Automation in Construction 182 (2026) 106712

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Automated and scalable BIM2BEM framework with zoning-based model simplification leveraging knowledge graph integration

Meng Wang *, Georgios N. Lilis, Dimitris Mavrokapnidis, Kyriakos Katsigarakis, Ivan Korolija, Dimitrios Rovas

Institute for Environmental Design and Engineering, University College London, London, UK

ARTICLE INFO

Dataset link: https://github.com/ucl-sbde

Keywords:
BIM2BEM
Knowledge graph integration
Zoning-based model simplification
Geometric data processing

ABSTRACT

Accurate and scalable generation of Building Energy Models (BEM) from Building Information Modelling (BIM) data is critical for performance-driven building design. However, existing methods are often constrained by data quality issues and rigid workflows, limiting automation. This paper proposes an automated and scalable BIM-to-BEM (BIM2BEM) framework enabled by knowledge graph integration, designed to support automation and scalability in model generation from imperfect BIM data. To manage model complexity, zoning-based mappings from BIM spaces to thermal zones are derived through multi-factor analysis of spatial relationships, functional usage, thermal load similarity, and HVAC configuration. Applied to a real-world complex building, the framework reduces simulation time by up to 70%, while maintaining energy use deviations within 3% and HVAC sizing variations up to 10%, compared with the full-model baseline. These findings indicate that the proposed framework can enhance BIM2BEM automation, supporting the scalable and flexible generation of simulation-ready models under practical data limitations.

1. Introduction

Building Information Modelling (BIM) provides rich spatial and semantic data that supports the automated generation of Building Energy Models (BEM), which are essential for the design and operation of energy-efficient buildings. This forms the foundation of BIM-to-BEM (BIM2BEM) workflows, aiming to improve modelling accuracy and reduce manual work. At the same time, semantic technologies, supported by ontologies and knowledge graphs, offer structured and consistent digital representations of buildings and their systems. These technologies enhance data integration and support the management of geometric, system-level, and device-level information. For building energy modellers, it is essential that BIM2BEM conversion results in simulation models with appropriate levels of detail and accuracy, including the correct mapping of BIM spaces to BEM thermal zones and the consistent incorporation of information on passive and active components. Integrating these elements into a cohesive BIM2BEM framework can significantly improve the flexibility, scalability, and reliability of modelling processes.

The feasibility of BIM2BEM methodologies has been extensively explored, with numerous studies demonstrating their potential to streamline building energy modelling in automated or semi-automated workflows. Bazjanac et al. [1] first showed that structured BIM data could be leveraged to support building performance simulations, laying the groundwork for subsequent developments. Gao et al. [2] reviewed automated BIM2BEM frameworks and highlighted their benefits in improving data exchange and simulation accuracy, particularly in sustainable design applications. Further reviews have categorised BIM-based energy modelling by key objectives such as performance prediction, operational management, and retrofit planning, as highlighted by Alhammad et al. [3] and Pezeshki et al. [4]. BIM has also proven valuable for computational fluid dynamics (CFD) applications, offering detailed geometric and system information for airflow and thermal modelling [5]. Despite these advancements, the vision of a fully automated and error-free BIM2BEM workflow remains unmet, largely due to persistent challenges in model completeness, data inconsistency, and semantic misalignment, as reported by Kamel and Memari [6]. These limitations underscore the need for ongoing advancements in automation, interoperability, and simulation readiness, calling for flexible and robust BIM2BEM processes that can effectively manage imperfect BIM data and accurately capture both spatial and system-level characteristics in complex building scenarios.

To address these challenges, this work aims to develop an automated and scalable framework for BIM2BEM conversion with model simplification, thereby improving the flexibility and reliability of model

* Corresponding author.
E-mail address: wang.meng@ucl.ac.uk (M. Wang).

https://doi.org/10.1016/j.autcon.2025.106712
Received 3 August 2025; Received in revised form 4 December 2025; Accepted 8 December 2025

generation for complex buildings. Specifically, the research investigates how semantic technologies and knowledge graph integration can support BIM2BEM conversion by managing imperfect BIM data and integrating building system information, while maintaining simulation accuracy and efficiency. Moreover, the simplified BEM models generated in this work are intended to support facility managers in assessing and optimising operational strategies, rather than for redesigning HVAC systems, particularly in large, complex buildings, where manual BEM development is highly time-consuming and error-prone.

Based on this objective, the proposed framework starts with the construction of a comprehensive knowledge graph that integrates semantic technologies with architectural and HVAC data to create a coherent and accurate representation of the building and its systems. Thermal zoning scenarios are then derived by querying and simplifying the knowledge graph, using a multi-factor analysis that considers spatial adjacency, functional usage, HVAC system configurations, and thermal load similarity. These scenarios guide the mapping of multiple IfcSpaces to thermal zones and support geometric simplification through space merging, resulting in a geometry-error-free model compatible with simulation requirements. Finally, a BIM2BEM workflow is established that combines geometric data processing with the integration of non-geometric information from the knowledge graph, ensuring that the simplified BEM models retain sufficient geometric precision, system detail, and modelling accuracy.

By applying the methodology to a newly constructed, large-scale building with a complex layout and intricate HVAC systems, the main contributions of this paper are as follows:

- From a scientific perspective, the framework integrates BIM2BEM conversion, knowledge graph technologies, and zoning-based model simplification into a unified workflow. This seamless integration significantly improves the scalability and flexibility of the BIM2BEM conversion process, enabling the generation of high-reliability models for building performance simulation.
- From a practical perspective, the framework can effectively handle imperfect BIM data and overly complex building geometry. By incorporating multi-factor zoning-based model simplification within BIM2BEM conversion, the resulting BEM model maintains a manageable level of complexity, enhancing simulation efficiency while preserving modelling accuracy.

The remainder of this paper is organised as follows. Section 2 reviews related literature. Section 3 introduces the methodology and framework. Section 4 presents the case study and zoning scenarios. Section 5 discusses the results, including the graph-based building representation and building performance under model simplification. Section 6 concludes and summarises the main findings.

2. Background and related work

The geometric conversion process in BIM2BEM has been the focus of extensive research. Industry Foundation Classes (IFC) is widely adopted as the primary input format due to its structured and comprehensive representation of architectural, mechanical, and electrical elements. IFC-based approaches have been implemented in various workflows. For instance, Ramaji and Memari [7] utilised IFC data to support energy model generation. Lilis et al. [8] proposed a method for producing second-level space boundary (2LSB) surface sets directly from IFC files, which is a critical yet technically demanding step in BIM2BEM workflows. Ying et al. [9] developed an algorithm to convert curved BIM geometries into polyhedral approximations, thereby improving geometric consistency and simulation efficiency. In addressing interoperability between IFC and simulation tools, Lobos et al. [10] introduced a framework for automated data exchange, enabling integration between BIM tools and national energy certification platforms.

In addition to IFC, gbXML has been used as a lightweight alternative to represent building information. Dena et al. [11] applied gbXML to support the generation of 2LSB, while Yang et al. [12] emphasised its advantage in reducing the effort required to reconstruct simulation models. Elnabawi et al. [13] demonstrated the integration of gbXML with EnergyPlus to facilitate energy modelling during early design stages. Alongside EnergyPlus, a widely adopted simulation engine, other studies have explored the use of Modelica in BIM2BEM workflows. Kim et al. [14] developed a Modelica library to support semi-automated conversion, while Jeong et al. [15] proposed a direct method for BIM2BEM conversion to support thermal simulation and system optimisation. More recently, Kiavarz et al. [16] investigated data-driven energy models interacting with IFC-based information for performance analysis. In the context of HVAC system modelling, Li et al. [17] and Chen et al. [18] used OpenStudio to map HVAC configurations, while Wang et al. [19] employed a graph-based approach to extract complete HVAC topologies for EnergyPlus simulations. These studies confirm the feasibility and adaptability of BIM2BEM workflows across various simulation engines and modelling approaches.

Besides these academic developments, several industry-adopted toolchains, such as IfcOpenShell, BlenderBIM, Ladybug and Honeybee, have also been extending their capabilities to support BIM2BEM conversion [20]. More recently, Pollination has emerged as a cloud-based platform that supports BIM2BEM conversion through geometry validation, model cleaning, and interoperable exports [21]. While these tools provide essential functionality, their performance is strongly dependent on the quality and completeness of the input IFC data [22]. In practice, missing attributes or inconsistent semantics often lead to incorrect outputs or even export failures, necessitating substantial manual intervention [23]. This dependency is particularly evident for building systems where the semantic links between equipment are frequently incomplete or inaccurate. Moreover, current tools are generally more advanced in geometric processing for architectural components than in HVAC modelling, which typically configure simplified systems through predefined templates [18]. However, their customisation options are often limited, restricting system-level and device-level analyses in complex buildings with diverse HVAC configurations. Moreover, some tools that embed thermal zoning approaches are primarily designed to subdivide large spaces into multiple thermal zones, whereas the automated aggregation of small spaces with similar functions remains limited and often requires manual adjustment [24]. Additionally, interoperability challenges across data formats compromise the robustness and scalability of these workflows, particularly when applied to large, complex buildings with imperfect BIM data.

Both academic methods and industry tools that support BIM2BEM have demonstrated feasibility and provided valuable functionalities, yet they remain somewhat constrained in terms of robustness and scalability. Most approaches still rely on one-to-one mappings between BIM spaces and thermal zones, often assume high-quality and complete input data, and are typically validated only on relatively simple cases. These limitations underscore the need for more flexible and error-tolerant solutions that can accommodate the complexity and imperfections inherent in real-world building models.

To overcome the limitations of imperfect BIM data and enable a digital representation of building components and their interrelationships, recent studies have introduced semantic technologies to integrate and manage BIM alongside other data sources. Ontology-based knowledge graphs structure BIM data into machine-interpretable formats, representing not only geometric and physical attributes but also semantic and topological relationships [25]. This graph-based approach enables advanced querying and reasoning, supporting the identification and correction of incomplete or inconsistent information in spatial and system-level data [26]. Existing building ontology schemas, such as Brick [27] and FSO [28], provide standardised vocabularies for interoperability, while Shapes Constraint Language (SHACL) enables rule-based validation of topological structure [29]. Werbrouck et al. [30] highlighted the use of semantic web technologies to enrich existing building geometry through graph-based representations,

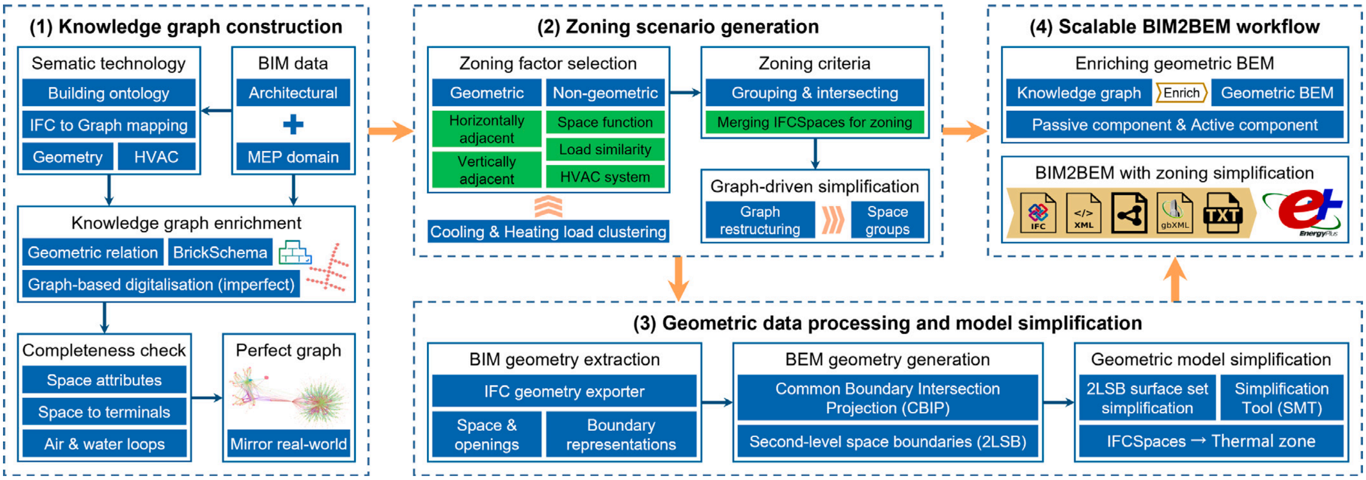


Fig. 1. Proposed framework.

addressing geometric uncertainties and enhancing the consistency and interpretability of reconstructed data. Similarly, Küçükavci et al. [31] demonstrated the effectiveness of these technologies in digitally representing HVAC systems and detecting data quality issues in BIM, thereby improving the accuracy and reliability of system information during the design phase. Beyond the integration of static data, Wang et al. [32] introduced a digital twin incorporating sensing and monitoring real-time data, and Boje et al. [33] further proposed a semantic construction that integrates BIM data with standard ontologies to enable lifecycle-based modelling and multi-source data exchange. These developments highlight the potential of creating a digital counterpart to serve as intelligent middleware, enhancing data interoperability and facilitating more flexible BIM2BEM conversions beyond rigid one-to-one mappings.

In large buildings with complex layouts, maintaining a strict one-to-one mapping between BIM spaces and BEM thermal zones often results in overly detailed models that significantly increase simulation time without a corresponding gain in accuracy. To address this, researchers have explored zoning optimisation techniques that intelligently aggregate spaces into thermal zones, aiming to balance simulation fidelity with computational efficiency [24]. Geometric simplification remains a fundamental step in this process. For example, Lilis et al. [34] reduced the complexity of 2LSB surface sets, while Georgescu et al. [35] applied Koopman operator theory to decompose building geometry into spatial modes, enabling zoning at different levels of granularity. These studies demonstrate the importance of managing geometric complexity as a basis for effective zoning.

Beyond building geometry, zoning strategies should also account for thermal loads and HVAC system configurations. Shin et al. [36] proposed a cluster- and load-based zoning method to enhance simulation accuracy in multi-zone buildings, while Chen et al. [37] analysed how different zoning schemes influence energy predictions across building stocks with varying HVAC systems. More recently, researchers have begun integrating zoning approaches into BIM2BEM workflows to improve modelling efficiency without sacrificing accuracy. Wu et al. [38] proposed an ontology-based BIM2BEM workflow with thermal zoning to achieve substantial reductions in modelling time, albeit in a simple single-floor case. Gourlis et al. [39] further investigated digital twin-based simplification guided by high-level HVAC system information. These studies demonstrate the feasibility of such approaches, suggesting that integrating zoning strategies into BIM2BEM workflows offers significant potential to enhance scalability, flexibility, and applicability. Despite recent progress, most existing studies are based on simplified cases and ideal data inputs, leaving their applicability to imperfect data and large, real-world buildings insufficiently explored.

While many studies have investigated BIM2BEM workflows and thermal zoning methods, most rely on simplified, illustrative cases and complete, error-free BIM data. Existing approaches often use rigid one-to-one mappings between BIM spaces and thermal zones, which limits scalability and flexibility when applied to large buildings with complex layouts and intricate building services systems. These methods may also fail to meet the practical needs of building energy modellers. Moreover, current zoning strategies tend to focus primarily on geometric simplification, with limited integration of information on system configurations, which is essential for generating reliable BEM models. The lack of seamless coordination between zoning strategies and BIM2BEM processes remains a significant challenge, underscoring the need for more robust and adaptable workflows that can effectively manage imperfect BIM data and accurately capture both spatial and system characteristics. Furthermore, the potential of semantic technologies to enhance the scalability and flexibility of BIM2BEM workflows also remains underexplored. Incorporating zoning-based model simplification supported by knowledge graph integration offers promising potential to improve simulation efficiency while maintaining accuracy, particularly in complex real-world building scenarios.

3. Methodology

This paper proposes an automated and scalable BIM2BEM framework with zoning-based model simplification leveraging knowledge graph integration. As illustrated in Fig. 1, this framework consists of four main components: (1) knowledge graph construction, (2) zoning scenario generation, (3) geometric data processing and model simplification, and (4) scalable BIM2BEM workflow.

First, a comprehensive knowledge graph is constructed by integrating semantic technologies with architectural and Mechanical, Electrical, and Plumbing (MEP) BIM data to represent building spaces, HVAC components, and their logical relationships. A rule-based validation process is then applied to ensure the topological completeness of the knowledge graph, resulting in a structurally coherent digital counterpart that accurately reflects the real-world system configuration.

Second, a multi-factor analysis is conducted to determine the key criteria for thermal zoning. These include geometric adjacency, functional usage, HVAC system configuration, and thermal load similarity. Spaces that meet all criteria are aggregated into candidate groups, forming the foundation for mapping IfcSpaces to thermal zones. Zoning scenarios are then generated by leveraging semantic technologies embedded in the knowledge graph to ensure consistency and traceability.

Third, geometric data processing converts architectural BIM data into geometry compatible with building performance simulations, starting with the extraction of volumetric representations via the IFC geometry exporter and the generation of second-level space boundaries

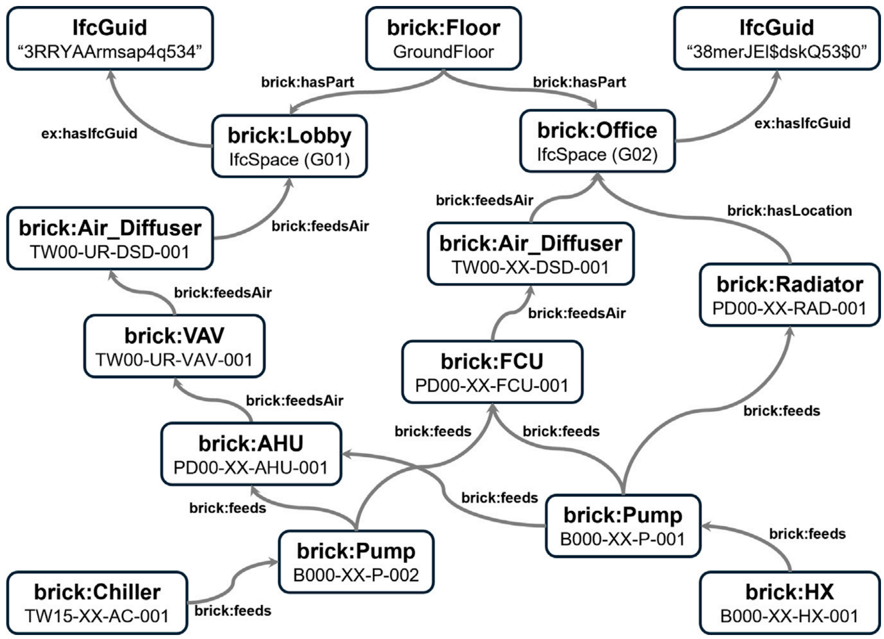


Fig. 2. Partial view of knowledge graph representing building components and relationships.

(2LSB). Based on the mapping defined by zoning scenarios, multiple IfcSpaces are merged to form corresponding thermal zones. In addition, to ensure the geometric model is suitable for simulation, a polygon simplification method is introduced to reduce surface complexity while maintaining topological consistency.

Finally, a scalable BIM2BEM workflow is developed to generate EnergyPlus-compatible BEMs automatically. Information extracted from the knowledge graph, including passive and active components, enriches both the full-model baseline and the simplified geometric BEM models. This workflow is a seamless and automated pipeline that integrates IFC, XML, gbXML, TTL, and IDF formats, ensuring data interoperability across domains. A comparative analysis is conducted to evaluate the impact of zoning-based model simplification on the accuracy and efficiency of the generated BEM models.

3.1. Knowledge graph construction

Knowledge graph facilitates seamless BIM2BEM conversion by unifying geometric, semantic, and topological data into a coherent, machine-readable structure. This section outlines a three-step development process comprising ontology-based semantic modelling, enrichment of connectivity within the graph, and rule-based validation of structural completeness. Together, these steps ensure that the resulting knowledge graph provides an accurate and well-structured digital counterpart, supporting the efficient transfer of consistent information from BIM to BEM.

3.1.1. Semantic technology and ontology

Semantic technologies are fundamental to constructing a comprehensive digital representation of buildings by integrating spatial geometry with device-level information. This work utilises ontology-based knowledge graphs to represent building spaces, HVAC components, and their attributes and relationships, thereby enabling a structured and interoperable representation. Rather than developing a project-specific ontology, existing ontologies are reused to ensure scalability and consistency across heterogeneous data sources.

Brick ontology, developed by the Brick Consortium, provides a standard vocabulary for representing spatial entities (e.g., spaces and zones) and HVAC equipment (e.g., air handling units, radiators, and VAV boxes), along with their associated semantic relationships. It supports data integration across BIM and building performance applications. In this work, IFC-derived information is automatically transformed into knowledge graph entities using the Knowledge Graph Generator (KGG), developed in the previous work [40]. KGG builds on IfcOpenShell and extends it into an automated ETL pipeline that supports multiple ontologies (including Brick, BOT, and FSO), thereby enabling richer semantic representation [41]. Each generated entity is explicitly linked to its original IFC global unique identifier (GUID) to ensure provenance, maintain traceability, and avoid naming conflicts.

3.1.2. Knowledge graph enrichment

While the initial graph construction captures component-level and spatial information using ontology classes, many interconnections remain incomplete due to limitations in BIM semantics and modelling inconsistencies. To address this, a hybrid enrichment approach is employed, combining semantic extraction and geometric inference. Functional relationships are first extracted from IFC entities such as `IfcRelConnectsPorts`, which define intended system connections between HVAC components. These are then supplemented through reasoning using the Geometric Relation Checker (GRC) [42], which identifies adjacency, clash, and containment based on the geometric configurations present in the BIM models.

By integrating these two sources, the knowledge graph incorporates explicit and inferred relationships, resulting in a more comprehensive representation of spatial and system configurations. The enriched graph accurately mirrors real-world building layouts and HVAC connectivity, providing a solid foundation for building digitisation. In addition to topological relationships, space-level attributes are required to support simulation tasks. Specifically, each space entity in the graph must be associated with its intended function (e.g., office, lab, toilet), as this directly affects internal heat gains and HVAC control logic. However, BIM models often lack or inconsistently define such functional classifications. Therefore, manual identification based on floor plans, design documents, or domain expert input is typically required for labelling, such as assigning space functions, identifying HVAC equipment types, or resolving missing and conflicting attributes.

The detailed methodology for constructing such a knowledge graph is described in the previous works [19,25]. Fig. 2 illustrates a representative example of the knowledge graph derived from BIM data.

Although the enriched knowledge graph improves coverage and connectivity, it may still contain incomplete or erroneous links due to data gaps or limitations in inference logic. Therefore, validation procedures are essential for assessing and refining the graph's structure before it is applied.

3.1.3. Graph completeness validation

Validating the structural completeness is essential to ensure that the knowledge graph accurately reflects the real-world system and supports reliable energy modelling. This work employs a rule-based validation approach using SHACL (Shapes Constraint Language) to evaluate the graph structure and verify that key functional and spatial relationships are explicitly defined. A domain-informed ruleset (see Table 1) is established based on the typical design logic of building service systems, specifying the required relationships among spaces, terminals, distribution devices, and energy sources. The goal is to detect missing, incorrect, or redundant connections that would hinder the graph's interpretability and completeness, thereby ensuring the accurate delivery of information to the BEM.

SHACL shapes are generated from the ruleset and applied to the established graph using a validation engine such as pySHACL. The validation process produces a detailed report that identifies non-conforming nodes or subgraphs, along with specific constraint violations. These results support targeted corrections directly on the knowledge graph, improving its structural integrity without modifying the original BIM data. When rule violations are identified, corrections are performed manually with reference to design drawings to ensure alignment with the intended system configuration. Further details on SHACL validation can be found in the previous work [29].

A structurally validated knowledge graph is a prerequisite for reliable BIM2BEM conversion. By ensuring that key system components and their relationships are consistently represented, the knowledge graph enables accurate transfer of information into simulation-ready BEMs, thereby maintaining alignment between design logic and performance analysis. The validation rules primarily address space-level equipment and typical HVAC system configurations, and therefore do not extend to every possible system element. Nevertheless, once validation passes and all defined relationships are confirmed, the knowledge graph can effectively support high-fidelity and trustworthy energy modelling.

3.2. Zoning scenario generation

Zoning, in the context of BIM2BEM, refers to the mapping process from IFC-defined building spaces (IfcSpace) to thermal zones used in BEM. As energy simulations typically require an abstracted thermal zoning structure, this mapping may involve a one-to-one or many-to-one relationship, depending on geometric relationships, functional usage, system configuration, thermal load similarity, and even modelling resources. This section introduces a structured approach to generate zoning scenarios that define how multiple IfcSpaces can be aggregated into a single thermal zone. It includes the selection of relevant zoning factors, the definition of zoning criteria, and graph-driven thermal zoning to guide the BIM2BEM conversion.

3.2.1. Zoning factor selection

The selection of zoning factors plays a crucial role in defining how BIM-defined spaces (IfcSpaces) are aggregated into thermal zones, directly influencing the realism and accuracy of building simulations. This work selects four key factors, including geometric relation, space function, HVAC system configuration, and thermal load similarity. These factors are chosen for their direct influence on thermal behaviour, control logic, and system operation, which are all critical to generating a reliable BEM model.

- Geometric relation: Adjacency is a prerequisite for merging multiple IfcSpaces into a thermal zone. Only horizontally or vertically adjacent spaces are eligible, as non-adjacent ones cannot support consistent wall merging or shared boundary generation in the geometric model.
- Space function: Spaces with the same functional usage typically share similar occupancy patterns, internal heat gains, and comfort requirements. Grouping spaces by function ensures that thermal zones reflect consistent usage profiles, which is fundamental for defining appropriate control schedules and boundary conditions in simulation.
- HVAC system configuration: HVAC terminal setup determines how spaces are conditioned and controlled. Spaces connected to the same ventilation system or served by similar terminals tend to operate under shared setpoints and control strategies. Recognising this system-level or equipment-level similarity is essential for creating zones that align with actual HVAC operation, particularly in buildings with mixed system types.
- Thermal load similarity: Even among functionally and system-wise similar spaces, variations in envelope conditions, orientation, storey, and internal loads can lead to divergent thermal demand profiles. By considering similarities in space-level cooling and heating loads obtained through Ideal Load Air System simulations, thermal zones can be formed to ensure uniform thermal behaviour, improving simulation accuracy and model robustness.

These four factors collectively support the creation of thermal zones that are physically meaningful, operationally aligned, and computationally effective for simulation-based performance analysis.

3.2.2. Identifying thermal load similarity

This paper evaluates the similarity of space-level cooling and heating loads using clustering analysis. First, a one-to-one mapping was established between BIM-defined spaces (IfcSpaces) and thermal zones. The geometric model and the conversion of passive components used to construct this full-resolution ideal-load BEM are developed by the proposed BIM2BEM workflow in Section 3.4. Based on this setup, each IfcSpace was simulated independently using the Ideal Load Air System to obtain its annual heating and cooling loads under ideal HVAC control.

To ensure consistent comparison across spaces of different sizes and heights, the simulated loads were normalised by each space's volume. This produced unit-volume indicators (in W/m^3 or kWh/m^3), which were the basis for assessing thermal load similarity. Besides, spaces were then classified into two categories: (a) those requiring both cooling and heating, and (b) those requiring heating only. This categorisation was based on the results of the ideal load simulation, informed by the HVAC system configuration and control setpoints specified in the building design manuals. For each category, clustering analysis used the standardised annual cooling and heating demands per cubic metre as input features.

Gaussian Mixture Model (GMM) clustering was adopted for its probabilistic foundation and its capacity to represent overlapping clusters [43]. To ensure an objective and data-driven selection of the number of clusters, the Bayesian Information Criterion (BIC) was evaluated across candidate results with varying component counts, and the configuration with the lowest BIC was selected. This approach enables objective clustering results that capture common patterns in thermal loads across different spaces.

The resulting clusters represent the underlying similarity in heating and cooling loads, serving as one of the key zoning factors in the subsequent model simplification. These cluster labels were embedded into the previously constructed knowledge graph by tagging each IfcSpace node with its corresponding load similarity cluster. This semantic annotation enhances the graph's capacity to support informed, consistent decisions when aggregating IfcSpaces into thermal zones.

Table 1
Validation ruleset for ensuring completeness of the generated knowledge graph.

Node checking	Constraints			
	Source	Edge	Sink	Card.
Brick:Air_Diffuser	[a]	Brick:feedsAir	Brick:Space	=1
Brick:FCU	[a]	Brick:feedsAir	Brick:Space	=1
Brick:Radiator	[a]	Brick:hasLocation	Brick:Space	=1
Brick:AHU	[a]	Brick:feedsAir	{Brick:Air_Diffuser; brick:CAV; brick:VAV}	≥1
Brick:VAV	[a]	Brick:feedsAir	Brick:Air_Diffuser	≥1
Brick:VAV	Brick:AHU	Brick:feedsAir	[a]	≥1
Brick:Air_Diffuser	{Brick:VAV; brick:AHU}	Brick:feedsAir	[a]	≥1
Brick:Water_Pump	[a]	Brick:feeds	{Brick:HX; brick:AHU; brick:Radiator; brick:FCU}	≥1
Brick:Water_Pump	{Brick:Boiler; brick:HX; brick:Chiller}	Brick:feeds	[a]	≥1
Brick:AHU	Brick:Water_Pump	Brick:feeds	[a]	≥1
Brick:FCU	Brick:Water_Pump	Brick:feeds	[a]	≥1
Brick:Radiator	Brick:Water_Pump	Brick:feeds	[a]	≥1
Brick:Boiler	[a]	Brick:feeds	{Brick:Water_Pump; brick:AHU; brick:Radiator; brick:HX; brick:FCU}	≥1
Brick:Chiller	[a]	Brick:feeds	{Brick:Water_Pump; brick:AHU; brick:FCU}	≥1

[a] Indicates the node serving as the subject or object when validated.

3.2.3. Graph-driven thermal zoning

Thermal zoning in this paper is governed by a set of multi-factor criteria that determine whether multiple IfcSpaces can be merged into a single thermal zone. These criteria are applied to the knowledge graph, integrating both geometric and semantic information. Specifically, four core conditions must be satisfied for space aggregation: (1) vertical or horizontal adjacency, determined through geometric relationships; (2) identical functional usage, such as office or lab; (3) similar thermal load characteristics, based on the load similarity clusters derived from the above subsection; and (4) consistent HVAC system association.

Among the zoning criteria, HVAC configuration also plays a critical role. The knowledge graph captures system topology by linking each space to its associated terminal units and ventilation systems. In typical HVAC configurations for mixed-use non-domestic buildings in Europe, these systems include air handling units (AHUs), fan coil units (FCUs), and radiators, which condition the indoor environment. Spaces are eligible for merging only if the same ventilation system serves them or they do not require mechanical ventilation, ensuring they can be controlled as a single thermal zone. Additionally, spaces equipped with FCUs or radiators may be grouped if they share a hydronic loop with consistent control settings. These relationships are represented as edges in the graph and assessed through semantic queries. Fig. 3 illustrates how zoning criteria are applied in a representative example, comparing the original knowledge graph with its simplified counterpart.

It is important to note that all zoning criteria must be satisfied simultaneously. These include geometric adjacency, space function, HVAC system configuration, and thermal load similarity. In practice, this means that merging decisions are based on the intersection of these conditions rather than their union. Only spaces that meet all requirements are considered suitable for aggregation. As a result, the full set of IfcSpaces is divided into many smaller candidate groups, within which merging can be evaluated. Besides, multiple combinations of merging decisions may be generated for each zoning scenario. This results in up to 2^N possible zoning configurations, where N is the number of space groups that are independently eligible for merging. The exponential growth in potential scenarios significantly increases computational demands, especially for large and complex buildings.

All zoning criteria are resolved within the knowledge graph, enabling zoning operations to be performed directly. Once groupings are determined, a structured mapping is created between each IfcSpace and its assigned thermal zone. This mapping is exported as a plain-text file using IFC GUID, which serves as the zoning definition to guide the BIM2BEM conversion process. Based on this mapping, the original BIM-derived geometric model is updated to reflect the new zoning configuration. A simplified geometric model with fewer thermal zones is then generated in gbXML format, which is subsequently converted and enriched to produce a simulation-ready BEM model.

3.3. Geometric data processing and model simplification

To generate the appropriate geometric content for BEM based on specific space-to-zone grouping rules, two geometric processing steps are required. These are described in the following subsections. In brief, the first step involves generating the geometry of the complete BEM model, where thermal zones are not yet defined, and each space in the building corresponds to an enclosure (a closed volume) formed by the building's structural elements (e.g., walls and slabs). This BEM geometry, referred to as the 2LSB surface set, is derived from the space volumes and does not account for the grade of the building's construction materials. The required thermal characteristics of the building materials are added as an enrichment step after the initial and simplified BEM models are generated. In the second step, simplified BEM geometries are derived from the full model by applying predefined space-to-zone grouping rules, as introduced in the previous work [34]. Additionally, a novel polygon simplification algorithm is developed to reduce geometric complexity while preserving the essential spatial characteristics.

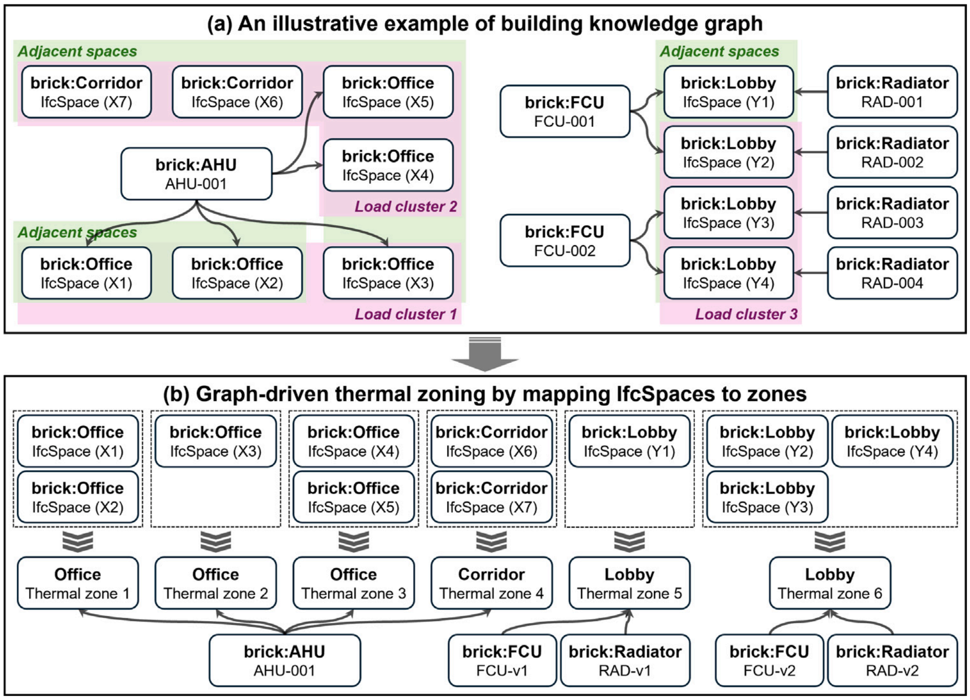


Fig. 3. Illustrative thermal zoning example for mapping IfcSpaces to thermal zones.

3.3.1. Geometric data processing for BIM2BEM

The geometric information required to generate the BEM geometry is initially extracted from the input BIM model, provided in IFC format, using the Geometry Exporter tool. This tool is part of the cloud-based toolkit, namely BIM-MP, which was presented in the previous work [44]. It retrieves 3D solid representations of internal spaces and, when available, the volumetric enclosures of openings such as windows and doors. These geometries are converted into boundary representations (Breps) that follow the outward normal convention, ensuring surface normals consistently point outward from solid volumes. The resulting data is stored in an intermediate XML format for further processing.

Subsequently, the obtained XML-based geometric data are processed using the Common Boundary Intersection Projection (CBIP) algorithm [8], to construct the BEM geometry. This step produces a comprehensive surface set that captures spatial adjacencies and consists of thermal exchange planar surfaces among the building spaces and the environment. Known as the 2LSB surface set, this structure encodes zone connectivity information and serves as the geometric backbone for simulation-ready models [45].

To satisfy the format requirements of building performance simulation tools, the 2LSB surface set undergoes an Extract-Transform-Load (ETL) process to be converted into a gbXML file. External surfaces are mapped directly, whereas each pair of internal surfaces is consolidated into a single representative surface. This is achieved by projecting both surfaces of the pair onto their median plane and computing the geometric intersection of these projections. As a result, the thickness of internal building constructions, originally defined by the distance between the planes of the paired internal 2LSB surfaces, is no longer needed in the generated gbXML and IDF models, since these models represent these surface pairs with single surfaces. Fig. 4 illustrates an example of a middle-plane polygon located within the slab between a building's floors, as represented in the output gbXML model.

This transformation produces a simplified yet topologically coherent gbXML representation, which is then converted into an EnergyPlus input data file (IDF format) using the OpenStudio SDK, thereby finalising the geometric conversion from BIM to BEM.

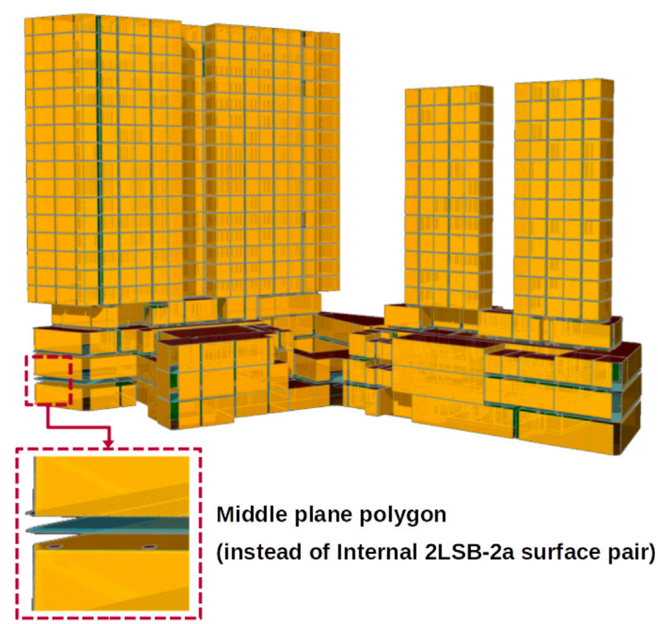


Fig. 4. Example of middle-plane polygon within building slab contained in output gbXML file.

3.3.2. Geometric model simplification

The geometric simplification for BIM2BEM involves two complementary procedures to improve computational efficiency while preserving spatial and simulation accuracy: (1) the merging of spaces into thermal zones based on zoning group definitions from the generated zoning scenarios, and (2) the simplification of polygonal surface geometry.

The first procedure applies predefined space-to-zone mappings to restructure the model geometry. These mappings, provided in a plain-text input file, define how individual IfcSpaces are aggregated into thermal zones. A dedicated tool called Simplification Tool (SMT), which was

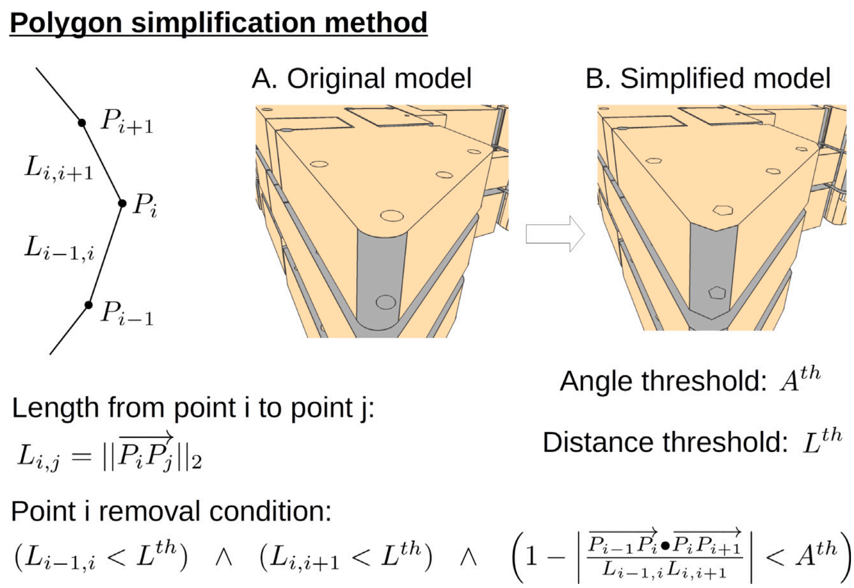


Fig. 5. Illustrative process of the proposed polygon simplification algorithm.

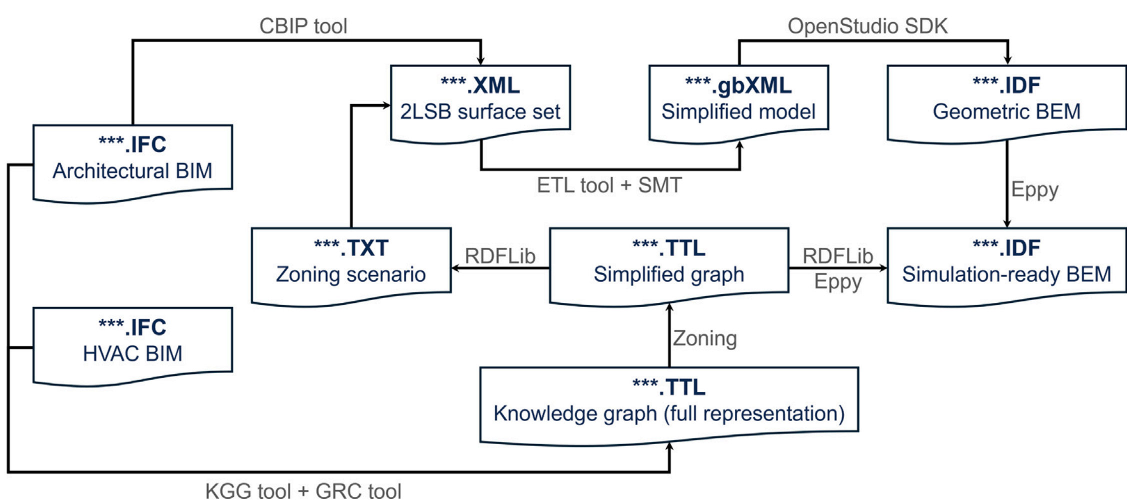


Fig. 6. Proposed BIM2BEM workflow with zoning-based model simplification. KGG = Knowledge Graph Generator; GRC = Geometric Relation Checker; CBIP = Common Boundary Intersection Projection; ETL = Extract-Transform-Load; SMT = Simplification Tool.

introduced and adopted in the previous works [34,46], is used to carry out this process. The SMT identifies adjacent spaces that belong to the same zone and merges their surfaces by constructing connecting planes across the volumetric gaps between them. This produces simplified gbXML geometries with fewer surfaces than the gbXML geometry of the initial unmerged BEM.

Furthermore, in the case of complex or large-scale buildings, the 2LSB surface set provides rich geometric detail; however, this level of complexity can hinder compatibility with simulation engines like EnergyPlus. In particular, curved edges represented as segmented polylines often result in excessive surface triangulation, generating a large number of small mesh elements. This, in turn, significantly increases computational load and can render the simulation process inefficient or even unfeasible. To address this challenge, a polygon simplification algorithm is applied to the boundary polygon points of all 2LSB surfaces, aiming to reduce geometric complexity while maintaining topological consistency. This is illustrated in Fig. 5, where the external and internal 2LSB surface polygons are displayed with orange and grey colours, respectively. As shown in Fig. 5, a boundary point P_i is removed from a polygon's perimeter if the following two conditions are satisfied:

(1) The length of the line segments $\overrightarrow{P_{i-1}P_i}$ and $\overrightarrow{P_i P_{i+1}}$ is smaller than a threshold L^{th}.
(2) The cosine of the angle formed by the line segments $\overrightarrow{P_{i-1}P_i}$ and $\overrightarrow{P_i P_{i+1}}$ differs from one by a quantity less than an angle threshold A^{th}.

As illustrated in Fig. 5, the polygon simplification process transforms geometrically complex contours, such as circular openings and curved slab edges, into simplified representations. In particular, rounded holes are approximated by regular polygons, and curved boundaries are substituted with a connected straight-line segment sequence. This simplification yields a simpler geometric model that aligns better with the requirements of building performance simulation.

3.4. Scalable BIM2BEM workflow

This paper presents an automated workflow for BIM2BEM conversion with zoning-based model simplification leveraging semantic technologies. The corresponding data flow is illustrated in Fig. 6. The primary objective is to generate simulation-ready BEM models for EnergyPlus from imperfect BIM inputs and complex service system

a) IfcSpaces in BIM

Function	Number
BMS	1
Cinema	2
Common	7
Corridor	80
Ensuite	556
Exhibition	1
Kitchen	32
Library	1
Lift	22
Lobby	10
Office	40
Plant	8
Retail	2
Store	11
Toilet	61
Total number	**834**

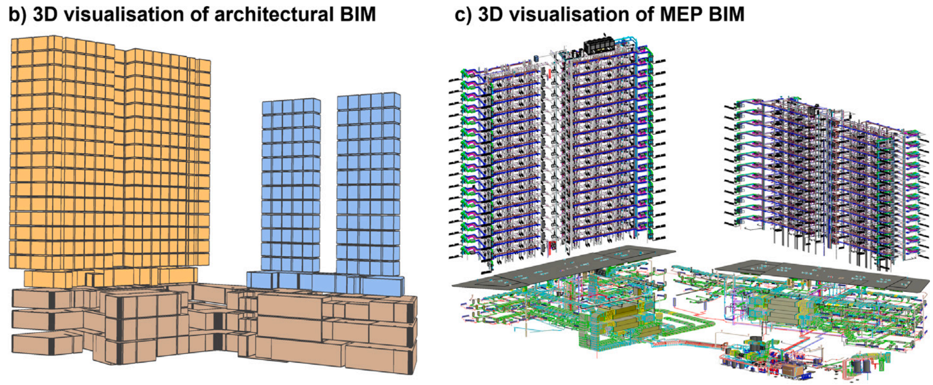


Fig. 7. Overview of the OPS building, including (a) IfcSpace inventory, (b) architectural BIM model, and (c) MEP BIM models.

configurations, while accounting for geometric and system variability. The proposed workflow comprises multiple transformation stages and operates on top of the knowledge graph. During the conversion process, thermal zoning is applied to aggregate multiple IfcSpaces into a single thermal zone, significantly reducing model complexity. Zoning scenarios are defined using semantic queries on the graph, informed by space attributes such as function, adjacency, HVAC system connections, and thermal load similarity, as detailed in the above subsections.

Building on the geometric processing outlined previously, a key feature of the latter stages of the conversion process is enriching the geometric model with simulation-relevant properties. This enrichment mainly consists of passive and active components.

- Passive component enrichment encompasses internal heat gain profiles, including occupancy density, lighting, and equipment loads. These parameters are typically derived from space function classifications and assigned using corresponding IDF object types such as People, Lights, and ElectricEquipment.
- Active component enrichment mainly incorporates HVAC system information based on the topology captured in the knowledge graph, as detailed in the previous work [19]. This task is not straightforward and requires additional graph processing to extract the HVAC topology, representing all logical relationships comprehensively. This ensures that water and air loops are accurately described through upstream and downstream dependencies, as well as primary and branch lines. The implementation is carried out in a Python environment, where RDFLib manages the graph and Eppy enables direct editing of IDF files. System-level configurations are generated using HVACTemplate objects (e.g., Zone, System, and Plant), with subsequent refinements to device types applied through post-processing if needed. Since operational schedules cannot be inferred from topology, they are obtained from operation manuals, with graph indices linking to stored schedule files. This process results in a simulation-ready BEM model.

The BIM2BEM conversion process is implemented using a hybrid C++ and Python environment that integrates several specialised tools. Most geometric data processing, including conversion and simplification, is conducted in C++, while semantic querying, zoning scenario generation, enrichment, and IDF file generation are handled in Python. The primary conversion pathway begins with the BIM model in IFC format, which is converted into XML and then to gbXML. Data quality checks are conducted at each step to ensure reliability. Thermal zoning operations are carried out during the transformation from XML to gbXML, after which the knowledge graph, stored in TTL format, is used to guide the enrichment of the gbXML-derived IDF file with both passive and active settings. This integrated approach results in a simulation-ready BEM model. The workflow is seamless and includes verification at each stage to ensure accuracy and enable traceability of potential errors.

Overall, the proposed BIM2BEM workflow enables the robust, scalable, and accurate model generation of building performance simulations. It is particularly effective in dealing with imperfect BIM data and complex buildings by combining semantic modelling, geometric simplification, and thermal zoning.

Additionally, the proposed workflow does not depend on complete or perfect IFC models. For the architectural BIM, only basic space information (e.g., volumes and boundaries) is required. For the HVAC BIM, geometric representation and essential classification are necessary, along with most equipment-level semantic links. The degree of data completeness primarily affects the level of automation. More complete models enable higher automation, while missing or inconsistent attributes can be supplemented within the knowledge graph. This ensures the workflow remains robust when applied to imperfect BIM inputs.

4. Case study

This work applies the proposed BIM2BEM framework to a real-world case, with this section structured around three core aspects: an overview of the building and systems, ideal load-based clustering analysis, and zoning scenarios for model simplification.

4.1. Overview

This work adopts One Pool Street (OPS), located on the UCL East Campus in London, as the case study building. OPS is a newly constructed, multi-purpose facility equipped with a complex HVAC infrastructure, managed by an integrated building management system. The building has a podium and two towers named Tower East and Tower West. The podium serves as a multi-use area, housing various types of rooms. The towers offer residential spaces, including accommodation units and shared kitchens. The HVAC system incorporates several energy technologies, such as an air-cooled chiller, a district heating connection, and multiple systems, including mechanical ventilation with heat recovery (MVHR) units, AHUs, FCUs, and radiators, to address cooling, heating, and ventilation requirements. Most of the HVAC components for cooling and heating purposes are located in the podium, which covers the ground to the third floor. Additional MVHR units and radiators have been installed in Tower East and Tower West to meet the heating, ventilation, and heat recovery needs of the residential areas.

The BIM data used in this paper are provided in IFC4 format and include both architectural and HVAC information required for BIM2BEM conversion. To manage the large file size and avoid memory issues during processing, the federated BIM model was divided into four separate IFC files, corresponding to architectural elements and MEP systems for the Podium, Tower East, and Tower West. Together, these models define the spatial layout and HVAC configuration. Fig. 7 presents the IfcSpace inventory and 3D visualisations of the architectural and MEP BIM models, where subfigures (a), (b), and (c) correspond to the space inventory, architectural model, and MEP model, respectively.

The architectural BIM model contains 834 IfcSpace entities, each mapped one-to-one to a thermal zone, forming the full-model baseline. The raw IFC data used in this paper, however, lacked explicit representations of external openings such as windows and doors. As a result, corresponding window elements could not be generated in the BEM, which reduces its geometric completeness and limits the physical realism of the model. The proposed framework, nevertheless, is capable of incorporating window information when such elements are present in the BIM model. Specifically, the BIM-MP tool embedded within this framework for geometric IDF generation has already been applied in previous studies [44,46], where it demonstrated the ability to process window information when provided in IFC inputs.

Moreover, the HVAC devices had both geometric representations and classification in the BIM, but many detailed performance parameters of individual components were missing. Despite this limitation, OPS remains an ideal case for evaluating the methodology. Its complex architectural layout and system configuration make complete one-to-one space-level manual modelling highly time-consuming, labour-intensive, and error-prone. In addition, detailed simulations of such a large and intricate model can be computationally expensive. These factors underscore the case's relevance for evaluating the effectiveness and scalability of the proposed BIM2BEM framework.

4.2. Ideal load-based clustering analysis

To facilitate a space-level load similarity analysis, an ideal load simulation was performed using the BEM model generated from the geometry-focused BIM2BEM conversion process. Each IfcSpace directly maps to a thermal zone in a one-to-one relationship. The purpose of this simulation was to independently determine the ideal heating and cooling loads for each space. As described in the preceding subsection, the resulting simulation model comprises 834 thermal zones, each precisely corresponding to one of the 834 IfcSpace instances. Moreover, the ideal load simulation was used solely for the clustering analysis of thermal load similarity in zoning-based scenario generation, whereas the subsequent BEM simulations incorporated detailed HVAC system configurations.

It should be noted that incomplete geometric data, particularly the absence of external openings such as windows, limits the representation of thermal dynamics in the model and thereby affects the accuracy of the ideal load simulation. Moreover, the thermal load of each space is influenced not only by internal heat gains and temperature control settings, but also by spatial factors such as orientation and position of the space within the building. Since the zoning approach spans multiple floors and the building features varying floor heights, the ideal load density was calculated based on unit volume rather than the more common unit floor area. As a result, while the simulation outcomes are sufficient for comparative load similarity analysis across spaces, the model is not suitable for detailed calibration or performance validation. Consequently, the influence of openings on the simulation results was not considered in this paper.

In addition to geometric incompleteness, detailed information on the construction envelope and electrical systems is not available. To address this, standard values were adopted based on building regulations aligned with the design characteristics and construction period of the case building. The thermal transmittance (U-values) used in the simulation are 0.18 W/m^2 K for roofs, 0.25 W/m^2 K for floor slabs, 0.26 W/m^2 K for external walls, and 0.30 W/m^2 K for internal walls. Moreover, the internal heat gain settings are primarily derived from the National Calculation Methodology [47], supplemented by other studies [48,49]. According to building operation manuals for

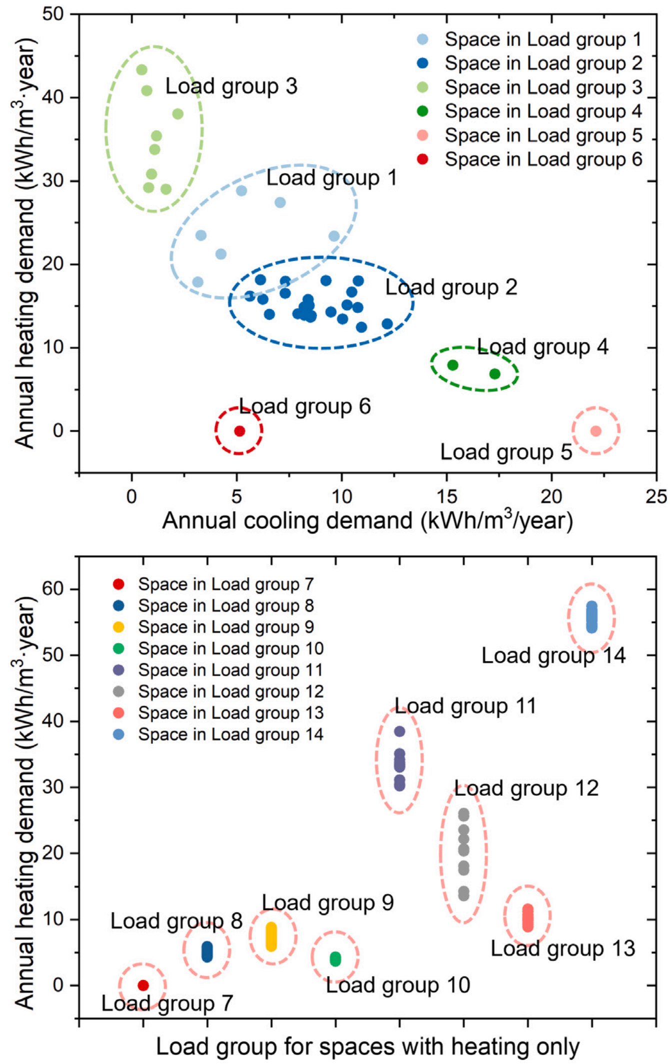


Fig. 8. Distribution and clustering results of spaces, excluding those without heating or cooling loads.

HVAC systems, primary function rooms such as offices and lobbies are typically maintained at 21 ± 1 °C, while secondary function rooms like toilets and store areas are usually supplied with air at 18 °C. The building management system (BMS) room and plant rooms operate with continuous cooling throughout the year.

For the clustering analysis of thermal load similarity, the 834 spaces were first categorised based on their load characteristics into three types: those with both heating and cooling loads, those with only heating loads, and those without any thermal loads. The first two types were analysed separately using the GMM algorithm, with the number of clusters determined according to BIC. This resulted in 6 clusters for the spaces with both heating and cooling demands, and 8 clusters for those with only heating demand. The remaining spaces, which exhibited neither heating nor cooling loads, were treated as a single group. Consequently, all 834 spaces were classified into 15 load groups based on their ideal load similarity. Fig. 8 illustrates the distribution of spaces with their heating and cooling loads, as well as their corresponding groupings according to load similarity.

4.3. Zoning scenarios for model simplification

A set of potential mergeable groups was identified based on the graph-driven thermal zoning strategies introduced earlier, including space function, HVAC configuration consistency, and thermal load similarity. While geometric adjacency is also considered, it serves more as a spatial constraint during geometric data processing rather than a direct factor in scenario generation. Since all three conditions must be met simultaneously to merge spaces, the resulting mapping from IfcSpaces to thermal zones includes various candidate groups. Specifically, 19 groups were identified based on HVAC configuration, 15 based on space function, and an additional 15 derived from the clustering analysis of thermal load similarity. A set of candidate groups was then generated by identifying the intersections across these three dimensions. Each candidate group was defined as the intersection of one HVAC-based group, one function-based group, and one load-based group. Groups in which the intersection included only a single IfcSpace were excluded, since merging requires at least two spaces. After this, 18 preliminary valid groups remained that satisfied all zoning criteria and were selected for model simplification.

Through further filtering, 5 unconditioned groups (spaces without temperature control and terminal devices) and 3 ventilation-only groups (spaces served only by simple ventilation devices) were merged by default, with the spaces inside each group being merged. Since these groups have no heating or cooling demand, they exert very limited influence on zoning outcomes compared with the conditioned spaces. Therefore, the remaining 10 conditioned groups were finally selected for model simplification, forming the basis for the zoning scenario generation. This reduction was necessary given the model complexity, as analysing all 2^{18} combinations (over 260,000 scenarios) would impose an unacceptable computational burden for such a complex building.

The full-model baseline maintains a one-to-one mapping between IfcSpaces and thermal zones, with no merging applied. To represent and manage different zoning scenarios, each scenario was encoded as a binary string in which each digit corresponds to a conditioned group. A digit of "1" indicates that the group is merged, while a digit of "0" indicates that it is not. This approach yields 2^N possible zoning scenarios, where N = 10. Under this scheme, the configuration "0000000000" corresponds to the case in which all unconditioned and ventilation-only groups are merged by default, while no conditioned groups are merged; this is distinct from the full-model baseline having a strict one-to-one mapping. Moreover, the configuration "1111111111" represents the case in which all conditioned groups are merged, in addition to the default merging of unconditioned and ventilation-only groups. A configuration with a single "1" specifies that only the spaces within the corresponding conditioned group are merged, while all other groups remain unmerged. In all scenarios, the unconditioned and ventilation-only groups are merged by default, regardless of the binary configuration, since their contribution to overall building performance is substantially smaller than that of conditioned spaces. This enables the analysis to focus on zoning scenarios that are more relevant to BEM outcomes.

Hence, this setup results in one baseline model and 1024 possible zoning scenarios. However, due to the high computational cost of running simulations for all scenarios, this paper further employs the Sobol sampling method [50] to explore the space of zoning scenario configurations efficiently. Through this method, the number of scenarios was reduced from 1024 to 256 representative samples. Additionally, the baseline, the "all-zero", the "all-one", and all single-merge configurations (those with only one "1") were included, while avoiding any duplicates already covered by the Sobol samples. In total, 256 (Sobol) + 2 ("all-one" and "all-zero") + 10 ("single-1") −1 (duplicates) = 267 unique zoning scenarios were selected to generate corresponding BEM models.

Finally, the zoning scenarios not only guided the SMT tool in simplifying the geometry-related components of the BEM model but also supported the corresponding simplification of the knowledge graph, ensuring consistency with the BEM geometry.

5. Results and discussion

This section illustrates the graph-based building representation, the full-model baseline simulation, and building performance under model simplification.

5.1. Graph-based building representation

This knowledge graph integrates both geometric and HVAC information within a structured node-edge framework, enabling scalable querying, reasoning, and model simplification. Unlike traditional geometric or schematic representations, it supports multi-layer abstraction and ensures data consistency between BIM and BEM, thereby facilitating seamless integration, interoperability, and performance analysis. The constructed knowledge graph, geometric representation, entity inventory, and their interconnections can be found in Fig. 9.

Fig. 9(a) illustrates the abstracted graph structure comprising spaces and HVAC components, where each node represents an entity (e.g., space, VAV, AHU), and each edge encodes a spatial or logical relationship, such as air supply or spatial containment. The geometric view beneath the knowledge graph serves as a visual reference, supporting interpretation of the spatial context and distribution of the elements.

Fig. 9(b) presents a chord diagram illustrating the interconnections among spaces and key HVAC entity types, such as Radiators, FCUs, AHUs, and Chillers. The thickness of each chord indicates the number of connections between two types, thereby highlighting dominant flow paths and subsystem structures. This visualisation reveals the modularity and heterogeneity of the HVAC systems within the OPS building, underscoring the prevalence of specific terminal types. This graph-based representation provides an accurate and structured overview of the system architecture and interconnections for large-scale buildings such as OPS with complex HVAC configurations.

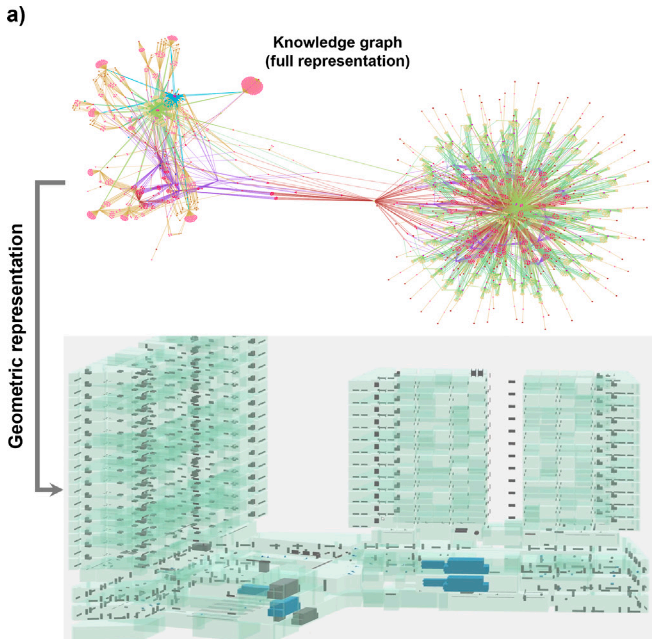
Fig. 9(c) provides a quantitative summary of entity counts extracted from the knowledge graph. The results highlight the diversity and large number of HVAC components, along with more than 800 spatial entities. This level of scale and complexity poses considerable challenges for generating accurate and high-quality BEM models, particularly in terms of transformation fidelity and computational cost. These findings further underscore the importance of the proposed BIM2BEM framework with model simplification in enabling scalable and efficient energy simulations.

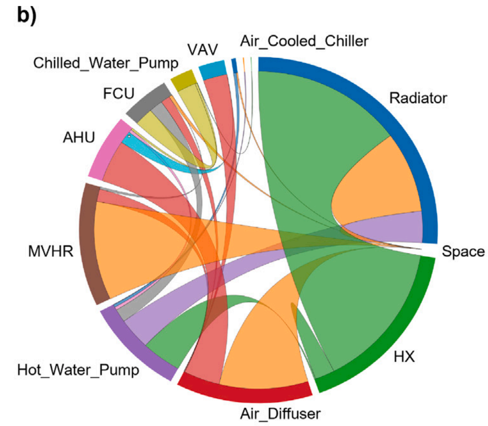
Additionally, the knowledge graph, acting as a back-end, offers significant potential for broader applications, such as real-time monitoring, control integration, and semantic querying. It integrates heterogeneous data, accurately captures complex relationships and supports a data-rich management architecture. It also ensures a consistent format across sources and enables straightforward validation through constraint-based reasoning. These features make it a robust and adaptable foundation for data interoperability and intelligent operation.

5.2. Baseline simulation of the full model

As a reference for evaluating the impact of model simplification, a baseline simulation was conducted using the full BEM model generated from the original BIM data and its knowledge graph without any zoning aggregation. In this configuration, each thermal zone maintains a one-to-one correspondence with its associated space in the BIM model, resulting in a detailed and high-resolution simulation setup. This full-model baseline serves as the benchmark for evaluating the impact of zoning simplification on energy performance and system behaviour.

To ensure consistency across zoning scenarios and to address the absence of detailed performance parameters for HVAC devices, the capacity of each HVAC component in the BEM model was configured using the "Autosize" setting in EnergyPlus. This practical compromise, adopted to handle incomplete input data, enables the simulation engine to automatically determine the appropriate sizing required to





c)

Entity type	Number	Entity type	Number
brick:Space	834	brick:VAV	75
brick:Air_Cooled_Chiller	1	brick:FCU	67
brick:Hot_Water_Pump	10	brick:Radiator	864
brick:Chilled_Water_Pump	4	brick:AHU	9
brick:HX	59	brick:MVHR	177
brick:Air_Diffuser	717	……	

Fig. 9. Overview of graph-based building representation for case study, including (a) knowledge graph and geometric representation, (b) interconnections among entities, and (c) the inventory of entities.

meet thermal loads and zone-level temperature setpoints. At the same time, it provides a consistent basis for comparing scenarios and allows examination of how zoning influences HVAC system sizing.

Given this autosizing setup, the simulation analysis in this subsection focuses on energy consumption, using Energy Use Intensity (EUI), defined in Eq. (1), as the primary performance metric. The EUI enables direct comparison across different zoning configurations, offering a consistent basis for evaluating the performance implications of zoning-based model simplification.

$$\mathrm{EUI} = \frac{\sum_{h=1}^{8760}(\mathrm{Load}_{\mathrm{h}}^{\text{electricity, district heating}} \times 1)}{\text{Floor Area}} \tag{1}$$

Fig. 10 illustrates the complete BIM2BEM process for generating the full-model baseline. Based on the current version of the BIM data, the building has a total floor area of approximately 13,500 m^2, including 11,300 m^2 of conditioned space and 2200 m^2 of unconditioned space. This is slightly smaller than the 17,300 m^2 indicated in the original design documents, primarily due to missing elements such as stairs, lifts, and an auxiliary service building. Additionally, external windows were not included in the original BIM data. While these omissions introduce some geometric discrepancies between the digital model and the actual structure, the case remains valid for testing the proposed methodology. They reflect data quality issues that may arise in practice, yet the model still retains the essential spatial and system information necessary for conducting the BIM2BEM conversion.

Based on the results of building performance simulation, the full-model baseline yielded an electricity EUI of 100.07 kWh/m^2/year, compared to the actual electricity use of approximately 120 kWh/m^2/year recorded by the building's meters for the year 2024. For district heating, the simulated heating EUI was 50.26 kWh/m^2/year, which closely aligns with the metred value of 47 kWh/m^2/year. A detailed breakdown of electricity use across end uses, such as chillers, pumps, ventilation, and lighting, is provided in the pie chart in the bottom right corner of Fig. 10. Despite the BIM model lacking certain geometric elements, such as windows and some internal spaces, the deviations of 16.6% in electricity and 6.9% in heating fall within the ranges reported in recent building energy simulation studies [51,52], in calibration review papers [53,54], and in ASHRAE Guideline 14 [55]. These results suggest that the full-model baseline provides a reliable basis for evaluating the impact of zoning-based model simplification on energy performance. More detailed calibration is achievable in future once more comprehensive data, including window information and complete on-site weather records, becomes available.

5.3. *Building performance under model simplification*

This section examines the impact of zoning-based model simplification on building performance from three key perspectives, including simulation efficiency, energy performance, and HVAC system sizing. The analysis is based on 267 zoning scenarios coupled with the BIM2BEM conversion, each representing a different level of spatial aggregation. The following subsections elaborate on how model simplification influences simulation time, energy consumption, and HVAC system sizing, offering insights into the behaviour of building performance simulations under varying levels of model complexity.

Fig. 11 presents the geometry of the baseline full-model BEM alongside 12 simplified models based on some representative zoning scenarios. These include the "all-zero" scenario, 10 "single-1" scenarios, and the "all-one" scenario. In this context, a value of "1" indicates that a specific group of spaces has been merged. This means that multiple IfcSpaces are merged into one or more thermal zones, taking into account spatial adjacency. The remaining 255 scenarios generated through Sobol sampling are not included here, as they represent different combinations of these 10 "single-1" scenarios.

5.3.1. *Simulation efficiency across zoning scenarios*

Fig. 12 illustrates the relationship between the number of thermal zones and simulation time across the 267 zoning scenarios. Generally, zoning-based model simplification reduces the number of thermal zones, thereby shortening simulation time. The full-model baseline, which maintains a one-to-one mapping between individual IfcSpaces and thermal zones, comprises 834 zones and takes approximately 1200 s to complete the simulation. In comparison, the most simplified scenario ("all-one"), where the spaces within all mergeable groups are merged accordingly, reduces the number of zones to fewer than 100 and lowers the simulation time to just under 400 s. This corresponds to a time reduction of around 70% relative to the full-model baseline, demonstrating the substantial efficiency gains enabled by zoning-based model simplification.

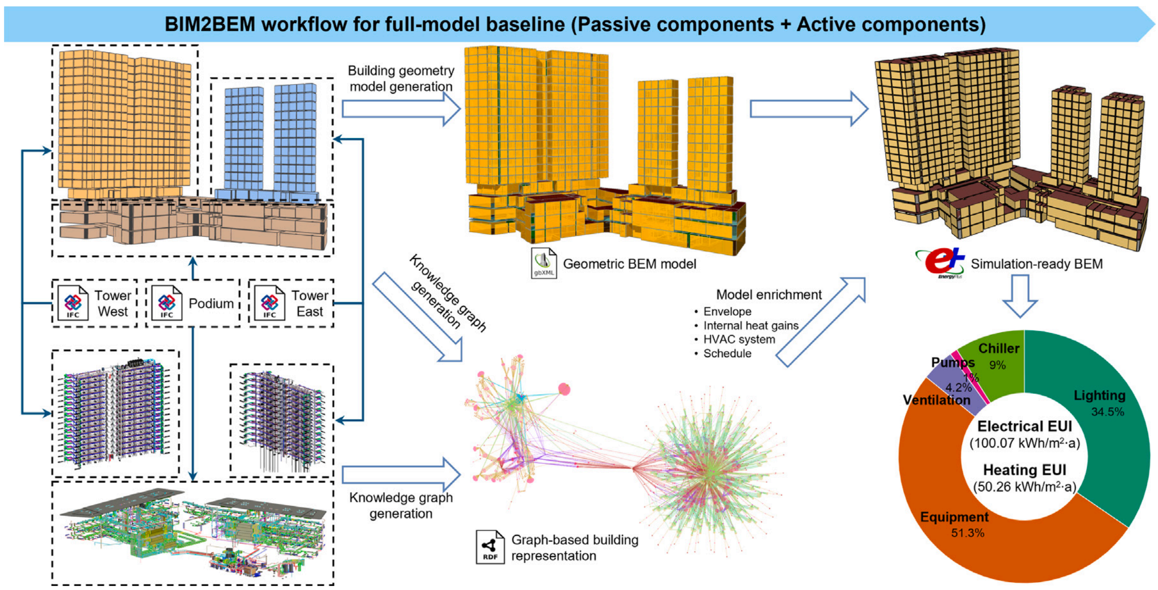


Fig. 10. BIM2BEM process for generating the baseline full-model building performance simulation.

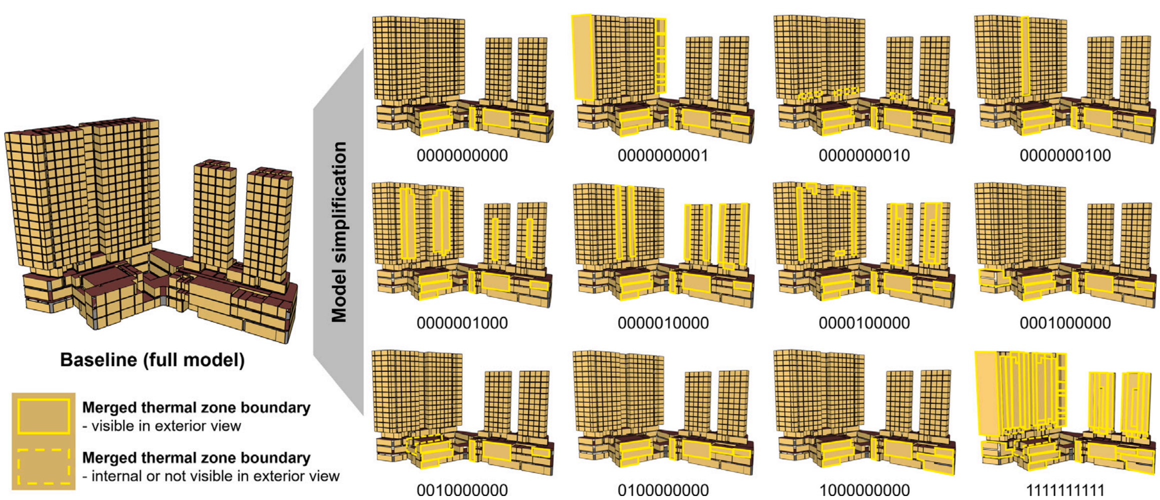


Fig. 11. Geometric representations of the baseline full-model BEM and 12 simplified models derived from representative zoning scenarios.

However, Fig. 12 also shows that simulation time does not always decrease in proportion to the number of zones. In some scenarios, merging spaces across different floors with complex geometric configurations can introduce additional computational overhead, particularly in zones that involve airflow or ventilation calculations. It can lead to large-volume thermal zones that slow down convergence during simulation, irregular zone shapes, and more intricate surface relationships, which may offset the expected efficiency gains. This effect is especially evident in some mergeable groups where merged zones span multiple levels or include diverse system types. Despite these exceptions, the overall trend confirms that zoning-based model simplification enhances simulation efficiency and is especially beneficial for large-scale building energy models with complex service system configurations.

5.3.2. Energy performance across zoning scenarios

Fig. 13 illustrates the distribution of HVAC-related electricity and heating energy use across all zoning scenarios. Each point represents a zoning scenario, with its position indicating the simulated electricity and heating energy use, and its colour reflecting the number of thermal zones. The results show that zoning-based model simplification, when guided by HVAC system configuration, space function, and thermal load similarity, has a limited impact on overall energy use. For HVAC-related electricity use, the baseline model yields approximately 191,000 kWh. Across all zoning scenarios, values for the simplified models fall within the range of 187,000 kWh to 196,000 kWh, corresponding to a deviation of under 3%. For heating energy use, the baseline model yields a value of approximately 674,000 kWh, while the highest value among the simplified models reaches 688,000 kWh, corresponding to a deviation of about 2%.

As the level of zoning-based model simplification increases, indicated by a darker colour gradient, deviations from the baseline become more noticeable, particularly in heating energy use. In contrast, electricity use displays a more scattered pattern, with values fluctuating above and below the baseline without a consistent trend. These findings indicate that even with substantial reductions in model resolution, the energy performance of the simplified models remains consistent and robust. While heating energy use tends to increase slightly in most simplified BEM models, and electricity use shows no systematic variation, both deviations remain small. These results suggest that the proposed zoning-based model simplification method can maintain accurate energy assessments even in large buildings with complex geometries and diverse HVAC configurations.

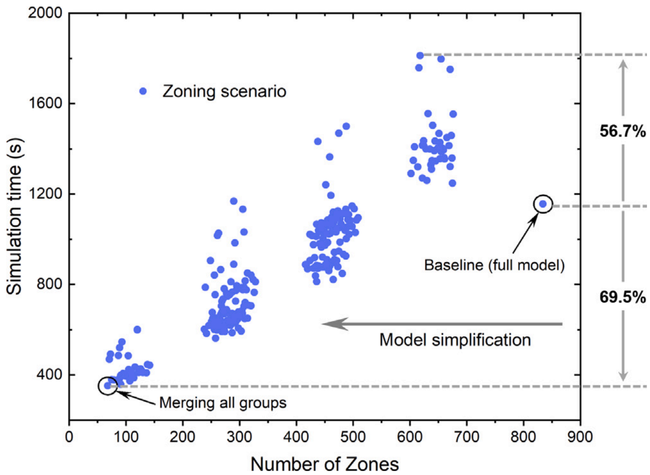


Fig. 12. Scatter plot of simulation time versus number of thermal zones across zoning scenarios.

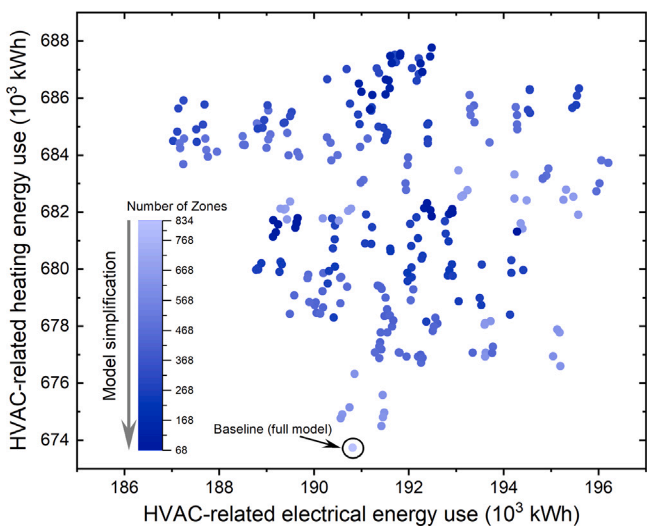


Fig. 13. Scatter plot of HVAC-related electricity and heating energy use across zoning scenarios.

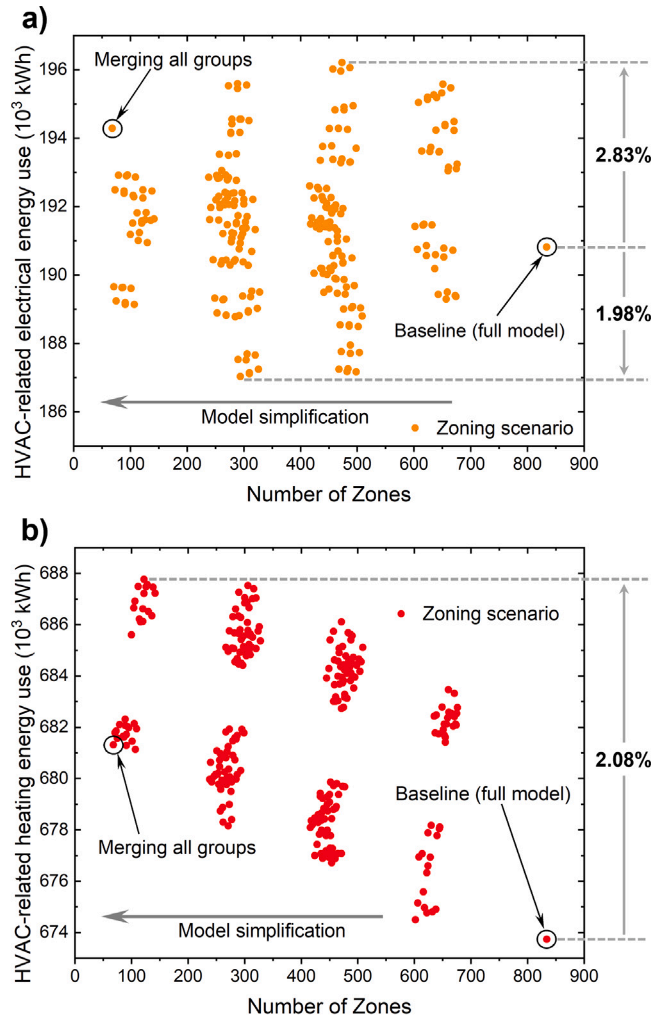


Fig. 14. Impact of model simplification on HVAC-related energy use for (a) electricity and (b) heating.

Fig. 14 further examines the relationship between the number of thermal zones and HVAC-related energy use, separating the results into (a) electricity use and (b) heating use. Compared to Fig. 13, this presentation offers a clearer view of how the degree of zoning-based model simplification influences each energy metric. The plots also explicitly show the deviation ranges introduced by zoning. The scatter patterns reveal that dots tend to cluster around specific zone counts, reflecting the structure of the predefined zoning groups. This indicates that the merging of intra-group spaces influences HVAC-related energy calculations to varying degrees.

The most simplified BEM model, in which all mergeable groups are merged into the minimum number of zones, does not produce the most significant deviation from the full-model baseline. This suggests that simplified models can still closely align with the full-model baseline when zoning decisions consider system configuration, space function, and load similarity. These findings reinforce that, when properly applied, zoning-based model simplification can preserve the accuracy of energy performance simulations while substantially improving computational efficiency.

In addition, Fig. 15 illustrates the impact of model simplification on solar heat gain on exterior surfaces. As shown in Fig. 11, zoning-based simplification slightly alters the exterior geometry because some wall elements are combined when the associated spaces are merged. This results in moderate variations in solar heat gain across the zoning scenarios. The full-model baseline, with the most detailed façade representation, yields the lowest annual solar heat gain, whereas the most simplified model records the highest value, about 10.6% above the baseline. This difference mainly arises from the combined effects of altered thermal coupling among interior zones and minor geometric changes to exterior walls introduced by the merging process. These factors affect the internal heat storage and transfer behaviour, slightly modifying the temperature distribution on exterior surfaces. Consequently, the simplified models tend to overpredict absorbed solar radiation, although the deviation remains small. However, the additional solar absorption is not fully utilised for space heating due to reduced thermal inertia and interzone heat transfer. This partly explains why heating energy use and solar heat gains vary in the same direction across zoning scenarios, as model simplification leads to a less accurate representation of heat redistribution within the building. Moreover, since the BIM data used in this case do not include windows, this deviation might become more pronounced if windows are considered. Overall, the zoning-based model simplification improves computational efficiency but introduces a minor loss of accuracy in estimating solar heat gains, which in turn affects the prediction of heating and cooling demand.

Nevertheless, while the full-model baseline lies close to the centre of the range for electricity use across zoning scenarios, it yields the lowest

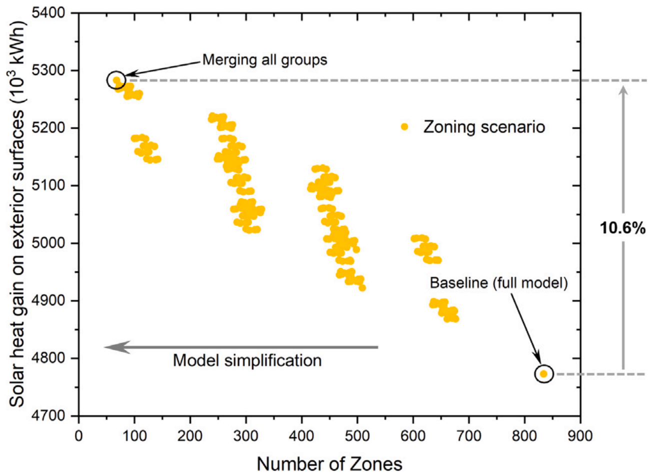


Fig. 15. Impact of model simplification on exterior surface solar heat gain.

heating energy use, with simplified models tending to overpredict. This may primarily result from the loss of load diversity after space aggregation. Although zoning groups were defined considering thermal load similarity, space merging may still reduce the temporal variability of thermal behaviour, coinciding peak loads and increasing heating demand. The loss of thermal inertia and reduced heat redistribution further reinforce this tendency. A secondary factor may involve the altered treatment of internal partitions, which in the full-model baseline helps offset heat gains and losses between adjacent spaces. Despite this, the deviations in energy use relative to the full-model baseline remain small, suggesting that the proposed approach can deliver reasonably reliable results under model simplification.

5.3.3. HVAC system sizing across zoning scenarios

To assess the impact of zoning-based model simplification on HVAC system sizing during simulation, this subsection examines the capacities of key components across all zoning scenarios. Four representative HVAC components are considered: central sources (chiller and district heating), AHUs, FCUs, and radiators. As shown in Fig. 16, the violin plots illustrate the distribution of their capacities across the 267 scenarios, with particular emphasis on the full-model baseline and the most simplified BEM model of the "all-one" scenario.

First, the results show that sizing variation across zoning scenarios is relatively limited. For all critical components, the distribution range remains within 10%, demonstrating the robustness of the proposed multi-factor zoning approach in the BIM2BEM conversion process. This is particularly evident in the central sources shown in Fig. 16(a). The chiller capacity, which serves only the public zones in the podium, remains closely aligned with the baseline. District heating, covering the entire building, exhibits a wider spread due to its broader service scope. However, the deviation from the baseline remains under 5% even in the most simplified model ("all-one" scenario). While the zoning-based model simplification slightly reduces the estimated chiller capacity, it tends to increase the capacity of district heating. This trend is specific to the case but remains within a narrow and acceptable range.

For the air-supplying subsystems, as shown in Fig. 16(b–c), zoning-based model simplification leads to larger thermal zone volumes, thereby increasing the airflow demands for both AHUs and FCUs. This increase is consistent across zoning scenarios, mainly driven by space merging across floors, though the magnitude remains limited. The coil capacities, however, show less consistent behaviour, while they may either increase or decrease depending on the spatial arrangement and functional usage of the merged spaces. Despite this variability, the estimated coil capacities remain within a reasonable range, supporting the applicability of the proposed zoning strategy. For radiators, as shown in Fig. 16(d), which operate through thermal radiation rather than air supply, a more apparent trend is observed. As zoning merges spaces, the associated increase in zone volume and fresh air requirements tends to result in consistently higher radiator sizing across zoning scenarios. In general, the full-model baseline produces the lowest sizing values for heating-related end-use components, consistent with its lower heating demand, which can be attributed to reduced load diversity and the treatment of internal partitions, as noted in the previous subsection.

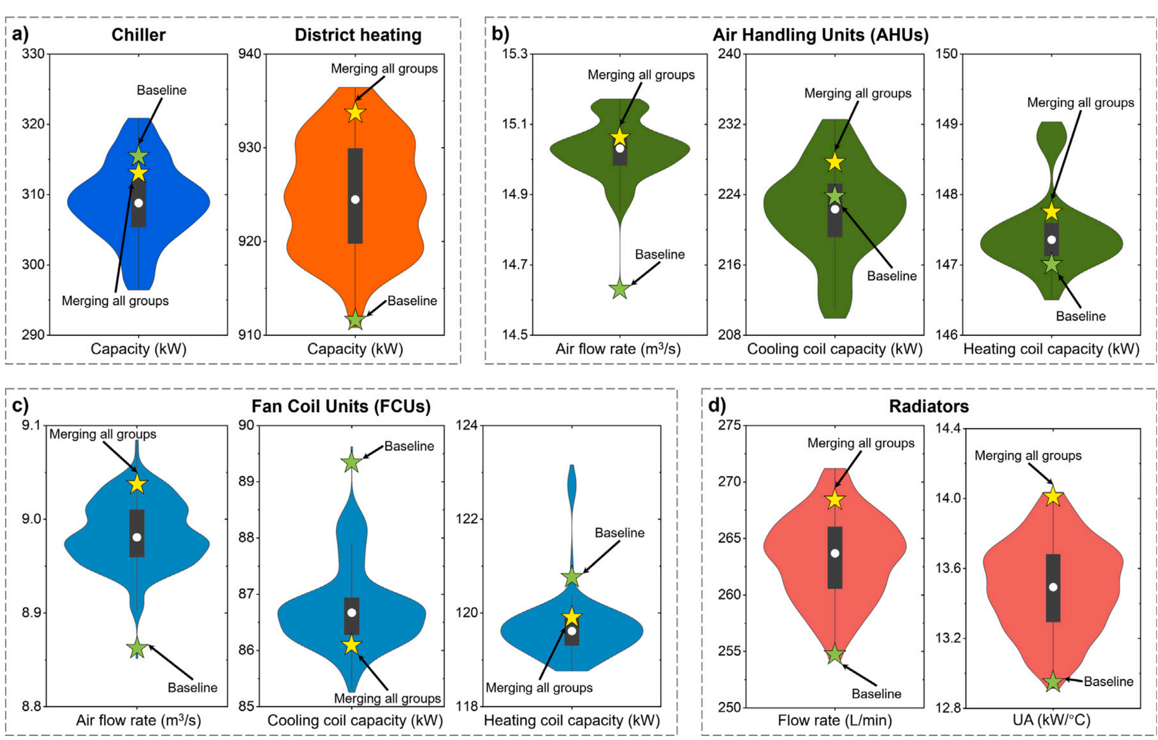


Fig. 16. Distribution of HVAC system sizing across zoning scenarios in terms of major equipment.

Overall, the impact of zoning-based model simplification on HVAC system sizing is limited, indicating that the proposed method successfully reduces simulation time and model complexity without compromising the reliability of sizing outcomes. The results confirm that the graph-driven thermal zoning enables consistent and accurate system sizing within the BIM2BEM conversion process, even under substantial spatial aggregation. Deviations in both energy performance and HVAC sizing remain minor and within acceptable limits, underscoring the robustness of the proposed methodology and its strong potential for broader application, particularly in large and complex buildings with intricate service systems.

6. Conclusions

This paper presented an integrated and synchronised framework that unifies BIM2BEM conversion, knowledge-graph integration, and zoning-based model simplification into a seamless workflow. When applied to a large, complex real-world building, this BIM2BEM framework helps address challenges related to imperfect BIM data and excessively detailed geometry, enabling the generation of well-structured and high-reliability building performance simulations. The results indicate that the proposed framework preserves the modelling accuracy of the generated BEMs while substantially improving simulation efficiency. The findings also highlight both the research and the practical relevance of embedding zoning-based model simplification and knowledge graph-based digitalisation within BIM2BEM conversion processes. The proposed framework offers a scalable and robust solution for delivering simulation-ready models that support performance-driven building design and the assessment of different operational strategies, which is particularly valuable for large and complex buildings where developing fully detailed BEMs is both time-consuming and error-prone. In summary, the main conclusions, limitations, and future research directions are summarised below.

(1) The proposed framework integrates BIM2BEM conversion, knowledge graph-based digitalisation, and zoning-based model simplification into a unified workflow, demonstrating the capability to handle imperfect BIM data and generate BEMs with appropriate levels of complexity.
(2) Geometric data processing methods are developed to extract 2LSB and simplify complex polygons, thereby reducing geometric complexity while preserving topological consistency and improving compatibility with simulation tools.
(3) A comprehensive knowledge graph is constructed to digitalise the building information and represent the relationships between spatial elements and HVAC components. Based on the graph structure, zoning scenarios are generated to map IfcSpaces to thermal zones by analysing multiple factors such as adjacency, function, thermal load similarity, and HVAC configuration.
(4) In the case study, the zoning-based model simplification improved simulation efficiency, achieving up to a 70% reduction in simulation time compared to the full-model baseline.
(5) Across the generated zoning scenarios, the zoning-based model simplification produced consistent energy performance and HVAC system sizing. Deviations in HVAC electricity and heating energy use were within approximately 3% of the full-model baseline, while system sizing variations reached up to about 10% in this case study.

Although the case study demonstrates the proposed framework's ability to support scalable and automated BIM2BEM conversion while accommodating imperfect BIM data, several limitations remain. First, while semantic technologies with knowledge graphs can identify and pinpoint data quality issues, repairing incomplete or low-quality BIM inputs still requires manual intervention, meaning that the level of automation decreases significantly when the source data are of very poor quality or insufficient detail. Second, the case study used in this work did not include external window elements in the BIM model. Hence, the applicability of the framework to buildings with complex openings still needs to be verified. This may also limit the reliability of evaluating energy performance differences under various thermal zoning scenarios. Third, the observed deviations of around 10% in HVAC system sizing highlight the need for a calibration module to ensure closer alignment between building performance simulations and real design requirements, thereby enhancing the reliability of the generated BEMs. Finally, despite the case building being complex in both geometry and HVAC configuration, the proposed framework has so far only been tested on a single project. Broader validation across a wider range of building types, climatic conditions, and operational contexts will be necessary to confirm the generalisability of the proposed framework.

Future work will therefore focus on several directions. First, beyond detecting data quality issues, future studies could investigate automated error correction, either by repairing missing or inconsistent information directly within the BIM model or within the digital representation (e.g., knowledge graphs). In particular, the integration of window-related details, when available, will be a priority to enhance the reliability of the generated BEMs and their simulation outcomes. Second, the framework should be applied to a wider range of building types and contexts to test its robustness under different design practices and conditions. Third, incorporating dynamic real-time IoT data would allow the transition from a static digital representation to a dynamic digitalisation process. This would move the work towards a building digital twin with bidirectional data flow and enable real-time calibration of the BEMs, thereby improving the alignment of simulation outcomes with actual building operation.

CRediT authorship contribution statement

Meng Wang: Writing – original draft, Methodology, Investigation, Conceptualization. **Georgios N. Lilis:** Writing – original draft, Software, Investigation. **Dimitris Mavrokapnidis:** Software, Resources. **Kyriakos Katsigarakis:** Software, Resources. **Ivan Korolija:** Writing – review & editing, Project administration. **Dimitrios Rovas:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT-4 in order to improve grammar, correct errors, and enhance readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Horizon Europe DigiBUILD project under contract No. 101069658 and co-funded by the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee grant number 10040988. It was also supported by the Horizon Europe BuildON project under contract No. 101104141 and co-funded by UKRI under the UK government's Horizon Europe funding guarantee grant number 10069441, and by UKRI / ESRC under grant ES/Z504671/1 (Smart ENergy data ServicE – SENSE).

Data availability

A portion of the derived data presented in this paper is publicly available on GitHub at: https://github.com/ucl-sbde.

References

[1] V. Bazjanac, IFC BIM-Based Methodology for Semi-Automated Building Energy Performance Simulation, Berkeley, CA (US), Ernest Orlando Lawrence Berkeley National Laboratory, 2008, URL https://www.osti.gov/biblio/938422.

[2] H. Gao, C. Koch, Y. Wu, Building information modelling based building energy modelling: A review, Appl. Energy 238 (2019) 320–343, http://dx.doi.org/10.1016/j.apenergy.2019.01.032, URL https://www.sciencedirect.com/science/article/pii/S0306261919300327.

[3] M. Alhammad, M. Eames, R. Vinai, Enhancing Building Energy Efficiency through Building Information Modeling (BIM) and Building Energy Modeling (BEM) Integration: A Systematic Review, Buildings 14 (3) (2024) http://dx.doi.org/10.3390/buildings14030581, URL https://www.mdpi.com/2075-5309/14/3/581.

[4] Z. Pezeshki, A. Soleimani, A. Darabi, Application of BEM and using BIM database for BEM: A review, J. Build. Eng. 23 (2019) 1–17, http://dx.doi.org/10.1016/j.jobe.2019.01.021, URL https://www.sciencedirect.com/science/article/pii/S2352710218309501.

[5] L. Zheng, W. Lu, L. Wu, Q. Zhou, A review of integration between BIM and CFD for building outdoor environment simulation, Build. Environ. 228 (2023) 109862, http://dx.doi.org/10.1016/j.buildenv.2022.109862, URL https://www.sciencedirect.com/science/article/pii/S0360132322010927.

[6] E. Kamel, A.M. Memari, Review of BIM's application in energy simulation: Tools, issues, and solutions, Autom. Constr. 97 (2019) 164–180, http://dx.doi.org/10.1016/j.autcon.2018.11.008, URL https://www.sciencedirect.com/science/article/pii/S0926580517304958.

[7] I.J. Ramaji, J.I. Messner, E. Mostavi, IFC-Based BIM-to-BEM Model Transformation, J. Comput. Civ. Eng. 34 (3) (2020) 04020005, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000880.

[8] G.N. Lilis, G.I. Giannakis, D.V. Rovas, Automatic generation of second-level space boundary topology from IFC geometry inputs, Autom. Constr. 76 (2017) 108–124, http://dx.doi.org/10.1016/j.autcon.2016.08.044, URL https://www.sciencedirect.com/science/article/pii/S0926580516301984.

[9] H. Ying, S. Lee, An algorithm to facet curved walls in IFC BIM for building energy analysis, Autom. Constr. 103 (2019) 80–103, http://dx.doi.org/10.1016/j.autcon.2019.03.004, URL https://www.sciencedirect.com/science/article/pii/S0926580518302358.

[10] D. Lobos Calquín, R. Mata, J.C. Vielma, J.C. Beaumont-Sepulveda, C. Correa, E. Nuñez, E. Forcael, D. Blanco, P. Pulgar, A simplified framework to integrate databases with building information modeling for building energy assessment in multi-climate zones, Sustainability 16 (14) (2024) http://dx.doi.org/10.3390/su16146123, URL https://www.mdpi.com/2071-1050/16/14/6123.

[11] A.J. Guízar Dena, M. del Carmen García Hípola, C. Fernández Bandera, Optimization testing for the modeling and characterization of three-dimensional elements to enhance interoperability from building information modeling (BIM) to building energy modeling (BEM), Energy Build. 317 (2024) 114394, http://dx.doi.org/10.1016/j.enbuild.2024.114394, URL https://www.sciencedirect.com/science/article/pii/S0378778824005103.

[12] Y. Yang, Y. Pan, F. Zeng, Z. Lin, C. Li, A gbXML reconstruction workflow and tool development to improve the geometric interoperability between BIM and BEM, Buildings 12 (2) (2022) 221, http://dx.doi.org/10.3390/buildings12020221.

[13] M.H. Elnabawi, Building information modeling-based building energy modeling: investigation of interoperability and simulation results, Front. Built Environ. 6 (2020) 573971, http://dx.doi.org/10.3389/fbuil.2020.573971.

[14] J.B. Kim, W. Jeong, M.J. Clayton, J.S. Haberl, W. Yan, Developing a physical BIM library for building thermal energy simulation, Autom. Constr. 50 (2015) 16–28, http://dx.doi.org/10.1016/j.autcon.2014.10.011, URL https://www.sciencedirect.com/science/article/pii/S0926580514002222.

[15] W. Jeong, J.B. Kim, M. Clayton, J. Haberl, W. Yan, Translating building information modeling to building energy modeling using model view definition, TheScientificWorldJournal 2014 (2014) 638276, http://dx.doi.org/10.1155/2014/638276.

[16] H. Kiavarz, M. Jadidi, A. Rajabifard, G. Sohn, An explainable & prescriptive solution for space-based energy consumption optimization using BIM data & genetic algorithm, J. Build. Eng. 92 (2024) 109763, http://dx.doi.org/10.1016/j.jobe.2024.109763.

[17] H. Li, J. Zhang, Improving IFC-Based Interoperability between BIM and BEM Using Invariant Signatures of HVAC Objects, J. Comput. Civ. Eng. 37 (2) (2023) 04022059, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0001063.

[18] Z. Chen, Z. Deng, A. Chong, Y. Chen, AutoBPS-BIM: A toolkit to transfer BIM to BEM for load calculation and chiller design optimization, Build. Simul. 16 (7) (2023) 1287–1298, http://dx.doi.org/10.1007/s12273-023-1006-4.

[19] M. Wang, G.N. Lilis, D. Mavrokapnidis, K. Katsigarakis, I. Korolija, D. Rovas, A knowledge graph-based framework to automate the generation of building energy models using geometric relation checking and HVAC topology establishment, Energy Build. 325 (2024) 115035, http://dx.doi.org/10.1016/j.enbuild.2024.115035, URL https://www.sciencedirect.com/science/article/pii/S0378778824011514.

[20] O. Spielhaupter, BIM to BEM Transformation Workflows: A Case Study Comparing Different IFC-Based Approaches, (Diplomarbeit), Technische Universität Wien, Vienna, Austria, 2021, http://dx.doi.org/10.34726/hss.2021.79545.

[21] IESVE, BIM-to-BEM Interoperability: Pollination's Integration with IES-VE, Revit and Rhino, 2025, URL https://www.iesve.com/discoveries/view/52643/interoperability-bim-to-bem (Accessed: 3 October 2025).

[22] G. Bastos Porsani, K. Del Valle de Lersundi, A. Sánchez-Ostiz Gutiérrez, C. Fernández Bandera, Interoperability between building information modelling (BIM) and building energy model (BEM), Appl. Sci. 11 (5) (2021) 2167, http://dx.doi.org/10.3390/app11052167.

[23] A. Andriamamonjy, D. Saelens, R. Klein, An automated IFC-based workflow for building energy performance simulation with Modelica, Autom. Constr. 91 (2018) 166–181, http://dx.doi.org/10.1016/j.autcon.2018.03.019, URL https://www.sciencedirect.com/science/article/pii/S0926580517308282.

[24] M. Shin, J.S. Haberl, Thermal zoning for building HVAC design and energy simulation: A literature review, Energy Build. 203 (2019) 109429, http://dx.doi.org/10.1016/j.enbuild.2019.109429, URL https://www.sciencedirect.com/science/article/pii/S0378778819312769.

[25] M. Wang, G.N. Lilis, K. Katsigarakis, D. Mavrokapnidis, I. Korolija, D. Rovas, Automatic HVAC Topology Generation Using BIM Geometry Checking and Knowledge Graph Technologies, in: Proceedings of the 2024 European Conference on Computing in Construction, in: Computing in Construction, 5, European Council on Computing in Construction, Chania, Greece, 2024, http://dx.doi.org/10.35490/EC3.2024.257, URL https://ec-3.org/publications/conference/paper/?id=EC32024_257.

[26] K. Forth, A. Borrmann, Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM, J. Build. Eng. 95 (2024) 110312, http://dx.doi.org/10.1016/j.jobe.2024.110312, URL https://www.sciencedirect.com/science/article/pii/S2352710224018801.

[27] B. Balaji, A. Bhattacharya, G. Fierro, J. Gao, J. Gluck, D. Hong, A. Johansen, J. Koh, J. Ploennigs, Y. Agarwal, M. Bergés, D. Culler, R.K. Gupta, M.B. Kjærgaard, M. Srivastava, K. Whitehouse, Brick : Metadata schema for portable smart building applications, Appl. Energy 226 (2018) 1273–1292, http://dx.doi.org/10.1016/j.apenergy.2018.02.091, URL https://www.sciencedirect.com/science/article/pii/S0306261918302162.

[28] V. Kukkonen, A. Kücükavci, M. Seidenschnur, M.H. Rasmussen, K.M. Smith, C.A. Hviid, An ontology to support flow system descriptions from design to operation of buildings, Autom. Constr. 134 (2022) 104067, http://dx.doi.org/10.1016/j.autcon.2021.104067, URL https://www.sciencedirect.com/science/article/pii/S0926580521005185.

[29] D. Mavrokapnidis, G.N. Lilis, K. Katsigarakis, I. Korolija, D. Rovas, Semi-automated extraction of HVAC system topology from imperfect Building Information Models, in: Proceedings of Building Simulation 2023: 18th Conference of IBPSA, in: Building Simulation, 18, IBPSA, Shanghai, China, 2023, pp. 964–971, http://dx.doi.org/10.26868/25222708.2023.1392, URL https://publications.ibpsa.org/conference/paper/?id=bs2023_1392.

[30] J. Werbrouck, P. Pauwels, M. Bonduel, J. Beetz, W. Bekers, Scan-to-graph: Semantic enrichment of existing building geometry, Autom. Constr. 119 (2020) 103286, http://dx.doi.org/10.1016/j.autcon.2020.103286, URL https://www.sciencedirect.com/science/article/pii/S0926580520300674.

[31] A. Kücükavci, M. Seidenschnur, P. Pauwels, M.H. Rasmussen, C.A. Hviid, Efficient management and compliance check of HVAC information in the building design phase using Semantic Web technologies, Semant. Web 15 (5) (2024) 1959–1989, http://dx.doi.org/10.3233/SW-243595, URL https://journals.sagepub.com/doi/abs/10.3233/SW-243595.

[32] T. Wang, V.J. Gan, D. Hu, H. Liu, Digital twin-enabled built environment sensing and monitoring through semantic enrichment of BIM with SensorML, Autom. Constr. 144 (2022) 104625, http://dx.doi.org/10.1016/j.autcon.2022.104625, URL https://www.sciencedirect.com/science/article/pii/S0926580522004952.

[33] C. Boje, A. Guerriero, S. Kubicki, Y. Rezgui, Towards a semantic Construction Digital Twin: Directions for future research, Autom. Constr. 114 (2020) 103179, http://dx.doi.org/10.1016/j.autcon.2020.103179, URL https://www.sciencedirect.com/science/article/pii/S0926580519314785.

[34] G.N. Lilis, G.I. Giannakis, K. Katsigarakis, D. Rovas, Space boundary topology simplification for building energy performance simulation speed-up, in: Proceedings of Building Simulation 2019: 16th Conference of IBPSA, International Building Performance Simulation Association (IBPSA), 2019, pp. 175–181, http://dx.doi.org/10.26868/25222708.2019.210693.

[35] M. Georgescu, I. Mezić, Building energy modeling: A systematic approach to zoning and model reduction using Koopman Mode Analysis, Energy Build. 86 (2015) 794–802, http://dx.doi.org/10.1016/j.enbuild.2014.10.046, URL https://www.sciencedirect.com/science/article/pii/S0378778814008913.

[36] M. Shin, J.S. Haberl, A procedure for automating thermal zoning for building energy simulation, J. Build. Eng. 46 (2022) 103780, http://dx.doi.org/10.1016/j.jobe.2021.103780, URL https://www.sciencedirect.com/science/article/pii/S2352710221016387.

[37] Y. Chen, T. Hong, Impacts of building geometry modeling methods on the simulation results of urban building energy models, Appl. Energy 215 (2018) 717–735, http://dx.doi.org/10.1016/j.apenergy.2018.02.073, URL https://www.sciencedirect.com/science/article/pii/S0306261918301983.

[38] Z. Wu, J.C. Cheng, Z. Wang, H.H. Kwok, An ontology-based framework for automatic building energy modeling with thermal zoning, Energy Build. 296 (2023) 113267, http://dx.doi.org/10.1016/j.enbuild.2023.113267, URL https://www.sciencedirect.com/science/article/pii/S0378778823004978.

[39] G. Gourlis, I. Kovacic, A holistic digital twin simulation framework for industrial facilities: BIM-based data acquisition for building energy modeling, Front. Built Environ. 8 (2022) 918821, http://dx.doi.org/10.3389/fbuil.2022.918821, URL https://www.frontiersin.org/journals/built-environment/articles/10.3389/fbuil.2022.918821.

[40] D. Mavrokapnidis, K. Katsigarakis, P. Pauwels, E. Petrova, I. Korolija, D. Rovas, A linked-data paradigm for the integration of static and dynamic building data in digital twins, in: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 2021, pp. 369–372, http://dx.doi.org/10.1145/3486611.3491125.

[41] D. Mavrokapnidis, Enabling scalable deployment of data-driven applications across building portfolios, in: Proceedings of the 2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference, in: Computing in Construction, vol. 4, European Council on Computing in Construction, Heraklion, Greece, 2023, http://dx.doi.org/10.35490/EC3.2023.342, URL https://ec-3.org/publications/conference/paper/?id=EC32023_342.

[42] G.N. Lilis, M. Wang, K. Katsigarakis, D. Mavrokapnidis, I. Korolija, R. Dimitrios, BIM-based semantic enrichment and knowledge graph generation via geometric relation checking, Autom. Constr. 173 (2025) 106081, http://dx.doi.org/10.1016/j.autcon.2025.106081, URL https://www.sciencedirect.com/science/article/pii/S0926580525001219.

[43] K. Li, Z. Ma, D. Robinson, J. Ma, Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering, Appl. Energy 231 (2018) 331–342, http://dx.doi.org/10.1016/j.apenergy.2018.09.050, URL https://www.sciencedirect.com/science/article/pii/S0306261918313606.

[44] K. Katsigarakis, G.N. Lilis, D. Rovas, A cloud IFC-based BIM platform for building energy performance simulation, in: Proceedings of the 2021 European Conference on Computing in Construction, in: Computing in Construction, vol. 2, European Council on Computing in Construction, Online Conference, 2021, pp. 350–357, http://dx.doi.org/10.35490/EC3.2021.177, URL https://ec-3.org/publications/conference/paper/?id=EC32021_177.

[45] H. Ying, S. Lee, Generating second-level space boundaries from large-scale IFC-compliant building information models using multiple geometry representations, Autom. Constr. 126 (2021) 103659, http://dx.doi.org/10.1016/j.autcon.2021.103659, URL https://www.sciencedirect.com/science/article/pii/S0926580521001102.

[46] H. Sayegh, G.N. Lilis, M. Bouquerel, T. Duforestel, K. Katsigarakis, D. Rovas, Automatic Modelica BEM generation from IFC BIM, in: 2024 IEEE International Workshop on Metrology for Living Environment, MetroLivEnv, 2024, pp. 437–441, http://dx.doi.org/10.1109/MetroLivEnv60384.2024.10615614.

[47] UK Government, National Calculation Method (NCM), 2024, URL https://www.uk-ncm.org.uk/ (Accessed: 23 June 2025).

[48] J. Zhou, P. Fennell, I. Korolija, P. Ruyssevelt, Percentage-based thermal zoning approach for enhanced stock-level building energy performance modelling, Energy Build. 329 (2025) 115231, http://dx.doi.org/10.1016/j.enbuild.2024.115231, URL https://www.sciencedirect.com/science/article/pii/S0378778824013471.

[49] I. Korolija, Y. Zhang, L. Marjanovic-Halburd, V.I. Hanby, Regression models for predicting UK office building energy consumption from heating and cooling demands, Energy Build. 59 (2013) 214–227, http://dx.doi.org/10.1016/j.enbuild.2012.12.005, URL https://www.sciencedirect.com/science/article/pii/S0378778812006548.

[50] M. Wang, H. Yu, X. Lin, R. Jing, F. He, C. Li, Comparing stochastic programming with posteriori approach for multi-objective optimization of distributed energy systems under uncertainty, Energy 210 (2020) 118571, http://dx.doi.org/10.1016/j.energy.2020.118571, URL https://www.sciencedirect.com/science/article/pii/S0360544220316790.

[51] Y. Yoon, S. Jung, P. Im, M. Salonvaara, M. Bhandari, N. Kunwar, Empirical validation of building energy simulation model input parameter for multizone commercial building during the cooling season, Renew. Sustain. Energy Rev. 188 (2023) 113889, http://dx.doi.org/10.1016/j.rser.2023.113889, URL https://www.sciencedirect.com/science/article/pii/S1364032123007475.

[52] S. Wijesuriya, H.A. Mazidi, R.A. Kishore, C. Booten, Energy performance of an operational government building retrofitted with ceiling phase change material tiles in a mixed-humid climate, Energy Build. 330 (2025) 115313, http://dx.doi.org/10.1016/j.enbuild.2025.115313, URL https://www.sciencedirect.com/science/article/pii/S037877882500043X.

[53] A. Chong, Y. Gu, H. Jia, Calibrating building energy simulation models: A review of the basics to guide future work, Energy Build. 253 (2021) 111533, http://dx.doi.org/10.1016/j.enbuild.2021.111533, URL https://www.sciencedirect.com/science/article/pii/S0378778821008173.

[54] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renew. Sustain. Energy Rev. 37 (2014) 123–141, http://dx.doi.org/10.1016/j.rser.2014.05.007, URL https://www.sciencedirect.com/science/article/pii/S1364032114003232.

[55] American Society of Heating, Refrigerating and Air Conditioning Engineers, Ashrae Guideline 14-2014: Measurement of Energy, Demand and Water Savings, ASHRAE guideline, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Atlanta, GA, USA, 2014, URL https://books.google.co.uk/books?id=zlJkAQAACAAJ.