target proportions of 85%, 10%, and 5% to mild, extremely large, and extremely small windows. Mild window-to-wall ratio (WWR) ranges depend on orientation: 0.10–0.25 north, 0.18–0.60 south, 0.12–0.40 east, and 0.12–0.35 west. Exterior walls with the same orientation within a building share a target range. The extreme window types extend coverage toward low and high glazing ratios.

Exterior shading is assigned by orientation, with occurrence probabilities of 60% north, 90% south, and 75% east and west. Shading depth is sampled within orientation-dependent ranges of 0.2–0.8,m for north-facing openings and 0.4–1.2,m for south-, east-, and west-facing openings. Interior shading is assigned independently to each window with a probability of 50%, and its coverage ratio is sampled from 0.3 to 1.0. This allows shading conditions to vary independently from window size and orientation.

Materials follow a hierarchy of building-level structural system, component-level construction layers, and surface-level properties. Small buildings mainly use light timber, cross-laminated timber (CLT), and brick. Larger buildings have higher proportions of concrete and CLT. Insulation materials are expanded polystyrene (EPS), mineral wool, and polyisocyanurate (PIR), with target proportions of 40%, 40%, and 20%.

The glazing systems include six types: single clear glazing (4%), double clear glazing (8%), double low-emissivity glazing with high solar gain (23%), double low-emissivity glazing with solar control (10%), triple low-emissivity glazing with high solar gain (40%), and triple low-emissivity glazing with solar control (15%). These combinations create variation in transmittance, solar gain, and heat capacity while retaining a small number of boundary cases.

The surrounding context is represented through sky-view factor (SVF) instead of explicit neighboring buildings. Each building is assigned to an open, suburban, urban, or sheltered context, with target proportions of 25%, 45%, 20%, and 10%. Wall SVF decreases from 0.80–0.95 in open settings to 0.10–0.30 in sheltered settings. Roof surfaces retain greater sky exposure. Wall SVF increases moderately with floor level to approximate reduced obstruction on higher floors.

Air exchange is the main dynamic operating boundary. Perimeter-zone ACH values follow four states: infiltration, mechanical ventilation, partial window opening, and full window opening. Representative ranges are 0.03–0.30, 0.50–1.50, 1.20–2.50, and 3.00–7.00 h^{-1}, respectively. Daytime schedules assign higher probability to window-opening states, while night-time schedules assign higher probability to infiltration and mechanical ventilation. Core zones use lower air-exchange ranges and omit the full-window-opening state. The final states are written as hourly ACH schedules for each zone.

Every valid building produces an HBJSON file. HBJSON includes zones, components, materials, boundary conditions, and operating schedules and serves as the common source for simulation and voxelization. Representative examples of the resulting building models are shown in Figure 3.3, illustrating the geometric diversity generated by the parametric framework.

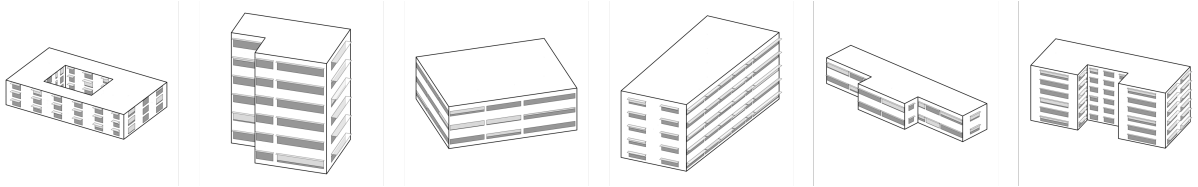


Figure 3.3: Representative building forms generated by the parametric framework. The examples illustrate variation in footprint geometry, story count, envelope configuration, and window distribution across the synthetic building data set.

3.3 Conversion from HBJSON to EnergyPlus

HBJSON serves as the source model for EnergyPlus simulations used to generate prediction labels. The building geometry, thermal zones, constructions, openings, boundary conditions, and schedules defined in HBJSON are converted into an EnergyPlus-compatible model through the Honeybee workflow. The resulting IDF preserves the spatial and physical properties required for zone-level thermal simulation. The HBJSON model is loaded through the Honeybee model interface, and the simulation period is defined from 1 January to 31 December. During conversion, zones, surfaces, openings, materials, constructions, boundary conditions, and schedules are mapped to the corresponding EnergyPlus objects. The generated IDF is then supplemented with additional environmental information required by the simulation setup.
For exterior surfaces, the sky-view factor stored in the building model is converted to a ground-view factor as

$$F_{\text{ground}} = 1 - F_{\text{sky}}. \tag{3.5}$$

The corresponding temperature schedules are derived from the EPW weather data, allowing the sky-view factor to influence long-wave radiative exchange and diffuse solar exposure.
The final IDF is combined with the prescribed schedules and meteorological inputs and simulated in EnergyPlus to generate hourly zone-level operative-temperature sequences for a standard meteorological year as shown in Figure 3.4. These simulated temperature sequences are used as the prediction labels for model training.

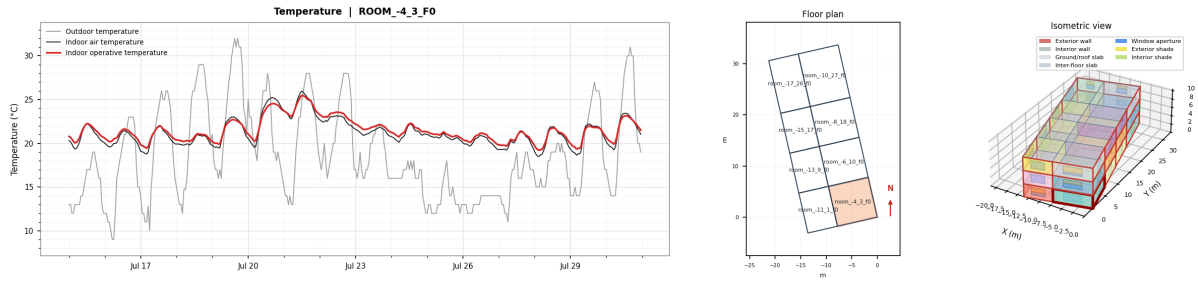


Figure 3.4: EnergyPlus simulation curves showing air temperature and operative temperature for a room from July 15 to July 30 of a standard year.

3.4 Sparse 3D building representation and zone indexing

3.4.1 HBJSON surface parsing and equivalent thermal properties

The geometry branch generates voxels directly from HBJSON component surfaces and millimeters are converted to meters. The parser includes eight component classes: exterior wall, window, roof, floor, intermediate slab, interior wall, exterior shading, and interior shading. Host walls are polygons with holes, with window openings subtracted from the wall region. Windows and shading devices are treated as panels with independent thickness, thermal properties, and semantic identifiers. For a wall or slab shared by adjacent zones, the parser records both the owning and neighboring zone while retaining a single global geometric object.

For a construction with N_L material layers, let d_l, λ_l, ρ_l, and c_l denote the thickness, thermal conductivity, density, and specific heat of layer l. Equivalent thickness and conductive resistance are

$$d_{\text{eq}} = \sum_{l=1}^{N_L} d_l, \qquad R_{\text{cond}} = \sum_{l=1}^{N_L} \frac{d_l}{\lambda_l}. \tag{3.6}$$

Equivalent thermal conductivity and overall thermal transmittance are

$$\lambda_{\text{eq}} = \frac{d_{\text{eq}}}{R_{\text{cond}}}, \qquad U = \frac{1}{R_{\text{si}} + R_{\text{cond}} + R_{\text{se}}}. \tag{3.7}$$

Volumetric heat capacity is the thickness-weighted capacity of the material layers:

$$C_{\text{vol}} = \frac{\sum_{l=1}^{N_L} \rho_l c_l d_l}{d_{\text{eq}}}. \tag{3.8}$$

Solar absorptance, solar transmittance, long-wave emissivity, boundary condition, and sky-view factor are extracted from the associated construction and surface properties.

3.4.2 Panel extrusion, occupancy tests, and conflict resolution

The main representation uses a regular voxel grid with edge length $L = 0.20\,\text{m}$. Each HBJSON face defines a local two-dimensional coordinate system from its origin, unit normal, and two in-plane basis vectors. The polygon is extruded symmetrically by its physical construction thickness d along the normal over $[-d/2, d/2]$, forming a thin 3D panel. The in-plane boundary is expanded by $d/2$ to cover discretisation gaps at component corners. Window holes remain removed from host walls, and window panels are voxelised independently.

Only voxels within the axis-aligned bounding box of each thin panel are enumerated. Intersection is tested from the projected voxel center and its normal distance to the face. Interior room air and empty exterior space are left inactive. Let s denote the signed normal distance between a candidate voxel center and the face plane. The overlap length between the panel and voxel along the normal direction is

$$\ell = \max\left[0, \min\left(\frac{d}{2}, s + \frac{L}{2}\right) - \max\left(-\frac{d}{2}, s - \frac{L}{2}\right)\right]. \tag{3.9}$$

The geometric fill ratio and component support ratio are

$$f = \frac{\ell}{L}, \qquad r = \frac{\ell}{\min(d, L)}. \tag{3.10}$$

The fill ratio f records the material share along the thickness direction and is stored for quality control. The support ratio r is measured relative to component thickness and determines activation. This separation retains thin windows and shading devices whose thickness is much smaller than the voxel edge length. The minimum support ratio is 0.15 for opaque components and 0.05 for windows and shading.

Corners, shared surfaces, and opening edges can place several candidate components in the same global voxel. For candidate component k in voxel i, the decision score is

$$S_{i,k} = r_{i,k} w_k, \tag{3.11}$$

and the attributes of the highest-scoring component are retained. The semantic weight w_k is 4 for

windows and interior shading, 3 for exterior shading, and 1 for walls, roofs, slabs, floors, and interior partitions. These weights retain openings and thin shading when they compete with host walls and produce one global record for each active position.
The overall conversion from the original building representation to the global voxel model is illustrated in Figure 3.5. As illustrated by the two example buildings, the resulting representation preserves the overall building form while retaining geometrically thin components such as windows and exterior shading devices.

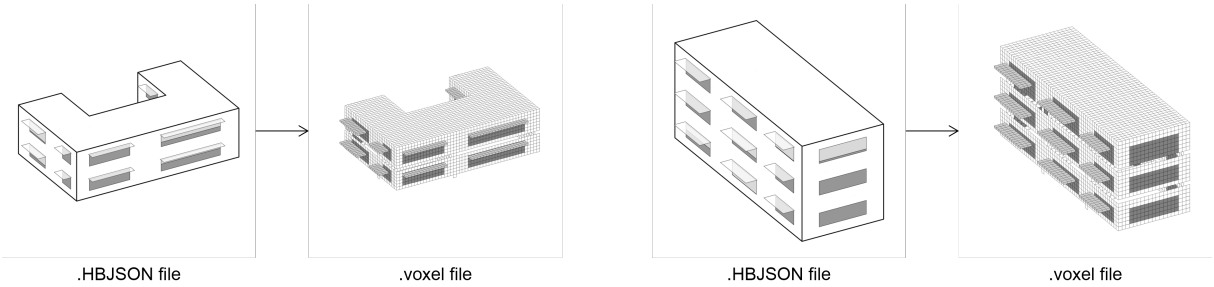


Figure 3.5: Schematic illustration of the voxelization workflow for two example buildings.

3.4.3 Voxel fields and zone indices

Voxelization produces `*_voxel.npz`. The `points` array has 15 columns:

$$[x, y, z, d, \lambda, U, C_{\text{vol}}, \alpha_{\text{sol}}, \tau_{\text{sol}}, \varepsilon_{\text{lw}}, F_{\text{sky}}, \text{bc}, q, m, f].$$

The first three columns give voxel-center coordinates. The next eight contain continuous thermophysical properties. The remaining columns contain boundary-condition code bc, component label q, face identifier m, and fill ratio f. Component label, face identifier, and fill ratio support conflict resolution and data auditing and are excluded from the default continuous neural input.
For active voxel i in building b, the default model feature is

$$\mathbf{x}_{b,i} = [\widetilde{\mathbf{a}}_{b,i} \| \mathbf{b}_{b,i}] \in \mathbb{R}^{12}, \tag{3.12}$$

where $\widetilde{\mathbf{a}}_{b,i} \in \mathbb{R}^8$ contains continuous physical properties standardized with training-set statistics, and $\mathbf{b}_{b,i} \in \mathbb{R}^4$ is a one-hot boundary-condition code. Table 3.1 lists the channels.

Table 3.1: Default sparse-voxel input features.

Channels	Feature group	Features
0–7	Continuous physical properties	Equivalent thickness (d_{eq}), equivalent thermal conductivity (λ_{eq}), overall thermal transmittance (U), volumetric heat capacity (C_{vol}), solar absorptance (α_{sol}), solar transmittance (τ_{sol}), long-wave emissivity (ε_{lw}), and sky-view factor (F_{sky}).
8–11	Boundary-condition one-hot code	`Outdoors`, `Ground`, `Surface`, and `Adiabatic`.

The NPZ file stores the ownership relationships between active voxels and thermal zones. Let $\mathcal{V}_b$ denote the active-voxel set of building b, and let $\mathcal{E}_b^{\text{zv}}$ denote the zone–voxel association set. The voxel set associated with zone z is

$$i \in \mathcal{V}_b \mid (z,i) \in \mathcal{E}_b^{\mathrm{ZV}}. \tag{3.13}$$

A voxel may be associated with multiple adjacent zones, for example when it represents a shared wall or slab. Each voxel is stored only once in the global voxel representation, avoiding duplicate spatial records.

3.5 Geometry-aware spatio-temporal prediction model

3.5.1 Overall model structure

The model consists of four consecutive stages: whole-building sparse three-dimensional geometric encoding, target-zone-driven extraction of whole-building spatial context, neighboring-zone dynamic-context aggregation, and causal autoregressive temporal prediction. For a target zone z in building b, the model first learns spatial tokens from the whole-building sparse voxel field and extracts local tokens belonging to the target zone according to the zone–voxel association. The local tokens act as Queries and retrieve target-zone-relevant spatial information from the global tokens of the same building, producing a static geometric representation of the zone. During dynamic prediction, currently known weather, solar, and operating conditions, together with the previous indoor state, form a temporal token. The zone geometric representation is projected into every time step, neighboring-zone attention then adds known surrounding-zone dynamics, and a causal decoder finally generates operative temperature at each hour.

The above computation can be written as

$$\mathbf{Z}_b = E_{\mathrm{cnn}}(\mathscr{X}_b), \tag{3.14}$$

$$\mathbf{Z}_{b,z} = R_z(\mathbf{Z}_b), \tag{3.15}$$

$$\mathbf{g}_{b,z} = F_{\mathrm{cross}}\left(\mathbf{Z}_{b,z}, \mathbf{Z}_b\right), \tag{3.16}$$

$$\mathbf{x}_{b,z,t}^{\mathrm{temp}} = F_{\mathrm{tok}}\left(\mathbf{d}_{b,z,t}, \widetilde{T}_{b,z,t-1}^{\mathrm{op}}, \Delta\widetilde{T}_{b,z,t-1}^{\mathrm{op}}, \Delta\mathbf{w}_t, m_t\right), \tag{3.17}$$

$$\mathbf{s}_{b,z} = F_g(\mathbf{g}_{b,z}), \tag{3.18}$$

$$\mathbf{x}_{b,z,t} = \mathbf{x}_{b,z,t}^{\mathrm{temp}} + \mathbf{s}_{b,z}, \tag{3.19}$$

$$\mathbf{c}_{b,z,t}^{\mathrm{nb}} = F_{\mathrm{attn}}^{\mathrm{nb}}\left(\mathbf{x}_{b,z,t}, \{\mathbf{d}_{b,j,t}^{\mathrm{nb}}, \mathbf{r}_{b,zj}\}_{j\in\mathscr{N}_b(z)}\right), \tag{3.20}$$

$$\mathbf{u}_{b,z,t} = F_{\mathrm{res}}^{\mathrm{nb}}\left(\mathbf{x}_{b,z,t}, \mathbf{c}_{b,z,t}^{\mathrm{nb}}\right), \tag{3.21}$$

$$\mathbf{h}_{b,z,t} = F_{\mathrm{causal}}\left(\mathbf{u}_{b,z,t}; \mathscr{C}_{b,z,t-1}^{\mathrm{KV}}\right), \tag{3.22}$$

$$\widehat{\mathbf{q}}_{b,z,t} = F_{\mathrm{out}}\left(\mathbf{h}_{b,z,t}\right). \tag{3.23}$$

where b, z, and t denote the building, target zone, and time step, respectively. $\mathscr{X}_b$ is the whole-building sparse-voxel input for building b; E_{cnn} is the whole-building sparse three-dimensional CNN; $\mathbf{Z}_b$ is the set of whole-building spatial tokens produced by it; R_z denotes target-zone token routing based on the zone–voxel association; $\mathbf{Z}_{b,z}$ is the set of local spatial tokens belonging to target zone z. F_{cross} denotes target-zone-to-whole-building cross-attention aggregation, and $\mathbf{g}_{b,z}$ is the resulting target-zone-conditioned geometric representation.

In the temporal branch, $\mathbf{d}_{b,z,t}$ denotes the weather, time, and target-zone operating conditions known at time t; $\widetilde{T}_{b,z,t-1}^{\mathrm{op}}$ denotes the previous indoor operative-temperature state available at the current prediction

step; $\Delta\widetilde{T}^{\mathrm{op}}_{b,z,t-1}$ denotes the operative-temperature change during the previous hour; $\Delta\mathbf{w}_t$ denotes the hourly change in selected weather variables; m_t indicates whether the current token belongs to the warm-up or autoregressive prediction stage. F_{tok} maps these variables to the base temporal token $\mathbf{x}^{\mathrm{temp}}_{b,z,t}$ $\mathbf{s}_{b,z} = F_g(\mathbf{g}_{b,z})$ is the static spatial condition obtained by projecting the geometric representation into the hidden dimension of the temporal decoder; $\mathbf{x}_{b,z,t}$ is the geometry-conditioned temporal token obtained by adding the base temporal token and the spatial condition. $\mathcal{N}_b(z)$ denotes the set of valid neighboring zones for target zone z; j denotes any one of these neighboring zones; $\mathbf{d}^{\mathrm{nb}}_{b,j,t}$ is the dynamic operating input available for neighboring zone j at time t; $\mathbf{r}_{b,zj}$ is the static spatial relationship between target zone z and neighboring zone j. $F^{\mathrm{nb}}_{\mathrm{attn}}$ aggregates neighboring-zone information from the current target-zone token and produces the neighboring-zone dynamic context $\mathbf{c}^{\mathrm{nb}}_{b,z,t}$ $F^{\mathrm{nb}}_{\mathrm{res}}$ denotes residual fusion that retains the target-zone representation, and its output $\mathbf{u}_{b,z,t}$ is the neighboring-zone-enhanced temporal token. F_{causal} is the causal temporal decoder, $\mathscr{C}^{\mathrm{KV}}_{b,z,t-1}$ denotes the Key/Value temporal cache retained by causal self-attention before the current time step, and $\mathbf{h}_{b,z,t}$ is the hidden state after causal decoding. F_{out} is the temperature output layer, $\widehat{\mathbf{q}}_{b,z,t} = [\widehat{q}_{0.1,b,z,t}, \widehat{q}_{0.5,b,z,t}, \widehat{q}_{0.9,b,z,t}]$ and represents the three operative-temperature quantile outputs. The median prediction $\widehat{q}_{0.5,b,z,t}$ is used as the point prediction and is fed back as the indoor state for the next time step during prediction.

It should be emphasized that $\widetilde{T}^{\mathrm{op}}_{b,z,t-1}$ is the state variable used by F_{tok} when constructing the current temporal token. Autoregressive feedback needs to change the temporal token at the next time step at first, then affects the next prediction through subsequent spatial conditioning, neighboring-zone aggregation, and causal temporal modeling.

The complete workflow of spatial encoding, neighboring-zone context aggregation, and causal autoregressive prediction is shown in Fig. 3.6.

3.5.2 Whole-building sparse encoding and target-zone cross-attention

The whole-building sparse three-dimensional CNN is executed once per building and outputs a set of tokens that retain their spatial distribution.

$$\mathbf{Z}_b = \left\{(\mathbf{z}_{b,j}, \mathbf{p}_{b,j})\right\}_{j=1}^{M_b}, \tag{3.24}$$

where $\mathbf{z}_{b,j} \in \mathbb{R}^{d_s}$ is the feature of the jth CNN token, $\mathbf{p}_{b,j} \in \mathbb{R}^3$ is its physical coordinate in meters, and M_b is the number of valid CNN tokens for building b. The spatial feature dimension in Our model is $d_s = 256$.

Target zone z inherits a set of local tokens from the same whole-building token field according to zone–voxel ownership:

$$\mathbf{Z}_{b,z} = \left\{(\mathbf{z}_{b,i}, \mathbf{p}_{b,i}) \mid i \in \mathscr{I}_{b,z}\right\}, \tag{3.25}$$

where $\mathscr{I}_{b,z}$ is the set of CNN-token indices inherited by target zone z after downsampling. The target-zone and whole-building representations are produced by one same encoder. They come from the same whole-building CNN output but have different selection ranges.

The local tokens of the target zone act as Queries, while the whole-building tokens act as Keys and Values:

$$\mathbf{Q}_{b,z} = W_Q\mathbf{Z}_{b,z}, \qquad \mathbf{K}_b = W_K\mathbf{Z}_b, \qquad \mathbf{V}_b = W_V\mathbf{Z}_b. \tag{3.26}$$

For local Query token i and whole-building Key token j, a metric distance bias based on the actual

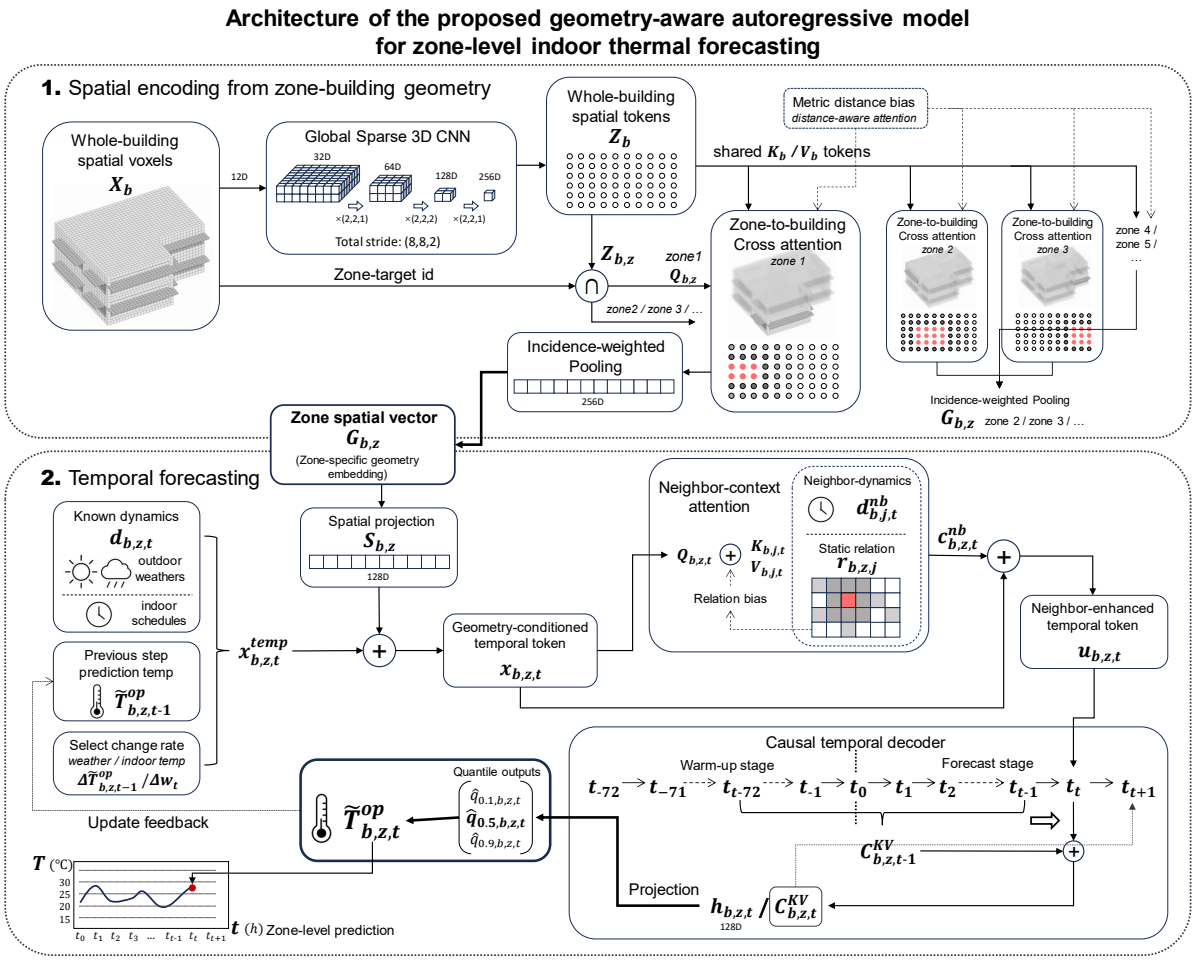


Figure 3.6: Overall architecture of the proposed geometry-aware autoregressive model for zone-level indoor thermal forecasting.

physical coordinates is further added:

$$B_{ij}^{\mathrm{sp}} = \mathrm{clip}\left(-\frac{\|\mathbf{p}_{b,i}-\mathbf{p}_{b,j}\|_2^2}{2\sigma_{\mathrm{sp}}^2}, -b_{\max}, 0\right), \tag{3.27}$$

where σ_{sp} is a learnable spatial scale, and $b_{\max}$ is the truncation magnitude of the distance bias. The cross-attention is therefore written as

$$\mathrm{CrossAttn}\left(\mathbf{Q}_{b,z}, \mathbf{K}_b, \mathbf{V}_b\right) = \mathrm{softmax}\left(\frac{\mathbf{Q}_{b,z}\mathbf{K}_b^{\top}}{\sqrt{d_h}} + \mathbf{B}_{b,z}^{\mathrm{sp}}\right)\mathbf{V}_b, \tag{3.28}$$

where d_h is the feature dimension of one attention head.

The distance bias increases the prior accessibility of spatially closer tokens but does not impose a hard spatial cutoff. Information from distant parts of the building can still be retrieved when it is content-relevant. After cross-attention, the target zone still corresponds to a set of updated local tokens. To obtain the single zone geometric representation required by the later temporal model, a weighted mean is calculated using the number of original target-zone voxels inherited by each CNN token:

$$\mathbf{g}_{b,z} = \frac{\sum_i \omega_{b,z,i}\,\widetilde{\mathbf{z}}_{b,z,i}}{\sum_i \omega_{b,z,i}} \in \mathbb{R}^{256}, \tag{3.29}$$

where $\widetilde{\mathbf{z}}_{b,z,i}$ is the target-zone token updated by cross-attention, and $\omega_{b,z,i}$ is the number of target-zone voxels inherited by this token during CNN downsampling. This weighted pooling compresses a variable number of target-zone tokens into a fixed 256-dimensional zone geometric representation.
Thus, each building requires only one sparse CNN execution, while different target zones form their own conditioned geometric vectors from the shared whole-building spatial representation.

3.5.3 Neighboring-zone dynamic context

The dynamic state of the target zone is also affected by operating boundaries and spatial connections of surrounding zones. This study does not use true or predicted neighboring-zone operative temperatures. It aggregates available exogenous neighboring-zone boundary conditions and static topology that are causally available at the prediction time.
For the neighbors of target zone z $j \in \mathcal{N}_b(z)$ their known dynamic representation at time t is

$$\mathbf{d}^{\mathrm{nb}}_{b,j,t} = \left[\mathrm{ACH}_{b,j,t}, E_{b,j,t}, O_{b,j,t}, H_{b,j,t}\right], \tag{3.30}$$

where E, O, and H denote the equipment, occupancy, and HVAC operating channels, these three channels are zero in the current free-running dataset but are retained in the model input interface.
The static relationship between the target zone and a neighboring zone is defined as

$$\mathbf{r}_{b,zj} = \left[\log(1+d_{b,zj}), \log\left(\frac{V_{b,j}}{V_{b,z}}\right), C_{b,zj}\right], \tag{3.31}$$

where $d_{b,zj}$ is the distance between the centroids of the two zones, $V_{b,j}/V_{b,z}$ is the volume ratio of the neighboring zone to the target zone, and $C_{b,zj}$ is the normalized contact ratio.
The static relationship $\mathbf{r}_{b,zj}$ participates in neighboring-zone aggregation through two paths. The first path jointly encodes neighboring-zone dynamics and spatial relationships as a neighboring-zone content representation:

$$\mathbf{e}^{\mathrm{nb}}_{b,zj,t} = F_{\mathrm{dyn}}\left(\mathbf{d}^{\mathrm{nb}}_{b,j,t}\right) + F_{\mathrm{rel}}\left(\mathbf{r}_{b,zj}\right), \qquad [\mathbf{k}_{b,zj,t} \| \mathbf{v}_{b,zj,t}] = W_{KV}\mathbf{e}^{\mathrm{nb}}_{b,zj,t}, \tag{3.32}$$

where F_{dyn} and F_{rel} are learnable projections of neighboring-zone dynamics and static relationships, respectively, and $\mathbf{k}_{b,zj,t}$ and $\mathbf{v}_{b,zj,t}$ are the Key and Value corresponding to neighboring zone j. Thus, even when two neighboring zones have the same dynamic inputs, their different distances, volume relationships, and contact relationships still produce different Key/Value representations.
The second path introduces the spatial relationship directly into the attention weights. The geometry-conditioned temporal token of the target zone $\mathbf{x}_{b,z,t}$ acts as the Query, and a relationship-dependent bias is added

$$B^{\mathrm{nb}}_{b,zj} = \mathrm{clip}\left(-\frac{d^2_{b,zj}}{2\sigma^2_{\mathrm{nb}}}, -b_{\max}, 0\right) + F_{\mathrm{topo}}\left(\log\frac{V_{b,j}}{V_{b,z}}, C_{b,zj}\right), \tag{3.33}$$

where σ_{nb} is a learnable neighboring-zone distance scale, the first term is the distance bias, and F_{topo} is a learnable topological bias generated from the volume ratio and contact ratio.
The aggregated neighboring-zone dynamic context can then be abbreviated as

$$\mathbf{c}^{\mathrm{nb}}_{b,z,t} = \mathrm{Attn}_{\mathrm{nb}}\left(W_Q\mathbf{x}_{b,z,t}, \{\mathbf{k}_{b,zj,t}, \mathbf{v}_{b,zj,t}\}_{j\in\mathcal{N}_b(z)}; \mathbf{B}^{\mathrm{nb}}_{b,z}\right). \tag{3.34}$$

The geometric temporal representation of the target zone, $\mathbf{x}_{b,z,t}$, is enhanced through a residual path. Its core information flow is

$$\mathbf{x}_{b,z,t} + W_O \mathbf{c}^{\mathrm{nb}}_{b,z,t} \longrightarrow \mathbf{u}_{b,z,t}, \tag{3.35}$$

where W_O is the output projection of neighboring-zone attention, and $\mathbf{u}_{b,z,t}$ is the neighboring-zone-enhanced temporal token after standard residual updating, normalization, and feed-forward transformation. When the target zone has no valid neighbors, this branch reduces to retaining only the target-zone geometric-conditioned temporal representation.

3.5.4 Temporal tokens and causal autoregressive decoding

A base temporal token is first constructed for the target zone at each time step. Define

$$\mathbf{d}_{b,z,t} \in \mathbb{R}^{17} \tag{3.36}$$

as the absolute dynamic input at time t, containing: six physical weather variables: outdoor dry-bulb temperature, relative humidity, global horizontal radiation, direct normal radiation, diffuse horizontal radiation, and wind speed; four periodic time encodings: sine and cosine encodings of the hour and day of the year; three solar-position variables: the sine and cosine of solar azimuth and the sine of solar altitude; and four operating channels for the target zone: ACH, equipment, occupancy, and HVAC. Hourly differences are further calculated for the first six physical weather variables $\Delta\mathbf{w}_t \in \mathbb{R}^6$. The currently available indoor-temperature change is also defined as

$$\Delta\widetilde{T}^{\mathrm{op}}_{b,z,t-1} = \widetilde{T}^{\mathrm{op}}_{b,z,t-1} - \widetilde{T}^{\mathrm{op}}_{b,z,t-2}. \tag{3.37}$$

The base temporal token is constructed as

$$\mathbf{x}^{\mathrm{temp}}_{b,z,t} = F_{\mathrm{tok}}\left(\left[\widetilde{T}^{\mathrm{op}}_{b,z,t-1} \| \Delta\widetilde{T}^{\mathrm{op}}_{b,z,t-1} \| \mathbf{d}_{b,z,t} \| \Delta\mathbf{w}_t \| m_t\right]\right) \in \mathbb{R}^{d_t}, \tag{3.38}$$

where m_t is the warm-up/prediction-stage indicator. With indoor-temperature change and weather differences included, the raw vector entering the token builder contains $1+1+17+6+1=26$ channels. F_{tok} maps this vector to a temporal hidden dimension of $d_t = 128$.
The output of the spatial branch, $\mathbf{g}_{b,z} \in \mathbb{R}^{256}$ is first mapped to the same temporal hidden dimension through a learnable projection:

$$\mathbf{s}_{b,z} = F_g(\mathbf{g}_{b,z}) \in \mathbb{R}^{128}. \tag{3.39}$$

Because the geometric condition remains constant within the same prediction segment, $\mathbf{s}_{b,z}$ needs to be calculated only once for each target zone and is added to all time steps of that zone:

$$\mathbf{x}_{b,z,t} = \mathbf{x}^{\mathrm{temp}}_{b,z,t} + \mathbf{s}_{b,z}. \tag{3.40}$$

The present study refers to $\mathbf{x}_{b,z,t}$ as the geometry-conditioned temporal token. The neighboring-zone module defined in the previous subsection then further generates $\mathbf{u}_{b,z,t}$, and sends it to the causal temporal decoder.
The temporal decoder uses causal self-attention, so the hidden state at time t can access only the current and previous temporal tokens. Each layer uses standard Pre-LayerNorm, self-attention residual, and feed-forward-network residual connections. This study uses a temporal hidden dimension of 128, four

attention heads, and four causal decoder layers.
The model first processes the warm-up sequence of length $W = 72$ h with parallel causal operations and establishes a Key/Value cache at each decoder layer. The hidden state at the end of the warm-up sequence is used to estimate the initial indoor state at the start of the prediction period:

$$\widehat{T}^{\mathrm{op}}_{b,z,0} = F_{T_0}\left(\mathbf{h}_{b,z,W}\right), \tag{3.41}$$

where F_{T_0} is the initial-temperature prediction head. The evaluation uses cold-start inputs so that the warm-up stage does not require true historical indoor temperatures. During training, cold-start and true historical states are mixed according to the strategy described in Section 3.8.
After entering the prediction stage, the model performs autoregressive decoding one time step at a time. After the three quantile outputs are obtained at step t, the median $\widehat{q}_{0.5,b,z,t}$ is used as the indoor state at the next time step:

$$\widetilde{T}^{\mathrm{op}}_{b,z,t} = \widehat{q}_{0.5,b,z,t}, \qquad \widehat{q}_{0.5,b,z,t} \rightarrow \mathbf{x}^{\mathrm{temp}}_{b,z,t+1}. \tag{3.42}$$

True future operative temperatures are not used during the prediction period. The previous prediction will participate in constructing the base temporal token at the next time step. It then forms a new causal input together with the static geometric condition and neighboring-zone dynamics at the corresponding time.
During prediction, each causal layer reuses the previously established Key/Value cache and uses a sliding temporal window of length W to limit the maximum historical context.

3.5.5 Output layer and training objective

The hidden state after causal decoding is mapped by the output layer to three operative-temperature quantiles:

$$\widehat{\mathbf{q}}_{b,z,t} = \left[\widehat{q}_{0.1,b,z,t}, \widehat{q}_{0.5,b,z,t}, \widehat{q}_{0.9,b,z,t}\right]. \tag{3.43}$$

All point-prediction metrics and autoregressive feedback in this study use the median output $\widehat{q}_{0.5}$
The training objective is based on quantile regression. For quantile τ, the pinball loss is defined as

$$\mathscr{L}_{\tau}\left(y, \widehat{y}_{\tau}\right) = \max\left[\tau(y - \widehat{y}_{\tau}), (\tau - 1)(y - \widehat{y}_{\tau})\right]. \tag{3.44}$$

To constrain temperature level, hourly changes, larger prediction errors, warm-up initialization, and quantile ordering simultaneously, the actual training objective is further written as

$$\mathscr{L} = \mathscr{L}_{\mathrm{pin}} + \lambda_{\mathrm{tr}}\mathscr{L}_{\mathrm{tr}} + \lambda_{\mathrm{focal}}\mathscr{L}_{\mathrm{focal}} + \lambda_{0}\mathscr{L}_{T_0} + \lambda_{\mathrm{cross}}\mathscr{L}_{\mathrm{cross}}, \tag{3.45}$$

where $\mathscr{L}_{\mathrm{pin}}$ is the pinball loss for the three quantiles; $\mathscr{L}_{\mathrm{tr}}$ constrains the hourly change between the predicted median and the true temperature; $\mathscr{L}_{\mathrm{focal}}$ increases the training penalty for larger temperature errors; $\mathscr{L}_{T_0}$ constrains the initial temperature recovered during warm-up; $\mathscr{L}_{\mathrm{cross}}$ suppresses quantile crossing. These additional losses use true temperatures to calculate supervised errors only during training and provide no additional future ground-truth temperatures to the prediction network.

3.6 Flattened representation baseline

A complete manually aggregated representation provides the main baseline, referred to as the **Flattened baseline**. Building- and zone-level descriptors encode geometry and thermal conditions through area, volume, WWR, position, materials, and related summary information. The baseline uses the same known dynamic boundaries, neighboring-zone exogenous inputs, and causal autoregressive temporal framework as the 3D model.
The comparison therefore concentrates on the representation of static building space. The proposed model learns from the spatial distribution of 3D components, while the baseline uses selected low-dimensional descriptors. This focused comparison tests the practical value of detailed spatial information for prediction on unseen buildings, controlled parameter response, and dynamic thermal behavior.

3.7 Dynamic boundaries and information availability

Dynamic inputs are divided by physical origin into outdoor weather, solar and radiative effects, air exchange, periodic time information, and optional operating states. Static shading and opening geometry are represented mainly by the geometry encoder, while time-varying weather, solar response, and prescribed ACH enter the temporal input. In current dataset, equipment, occupancy, and HVAC channels remain zero. These channels are retained in the model interface for architectural consistency and future extension, but their non-zero operational effects are not evaluated in the present experiments.
The availability of different information during the warm-up period and prediction horizon is shown in Table 3.2. During training, the model randomly switches between true historical states and cold-start inputs to improve adaptation to different initialization conditions. In cold start, historical indoor temperature from the target building is not used. The validation and test sets use this cold-start setting consistently and therefore do not require indoor-temperature measurements.

Table 3.2: Availability of different information during warm-up and the prediction horizon.

Input information	**Warm-up period**	**Prediction horizon**
Static building and zone geometry	Available	Available
Weather and solar conditions	Available	Available
Prescribed ACH schedule	Available when the variables are known	Available when the schedule is known
Equipment, occupancy, and HVAC	Zero in the dataset	Zero in the dataset
Target-zone temperature	Mixed during training; cold-start proxy in the evaluation	Autoregressively predicted by the model
neighboring-zone dynamics and topology	Available when the variables are known	Available when the variables are known
True neighboring-zone temperature	Not used	Not used

3.8 Training and inference protocol

The dataset is divided at the building level into 2,000 training buildings, 400 validation buildings, a separate parameter-validation set, and 200 test buildings. The same target definition, input standardization, cold-start setting, and evaluation procedures are used throughout the study. Statistics for standardizing continuous variables are fitted only on the training buildings and fixed for all subsequent datasets.

During training, buildings are the sampling units. The prediction start, prediction segment, and up to 10 target zones are selected randomly within each building. The warm-up window is fixed at 72 h. Training forecast segments are sampled from May to October, with segment lengths randomly selected from 72–240 h to cover different forecast horizons and dynamic states. For all controlled evaluation experiments, a fixed building set, all valid zones, and three fixed 168 h forecast windows are used to ensure that results are compared on exactly the same samples.

For architectures with sparse three-dimensional geometric encoding, the whole-building sparse CNN is executed only once per building. All target zones then share the global spatial-token field of that building while separately performing target-zone context extraction and temporal prediction. During the prediction period, the autoregressive model uses the median output $q_{0.5}$ as the target-zone state at the next time step and does not access true future temperatures. The neighboring-zone module uses only external dynamics and static topology available at the prediction time.

Table 3.3 summarizes the key training and inference settings.

Table 3.3: Key training and inference configuration.

Setting	**Configuration**
Warm-up window	72 h
Training forecast length	72–240 h, randomly sampled
Target zones during training	Maximum 10 per building
Target zones during inference	All valid zones
Maximum neighbors	15 per target zone
Training neighbor dropout	Maximum 5; at least 1 valid neighbor retained
Training cold-start probability	0.75
Validation/test cold-start probability	1.00
Optimizer	AdamW
Initial learning rate	1×10^{-4}
Weight decay	1×10^{-4}
Learning-rate schedule	5-epoch linear warm-up + cosine annealing
Minimum learning rate	1×10^{-6}
Gradient clipping	1.0
Primary training range	100 epochs
Checkpoint criterion	Minimum random-validation loss
Output quantiles	0.1, 0.5, 0.9
Point prediction	$q_{0.5}$
Implementation	PyTorch

3.9 Evaluation framework

The evaluation covers five capabilities required for rapid thermal-environment screening:

(1) **Temperature accuracy on unseen buildings:** the ability to generate useful zone-level operative-temperature trajectories across buildings;

(2) **Building-parameter response:** agreement with EnergyPlus under changes in WWR, shading, orientation, glazing, and envelope construction;

(3) **Short-term dynamic response:** representation of hourly warming, cooling, and rapid transitions;

(4) **High-temperature states and climate domain:** temperature-band accuracy, and performance under climate shifts; and

(5) **Computational efficiency:** reduction in repeated-scenario cost through static geometry caching.

Point predictions use $q_{0.5}$. Overall accuracy is measured mainly by mean absolute error (MAE) and mean bias:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{T}_i - T_i|, \qquad \text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(\widehat{T}_i - T_i). \tag{3.46}$$

The analysis also reports the proportion of absolute errors above 2.0,$^\circ$C, long-horizon error accumulation, and the balance of prediction performance across zones.
Static-parameter experiments use 18 reference buildings. Each variant changes one target parameter. The current data contain 33 static parameter levels and $33 \times 18 = 594$ parameter variants.
Dynamic-response analysis evaluates changes between adjacent time steps. Climate-transfer experiments apply a model trained only in Gothenburg directly to Copenhagen, Munich, and Tromsø without target-climate fine-tuning. The efficiency experiment selects 50 test buildings and evaluates ten independent dynamic scenarios for each building.

Chapter 4

Results

4.1 Training and validation performance

The proposed model and the Flattened baseline are trained under the same data split, dynamic boundary conditions, temporal prediction setting, and checkpoint-selection criterion. This section reports observations from the Training and Validation sets only. Results on fully held-out Test buildings are presented separately in Section 4.4.
With 2,000 training buildings, the proposed model reaches a minimum Validation loss of 0.567 and a Validation MAE of 0.80 °C. The Flattened baseline reaches a Validation loss of 0.674 and a Validation MAE of 0.94 °C. The proposed model also has a lower share of absolute errors above 2.0 °C, a lower MAE after the first 48 forecast hours, and a lower zone-level P90 MAE. These results show a consistent validation advantage when detailed three-dimensional spatial information is used instead of manually aggregated static descriptors.

Table 4.1: Training and validation performance of the proposed model and the Flattened baseline with 2,000 training buildings.

Model	**Train loss**	**Train MAE** (°C)	**Val. loss**	**Val. MAE** (°C)	**Bias** (°C)	$\lvert e\rvert > 2.0$°C (%)	**48 h+ MAE** (°C)	**Zone P90** (°C)
Proposed model	0.299	0.47	0.567	0.80	-0.16	7.4	0.80	0.79
Flattened baseline	0.629	0.86	0.674	0.94	-0.01	10.2	0.93	0.92

4.2 Response to static building-parameter perturbations

The static-parameter experiments test whether the two models reproduce temperature responses caused by controlled changes in building geometry, envelope properties, and exposure conditions. The data contain 18 reference buildings. Each variant is generated from the same reference HBJSON while changing only one target factor. The experiment covers 33 static parameter levels and $33 \times 18 = 594$ variants. ACH and other dynamic operating boundaries are excluded from this analysis.
For a reference building, the EnergyPlus and model responses to parameter level p are

$$\Delta T_{EP,i}(p) = T_{EP,i}(p) - T_{EP,i}(\text{ref}), \tag{4.1}$$

$$\Delta T_{m,i}(p) = \widehat{T}_{m,i}(p) - \widehat{T}_{m,i}(\text{ref}). \tag{4.2}$$

The EnergyPlus mean absolute response magnitude is

$$A_{EP}(p) = \frac{1}{N}\sum_i |\Delta T_{EP,i}(p)|, \tag{4.3}$$

and the model response error is

$$E_{\text{resp},m}(p) = \frac{1}{N}\sum_i |\Delta T_{m,i}(p) - \Delta T_{EP,i}(p)|. \tag{4.4}$$

A lower E_{resp} indicates closer agreement with the EnergyPlus response to the same controlled change. Across the complete static-parameter set, absolute-temperature MAE is 0.64 °C for the proposed model and 0.82 °C for the Flattened baseline. These errors are lower than those observed in some of the other evaluation sets because the static-parameter set contains a larger proportion of mild perturbations. Table 4.2 reports factor-level response errors together with the corresponding EnergyPlus response magnitude. WWR produces the largest mean absolute response, followed by glazing, exterior shading, and orientation. Material and construction-thickness changes produce smaller mean responses.

Table 4.2: EnergyPlus mean absolute response magnitude and parameter-response errors of the proposed model and the Flattened baseline for major static building factors.

Factor	A_{EP}	**Proposed model** (°C)	**Flattened baseline**
WWR	3.57	0.96	0.86
Glazing	1.63	0.52	0.43
Exterior shading	1.19	0.43	0.61
Orientation	0.92	0.48	0.68
Interior shading	0.29	0.29	0.38
Ext.-wall insulation thickness	0.17	0.39	0.26
Ext.-wall structure thickness	0.14	0.31	0.15
SVF	0.14	0.14	0.23
Ext.-wall structure material	0.09	0.11	0.10
Ext.-wall insulation material	0.04	0.08	0.09

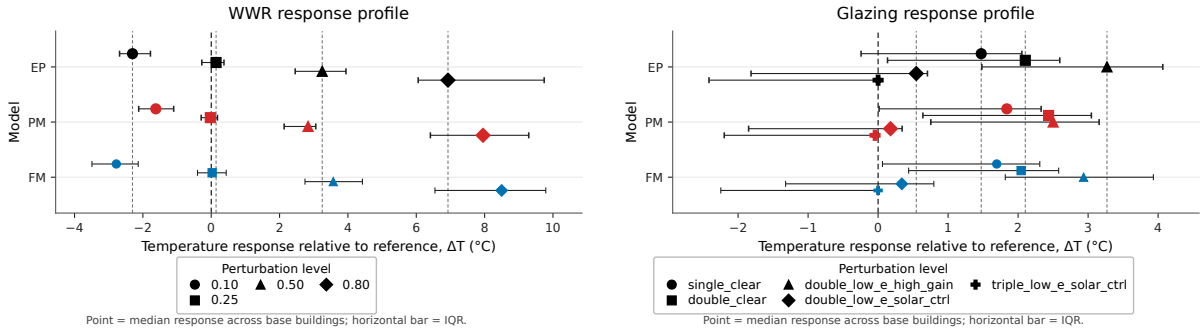
The lowest factor-level response error varies between the two models. The Flattened baseline performs well for WWR and glazing, both of them can be described directly by compact scalar variables. The proposed model produces lower response errors for exterior shading and orientation, where component direction, position, and spatial combination play a larger role. For weak perturbations such as material-thickness changes, the response error can approach or exceed the EnergyPlus response magnitude, so these cases mainly indicate response stability.

4.2.1 Window-to-wall ratio and glazing

WWR and glazing produce some of the strongest static responses in the perturbation experiments. WWR shows a clear monotonic relationship with zone temperature across the tested range. Based on the mean signed response relative to the paired reference cases, the EnergyPlus response changes from -2.272 °C at WWR 0.10 to $+7.885$ °C at WWR 0.80. Both learned models reproduce the same overall progression from cooling at low WWR to warming at high WWR, as shown in Figure 4.1a. The proposed model generally shows a smaller response amplitude, whereas the Flattened baseline produces

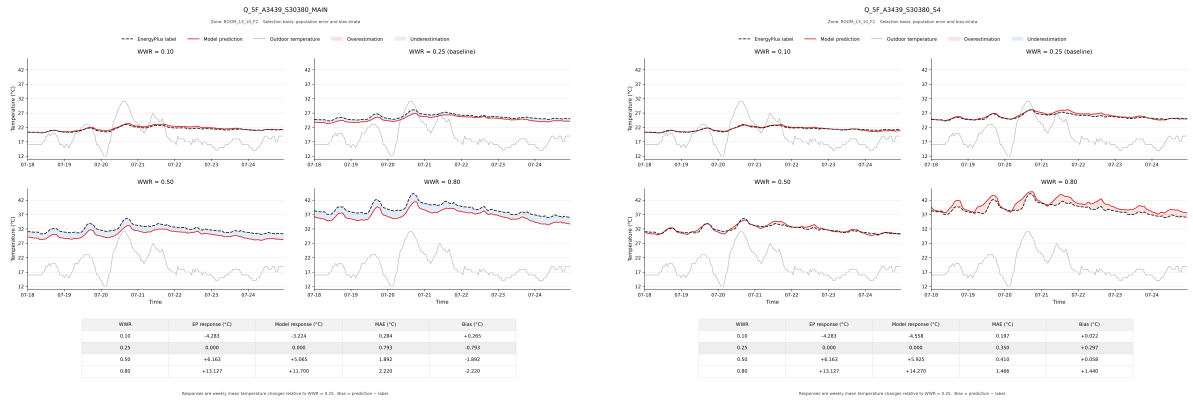
a stronger response at the higher WWR levels. Figure 4.2 provides one representative zone-level example of this contrast.

Glazing also produces a comparatively large thermal response, although the perturbation is categorical rather than ordered continuously. Across the tested glazing systems, the EnergyPlus mean absolute response is 1.6319 °C. Both models reproduce the broad separation between glazing configurations associated with stronger warming and those associated with reduced or negative temperature responses (Figure 4.1b). For this factor, the Flattened baseline gives a lower overall response error than the proposed model.



(a) Window-to-wall ratio (b) Glazing system

Figure 4.1: Static temperature-response profiles for WWR and glazing perturbations. EP denotes EnergyPlus, PM the proposed model, and FM the Flattened baseline. Points show the median temperature response across paired base buildings and horizontal bars show the interquartile range. The black dashed line indicates zero response, while the gray dashed lines indicate the corresponding EnergyPlus median responses for the tested perturbation levels.



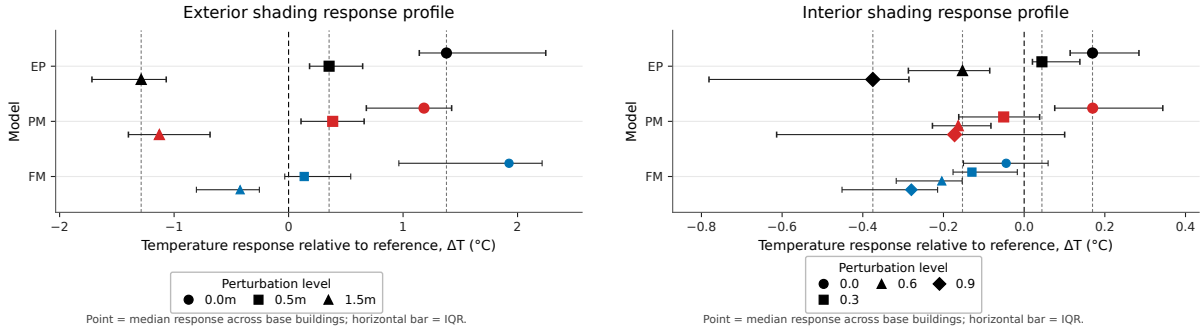
(a) Proposed model (b) Flattened baseline

Figure 4.2: Representative zone-level WWR-response example for the proposed model and the Flattened baseline. Each panel shows the predicted and EnergyPlus temperature trajectories under different WWR perturbation levels relative to the same baseline case.

4.2.2 Exterior and interior shading

Shading perturbations produce directional responses that depend on the presence, depth, and spatial configuration of the shading elements. For exterior shading, removing the shading produces a warming response, increasing the shade depth progressively shifts the response toward cooling. The mean signed EnergyPlus response changes from $+1.604\,^{\circ}\mathrm{C}$ with no exterior shading to $-1.420\,^{\circ}\mathrm{C}$ at a shade depth of 1.5 m. Both models reproduce this directional change (Figure 4.3a). At the strongest shading level,

the proposed model gives a larger cooling response than the Flattened baseline and is closer to the EnergyPlus response.
Interior shading shows the same general transition from warming when shading is absent toward cooling as the shading level increases (Figure 4.3b). Both reproduce the cooling direction at the higher shading levels, the proposed one gets closer response to reference.



(a) Exterior shading (b) Interior shading

Figure 4.3: Static temperature-response profiles for exterior and interior shading perturbations. EP denotes EnergyPlus, PM the proposed model, and FM the Flattened baseline. Points show the median temperature response across paired base buildings and horizontal bars show the interquartile range. The black dashed line indicates zero response, while the gray dashed lines indicate the corresponding EnergyPlus median responses for the tested perturbation levels.

4.2.3 Other static factors

The remaining static perturbations generally produce smaller aggregate temperature responses than WWR, glazing, and shading. Orientation is the clearest example in which the mean signed response is small despite non-negligible responses at the building and zone levels. For rotations of $+90°$ and $-90°$, EnergyPlus gives mean signed responses of -0.034 and $+0.114\,°C$, respectively, while the corresponding mean absolute response reaches $0.9173\,°C$. The proposed model gives an orientation-response error of $0.48\,°C$, compared with $0.68\,°C$ for the Flattened baseline.
Perturbations to window position, floor index, centroid radius, material properties, construction thickness, and sky-view-related parameters produce smaller or more spatially heterogeneous aggregate responses. For these factors, positive and negative changes across buildings, zones, and hours can partially cancel in the signed averages. Both architectures retain the principal response direction for most tested perturbations, while the magnitude of the response varies by factor. Across the complete static-response experiment, the proposed model retains the lower absolute-temperature prediction error.

4.3 Temporal-response fidelity under changing thermal conditions

Temperature-level error alone does not describe whether the predicted trajectory reproduces the direction and magnitude of short-term thermal changes. The temporal-response analysis here evaluates consecutive hourly temperature changes for the proposed model and the Flattened baseline.
For adjacent times within the same building, zone, and continuous prediction window,

$$\Delta T_t = T_t - T_{t-1}, \tag{4.5}$$
$$\Delta \widehat{T}_t = \widehat{T}_t - \widehat{T}_{t-1}. \tag{4.6}$$

Hourly temperature-change error is

$$\mathrm{MAE}_{\Delta T} = \frac{1}{N}\sum_{t=1}^{N}\left|\Delta \widehat{T}_t - \Delta T_t\right|. \tag{4.7}$$

Pearson correlation $r_{\Delta T}$ measures agreement in the temporal change pattern. For steps with $|\Delta T| > 0.1\,^{\circ}\mathrm{C/h}$, direction accuracy measures whether predicted warming and cooling have the same sign as the EnergyPlus reference. Dynamic-response magnitude is measured by

$$G = \frac{\sum_t \Delta T_t \Delta \widehat{T}_t}{\sum_t (\Delta T_t)^2}. \tag{4.8}$$

A gain of $G = 1$ matches the overall EnergyPlus response amplitude, while $G < 1$ indicates a certain degree of attenuation.

4.3.1 Overall and strong-transition response fidelity

The two architectures are evaluated on the same Test buildings. Table 4.3 reports the overall temporal-response metrics and the subset of Strong transitions, defined here as $|\Delta T| \geq 0.6\,^{\circ}\mathrm{C/h}$. The proposed model has a lower overall $\mathrm{MAE}_{\Delta T}$, a higher change correlation, and a higher direction accuracy. The difference between the two architectures increases during Strong transitions.
Across the overall test set, the proposed model shows lower hourly-change error, higher change correlation, higher direction accuracy, and a response gain closer to unity than the Flattened baseline. These differences increase during Strong transitions, particularly for hourly-change error and response gain. These observation as also illustrated in Figure 4.4, where the proposed model remains closer to the one-to-one response over the tested range.

Table 4.3: Overall and strong-transition temporal-response fidelity of the proposed model and the Flattened baseline on the independent Test buildings.

Model	**Temp. MAE**	**Overall**				**Strong transitions**			
	(°C)	$\mathrm{MAE}_{\Delta T}$ (°C/h)	$r_{\Delta T}$	**Dir. acc.** (%)	**Gain**	$\mathrm{MAE}_{\Delta T}$ (°C/h)	$r_{\Delta T}$	**Dir. acc.** (%)	**Gain**
Proposed model	0.92	0.087	0.8720	94.01	0.8132	0.242	0.8528	99.09	0.8154
Flattened baseline	1.03	0.114	0.8163	91.07	0.7502	0.343	0.7902	97.86	0.6978

4.3.2 Zone-level paired response comparison

A zone-level paired comparison is used to examine how consistently the proposed model reduces hourly-change error across individual thermal zones. A zone is defined by a unique combination of building and zone identifiers,

$$z = (\text{building_id}, \text{zone_name}), \tag{4.9}$$

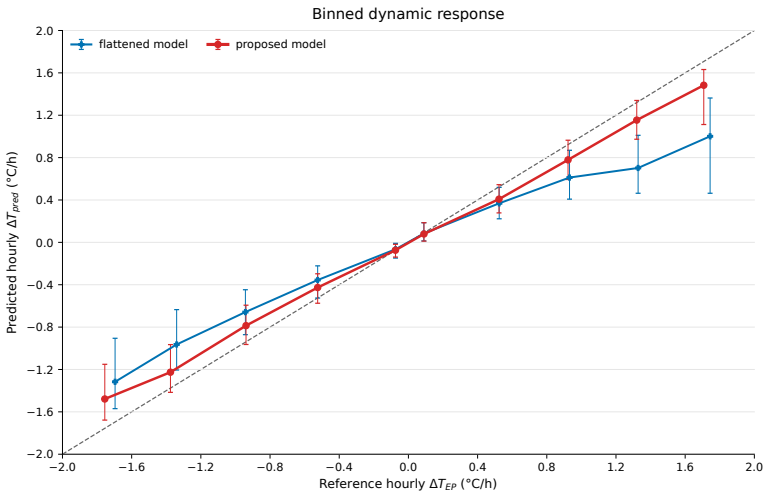


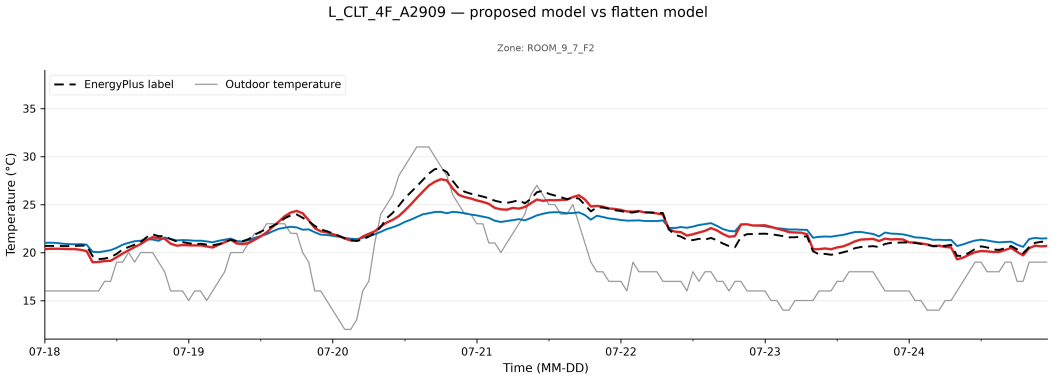
Figure 4.4: Binned comparison between the reference hourly temperature change and the predicted hourly response of the proposed model and the Flattened baseline. The dashed diagonal indicates a one-to-one response. The reported response gain summarizes the retained amplitude of the hourly thermal response.

For each valid zone-hour pair, the absolute hourly-change errors of the two architectures are compared. A win is assigned to the proposed model when

$$\left|\Delta T_{\text{Proposed}} - \Delta T_{\text{EP}}\right| < \left|\Delta T_{\text{Flattened}} - \Delta T_{\text{EP}}\right|. \tag{4.10}$$

The zone-level win rate is then calculated as the proportion of valid hourly transitions within a zone for which this condition is satisfied.

As shown in Figure 4.6. The median zone-level win rate of the proposed model is 58.9%. For the more restrictive $|\Delta T_{\text{EP}}| > 0.8\,^{\circ}\text{C/h}$ subset, the median zone-level win rate increases to 73.4%. The distribution of d_z also shifts further toward positive values for the high-change subset, indicating a larger zone-level error reduction during stronger thermal transitions. Figure 4.5 shows one representative zone-level example in which the proposed model more closely follows the EnergyPlus trajectory during a pronounced warming episode, while the Flattened baseline underestimates the response amplitude.



Color	Architecture	MAE (°C)	Bias (°C)	|Error| > 2°C
	proposed model	0.435	-0.062	0.0%
	flatten model	1.065	-0.169	11.3%

Figure 4.5: Representative zone-level dynamic prediction example comparing the proposed model and the Flattened baseline. The dashed black line denotes the EnergyPlus reference, the red line denotes the proposed-model prediction, the blue line denotes the Flattened-baseline prediction, and the gray line shows the outdoor temperature.

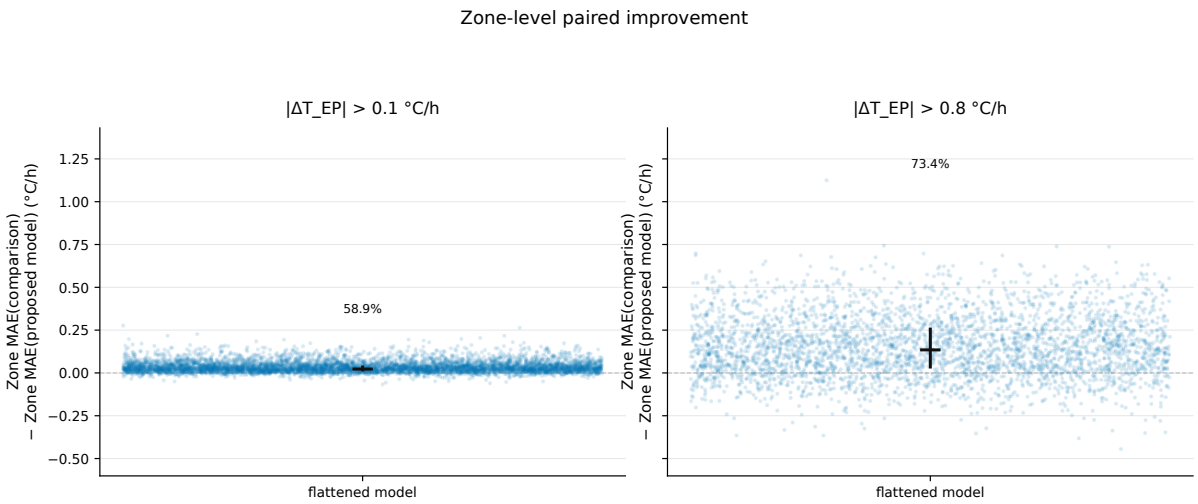


Figure 4.6: Zone-level paired comparison of hourly temperature-change prediction between the proposed model and the Flattened baseline. The left panel includes transitions with $|\Delta T_{\mathrm{EP}}| > 0.1\,^{\circ}\mathrm{C/h}$, while the right panel includes transitions with $|\Delta T_{\mathrm{EP}}| > 0.8\,^{\circ}\mathrm{C/h}$. Positive paired differences indicate lower zone-level $\mathrm{MAE}_{\Delta T}$ for the proposed model.

4.3.3 Recursive error accumulation over the forecast horizon

The temporal comparison is further extended across the full 168-hour recursive forecast to test whether prediction error systematically accumulates with forecast length. For each 24-hour endpoint, the temperature MAE is expressed relative to the MAE during the initial 0–24 h period.

As shown in Figure 4.7, neither architecture exhibits a monotonic increase in temperature MAE over the forecast horizon. The relative error changes remain close to zero for both models throughout the seven-day prediction window. The proposed model shows smaller variation around the initial-period error level, while the Flattened baseline exhibits larger positive and negative deviations at later forecast horizons.

The largest increase occurs around 72 h for both architectures, after which the relative MAE decreases again. At 144 and 168 h, neither model shows an increase above its initial 0–24 h error level. These results show no systematic recursive error accumulation over the evaluated 168-hour horizon, with the proposed model maintaining the more stable error profile across successive forecast periods.

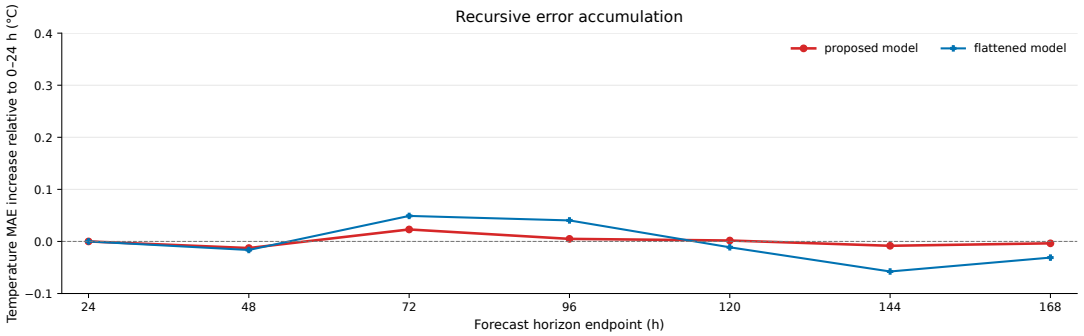


Figure 4.7: Change in temperature MAE across the 168-hour recursive forecast relative to the initial 0–24 h period. Values close to zero indicate that prediction error remains comparable to the initial forecast period, while positive values indicate increased error.

4.4 Generalization across unseen buildings and climates

The generalization analysis separates two conditions. The first tests new buildings generated from the same climate source as the Training set. The second changes the climate source while keeping the trained model fixed.

4.4.1 Unseen buildings under the source climate

The standard Test set contains 200 buildings that are fully excluded from training and checkpoint selection. All valid zones are evaluated under the common cold-start protocol using three fixed 168 h forecast windows. Under this setting, the proposed model achieves an operative-temperature MAE of 0.92 °C, compared with 1.03 °C for the Flattened baseline. The proposed model also has the lower absolute-temperature error when the building geometry is fully unseen.

Table 4.4: Cross-building generalization on the independent Test set under the source-climate condition.

Model	**Test MAE** (°C)	**Bias** (°C)	$\lvert e\rvert > 2$°C (%)	**48 h+ MAE** (°C)	**Zone P90** (°C)
Proposed model	0.92	-0.19	9.4	0.93	1.72
Flattened baseline	1.03	-0.15	12.4	1.03	1.89

The temporal-response results in Table 4.3 use the same held-out buildings and provide a complementary view of generalization. The absolute-temperature result establishes cross-building transfer of the predicted temperature level, while the dynamic metrics show whether the hourly thermal response remains consistent after transfer to unseen geometry.

4.4.2 Transfer to different climate sources

The model trained on Gothenburg data is applied directly to Copenhagen, Munich, and Tromsø without target-climate fine-tuning. The main comparison uses summer prediction windows consistent with the season represented during training. To examine whether additional training diversity improves climate transfer, the model trained on 2,000 Gothenburg buildings is compared with the same architecture trained on 4,000 buildings.

Figure 4.8 summarizes the summer transfer results. For both training-set sizes, transfer performance remains close to the Gothenburg baseline in Copenhagen but deteriorates more clearly in Munich and especially Tromsø. Increasing the training set from 2,000 to 4,000 buildings improves MAE in Gothenburg, Copenhagen, and Munich, while providing little benefit in Tromsø. The bias pattern is also similar for the two models, with increasingly positive bias toward the higher-latitude Tromsø case.

The increase in transfer error is accompanied by changes in the climate-input distribution. Although most individual weather variables remain within the ranges represented by the Gothenburg training data, their joint distributions differ across climates. The shift becomes more pronounced when solar-position and calendar-related variables are included, particularly for Munich and Tromsø, as shown in Table 4.5.

Forecast-horizon diagnostics show that the large transfer error in Tromsø is already present at the beginning of the recursive prediction rather than accumulating gradually over time. For MAIN-2000, the summer bias is approximately $+2.59$ °C at the first prediction hour and remains approximately $+2.48$ °C at hour 168. Using the true 72-hour indoor-temperature history as an ideal warm start reduces MAE by only about 3.4%, while reinitializing the model with true history every 24 hours produces no

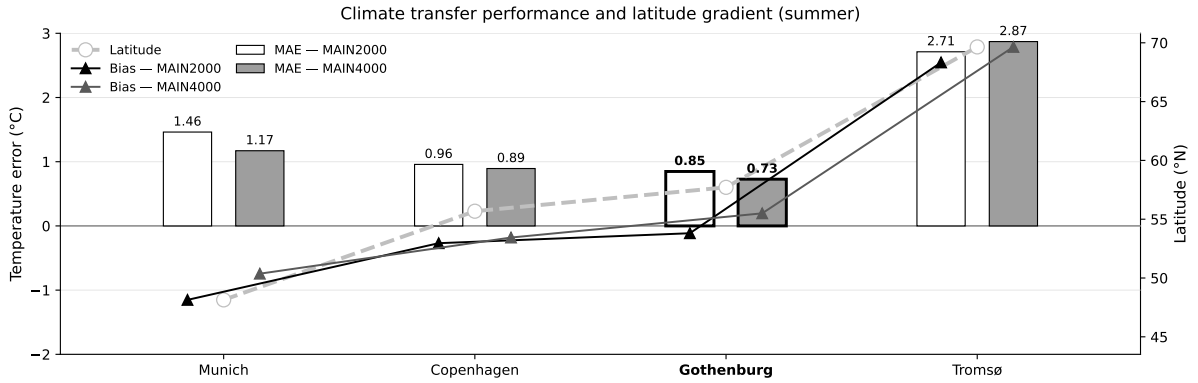


Figure 4.8: Summer climate-transfer performance of models trained on 2,000 and 4,000 Gothenburg buildings. Bars show MAE, triangles show mean prediction bias, and the dashed gray line indicates the latitude of each climate location.

Table 4.5: Joint input-distribution shift relative to the Gothenburg training domain. Values are the proportions of target samples with nearest-neighbour distance above the 99th percentile of the training distribution.

Climate	Physical weather 6D	Full EPW token 13D
Copenhagen	3.9%	8.1%
Munich	12.0%	21.5%
Tromsø	9.5%	48.8%

further improvement. The observed transfer error is therefore not associated with progressive autoregressive error accumulation over the forecast horizon.

As a stronger out-of-distribution stress test, the models are additionally evaluated under winter conditions, which are absent from the training data. For MAIN-2000, winter MAE reaches 3.08, 1.85, 3.50, and 7.85 °C in Gothenburg, Copenhagen, Munich, and Tromsø, respectively, with positive bias in all four climates. The largest deviation occurs in Tromsø, where winter dry-bulb temperatures frequently fall outside the range represented during training and the mean prediction bias reaches approximately $+7.84$ °C. These winter results are used only as an out-of-range stress test rather than as part of the main summer transfer evaluation.

4.5 Sensitivity to training-building coverage and voxel resolution

4.5.1 Number of training buildings

To evaluate the effect of training-data scale on model performance, the proposed model is trained with 1,000, 2,000, and 4,000 buildings while keeping the architecture unchanged. Panel (a) of Figure 4.9 summarizes the validation performance at each training-data scale.

As the number of training buildings increases, the validation error decreases consistently. Validation MAE declines from 0.90 °C at 1,000 buildings to 0.80 °C at 2,000 buildings and further to 0.71 °C at 4,000 buildings. The proportion of errors above 2°C follows the same trend, decreasing from 9.3% to 7.4% and then to 4.3%.

Within the current range of 1,000–4,000 buildings, saturation with data scale is not observed. The improvement from 2,000 to 4,000 buildings remains substantial, indicating that the model still benefits

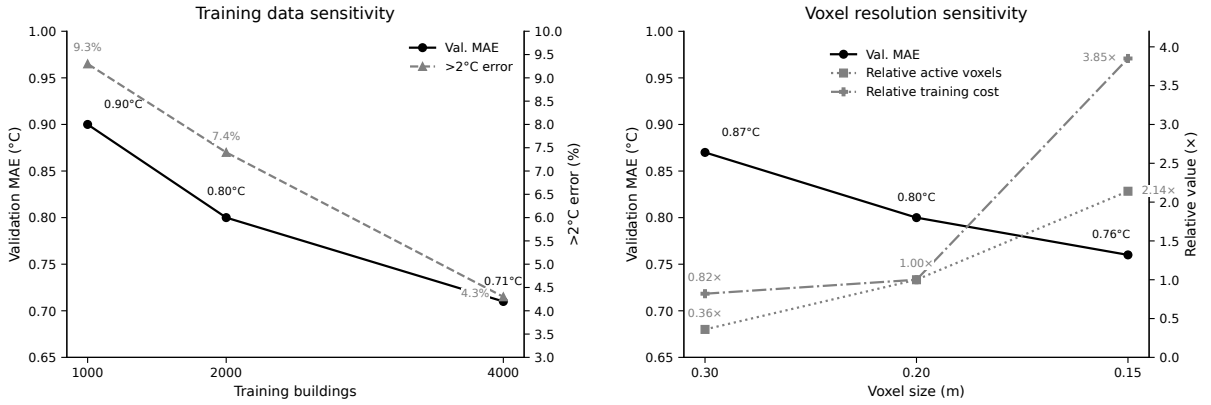
from additional training diversity within the tested range. Further expansion of the training-building set may continue to improve generalization to unseen buildings, although this study does not evaluate scales beyond 4,000 buildings.

4.5.2 Voxel resolution

To evaluate the effect of spatial-discretization resolution on model performance and computational cost, voxel sizes of 0.15 m, 0.20 m, and 0.30 m are compared using the same model architecture, input features, training and test buildings, forecast windows, and evaluation metrics. The 0.20 m grid is the default resolution used in the current experiments. This experiment examines the relationship between prediction benefit from finer geometric discretization and the corresponding increase in sparse-representation size and computational cost.

Panel (b) of Figure 4.9 summarizes the results. Refining the grid from 0.30 m to 0.20 m reduces Validation MAE from 0.87 to 0.80 °C, indicating that the coarser resolution loses part of the thermally relevant spatial information. Further refinement to 0.15 m reduces Validation MAE to 0.76 °C, but this also increases the number of active voxels to 2.14 times the 0.20 m reference and raises the observed training cost to 3.85 times the reference.

Within the tested range, the 0.20 m grid provides the selected balance between prediction accuracy and computational cost. It improves clearly over the 0.30 m grid while avoiding most of the computational increase associated with the 0.15 m setting.



(a) Training-data sensitivity (b) Voxel-resolution sensitivity

Figure 4.9: Sensitivity analyses of training-data scale and voxel resolution. (a) Validation MAE and the proportion of predictions with absolute error above 2°C for different numbers of training buildings. (b) Validation MAE, relative number of active voxels, and relative observed training cost for different voxel resolutions; relative values are normalized to the 0.20 m setting.

4.6 Geometry reuse and computational cost

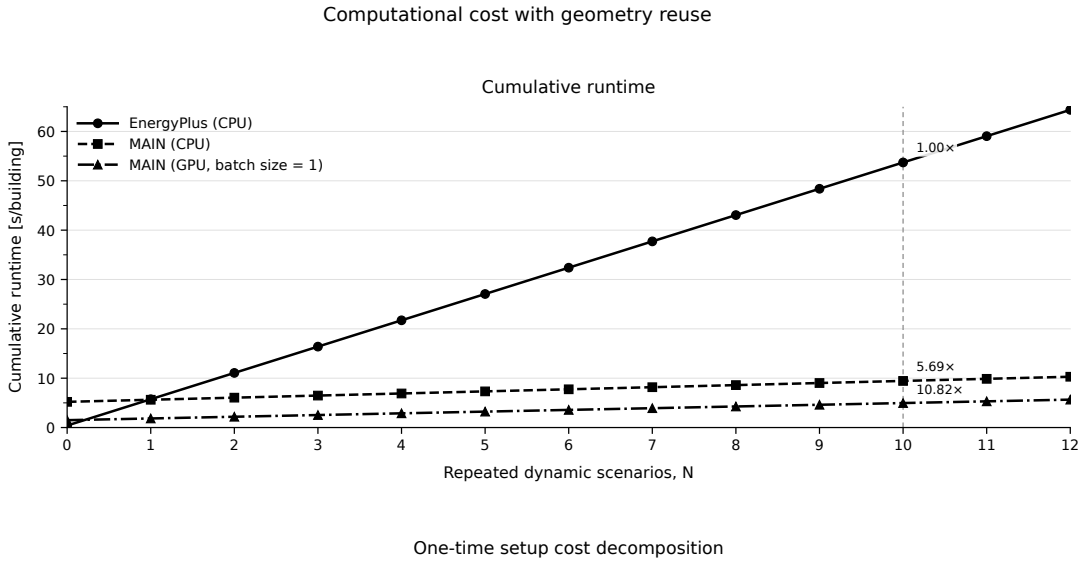
The efficiency experiment measures repeated-scenario computation. Fifty buildings are evaluated under ten independent temporal scenarios, producing 500 building–scenario samples. Each model scenario contains a 72 h warm-up and a 168 h forecast. The corresponding EnergyPlus run solves the complete 240 h period. The hardware is an AMD Ryzen 9 7945HX CPU and an NVIDIA RTX 4060 Laptop GPU. EnergyPlus and the CPU configuration of the proposed model use the same CPU.

The first pass through a building converts HBJSON to voxels, performs static geometry encoding, and creates a geometry cache. New weather or operating scenarios reuse this cache and recompute only

dynamic inputs, neighboring-zone aggregation, and temporal prediction. EnergyPlus reuses the IDF but runs a new dynamic simulation for every scenario.

Figure 4.10 indicates the computational-cost decomposition and the cumulative runtime under repeated scenarios. For EnergyPlus, the one-time IDF preparation cost is 0.405 s and the median dynamic simulation cost is 5.331 s per scenario. For the proposed model, voxelization requires 0.999 s, and geometry encoding requires 4.207 s on the CPU or 0.493 s on the GPU. After cache construction, the median dynamic inference cost is 0.424 s per scenario on the CPU and 0.347 s per scenario on the GPU. These values correspond to marginal speed-ups of 12.57× on the CPU and 15.36× on the GPU relative to EnergyPlus when only repeated dynamic scenarios are considered. When one-time setup costs are included, the cumulative runtime still grows much more slowly for the cached model than for repeated EnergyPlus simulation. At $N = 10$ repeated scenarios, the cumulative speed-up relative to EnergyPlus reaches 5.69× for the CPU configuration and 10.82× for the GPU configuration.

Overall, this experiment verifies the reusability of the static geometric representation. Once voxelization and geometry encoding have been completed, the same building does not need to repeat three-dimensional spatial encoding for a new weather or operating scenario. As a result, the cost of evaluating additional dynamic scenarios is substantially lower than repeated EnergyPlus simulation, and the one-time geometric preprocessing cost can be amortized over a limited number of repeated scenarios.



One-time setup cost decomposition

Method	Setup cost composition	Total setup	Marginal scenario cost
EnergyPlus (CPU)	IDF preparation: 0.405 s	0.405 s	5.331 s/scenario
MAIN (CPU)	Voxelization: 0.999 s + Geometry encoding: 4.207 s	5.206 s	0.424 s/scenario
MAIN (GPU, batch size = 1)	Voxelization: 0.999 s + Geometry encoding: 0.493 s	1.492 s	0.347 s/scenario

Figure 4.10: Computational cost with geometry reuse. The upper panel shows cumulative runtime as the number of repeated dynamic scenarios increases. The lower panel decomposes the one-time setup cost and the marginal scenario cost for EnergyPlus, Proposed model (CPU), and Proposed model (GPU, batch size = 1).

Chapter 5

Discussion

This chapter interprets the results in relation to the four research questions of the thesis. The discussion focuses on the value of detailed spatial representation for cross-building prediction, the fidelity of the predicted thermal response, the boundaries of generalization, and the computational role of a reusable static building representation. Together, these results indicate that detailed three-dimensional building information can support accurate and reusable zone-level thermal prediction within a defined simulation domain, while the reliability of this capability remains dependent on the coverage of building and climate conditions represented during training.

5.1 Cross-building prediction and the value of detailed spatial representation

RQ1 asks whether a cross-building surrogate can predict zone-level operative-temperature dynamics on unseen buildings and whether detailed 3D spatial information provides value beyond manually aggregated low-dimensional descriptors. The results support both parts of this question within the sampled building domain. The proposed model reaches a Validation MAE of 0.80 °C compared with 0.94 °C for the Flattened baseline.

The comparison is particularly informative because the Flattened baseline already contains essential building information. Area, volume, window-to-wall ratio, position, material properties, and other manually aggregated variables provide compact descriptions of thermal attributes. In a parametrically generated design space, these descriptors can also be strongly related with the potential geometry. The performance improvement of the proposed model can suggest that the 3D representation provides additional information about spatial configuration and geometric relationships beyond what can be captured by aggregated descriptors alone.

The static perturbation experiments distinguish between strong-effect parameters and parameters whose thermal influence emerges through interactions with building geometry. WWR and glazing can be represented directly by low-dimensional variables, and the Flattened baseline gives slightly lower response error for these factors. This is mainly because these factors appear as strong and explicit signals in the flattened parameter set, providing a more direct input to prediction than their spatial configuration, as shown in Figure 5.1. Exterior shading and orientation depend more strongly on component position, direction, and the spatial combination of windows, facades, and solar exposure. For these factors, the proposed model gives lower response error. The benefit of the 3D representation

appears when thermal response depends on where building components are located and how they are arranged, while a compact scalar can remain efficient when a physical change already has a direct low-dimensional description.

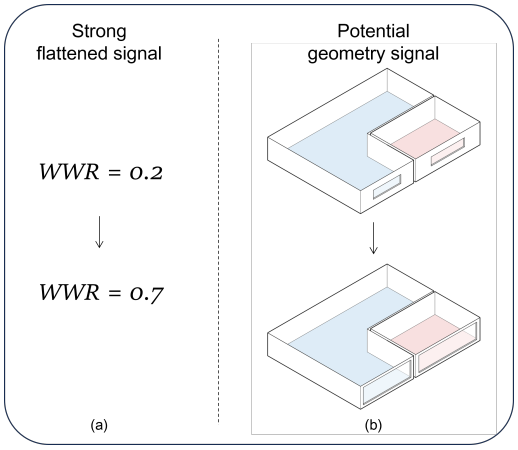


Figure 5.1: Comparison of flattened and geometry-aware input representations. Flattened inputs provide direct low-dimensional signals for individual building factors, whereas geometry-aware inputs encode these factors less directly through component geometry while retaining richer spatial relationships.

This interpretation is consistent with the broader development of geometry-aware building-performance surrogates, where learnable representations have been used to retain spatial information that is difficult to reconstruct from predefined scalar descriptors (Singh & Smith, 2023). The perturbation experiments further show that the two representations respond differently to different types of parameters. For strong scalar indicators such as WWR and glazing-related properties, the Flattened baseline sometimes produces larger and more explicit response magnitudes, suggesting that these effects can be learned effectively from directly provided aggregated variables. However, such isolated and extreme parameter changes represent only part of practical building variation. In more typical design conditions, thermal responses are often shaped by the combined influence of several geometric and physical factors. Under these coupled perturbations, the proposed geometry-aware architecture shows more stable and consistent responses, indicating an advantage in representing interactions that cannot be reduced to individual scalar indicators.

Prediction accuracy improves as the voxel resolution becomes finer, but the computational cost increases more significantly. This indicates that finer discretization retains more geometric detail, although its practical value depends on the scale of spatial features relevant to the application. The building samples in this study mainly represent features such as window openings, wall segmentation, and zone-level spatial organization, rather than detailed component construction or workmanship. The selected $0.20\,\mathrm{m}$ resolution here provides sufficient geometric detail for the present modeling scope without introducing excessive computational cost.

Taken together, the results answer RQ1 by showing that cross-building zone-level prediction is feasible within the current domain and that detailed 3D information provides a consistent additional source of predictive information.

5.2 Thermal-response fidelity and temporal stability

RQ2 concerns whether the framework preserves building thermal responses under changes in physical parameters and dynamic boundary conditions. The results show that this fidelity extends beyond

absolute temperature accuracy, with the model capturing the direction and magnitude of temperature changes, maintaining accuracy during periods of large hourly variation, and remaining stable over long recursive prediction horizons.

The controlled static perturbations provide the first level of evidence. Across the major parameter groups, the proposed model reproduces the principal warming and cooling relationships observed in EnergyPlus. Increasing WWR produces warming, deeper exterior shading produces cooling, and changes in orientation generate spatially dependent responses. The model also achieves a lower absolute-temperature MAE across the complete perturbation set. The preceding section has discussed how response accuracy differs between directly represented scalar parameters and geometry-dependent parameters. For RQ2, the main conclusion is that the proposed framework responds correctly to perturbations in different building parameters, reproducing both the direction and magnitude of the resulting thermal responses.

The temporal-response results further show that the model preserves the main characteristics of continuous warming and cooling. Across the Test buildings, it reproduces the direction of hourly temperature change with 94.01% accuracy and retains most of the response amplitude, with a dynamic gain of 0.8132. Together with the high change correlation, these results indicate that the predicted temperature trajectories generally follow both the timing and magnitude of the reference thermal response. The higher dynamic gain than the Flattened baseline also suggests that detailed spatial representation helps reproduce response fidelity.

This advantage enlarges during Strong transitions, where the thermal state changes more rapidly. The proposed model reduces the Strong-transition error from 0.343 to 0.242,$^{\circ}\mathrm{C},\mathrm{h}^{-1}$ and achieves lower error on 73.4% of the Test buildings. Its direction accuracy remains close to 100%, showing that rapid warming and cooling events are mostly captured. The improvement across most buildings also indicates that this enhanced dynamic fidelity is a consistent model advantage rather than the only a small number of specific cases.

The autoregressive predictions also remain stable over the 168 h forecast horizon. On the Validation set, the error during the later forecast period remains nearly unchanged from the starting MAE, providing little evidence of progressive error accumulation. The Tromsø climate-transfer experiment supports the same interpretation from a more demanding setting: a huge bias appears from the beginning of the forecast but remains quite stable throughout the 168 h horizon, while reinitialisation with true temperature history provides only limited improvement. These results indicate that extending the prediction horizon does not itself lead to a progressive deterioration in model performance.

Taken together, the results answer RQ2 by showing that the framework preserves the principal responses to building-parameter changes, follows the direction and temporal pattern of hourly warming and cooling, retains comparatively high fidelity during strong temperature transitions, and remains stable across the full prediction horizon.

5.3 Generalization across building and climate domains

For the third research question, the current experiments show that model generalization has at least three related dimensions with different boundaries: building space, climate state, and forecast horizon. The building-level test set demonstrates that the spatial–thermal-response relationships learned during training can transfer to new unseen buildings. The training-scale and cross-climate experiments show that this transfer capability continues to depend on the coverage of the corresponding training data.

The experiment varying the number of training buildings clearly shows the importance of building-level data coverage. As the number of training buildings increases from 1,000 to 4,000, the validation MAE decreases continuously, and the current training scale still has considerable room for data expansion. In this study, each building represents a distinct combination of geometric and spatial conditions. A large number of time steps can increase coverage of dynamic states, whereas the number of independent buildings determines coverage of spatial and building-physics combinations. Potential building states are formed by multiple continuous and discrete variables, and their combinations are practically close to inexhaustible. Therefore, increasing the number and diversity of training buildings continues to improve performance, with no clear saturation observed within the current dataset scale.

The cross-climate experiment reveals another independent generalization axis. The model maintains good performance when transferred from Gothenburg to Copenhagen, while systematic error increases clearly when transferred to Munich and Tromsø. Increasing the number of Gothenburg buildings continues to improve prediction in the source climate and similar conditions such as Copenhagen, but does not simultaneously improve the shift error in extreme extrapolative climates. This distinguishes building-space coverage from climate-state coverage. Improving climate transfer may require greater diversity of climate conditions in training data.

Further climate-transfer diagnostics show that the joint state space can be the key to this boundary. In Tromsø summer weather tokens, 48.8% exceed the 99th-percentile threshold of the distribution in the Gothenburg training data. The model therefore faces many combinations of weather, time, and solar conditions that occur infrequently during training. The underprediction of high temperatures in Munich and overprediction of low temperatures in Tromsø also show a common trend: predictions tend to shift toward indoor thermal conditions that are more common in the source climate.

The winter stress test further demonstrates this extrapolation boundary. When weather conditions become evidently out of distribution, prediction error and systematic Bias increase rapidly. These climate transfer experiments demonstrate that zero-shot transfer becomes less reliable and more biased when target conditions are poorly covered by the training data.

The forecast horizon forms the third generalization dimension. All architectures using a causal decoder maintain generally stable errors at different prediction stages. Long-horizon prediction does not show continuously increasing MAE or systematic Bias with forecast horizon, indicating that the model can maintain stable state evolution during long autoregressive prediction.

Overall, these results show that the generalization capability of a cross-building thermal-dynamics model has a clear multidimensional structure. Training coverage of building geometry and physical parameters determines spatial generalization to new building combinations. Joint coverage of weather, solar position, and thermal states determines cross-climate transfer. The stable long-horizon predictions observed here are likely supported by both the representation of building geometry, thermal mass and the autoregressive prediction process.

5.4 Reusable spatial representation and the computational role of the surrogate

For the fourth research question, the computational experiments show that explicit three-dimensional geometric representation introduces additional voxelization and spatial-encoding cost, but this cost is for one-time. For repeated evaluation with an unchanged static building state, geometric-encoding results can be cached, while dynamic inputs such as weather, operating boundaries, and prediction

windows continue to be updated independently. The model computation can be divided into one-time spatial-condition construction and repeatable dynamic prediction. In the tests, the single-scenario inference cost after caching is much lower than repeated EnergyPlus simulation. Even after the initial spatial-processing cost is included, cumulative computational benefit continues to form as the number of scenarios evaluated for the same building increases.

This computational structure corresponds well to multi-scenario tasks in building-performance analysis. During early design, a candidate geometric scheme often needs repeated evaluation under different weather, ventilation, or operating assumptions. Each candidate scheme only needs to establish the spatial representation once and can be reused for multiple dynamic scenarios. In existing-building renovation analysis, the same representation can be reused for operational, climate, or control scenarios when geometry and static properties remain unchanged. If the geometric conditions change, the spatial representation must be regenerated. The reuse benefit focuses on applying the same static building representation to multiple dynamic scenarios.

Building-stock analysis provides another form of reuse. Each building can establish and store its own spatial representation, after which batch dynamic screening can be performed for diverse climate scenarios, or operating assumptions. Cached representations can be evaluated in parallel across multiple buildings and scenarios, allowing more efficient use of computational hardware and reducing the cost of large-batch evaluation. This makes the approach more suitable for urban-scale building screening.

The computational advantage shown in this study resolves the issue of converting computationally expensive but time-invariant spatial information into reusable conditions. This structure may not have the greatest computational advantage for a one-time single-building evaluation. When the same building state is evaluated repeatedly, or when many buildings are screened under different dynamic conditions, the one-time geometric encoding provides a considerable marginal-cost advantage. This directly addresses the fourth research question of this study.

5.5 Scope, limitations, and research outlook

The conclusions of this study are based on a controlled simulation and data range. First, the training, validation, and test data all come from parametrized building models and EnergyPlus simulations. The current experiments mainly validate cross-building representation learning within the *simulation domain*. Real buildings also contain construction deviations, material aging, model-parameter uncertainty, sensor noise, and operational states that are difficult to observe completely. These effects are simplified in the current study.

Second, although the current number of buildings reaches several thousand, their geometries still come from a limited set of parametrized generation rules. Regular spatial patterns, limited zone topologies, and predefined opening combinations strengthen the statistical association between low-dimensional descriptors and three-dimensional geometry and provide effective information-compression paths for flattened architectures. The space of possible building configurations is nearly unlimited, and real buildings also contain a wide range of irregular geometries that are not represented in the current dataset.

Third, the spatial scope of this study focuses on a single building and its internal zones. The surrounding urban environment is currently compressed through the sky-view factor and related external-exposure parameters, so external obstruction relationships remain low-dimensional environmental conditions. At larger scales, neighboring-building geometry, urban-canyon form, mutual

facade obstruction, and precise inter-building solar shading have clear spatial structure. Future work can introduce an urban-context representation on top of the current zone–building level, allowing the target zone to aggregate further spatial information related to radiation and obstruction from the three-dimensional environment around the building.

Fourth, this study considers operative temperature as the key indicator of the indoor thermal environment, while other factors such as indoor humidity and air movement are not included. Outdoor-air exchange is represented by prescribed and uniform ACH schedules, without considering spatial differences in airflow related to the distance and position relative to openings, opening states, wind pressure, and stack effects.

A similar limitation applies to dynamic shading. The current geometric encoding represents only fixed shading conditions, while automatically controlled shading systems are not considered in this study.

Fifth, the dataset actively controls occupants, lighting, equipment internal gains, and HVAC systems so that temperature changes more directly reflect the combined effects of geometry, envelope, weather, air exchange, and thermal inertia. This controlled free-running setting improves the identifiability of comparisons between geometric representations and is also an important experimental condition for studying the value of geometric representation. Real building operation also contains strongly time-varying factors such as occupant activity, equipment loads, set-point temperatures, mechanical ventilation, and closed-loop HVAC control. Future studies can gradually restore these operating variables, broaden building-form diversity, and expand training across multiple climate domains.

Chapter 6

Conclusion

This study develops a framework for rapid cross-building prediction of zone-level operative-temperature dynamics by combining sparse three-dimensional geometric encoding, target-zone and whole-building spatial interaction, neighboring-zone information, and causal hourly prediction. The work is motivated by the need for computationally efficient indoor-temperature assessment across large and heterogeneous building sets, particularly for summer thermal-condition and overheating analysis in the European building stock. Within this context, the study investigates whether detailed three-dimensional building information can support accurate and reusable thermal prediction across unseen buildings, changing boundary conditions, and repeated operating scenarios, providing a faster intermediate approach between simplified screening indicators and detailed building-energy simulation.

The experimental results show that learnable three-dimensional geometric representation can consistently improve zone-level operative-temperature prediction. Architectures retaining three-dimensional geometric encoding are obviously better than models relying entirely on flat static descriptions. Static-parameter perturbations further show that the proposed architecture maintains relatively stable response capability across different building factors. The voxel-resolution experiment also shows that finer voxel resolution can improve prediction performance, but the gain needs to be balanced against computational cost. In this study, 0.20m provides a suitable trade-off between the two. Ablation results show that building-scale geometry, target-zone information, and neighboring-zone dynamics provide complementary contributions to zone-level prediction. Hybrid architectures show that part of the geometric information can be efficiently represented by compact scalar descriptors within the current parametric design space, while the complete model provides stronger comprehensive fidelity. Target-zone-driven geometric interaction allows different zones to extract information related to themselves from the same building representation, while known neighboring-zone dynamics further supplement the operating state of surrounding zones. Temporal-response experiments show that the proposed model can reproduce building thermal responses with good dynamic fidelity, including during periods of rapid temperature change.

Generalization experiments show that the framework can transfer to new unseen buildings within this study's boundary. As the number of training buildings increases from 1,000 to 4,000, prediction performance continues to improve, with no clear saturation at the current data scale.

Building-design-space coverage still remains an important limitation on cross-building generalization. Cross-climate experiments show that building-geometry generalization and climate extrapolation require different data coverage. Larger shifts in the climate joint state will reduce its accuracy. At the same time, no clear continuous error accumulation is observed during long autoregressive prediction,

indicating that prediction performance remains stable over the current forecast horizon.
Computational experiments verify the repeated-reuse capability of static geometric representation. After geometric encoding is completed, the cached representation can be reused for a new weather or operating scenario in the same building. Dynamic inference provides a substantial computational speed-up over EnergyPlus, and this advantage remains clear in repeated multi-scenario evaluations even after the one-time geometric-processing cost is included, and its benefit continues to accumulate as the number of scenarios increases.
The conclusions of this study are mainly based on parametrized buildings and EnergyPlus simulation data. Current building geometry remains constrained by prescribed generation rules, the surrounding urban environment is represented with simplified parameters. In the current study, dynamic features such as natural ventilation, dynamic openings, occupants, equipment, and active HVAC control are simplified to some extent. Future work can extend the study to real and irregular buildings, multi-climate training, and actual operational data, while adding more complete external-building obstruction and operation mechanisms to test the applicability of the learned spatial representation under more complex conditions.
In summary, this study demonstrates that three-dimensional building information can provide a more complete representation of the spatial conditions governing zone-level thermal dynamics. When combined with target-zone geometry and dynamic zone information, the proposed framework achieves accurate and temporally stable operative-temperature prediction for unseen buildings within the investigated design space, while geometry caching substantially reduces the cost of repeated scenario evaluation. These capabilities support rapid indoor-temperature assessment for identifying buildings and zones with extreme summer temperatures, comparing retrofit and design scenarios, and assessing existing European building stocks at scale. Therefore, the study provides a methodological basis for geometry-aware thermal surrogates that can complement detailed simulation in future overheating and building-stock assessment workflows.

References

Anthropic. (2026). *Claude* [Large language model-based AI assistant]. Retrieved September 21, 2026, from https://claude.ai/

Arriazu-Ramos, A., Bes-Rastrollo, M., Sánchez-Ostiz Gutiérrez, A., & Monge-Barrio, A. (2023). Building parameters that influence overheating of apartment buildings in a temperate climate in southern Europe. *Building and Environment*, *228*, 109899. https://doi.org/10.1016/j.buildenv.2022.109899

C3S/ECMWF & World Meteorological Organization. (2026). *C3S–WMO European State of the Climate 2025*. Copernicus Climate Change Service and World Meteorological Organization. https://doi.org/10.24381/zy93-sb27

Canli Akyol, I., Karadag, Y. M., Ucar, S., Talaz, I., Gursoy, F. E., Gursel Dino, I., & Kalkan, S. (2025). Transfer learning and parameter-efficient fine-tuning for heating energy consumption prediction using urban building energy models (UBEM). *Advanced Engineering Informatics*, *68*, 103576. https://doi.org/10.1016/j.aei.2025.103576

Cao, J., Zhang, H., Savov, A., Hall, D. M., & Dillenburger, B. (2022). Energy-aware design: Predicting building performance from layout graphs. *2022 European Conference on Computing in Construction*. https://doi.org/10.35490/EC3.2022.210

Choy, C., Gwak, J., & Savarese, S. (2019). 4d spatio-temporal convnets: Minkowski convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3075–3084. https://doi.org/10.1109/CVPR.2019.00319

Copernicus Climate Change Service (C3S). (2026). Copernicus: August was the world's joint hottest month on record, pushing global temperatures back above 1.5C above pre-industrial levels [Published 10 September 2026. Accessed 20 September 2026]. https://climate.copernicus.eu/copernicus-august-was-worlds-joint-hottest-month-record-pushing-global-temperatures-back-above

Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., Strand, R. K., Liesen, R. J., Fisher, D. E., Witte, M. J., & Glazer, J. (2001). EnergyPlus: Creating a new-generation building energy simulation program. *Energy and Buildings*, *33*(4), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6

Dai, T.-Y., Niyogi, D., & Nagy, Z. (2025). CityTFT: A temporal fusion transformer-based surrogate model for urban building energy modeling. *Applied Energy*, *389*, 125712. https://doi.org/10.1016/j.apenergy.2025.125712

DeepSeek. (2026). *DeepSeek* [Large language model-based AI assistant]. Retrieved September 21, 2026, from https://chat.deepseek.com/

Di Natale, L., Svetozarevic, B., Heer, P., & Jones, C. N. (2022). Physically consistent neural networks for building thermal modeling: Theory and analysis. *Applied Energy*, *325*, 119806. https://doi.org/10.1016/j.apenergy.2022.119806

Di Natale, L., Svetozarevic, B., Heer, P., & Jones, C. N. (2023). Towards scalable physically consistent neural networks: An application to data-driven multi-zone thermal building models. *Applied Energy*, *338*, 121071. https://doi.org/10.1016/j.apenergy.2023.121071

Drgoa, J., Tuor, A. R., Chandan, V., & Vrabie, D. L. (2021). Physics-constrained deep learning of multi-zone building thermal dynamics. *Energy and Buildings*, *243*, 110992. https://doi.org/10.1016/j.enbuild.2021.110992

Edwards, J. R., De Roos, A. J., Hampo, C. C., Huang, W., Lincoln, E., Hoque, S., & Schinasi, L. H. (2025). Residential indoor temperatures and health: A scoping review of observational studies. *Science of the Total Environment*, *979*, 179377. https://doi.org/10.1016/j.scitotenv.2025.179377

European Environment Agency. (2022). *Cooling buildings sustainably in Europe: Exploring the links between climate change mitigation and adaptation, and their social impacts* (tech. rep. No. Briefing No. 20/2022). European Environment Agency. https://doi.org/10.2800/61355

European Parliament and Council of the European Union. (2024). Directive (eu) 2024/1275 of the European parliament and of the council of 24 april 2024 on the energy performance of buildings (recast). Retrieved July 18, 2026, from https://data.europa.eu/eli/dir/2024/1275/oj

Fang, Z., Crimier, N., Scanu, L., Midelet, A., Alyafi, A., & Delinchant, B. (2021). Multi-zone indoor temperature prediction with LSTM-based sequence to sequence model. *Energy and Buildings*, *245*, 111053. https://doi.org/10.1016/j.enbuild.2021.111053

Farahani, A. V., Jokisalo, J., Korhonen, N., Jylhä, K., Ruosteenoja, K., & Kosonen, R. (2021). Overheating risk and energy demand of Nordic old and new apartment buildings during average and extreme weather conditions under a changing climate. *Applied Sciences*, *11*(9), 3972. https://doi.org/10.3390/app11093972

Google. (2026). *Gemini* [Large language model-based AI assistant]. Retrieved September 21, 2026, from https://gemini.google.com/

Graham, B., Engelcke, M., & van der Maaten, L. (2018). 3d semantic segmentation with submanifold sparse convolutional networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9224–9232. https://doi.org/10.1109/CVPR.2018.00961

Intergovernmental Panel on Climate Change. (2022). *Climate change 2022: Impacts, adaptation and vulnerability: Contribution of working group ii to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. https://doi.org/10.1017/9781009325844

Jano, T., Quijal-Zamorano, M., Shartova, N., Gallo, E., Méndez Turrubiates, R. F., Beltrán Barrón, N. D., Peyrusse, F., & Ballester, J. (2025). Heat-related mortality in Europe during 2024 and health emergency forecasting to reduce preventable deaths. *Nature Medicine*, *31*, 4065–4074. https://doi.org/10.1038/s41591-025-03954-7

Kuczyski, T., Staszczuk, A., Gortych, M., & Stryjski, R. (2021). Effect of thermal mass, night ventilation and window shading on summer thermal comfort of buildings in a temperate climate. *Building and Environment*, *204*, 108126. https://doi.org/10.1016/j.buildenv.2021.108126

Nguyen, A.-T., Reiter, S., & Rigo, P. (2014). A review on simulation-based optimization methods applied to building performance analysis. *Applied Energy*, *113*, 1043–1058. https://doi.org/10.1016/j.apenergy.2013.08.061

OpenAI. (2026). *Codex* [AI coding agent]. Retrieved September 21, 2026, from https://openai.com/codex/

OpenAI. (n.d.).

Rahif, R., Amaripadath, D., & Attia, S. (2021). Review on time-integrated overheating evaluation methods for residential buildings in temperate climates of Europe. *Energy and Buildings*, *252*, 111463. https://doi.org/10.1016/j.enbuild.2021.111463

Roman, N. D., Bre, F., Fachinotti, V. D., & Lamberts, R. (2020). Application and characterization of metamodels based on artificial neural networks for building performance simulation: A systematic review. *Energy and Buildings*, *217*, 109972. https://doi.org/10.1016/j.enbuild.2020.109972

Schade, J., Ylmén, P., Wallenberg, N., & Thorsson, S. (2025). Overheating in a common Swedish residential multi-story building under current and future climate: Integrating trees to combat overheating. *City and Environment Interactions*, *27*, 100206. https://doi.org/10.1016/j.cacint.2025.100206

Singaravel, S., Suykens, J., & Geyer, P. (2018). Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. *Advanced Engineering Informatics*, *38*, 81–90. https://doi.org/10.1016/j.aei.2018.06.004

Singaravel, S., Suykens, J., & Geyer, P. (2019). Deep convolutional learning for general early design stage prediction models. *Advanced Engineering Informatics*, *42*, 100982. https://doi.org/10.1016/j.aei.2019.100982

Singh, M. M., & Smith, I. F. C. (2023). Convolutional neural network to learn building-shape representations for early-stage energy design. *Energy and AI*, *14*, 100293. https://doi.org/10.1016/j.egyai.2023.100293

Taylor, J., McLeod, R., Petrou, G., Hopfe, C., Mavrogianni, A., Castaño-Rosa, R., Pelsmakers, S., & Lomas, K. (2023). Ten questions concerning residential overheating in central and northern Europe. *Building and Environment*, *234*, 110154. https://doi.org/10.1016/j.buildenv.2023.110154

Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, *49*, 560–567. https://doi.org/10.1016/j.enbuild.2012.03.003

van Loenhout, J. A. F., le Grand, A., Duijm, F., Greven, F., Vink, N. M., Hoek, G., & Zuurbier, M. (2016). The effect of high indoor temperatures on self-perceived health of elderly persons. *Environmental Research*, *146*, 27–34. https://doi.org/10.1016/j.envres.2015.12.012

Wang, X., Harrison, J., Teigland, R., & Hollberg, A. (2024). Machine learning (ML) as a surrogate model for early-stage heating demand optimization. *SimBuild Conference Proceedings: IBPSA-USA SimBuild 2024 Conference*, *11*, 454–465. https://publications.ibpsa.org/conference/paper/?id=simbuild2024_2143

Wang, X., Wang, X., Yin, X., Li, K., Wang, L., Wang, R., & Song, R. (2024). Distributed LSTM-GCN based spatial-temporal indoor temperature prediction in multi-zone buildings. *IEEE Transactions on Industrial Informatics*, *20*(1), 482–491. https://doi.org/10.1109/TII.2023.3268467

Westermann, P., & Evins, R. (2019). Surrogate modelling for sustainable building design: A review. *Energy and Buildings*, *198*, 170–186. https://doi.org/10.1016/j.enbuild.2019.05.057

Wu, Z., Li, M., Liu, W., Cheng, J. C. P., Wang, Z., Kwok, H. H. L., Huang, C., & Hou, F. (2025). Developing surrogate models for the early-stage design of residential blocks using graph neural networks. *Building Simulation*, *18*, 679–698. https://doi.org/10.1007/s12273-025-1237-7

Xing, T., Sun, K., & Zhao, Q. (2023). MITP-Net: A deep-learning framework for short-term indoor temperature predictions in multi-zone buildings. *Building and Environment*, *239*, 110388. https://doi.org/10.1016/j.buildenv.2023.110388

Yang, Z., Gaidhane, A. D., Drgoa, J., Chandan, V., Halappanavar, M. M., Liu, F., & Cao, Y. (2024). Physics-constrained graph modeling for building thermal dynamics. *Energy and AI*, *16*, 100346. https://doi.org/10.1016/j.egyai.2024.100346

Student Background and an Evolving Approach to Architecture and Urban Design

I received my Bachelor of Architecture from Southeast University, China, and am currently completing the Master's Programme in Architecture and Urban Design at Chalmers University of Technology. My studies at Chalmers gradually expanded the way I understand both architecture and the role of architectural research.

Several courses were particularly important in establishing the technical foundation of my current research direction. *Building Physics, Advanced Course* (VBF021) introduced me to the physical processes that govern indoor environments and building performance. *Deep Machine Learning* (SSY340) provided the computational foundation for working with neural networks and data-driven models. *Life Cycle Engineering* (BOM251) further broadened this technical perspective by connecting the performance of buildings and technical systems to larger questions of resources, environmental impact, and long-term decision-making. Together, these courses created the basis for the interdisciplinary direction that eventually developed into this thesis.

Equally important, however, was the academic freedom that I experienced at Chalmers. The programme allowed considerable room to explore questions outside the conventional boundaries of architectural design and encouraged students to develop their own intellectual positions. The combination of this freedom with technically ambitious and forward-looking courses made it possible for me to approach architecture from a wider perspective and to investigate questions that extend beyond the production of drawings or individual buildings.

Two of my first courses at Chalmers, *Sustainable Development and the Design Professions* (ACE380) and *Architecture in the Anthropocene* (ACE400), were especially important in this respect. They introduced sustainability not simply as a design criterion, but as a much larger architectural and societal question involving climate adaptation, environmental responsibility, resource use, and questions of justice. For me, these courses became an important starting point for asking what architects and urban researchers can contribute beyond design.

The two studios that I joined, *Material and Technique* (ACE475) and *Mediated Material Interfaces* (ACE510), further reinforced this direction. Both studios explored architectural questions through methods and media that extended beyond conventional design practice. They demonstrated that architectural inquiry can operate through experimentation, computation, material systems, technical analysis, and new forms of representation. During my two years at Chalmers, I did not complete a conventional building-design studio in the traditional sense. I increasingly came to see this as an opportunity to investigate what may matter for the future of architecture and urban research.

This thesis emerged directly from that process.
During the summer of 2025 and 2026, I spent a considerable amount of time in a south-facing room that was pleasant and warm during the Swedish winter but became increasingly uncomfortable during hot summer days. The experience was simple, but it raised a fundamental architectural question. A building may have an elegant form, a carefully composed facade, and a convincing architectural concept, but it ultimately acts upon the people who inhabit it.
As I began reading more extensively about overheating, climate adaptation, and the changing European climate, this issue became much more concrete. Climate adaptation could no longer be understood only as a broad sustainability ambition. Across Northern and Central Europe, large numbers of dwellings were developed for climates in which winter performance was historically the dominant concern. Many remain free-running or have limited active cooling, while increasingly warm summers are changing the conditions under which they operate. These silent buildings do not directly report whether a particular room is becoming uncomfortable, but the buildings exist, the spatial and physical conditions exist, and the people affected by them exist. Architects and urban designers therefore need to ask how these conditions can be identified, understood, and addressed. This is not separate from architectural responsibility, it is increasingly part of it.
The interdisciplinary education at Chalmers made it possible for me to turn this question into a concrete research method. Building physics provided a way to understand the thermal problem. Building-performance simulation provided a means of producing physically structured evidence. Machine learning provided a way to investigate whether this information could be reused across many buildings. Architectural knowledge remained important because the problem itself is fundamentally spatial: orientation, windows, shading, floor position, zoning, adjacency, and three-dimensional configuration all influence how thermal conditions develop.
The resulting thesis therefore represents more than the application of machine learning to a building-performance problem. For me, it reflects a broader change in how I understand architecture and urban design. Design remains central to the discipline, but the built environment also demands methods for observation, prediction, evaluation, and large-scale decision support. This is becoming increasingly important as European cities face simultaneous pressures from climate change, aging building stocks, energy transition, and the need to provide healthy and resilient living environments.
This thesis marks the end of my master's studies, but not the end of this research direction. Questions concerning building energy, indoor environmental quality, climate adaptation, spatial modeling, and urban-scale decision-making are becoming increasingly interconnected. I hope to continue exploring how architectural knowledge, building physics, and computational methods can be combined to address these questions more directly.
At last, I am particularly grateful for the academic freedom offered by Chalmers and for the openness of the School of Architecture toward emerging research topics and interdisciplinary technical methods. The opportunity to move between architectural thinking, environmental questions, physical modeling, and new technology has become an essential foundation for how I now approach research in architecture and the built environment.

Appendix A

Generative AI Use Statement

Generative AI tools were used as supporting tools during this thesis, including programming assistance, language and document refinement, and information retrieval. The research questions, methodological framework, model architecture, experimental design, interpretation of results, and final academic decisions were developed and controlled by the author.

Programming assistance. Generative AI tools, primarily OpenAI Codex (OpenAI, 2026), DeepSeek (DeepSeek, 2026), and Claude (Anthropic, 2026), were used to support programming tasks. Main applications included computational-performance optimization, debugging and error correction, and refinement of data visualizations. Programming was conducted primarily in Python within Visual Studio Code. The overall software architecture, core algorithmic logic, research workflow, and design of comparative experiments were defined and controlled by the author. AI-generated or AI-modified code was reviewed, tested, and revised before being incorporated into the research workflow.

Language and document preparation. ChatGPT (OpenAI, n.d.) and Gemini (Google, 2026) were used to support the refinement of some technical terminology and academic expressions. The structure of the thesis, its argumentation, scientific reasoning, and research conclusions were developed by the author. The thesis was prepared in LaTeX using Overleaf. Generative AI was also used to assist with the refinement and troubleshooting of LaTeX formatting code, while decisions regarding page layout, figure composition, and overall visual presentation remained under the author's control.

Literature and background-information retrieval. Generative AI-assisted search was used for the preliminary retrieval and screening of institutional information, and local news reports relevant to the background of the study. This was particularly used to support the collection and comparison of background statistics and contextual information. All cited webpages and source URLs were manually verified by the author. Academic publications included in the thesis were independently read and assessed by the author before citation.

Author responsibility and verification. Generative AI outputs were treated as intermediate supporting material. Code, technical statements, bibliographic information, numerical values, and source attribution were checked before use. Responsibility for the research process, analysis, interpretation, written argument, and final thesis remains entirely with the author.