Title
Enhancing occupant behavior representation for interoperability between building information modeling and building energy modeling

Journal
Building Simulation, 18(9)

ISSN
1996-3599

Authors
Chung, Jihoon
Hong, Tianzhen
Malik, Jeetika
et al.

Publication Date
2025-09-01

DOI
10.1007/s12273-025-1313-z

BUILD SIMUL (2025) 18: 2435–2458
https://doi.org/10.1007/s12273-025-1313-z

Research Article

Enhancing occupant behavior representation for interoperability between building information modeling and building energy modeling

Jihoon Chung[1] (✉), Tianzhen Hong[2], Jeetika Malik[2], Dennis Shelden[1]

1. School of Architecture, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180, USA
2. Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, CA 94720, USA

Abstract

Building Performance Simulation (BPS) has been adopted as an essential tool for designing, operating, and retrofitting buildings to optimize energy efficiency throughout the building life cycle. The Green Building XML (gbXML) schema facilitates seamless data exchange between Building Information Modeling (BIM) and Building Energy Modeling (BEM) software tools. However, limited occupant behavior (OB) representation in BIM often leads to inconsistent and inaccurate energy simulation in BEM software. This paper presents 154 systematic enhancements to the existing occupant behavior XML (obXML) schema v1.3.4, initially developed for standardizing OB representation for BEM, to address existing limitations and improve interoperability with BIM models. The enhancements encompass improved integration with BIM models through extended building representations and system operations, expanded support for advanced OB models with additional environmental parameters and mathematical capabilities, and implementation of a standardized model documentation framework. To facilitate seamless data transformation between gbXML and obXML schemas, we developed a publicly available gb-obXML Schema Converter. Three case studies demonstrate the enhanced schema's capabilities: representation of building information using a two-story office building model, documentation of a window operation behavior model, and validation of the schema converter's functionality. The enhanced obXML schema v1.4 enables sophisticated modeling of occupant-building interactions while maintaining consistency with industry-standard BIM schemas. The standardized documentation framework facilitates reproducibility and knowledge sharing in the OB research community, while the schema converter automates the integration of building information into OB simulation workflows. These enhancements establish a foundation for more accurate building performance simulation by supporting sophisticated representation of occupant behavior within the BIM-to-BEM simulation workflows.

Keywords

occupant behavior
obXML
occupant behavior modeling
building information modeling
gbXML

Article History

Received: 18 February 2025
Revised: 19 May 2025
Accepted: 02 June 2025

Architecture and Human Behavior

1 Introduction

In the Architecture, Engineering, and Construction (AEC) industry, interoperability has been considered as one of the most problematic issues that hinder the collection and integration of information from design to construction and operational phases (Teicholz 2013). National Institute of Standards and Technology (NIST) reported that inefficient interoperability led to an increase in construction cost of $15.8 billion (NIST 2004). To tackle this barrier, Building Information Modeling (BIM) has been developed as a multidisciplinary data repository from design phase to the engineering and construction phases by delivering project information with high-fidelity representation, including 3D geometries, spatial location, semantic properties, functional characteristics, cost, and construction schedules (Lu et al. 2021; Volk et al. 2024). Furthermore, the integration of BIM with Internet of Things (IoT) sensors allows for real-time data collection and analysis, leading to more effective facility management and improved operational

E-mail: jihoonchung.research@gmail.com

efficiency as well as construction on-site monitoring (Tang et al. 2019). Consequently, this integration enabled the realization of the digital twin paradigm by synchronizing the BIM model with diverse physical devices in real time, significantly enhancing BIM's potential as a data repository for the building operation and maintenance phases (Dave et al. 2018; Boje et al. 2020).

The federation of BIM with Occupant Behavior (OB) modeling is becoming increasingly crucial for developing comprehensive digital twins that accurately reflect the dynamic nature of human-building interactions. As buildings become increasingly equipped with IoT sensors, including occupancy sensors, indoor environmental quality monitors, and energy meters, these technologies generate continuous streams of data regarding occupants' presence, movement patterns, and interactions with building systems (Dong et al. 2019; Tan et al. 2022). This sensor-based monitoring infrastructure enables real-time occupancy detection, prediction of movement patterns, and analysis of human-building interactions, transforming how buildings respond to human needs (Azimi and O'Brien 2022). This data provides essential inputs for calibrating simulation models that capture the stochastic nature of human-building interactions (Jiang et al. 2025). Enhanced interoperability between BIM and OB models aligns with the industry's trajectory toward high-fidelity virtual representations that not only mirror physical and operational characteristics but also incorporate the human behavioral factors and their impact on building performance (Boje et al. 2020; Kjærgaard et al. 2020). Without standardized data exchange mechanisms between them, the potential for occupant-centered simulation within digital twin environments remains significantly constrained by fragmented workflows and inconsistent model representations across domains.

Despite the significant potential of BIM's integration with OB and BPS, it has been hindered by interoperability issues with Building Energy Modeling (BEM). The limited compatibility between BIM and BEM software necessitates repetitive manual work to create BEM models from BIM data with a non-standardized and inconsistent process (Bastos Porsani et al. 2021; González et al. 2021). The fragmented and inefficient workflow impedes productivity and the accuracy of building performance analysis, thereby reducing the adoption of BIM for BPS (Li et al. 2022; Ciccozzi et al. 2023). Consequently, these interoperability challenges restrict the ability to leverage BIM data for optimizing building design and operation for energy efficiency and sustainability.

Green Building XML (gbXML) schema has emerged as a solution for improving interoperability between BIM and BEM. BIM models are usually represented in the Industry Foundation Classes (IFC) to deliver building project information to other stakeholders across the building life cycle, which leads to complex semantics of the building data (Pauwels et al. 2017). In contrast, the gbXML schema adopts nonproprietary and user-friendly file format, eXtensible Markup Language (XML), with fewer layers of complexity to facilitate the exchange of building information between BIM and BPS tools, enabling the reuse of the detailed information available in BIM models (Ham and Golparvar-Fard 2015; González et al. 2021). This enhanced workflow contributes to improving productivity and ensuring sound accuracy of building information for BPS with minimal human errors. As a result, the gbXML has been adopted in more than 50 engineering modeling tools in the industries and became one of the most widely used data exchange formats between BIM and BEM (Ham and Golparvar-Fard 2015; Luo et al. 2021).

In the building simulation community, occupant behavior has received significant attention as one of the most influential factors for understanding building energy consumption (Hong et al. 2017; Yan et al. 2017; Yoshino et al. 2017). Unlike the assumption of conventional building performance simulation, occupant status is rarely deterministic and static with one or few representative features (Jia et al. 2017; Kjærgaard et al. 2020). The interaction between occupant behavior, indoor environment, and building systems is dynamically bidirectional and complex because occupants mainly interact with building systems for their desired thermal, visual, and acoustic comfort (Hong et al. 2015a; O'Brien et al. 2020). Additionally, high variability and uncertainty from the stochastic nature of human behavior significantly contribute to the discrepancy between actual and simulated energy consumption (Yang et al. 2016). In this context, the Annex 66 project, definition and simulation of occupant behavior in buildings, under the International Energy Agency (IEA) Energy in Buildings and Communities (EBC) Programme was initiated to mitigate the uncertainty and suggested methodologies for quantitatively modeling OB and integrating it with BPS programs (IEA 2017). As one of Annex 66's main outcomes, occupant behavior XML (obXML) schema was developed to standardize OB model representation and to capture detailed OB information in a consistent way during simulation process (Hong et al. 2015b; Hong et al. 2018).

Currently, representations of occupant behaviors are still limited in existing BIM schemas, such as IFC and gbXML despite recent research activities that highlight the importance of capturing human interaction with building systems (Hong et al. 2018; Luo et al. 2021; Dong et al. 2022). This limitation restricts the ability to simulate the dynamic and complex nature of occupant behavior using BIM models, which is essential for achieving realistic and sophisticated building performance simulations. Enhancing the representation of occupant behavior within the BIM schemas is critical for improving the accuracy of simulation results.

On the other hand, the obXML schema is designed to incorporate OB models into BPS tools through standardized model exchange and co-simulation. The previous research works for OB model representation has emphasized the need for linkage between obXML and BIM for mitigating difficulties of data reuse without extra effort (Hong et al. 2015b; Deme Belafi et al. 2019; Luo et al. 2021); nevertheless, the compatibility and workflow between them have been barely explored because of the different intentions of the data standards. As the building industry is shifting towards the digital twin paradigm, interoperability between different schemas across domains has become crucial to achieve high-fidelity simulation virtual environments (Boje et al. 2020).

This study presents an essential set of enhancements to the existing obXML schema v1.3.4, referred to hereafter as existing obXML schema, for improving interoperability between BIM and BEM. Key contributions include: (1) improved integration with BIM models, (2) extended support for advanced OB models and standardized model documentation framework, and (3) development of gb-obXML Schema Converter. Three case studies demonstrate the enhanced obXML schema (v1.4)'s capabilities and the functionality of the gb-obXML converter which facilitates seamless incorporation of OB models into BIM-to-BEM simulation workflow, and vice versa.

2 Literature review

2.1 Overview of gbXML schema for data exchange between BIM and BEM

Over the past decade, many researchers have paid attention to integration of BIM with BEM due to the following potential benefits: (1) automating energy modeling process, (2) visualizing simulation output on BIM models, (3) storing energy-related building information in a standardized data repository, and (4) enhancing existing libraries concerning building materials for energy modeling (Kamel and Memari 2019). Nevertheless, the different nature of information leads to semantic and geometric gaps, which has been considered as a substantial obstacle to achieve seamless data exchange between BIM and BEM tools (Di Biccari et al. 2022). For example, some information such as boundary conditions are not embedded in BIM models. Also, not all information in BIM needs to be translated into BEM for energy simulation (Kim et al. 2015). In the case of the geometric gap, BIM supports boundary representation (B-rep), constructive solid geometry, and sweep volume, while BEM only accepts surface representation with boundary loops to calculate heat flows between thermal zones (Yang et al. 2022; Alhammad et al. 2024).

To facilitate the transfer of BIM models from Computer Aided Design (CAD) tools to BPS tools, the Green Building eXtensible Markup Language (gbXML) was developed by Autodesk Green Building Studio in 2000, aiming to support seamless collaboration between architects, engineers, and energy modelers for designing energy efficient buildings (Ramaji et al. 2020; gbXML 2024). This schema represents building data in XML format to take advantage of human- and machine-readable format as well as platform-independent language, which enables integration with diverse software environment including web applications (Khare and Rifkin 1997). Furthermore, the bottom-up approach of the gbXML allows flexible, open-source, and a straightforward data structure, compared to top-down approach such as IFC schema that yields complex data structure and large size of the data files (Jalaei and Jrade 2014; Kamel and Memari 2019). Based on these advantages, gbXML-based energy modeling methods have shown better performance for building information transformation including geometries, material properties, and thermal zones, compared to IFC-based energy modeling methods (Jalaei and Jrade 2014; Ivanova 2015; Gao et al. 2019). For instance, Osello et al. (2011) conducted case studies to compare interoperability from Revit to BPS tools, such as Ecotect Analysis, Transys17, and IES-VE, through IFC, gbXML, and DXF data format. The results showed that the gbXML schema enabled correct transfer of building component, building location, and building type information for BPS unlike the other formats. Additionally, among the XML-based schemas including gbXML, ifcXML, and aecXML, the gbXML is the most common data format for building performance analysis in the AEC industry and has been funded by diverse organization such as the U.S. Department of Energy, American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), Autodesk, Bentley Systems, and others (U.S. GSA 2012; El Asmi et al. 2015; Ham and Golparvar-Fard 2015; gbXML 2024).

2.2 OB representation in different schemas

Recently, as more researchers and experts have paid attention to OB and its importance in the AEC industries, OB representation has been adopted in various data schemas, such as IFC, Energy ADE, BuildingSync, Brick Schema, and Annex 66 Ontology. Table 1 provides a comprehensive comparison of these schemas' capabilities in representing different aspects of OB, including occupant characteristics,

Table 1 Coverage of occupant behavior representation by existing schemas in building industries

Categories	Concept	IFC 4.3	Energy ADE	Building Sync	Brick	gbXML v7.04	Enhanced obXML (v1.4)
Occupant	Occupant type	✓	✓	✓		✓	✓
	Occupancy	✓	✓	✓	✓	✓	✓
	Activity			✓	✓	✓	✓
	Socio-economic			✓	✓	✓	✓
	Geographic location				✓	✓	✓
	Subjective values				✓	✓	✓
	Clothing level				✓		✓*
Environment	Room air temperature	✓			✓	✓	✓
	Room CO_2 concentration	✓			✓	✓	✓
	Room workplane Daylight illuminance	✓			✓	✓	✓
	Room lights power density	✓			✓	✓	✓
	Room wind speed	✓			✓		✓*
	Outdoor dry-bulb temperature	✓	✓		✓	✓	✓
	Outdoor rain indicator	✓			✓	✓	✓
Comfort	Thermal comfort				✓	✓	✓
	Visual comfort					✓	✓
	Indoor air quality					✓	✓
Interaction	HVAC operation	✓	✓	✓	✓	✓	✓
	Lighting switch	✓	✓	✓	✓	✓	✓
	Appliance use	✓	✓	✓	✓	✓	✓
	Thermostat adjustment	✓	✓	✓	✓	✓	✓
	Fan operation	✓	✓	✓	✓		✓*
	Window operation	✓			✓	✓	✓
	Shading operation	✓			✓	✓	✓
	Door operation	✓			✓		✓*
Model Documentation	Type of data collection						✓
	Space granularity						✓*
	Time granularity						✓*
	Model development						✓*
	Model implementation						✓*
	Model evaluation						✓*

Notes: ✓ = Covered; Blank = Not Covered; ✓* = Newly added concepts in the enhanced obXML schema (v1.4)

environmental conditions, occupant comforts, system interactions, and model documentation. The following subsections describe the main features and limitations of these schemas.

IFC: In 1994, the development of the Industry Foundation Classes (IFC) was initiated by International Alliance for Interoperability (IAI) for open data standard in AEC industry, aiming to uniformly represent building data and seamlessly share the data with stakeholders across the building life cycle (Suermann and Issa 2010). The IFC is defined based on the EXPRESS language following ISO 10303-21 and STEP physical file format to standardize data model with terms, such as types, entities, properties, using information-attributed and object-oriented approaches (Pauwels and Terkaj 2016). To enhance interoperability, accessibility, and applicability of the IFC schema in different domains, alternative data formats have been proposed, such as ifcXML, ifcOWL, and ifcJSON (Afsari et al. 2017; Pauwels et al. 2017). Furthermore, buildingSMART International has developed buildingSMART Data Dictionary (bSDD) based on ISO 12006-3, ISO 23386, and Linked Data Standards to provide extensive terms and concepts in

BIM models, containing 1000 terminologies in 108 domains (BuildingSMART 2024). Recently, the IFC schema has adopted the concepts of occupant presence and occupant types, such as *IfcOccupant* and *IfcOccupantTypeEnum* classes; nonetheless, description of occupant behavior is still limited to deliver detailed OB information from BIM software to BPS tools (Dong et al. 2022).

Energy ADE: The Energy Application Domain Extension (Energy ADE) is an extension to the CityGML 2.0 standard with energy-relevant properties and new feature types, aiming to enhance the capability of semantic 3D city models in the domain of urban energy modeling and analysis (Benner 2018). As a standard-based data model, it allows bridging the gap between detailed single-building energy simulation and city-wide, bottom-up energy assessment, focusing on the building sector. In terms of application stage, this data model can be utilized for building usage analysis at urban scale, such as energy demand diagnostics, solar radiance analysis, or simulation of decarbonization strategies.

The Energy ADE is structured into six interconnected modules to support modeling building physics, occupant behavior, materials and construction, energy systems, and supporting elements like time series and schedules (EnergyADE 2024). Although the occupant behavior module is closely related to OB representation, it mainly contains the detailed characterization of building's usage (*UsageZone*) or ownership information (*BuildingUnit*), instead of detailed behavioral information. Moreover, the *Occupants* object represents a homogeneous group of occupants with an occupant type (*occupantType)* as well as number of people (*numberOfOccupants*) and time schedule (*occupancyRate*), and heat dissipation (*heatDissipation*). Nevertheless, this level of detail for OB is insufficient to represent occupant heterogeneity and the interaction between occupants and building systems (Malik et al. 2022).

BuildingSync: BuildingSync was developed to standardize languages to exchange commercial building energy auditing data between different software and databases, based on the extensive collaboration between the U.S. Department of Energy (DOE) and industry experts. This schema was designed based on standard terminology of Building Energy Data Exchange Specification (BEDES) for minimizing redundant effort for subsequent auditing and achieving higher energy efficiency. BuildingSync aims to overcome challenges such as data inconsistency between different auditors by enabling easier aggregation of energy audit data across software tools from different companies and ensuring a consistent format of the historical audit for large scale analysis (DeGraw et al. 2018).

Like gbXML, BuildingSync is represented in XML format for facilitating easier data transfer across multiple industries. The later version directly imported gbXML's namespaces and elements to avoid duplication of effort. The BuildingSync schema primarily consists of two sub-elements, *Programs* and *Facilities*. The *Facilities* element represents buildings that are evaluated and described using the following sub-elements: *Tenants*, *Contacts*, *Reports*, *Measures*, *Schedules*, *Systems*, and *Sites*. This schema includes concepts of tenants and owners to handle energy audits of multi-tenant buildings; however, it doesn't support detailed OB representation besides occupancy schedule.

Brick Schema: The Brick is an open-source metadata schema that represents buildings in a uniform and portable manner for building management, developed by a consortium of researchers and industry experts to facilitate the deployment of smart building applications (Brick Ontology 2024). It employs tag-based representation for flexibility similar to Project Haystack and offers standardized semantic descriptions of building infrastructure, including equipment, sensors, locations, and their interrelationships (Balaji et al. 2016; Project Haystack 2024). To ensure interoperability across various building information systems, Brick was developed based on terminologies from commercial buildings and adopts the semantic data standard, Resource Description Framework (RDF), representing knowledge as triples consisting of subject, predicate, and object (Balaji et al. 2018).

Recently, an extension to the Brick schema has been proposed to incorporate contextual, behavioral, and demographic information about building occupants (Luo et al. 2022). This extension introduces new classes and subclasses for representing occupants, windows, doors, personal thermal comfort devices, occupant sensing, and level of controllability. The extension's data coverage was validated using 34 datasets from ASHRAE Global Occupant Database, demonstrating that the extended schema could represent, on average, 94% of data, an improvement of approximately 23% over the previous version. Brick mainly focuses on representing building operation data, using physical sensors and the relationships with their locations and measurement; however, there is still lack of detailed documentation for OB models (Dong et al. 2022).

Annex 66 Ontology: in the context of the IEA Annex 66, Mahdavi and Taheri (2017) proposed an ontology to address the lack of structured approaches in collecting, storing, analyzing, and sharing monitored building data by incorporating multiple layers of data from building monitoring systems. The authors identified and structured six basic data categories including inhabitants, indoor conditions, external conditions, control systems, equipment, and energy. Based on the categories and respective subcategories, data streams from sensors or meters can be represented with their values, sources, and possible actors (Huebner and Mahdavi 2019). While this ontology has potential for extensibility to capture detailed information of human-building interaction, this ontology is currently highly theoretical and less expressive with pilot implementation for a few datasets, compared to other data schemas (Luo et al. 2022). With a few concepts, it can capture the limited scope of contextual information about buildings and occupants, and few commercial applications have adopted this ontology (Luo et al. 2021). Due to the limited level of detailed information, this ontology is not included in Table 1.

In summary, significant efforts have been made to integrate BIM with BEM to streamline energy simulation processes. gbXML has been recognized as a promising data schema for seamless data exchange between BIM and BPS tools through its flexible XML format and simplified data structure compared to IFC. On the other hand, due to the significant impact of occupant behavior on building energy consumption, many data schemas in the AEC industry have incorporated OB-related elements into their data structures. However, as illustrated in Table 1, these schemas exhibit limitations in fully capturing detailed occupant behavior information. The obXML schema, grounded in the Driver-Needs-Actions-Systems (DNAS) framework, was introduced to address these challenges by standardizing energy-related OB representations and enabling detailed simulation of occupant-building interactions. However, despite the recognized potential synergies between OB representation and BIM, the practical interoperability between these schemas has not been significantly explored. This study aims to enhance the existing obXML schema for improving interoperability with BIM models while maintaining comprehensive OB representation capabilities, thereby facilitating seamless integration of OB modeling into BIM-to-BEM simulation workflows.

2.3 Recent efforts on integrating OB representation in BIM to BEM workflows: ASHRAE RP-1815

As part of the research project, RP-1815 funded by ASHRAE, the gbXML schema has been significantly updated to integrate OB representation for more seamless BIM to BEM workflows. A major part of this work involved incorporating elements of the obXML schema into gbXML, based on comprehensive review of existing technologies' support for OB representation and 16 stakeholder interviews for assessing the needs of improvement (Roth et al. 2024a). The key updates of the gbXML v7.04, include the introduction of new elements and enumerations, such as *TimeofDays*, *SeasonType*, *Gender*, and *IncomeGroup* to better represent temporal and socio-economic factors of OB models in building energy simulations, as shown in Appendix A, which is available in the Electronic Supplementary Material (ESM) of the online version of this paper. In particular, the *BuildingSystems* element was directly added to *Space* element, which is simplified building system representation identical to the *System* element in the existing obXML schema (v1.3.4). Additionally, the schema was updated to improve structural consistency, with over 200 updates ensuring better organization and validation, such as converting child elements into references and addressing naming conflicts. These updates aim to refine gbXML's capacity for representing occupant behavior, making it a more robust data format for energy and performance simulations.

Along with the gbXML updates, the research project RP-1815 also proposed 132 updates to the existing obXML schema to improve compliance with the latest XML specifications for seamless gbXML schema integration without validation issues. The proposed updates include conversion of instantiated child elements into referenced child elements, ensuring a clear hierarchy of elements, proper nesting, and consistent use of parent-child relationships. To ensure compliance with the XML standard, validation testing was conducted using an XML validation tool. Additionally, obXML data dictionary was developed based on the proposed changes and published to the buildingSMART Data Dictionary (bSDD), containing obXML elements corresponding to existing IFC classes and properties to facilitate IFC-based data exchange with OB representation (Roth et al. 2024b).

2.4 Research gaps

While ASHRAE RP-1815 project made significant contributions to incorporate occupant behavior elements within gbXML v7.04, this research focuses on a gbXML-to-obXML conversion approach through obXML schema enhancement, maintaining the specialized focus of each schema. The obXML schema is optimized for detailed representation of occupant behavior through the DNAS framework, while gbXML primarily provides data exchange between BIM and BEM. Maintaining separate schemas preserves this domain-specific focus while enabling interoperability. Incorporating all obXML elements directly into gbXML would increase schema complexity and potentially compromise its primary purpose. As gbXML files typically contain substantial geometry data not essential for occupant behavior modeling and simulation, conversion to obXML schema offers data efficiency advantages. Furthermore, this approach maintains compatibility with existing obXML-based tools such as obFMU, allowing them to leverage building information from gbXML without requiring modifications to their core structures.

To identify the critical gaps in the existing obXML schema, we conducted comprehensive review of the three main sources: (1) requests and feedback from obXML schema developers and OB researchers over the past decade, (2) technical reports from ASHRAE RP-1815, IEA EBC Annex 66 and Annex 79, and (3) research articles on obXML schema and occupant behavior modeling. Furthermore, detailed limitations are analyzed to derive enhancement plan using the existing obXML schema document. The following subsections detail the identified three key limitations in the current schema.

2.4.1 *Limited integration with BIM models*

The ASHRAE RP-1815 report highlights significant limitations in the existing obXML schema's building representation and interoperability with other data standards like IFC and gbXML (Roth et al. 2024a). Through stakeholder interviews, the report identified that no commercial software applications currently support obXML beyond Lawrence Berkeley National Laboratory (LBNL)'s applications like occupant behavior simulator (obFMU), significantly affecting its integration into BIM-to-BEM simulation workflows. The existing obXML schema exhibits limitations in its data structure that hinders the transfer of building information from BIM models to OB documents. A key structural misalignment exists between gbXML's zone-level management of thermal systems and obXML's space-level system handling (Figure 1), constraining schema transformation scalability for large-scale and complex buildings. The schema's lack of essential building-related elements, including building story, space area, and operational schedules, hinder effective BIM data reuse for precise OB simulation. This limitation is particularly problematic in co-simulation scenarios, where obXML's lack of system operational schedules forces an oversimplified assumption of complete manual control by occupants, potentially leading to simulation inaccuracies in buildings with preset HVAC and thermostat scheduling. Furthermore, the absence of reference attributes for building systems necessitates laborious manual updates when BIM models undergo design changes with new space configurations or system specifications, increasing the risk of data inconsistencies across modeling domains. These interoperability challenges impede the incorporation of OB models into BIM-to-BEM workflows, despite the recognized potential for significant synergies across domains (Hong et al. 2015b; Yan et al. 2017; Chen et al. 2018).

2.4.2 *Insufficient capabilities to support advanced OB models*

The obXML user feedback and Deme Belafi et al. (2019) highlight the need for expanded mathematical capabilities beyond trivariate models to implement advanced behavioral models that handle multiple environmental parameters and their non-linear relationships with occupant behaviors. Specifically, the existing obXML schema remains limited to univariate Weibull, quadratic, trivariate linear and logistic regression models, with only six environmental parameters: room air temperature, room CO_2 concentration, room workplane daylight illuminance, room lights power density, outdoor dry bulb temperature, and outdoor rain indicator. These constraints, emphasized in both the IEA EBC Annex 66 final report, make it challenging to implement recent OB models that account for a broader range of environmental factors influencing occupant comfort and behavior. For example, thermal comfort models need to consider both indoor air temperature and relative humidity (Takasu et al. 2017; Kim et al. 2018), while window operation models often incorporate four or more environmental factors simultaneously (Calì et al. 2016; Belazi et al. 2019), demonstrating the need for expanded mathematical capabilities and environmental parameters.

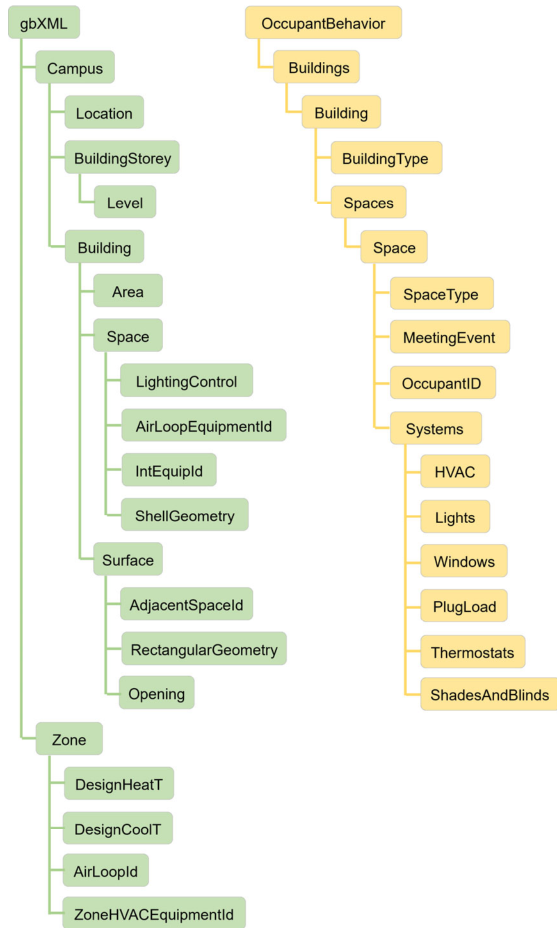

Fig. 1 Schematic representation of the hierarchical structure comparison of building elements between gbXML (left) and obXML (right) schemas

2.4.3 *Lack of standardized documentation framework*

Recent analyses from both the IEA EBC Annex 79 final report and academic literature identify a critical gap in standardized documentation frameworks for OB models (O'Brien et al. 2020; Dong et al. 2022). Most published OB models lack crucial documentation regarding their development and implementation details, hindering both reproducibility of research findings and effective knowledge transfer within the scientific community (O'Brien et al. 2020). As illustrated in Table 1, current data schemas in building industries lack a standardized framework for documenting OB models with consistent terminologies, methodological details, and validation metrics. This aligns with the study of Deme Belafi et al. (2019), which highlighted obXML's limitations on insufficient metadata representation, constraining consistent and reproducible model development and implementation. This limitation is particularly problematic for advanced OB models employing black-box approaches or sophisticated building control applications, where detailed documentation of data collection methodologies, preprocessing steps, and hyperparameter settings is essential for reproducibility. While recent efforts through the IEA EBC Annex 79 have established comprehensive guidelines for simulation and control applications, these frameworks have not been adopted into existing data schemas, including the existing obXML schema.

3 Research method

This research employs a systematic four-phase methodology, as illustrated in Figure 2: (1) identification of critical gaps through comprehensive analysis of user requests, technical reports, and academic literature, as described in Section 2.4; (2) enhancement to the existing obXML schema for addressing the identified limitations by improving BIM integration, supporting advanced OB models, and implementing standardized documentation frameworks (Section 3.1); (3) gb-obXML Schema Converter for automated transformation between gbXML and obXML formats (Section 3.2); and (4) demonstration through three case studies examining expanded building representation, model documentation, and schema conversion capabilities (Section 5).

3.1 Systematic enhancement of the obXML Schema

Building on the comprehensive review of research gaps identified in Section 2.4, this study implements systematic enhancements to the existing obXML schema in two main areas: (1) improving BIM integration through extended building representation and system operations, (2) extending support for advanced OB models through additional environmental parameters and mathematical capabilities, as well as implementing standardized documentation frameworks aligned with recent guidelines. The following subsections detail the methodological approaches for each enhancement component.

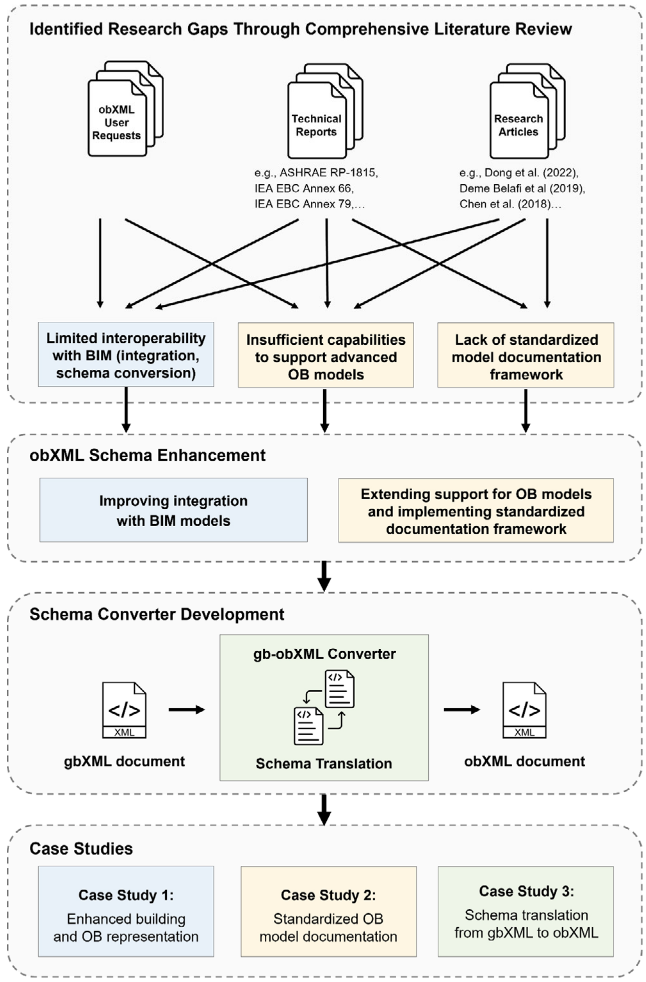

Fig. 2 Overall workflow of the obXML schema enhancement research

To enhance BIM integration capabilities, we incorporate the schema elements and attributes from the latest gbXML schema v7.04, updated under the ASHRAE RP-1815 project, to enable comprehensive representation of building information from BIM models within OB simulation. The enhancements encompass expansion of building representation elements to capture detailed spatial and building system information, integration of extensive operational schedule elements for building systems that occupants interact with, and incorporation of reference attributes for zone, building system and operational schedules aligned with gbXML schema. These enhancements establish the foundation for incorporating OB models into BIM-to-BEM workflows while maintaining data consistency across different modeling domains. Detailed implementation and specific schema elements are documented in Section 4.1.1.

To support more sophisticated behavioral models, we enhance the schema by expanding environmental parameters to include additional indoor and outdoor environmental parameters that influence occupant behaviors and extending the mathematical framework to support higher-dimensional mathematical model structures beyond trivariate linear and logistic regression models. These expanded environmental parameters and mathematical capabilities enable the schema to capture more nuanced relationships between environmental conditions and occupant actions while supporting advanced statistical methods for behavior prediction. Furthermore, to establish comprehensive documentation standards, we extend the schema's metadata framework based on the standardized documentation guidelines proposed by Dong et al. (2022). The enhancements focus on

- model description for capturing comprehensive contextual information including data collection circumstances and spatio-temporal granularity specifications to assess model applicability;
- model development for documenting methodological approaches and data processing procedures;
- model evaluation for documenting validation approaches and performance metrics to enable systematic comparison across implementations;
- model implementation for specifying computational requirements and experimental setups. These structured documentation frameworks establish a foundation for reproducible OB research while model sharing across the building simulation community.

The specific environmental parameters, mathematical models, extended metadata framework, and implementation details are described in Section 4.1.2.

3.2 Design of gb-obXML schema converter

To facilitate seamless integration of OB models into BIM-to-BEM simulation workflows, this study designs a gbXML-to-obXML schema translator that reuses BIM models from gbXML files to generate obXML documents for implementing OB simulation. The converter's architecture focuses on two primary design principles: robust schema validation and systematic data transformation. The validation component ensures structural compliance with both gbXML and obXML schema specifications, while the transformation component implements designed mapping rules to maintain essential relationships between building elements in each data format. The converter's mapping rules enable direct mapping of gbXML elements and attributes with obXML's ones, focusing on spatial elements, building systems, and operational schedules. For instance, space-level transformations maintain connections to thermal zones and building stories through reference attributes, while building system conversions preserve equipment specifications and schedule references. This systematic preservation of building information enables seamless incorporation of BIM data into OB simulation workflows while minimizing manual effort and potential errors in creating obXML documents.

4 Results

4.1 The enhanced obXML schema

The enhanced obXML schema incorporates 154 significant updates to improve interoperability with BIM models, while advancing OB modeling capabilities. These enhancements collectively strengthen the schema's capability to represent complex building information, support sophisticated behavioral models, and enable standardized model documentation across the building-simulation community. The following subsections detail each enhancement area with the corresponding schema elements, attributes, and relationships implemented through these enhancements.

4.1.1 *Improved integration with BIM models*

Extension of building system representations: The extension of building system representations introduces new elements to capture more detailed occupant-building interactions. As depicted in Figure 3, the *Systems* element is expanded to include child elements for *Fans* and *Doors*, as several research works demonstrate door opening's significant influence on air exchange, relative humidity, and CO_2 concentration (Cali et al. 2016; Mousavi et al. 2018; Liu et al. 2022a), while emphasizing frequent fan use with widening neutral temperatures and saving energy during summer (Sun et al. 2018; He et al. 2020; Imagawa et al. 2020). A new *BuildingStorey* element is added with *StoreyLevel* element to enable the schema to represent multistory buildings, as recent research works have explored modeling occupant movement between stories (Hou et al. 2020; Zhang et al. 2024). The *Area* element is introduced with *AreaUnit* attribute of type *AreaUnitType*, supporting various unit enumerations from *SquareKilometers* to *SquareInches*. Additionally, *HVACConditionType* was added to the *Space* element with enumerations (*Heated*, *Cooled*, *HeatedAndCooled*, *Unconditioned*, *Vented*, and *NaturallyVentedOnly*) to better characterize occupant behavior under specific space conditioning states, such

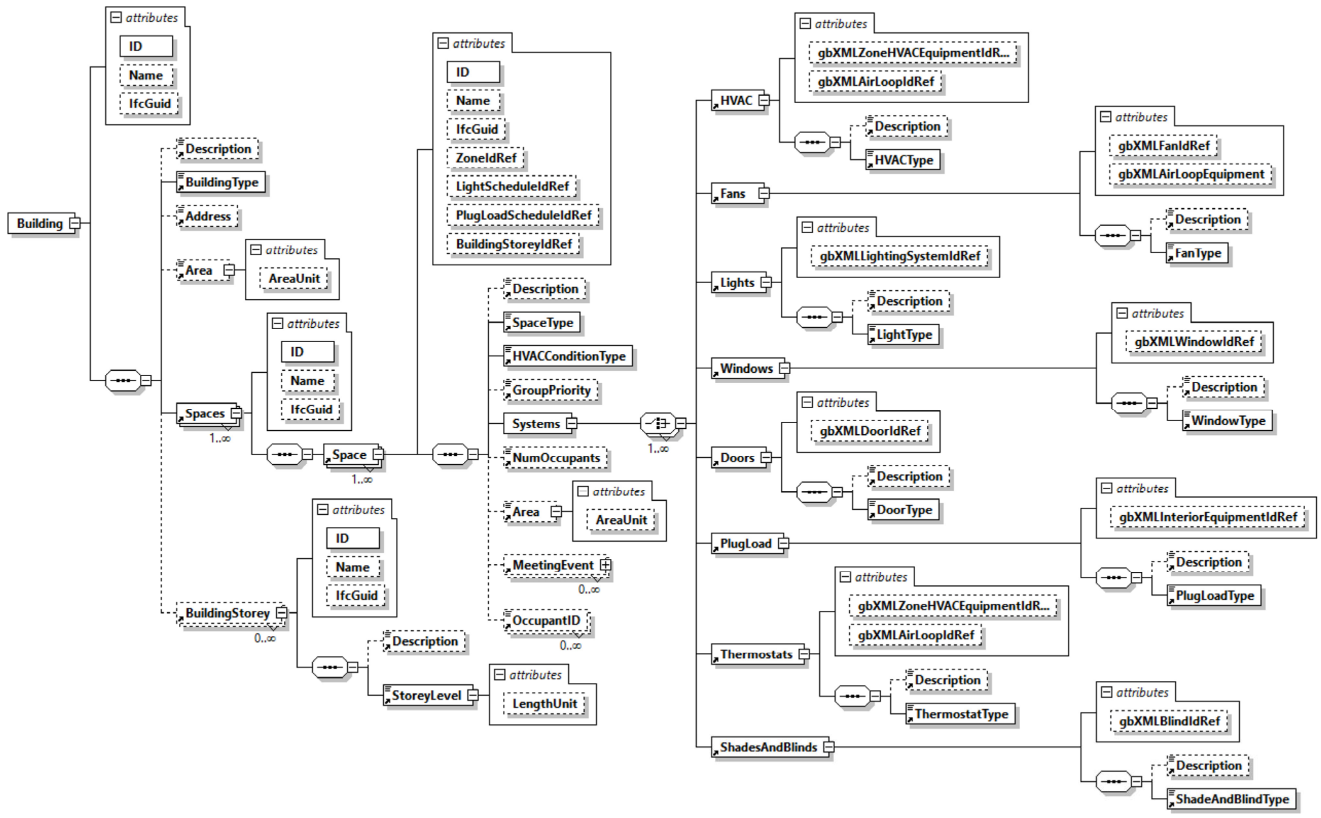

Fig. 3 Tree diagram of the enhanced obXML schema showing *Building*, *Space*, and *Systems* elements

as heating, cooling, or naturally ventilated conditioning (Takasu et al. 2017; Jeong et al. 2024). These enhancements enable more precise representation of building systems, facilitating more accurate simulation of human-building interactions and their impact on building performance.

The integration of operational schedules entails a comprehensive hierarchy to represent temporal patterns of building operations, as shown in Figure 4. To ensure seamless interoperability between BIM and occupant behavior models, the hierarchical structure of the schedules was directly adopted in the obXML schema with minimal modification. For instance, the *Schedule* element includes child elements *YearSchedule* that contains *BeginDate*, *EndDate*, and *WeekScheduleId* elements. The *WeekSchedule* element includes child element *DayScheduleId*; *DaySchedule* element includes child element *ScheduleValue* and attribute *ScheduleType* that provides enumerations for *Temp*, *Fraction*, *OnOff*, *ResetTemp*, and *ActivityLevel*. Moreover, the *Zone* element is newly introduced with references to HVAC-related schedules (*HeatScheduleIdRef*, *CoolScheduleIdRef*, and *FanScheduleIdRef*) as well as design temperatures and relative humidity (*DesignHeatT*, *DesignCoolT*, *DesignHeatRH*, and *DesignCoolRH*) to align the obXML schema with thermal analysis practices in BIM and BEM, where thermal loads and schedules are managed at the zone level rather than individual space level. This incorporation of operational schedules and settings not only enhances interoperability between the schemas but also enables sophisticated modeling of system-occupant interactions by considering both automated operations and occupant interventions, facilitating more realistic co-simulation of building control scenarios.

Incorporation of reference attributes: The incorporation of reference attributes establishes a robust mechanism for maintaining relationships between building elements while allowing direct mapping between obXML and gbXML schemas. Rather than creating complex parent-child hierarchies, these attributes enable flexible and simple data structures by preserving essential relationships between spaces, zones, schedules, and building systems. For instance, *Space* elements maintain connections to their thermal zones (*ZoneIdRef*), operational schedules (*LightScheduleIdRef* and *PlugLoadScheduleIdRef*), and vertical position (*BuildingStoreyIdRef*), while *Zone* elements track their HVAC-related schedules (*HeatScheduleIdRef*, *CoolScheduleIdRef*, and *FanScheduleIdRef*). The schema preserves connections to detailed system specifications in the original BIM model through gbXML-specific reference attributes (e.g., *gbXMLZoneHVACEquipmentIdRef* for HVAC systems, *gbXMLLightingSystemIdRef* for lighting systems). This referencing approach deliberately avoids duplicating physical properties such as envelope material attributes

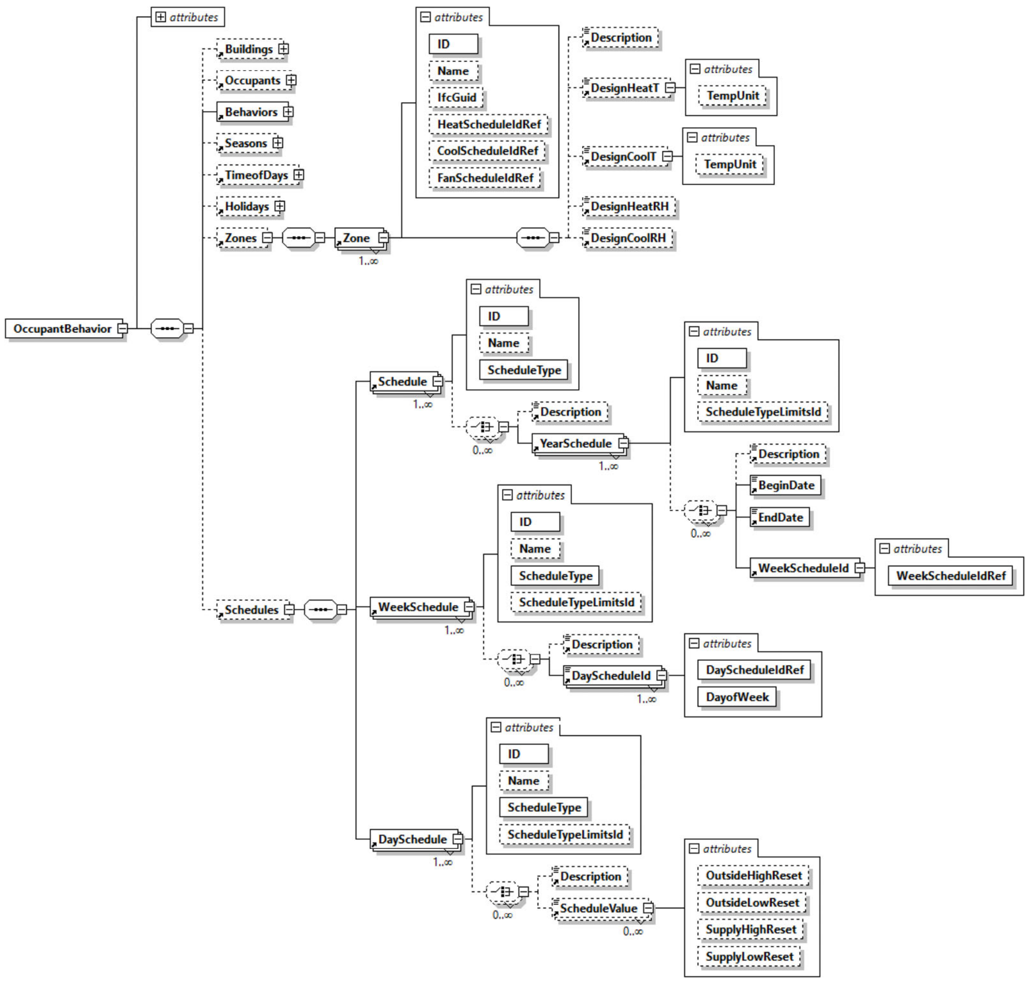

Fig. 4 Tree diagram of the enhanced obXML schema showing *Zone*, *Schedule*, *WeekSchedule*, and *DaySchedule* elements

and equipment efficiency parameters, which are essential for energy simulation but have minimal direct impacts on occupant behavior modeling. Instead, these properties remain in the source gbXML file and can be accessed by BEM tools directly, preventing data redundancy while maintaining the obXML schema's focus on OB representation. At the *OccupantBehavior* level, metadata attributes (*gbXMLProjectGuid*, *gbXMLFilename*, *gbXMLOriginatingSystem*, *gbXMLCreationDate*, and *gbXMLAuthor*) maintain traceability to the source BIM model. This comprehensive referencing system enables efficient updating of obXML documents for data consistency when BIM models change while keeping the obXML structure focused on behavioral modeling rather than duplicating detailed building system specifications.

These enhancements significantly improve the schema's ability to represent spatial and temporal aspects of occupant behavior while maintaining consistency with the industry-standard BIM schemas for BPS. The improved integration capabilities enable more seamless data exchange between occupant behavior models and building information models throughout the building lifecycle, from design to operation and retrofit phases.

4.1.2 *Extended support for advanced OB models*

Expansion of environmental parameters: The expansion of environmental parameters significantly broadens the schema's ability to represent various indoor and outdoor conditions that influence occupant behavior, based on the review of the recent OB models. As shown in Table 2, the *Parameter* element under *Environment* was expanded to include multiple new indoor environmental variables: *RoomRelativeHumidity*, *RoomVolatileOrganicCompounds*, *RoomWindSpeed*, *RoomSoundPressureLevel*, *RoomParticulateMatter*, *RoomCeilingDaylightIlluminance*, and *RoomOperativeTemperature*. Similarly, outdoor environmental parameters were enhanced with *OutdoorRelativeHumidity*, *OutdoorWindSpeed*, *OutdoorWindDirection*, *OutdoorPrecipitationDepth*, *OutdoorParticulateMatter*, *OutdoorCO2Concentration*, *OutdoorSolarIrradiance*, and *OutdoorSolarAltitudeAngle*. These additions reflect recent research findings showing that occupant window-opening behaviors

Table 2 Definitions of the environmental parameters in the enhanced obXML schema

Category	Parameter	Description	Reference
Indoor environment	RoomRelativeHumidity	Relative humidity level in the indoor space	Haldi et al. (2017), Zhang et al. (2022), Farjadnia et al. (2023)
	RoomVolatileOrganicCompounds	Concentration of volatile organic compounds in the room air	Kim et al. (2019), Pandey et al. (2023), Park et al. (2023)
	RoomWindSpeed	Air velocity in the indoor space	Schiavon et al. (2013), Sansaniwal et al. (2021), Zhao et al. (2023)
	RoomSoundPressureLevel	Sound pressure level in the room	Banihashemi et al. (2024), Torresin et al. (2024)
	RoomParticulateMatter	Concentration of particulate matters in the room	Gu et al. (2021), Liu et al. (2022b), Nguyen et al. (2022)
	RoomOperativeTemperature	Operative temperature combining air temperature and mean radiant temperature	Schweiker and Wagner (2016)
	RoomCeilingDaylightIlluminance	Illuminance level from daylight at the room ceiling	Borile et al. (2017), Gunay et al. (2017), Gilani and O'Brien, (2018)
Outdoor environment	OutdoorRelativeHumidity	Relative humidity level of the outdoor air	Belazi et al. (2019), Gu et al. (2021), Zhang et al. (2022)
	OutdoorWindSpeed	Air velocity in the outdoor environment	Shi and Zhao (2016), Yao and Zhao (2017), Pandey et al. (2023)
	OutdoorWindDirection	The direction from which the outdoor wind is blowing	Shi et al. (2018), Nguyen et al. (2022), Pandey et al. (2023)
	OutdoorPrecipitationDepth	The amount of liquid precipitation on the site, measured in millimeters (mm)	Jones et al. (2017), Jeong et al. (2022), Liu et al. (2022b)
	OutdoorParticulateMatter	Concentration of particulate matters in the outdoor air	Pan et al. (2018), Wei et al. (2019), Gu et al. (2021)
	OutdoorCO2Concentration	Carbon dioxide concentration level in the outdoor environment	Belazi et al. (2019)
	OutdoorSolarIrradiance	Power per unit area received from the Sun in the outdoor environment, measured in watts per square meter (W/m^2).	Belazi et al. (2019), Gu et al. (2021), Sansaniwal et al. (2021)
	OutdoorSolarAltitudeAngle	The angular elevation of the sun above the horizon, measured in degrees	Zhang and Barrett (2012), Roccotelli et al. (2021), Korsavi et al. (2022)

are influenced by multiple outdoor environmental factors, while indoor humidity levels significantly impact thermostat adjustments (Shi et al. 2018; Belazi et al. 2019; Gu et al. 2021). These new environmental parameters within the *ParameterType* element enables more comprehensive modeling of environmental factors that drive occupant behavior.

The majority of the expanded environmental parameters are well-supported by existing simulation engines such as EnergyPlus, one of the most widely adopted open-source BEM software (Hong et al. 2016). Parameters like *RoomRelativeHumidity*, *RoomWindSpeed*, *RoomOperativeTemperature*, *OutdoorRelativeHumidity*, *OutdoorWindSpeed*, *OutdoorWindDirection*, *OutdoorPrecipitationDepth*, *OutdoorSolarAltitudeAngle*, *OutdoorCO2Concentration*, and *OutdoorSolarIrradiance* are directly accessible through EnergyPlus' standard outputs. While some specialized parameters (e.g., *RoomVolatileOrganicCompounds*, *RoomParticulateMatter*, *OutdoorParticulateMatter,* and *RoomSoundPressureLevel*) are not natively supported by EnergyPlus, researchers have implemented co-simulation approaches to bridge this gap, using specialized simulation tools such as CONTAM for indoor and outdoor air pollutants (Justo Alonso et al. 2022; Wang et al. 2025), ANSYS FLUENT for air flow (Shan et al. 2020), Radiance for daylight (Giovannini et al. 2021), MATLAB and JOIN for noise (Schaumann et al. 2020; Gonzalez-Caceres et al. 2024). Our enhancement of the obXML schema aims to provide comprehensive coverage of environmental factors identified from recent OB research, ensuring the schema can represent them while maintaining flexibility for researchers to implement OB models with these factors within co-simulation environments.

Extension of mathematical framework: The mathematical framework of the schema was extended to accommodate more sophisticated behavioral models. The previous version (v1.3.4) was limited to trivariate linear or logistic regression models with six environmental parameters. The enhanced schema now supports higher-dimensional mathematical representations by introducing new elements including *Logit4D* though *Logit9D*, *Linear4D* through *Linear9D*, *Quadratic2D*, *Logit2DQuadratic*, and *Cubic1D*. For example, the *Logit9D* element allows for modeling complex window-opening behaviors that consider nine environmental factors simultaneously, such as temperature, humidity, CO_2, particulate matters, wind speed, and solar radiance (Gu et al. 2021). Each mathematical element includes *Description*, *CoefficientA* through *CoefficientJ*, and up to nine *ParameterID* references to accommodate more complex behavioral models. Additionally, *Quadratic2D, Logit2DQuadratic,* and *Cubic1D* models were introduced to capture non-linear relationships between environmental conditions and occupant actions, particularly relevant for thermal comfort studies that explore modeling occupants' comfort temperature and clothing levels (Asif et al. 2022; Achchige et al. 2024). Accordingly, the *Formula* element under *Actions* was restructured to support both these advanced mathematical representations as well as existing regression models from the previous schema version, obXML v1.3.4, enabling more nuanced occupant behavior models for better reflecting the complexity of human-building interactions observed in recent field studies.

The expanded mathematical capabilities and environmental parameters together provide a more robust framework for representing advanced occupant behavior models. This enhancement allows researchers and practitioners to implement more sophisticated behavioral models that capture the multifaceted nature of occupant comfort and decision-making processes, leading to more accurate building performance simulations.

The selection of these specific mathematical extensions prioritized model transparency and reproducibility within occupant behavior simulators, such as obFMU. Although advanced machine learning models (Random Forest, Support Vector Machine, or eXtreme Gradient Boosting) have demonstrated excellent performance in OB prediction (Zhang et al. 2018; Saha et al. 2019), their black-box nature presents significant challenges to describe in standardized schema formats and their reuse across different environments without identical datasets. Integrating such models would require simultaneous enhancement of occupant behavior simulators (e.g., obFMU), which extends beyond this study's scope. This research instead focuses on expanding white-box model coverage with interpretable parameters.

Extension of model metadata: As illustrated in Figure 5, the extension of model metadata introduces a comprehensive *ModelDescription* element as a parent element for structured model documentation under the *Behavior* element and reorganizes existing OB metadata elements into this new structure, such as *PaperTitle, PaperAuthors, PaperYear*, *SystemType*, and *ActionType*. Moreover, the schema enhancement replaces the broad attribute *DataCollectionRegion* with more granular attributes *DataCollectionCountry* and *DataCollectionCity* to refine the geographical context where OB models were developed while introducing *DataCollectionClimateZone* to explicitly capture the climatic representation in the Köppen–Geiger classification system (Köppen and Geiger 1930; Carlucci et al. 2020). Temporal aspects of data collection are captured through *StartDateOfDataCollection*, *EndDateOfDataCollection*, and *DurationOfDataCollection* attributes. The *TypesData-CollectedList* simpleType is expanded to include new values such as *ElectricityMeter*, *FieldVisit*, *Interview*, *Occupancy*, *SubjectObservation*, *WindowStateLog*, *BlindStateLog*, *LightStateLog*, *ThermostatControlLog*, and *WeatherStation*. Following the guidelines of documenting OB models proposed by Dong et al. (2022), new elements are introduced to describe spatio-temporal dimensions of the OB models. For instance, *TimeGranularity* attribute provides enumerations (*LessThan1*, *1to9*, *10to19*, and *30andMoreThan59*) to specify the shortest time window for behavioral prediction, while *SpaceGranularity* includes options for *LabBased*, *Room*, *Floor*, and *Building* levels. The *PredictiveHorizon* (values: *LessThan1*, *1*, *3*, *5*, *15*, and *24*) defines the time horizon for the OB modeling, and *ControlHorizon* (values: *LessThan1*, *1*, *3*, *5*, *15*, and *24*) to specify the time horizon for control variable modeling.

Standardized Documentation: The standardized documentation framework incorporates detailed model development and implementation elements. The *Model-Development* element includes *DataPreparation* with child elements for *DataCleaning*, *DimensionReduction*, *DataScaling*, *FeatureCreation*, *DataPartitioning*, and *DataMerging*. The *ModelingCategory* element under the *ModelDevelopment* classifies models into *RuleBasedModel*, *Statistical-StochasticModel*, and *DataDrivenModel* with 20 subcategories. The *ModelEvaluation* element incorporates *Performance-Metric* with child elements for nine types of *AbsoluteMetric*, aggregated and interval-by-interval types of *DomainMetric*, and comfort-related and energy-related types of *IndirectPerformanceMetric*. On the other hand, the *ModelImplementation* element includes *Computational-Environment* element (*ProgrammingLanguage*, *Software-PackagesTool*, and *OperatingSystem*) and *ComputationTime* element (*InferenceRuntime*, *InferenceMemoryRequirement*, *TotalRequiredTrainingRuntime*, and *ModelScalability*). Additionally, *ExperimentSetup* element captures *sensor* information with detailed *SensorType* enumerations and hierarchical location specification through *SpaceLevelof-SensorLocation* and *BuildingLevelofSensorLocation*.

These enhancements establish a comprehensive framework for documenting occupant behavior models, enabling researchers and practitioners to fully understand model development, implementation, and evaluation processes.

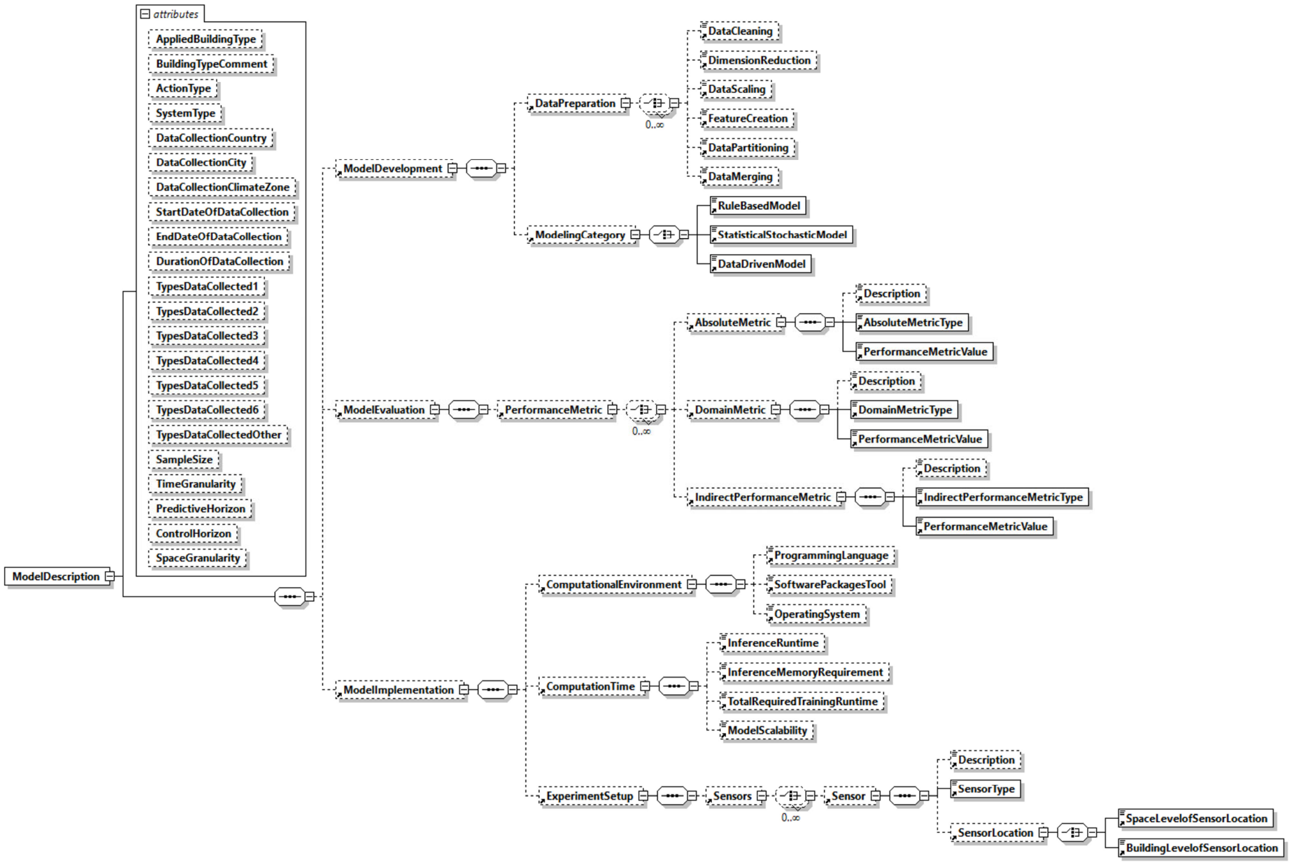

Fig. 5 Tree diagram of the enhanced obXML schema showing *ModelDescription*, *ModelDevelopment*, *ModelEvaluation*, and *ModelImplementation*

The structured documentation supports reproducibility and facilitates appropriate model selection for specific applications.

4.2 The gb-obXML Schema Converter

The gb-obXML converter was developed as a Python-based tool that processes gbXML files through a three-step workflow, as illustrated in Figure 6. First, the tool validates the input gbXML file against the latest version of gbXML v7.04 schema specification from ASHRAE RP-1815, using the 'xmlschema' Python library to ensure data consistency. Second, it transforms the building information according to predefined mapping rules that convert gbXML elements into corresponding obXML elements. Finally, the converter validates the generated obXML document against the enhanced schema specifications to ensure structural compliance. The implementation provides robust error

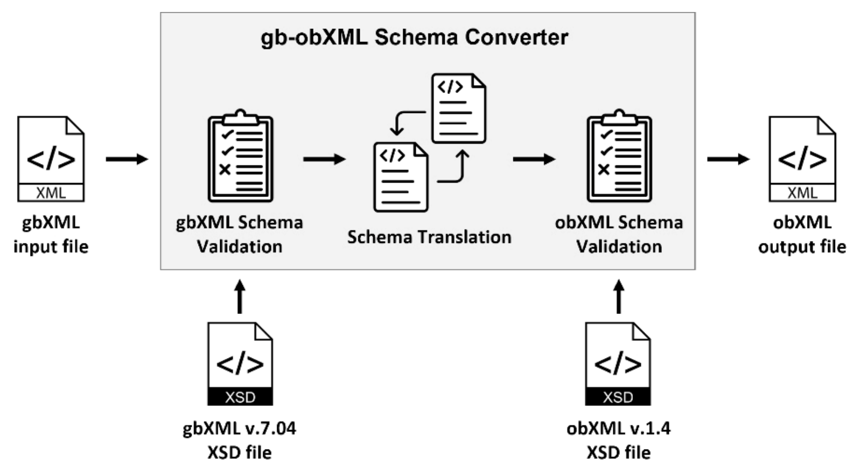

Fig. 6 Workflow of the gb-obXML converter for schema translation between gbXML and obXML

handling for schema validation and transformation processes, with detailed error reporting to facilitate debugging of invalid XML documents. The tool accepts command-line arguments for input/output file paths and schema locations, enabling users to validate files against different versions of gbXML and obXML schemas.

The converter adopts a modular architecture where the core conversion logic is encapsulated in a standalone, platform-independent module. This design approach enables the translation capability to be readily integrated into other Python-based applications, providing flexibility for different implementation contexts. The portable transformation rules ensure that the schema conversion logic can be maintained separately from the implementation code, facilitating updates to the transformation rules without requiring changes to the underlying application code. The gb-obXML Schema Converter and associated resources are publicly available with example files through a GitHub repository (https://github.com/JChung-Research/gb-obXML-Converter), facilitating community adoption and enhancement of the schema translation capabilities.

5 Case studies

To validate the enhanced obXML schema's capabilities, three case studies were conducted. The first case study illustrates the representation of building information and occupant behavior using a two-story office building model, focusing on spatial hierarchy, building systems, and operational schedules, while the second case study demonstrates the standardized documentation framework through a window operation behavior model. The third case study validates the gb-obXML Schema Converter's functionality using another office building model. These studies collectively evaluate the schema's effectiveness in representing building characteristics, documenting occupant behavior models, and enabling schema transformation.

5.1 Case study 1: Building information and OB representation in the enhanced obXML schema

To demonstrate the capabilities of the enhanced obXML schema for representing building information and supporting occupant behavior modeling, a case study was conducted using an office building model from the GreenBuildingXML repository. The model represents a two-story office building located in Boston, MA, with a total floor area of 19,817 square feet (1,841 m^2). The building contains multiple space types including conference rooms, offices, and auxiliary spaces, making it suitable for demonstrating the schema's ability to handle diverse spatial configurations and building systems.

As illustrated in Appendix B of the ESM, the enhanced obXML schema demonstrates robust capabilities for representing complex building hierarchies through its expanded spatial elements, aligned with the gbXML format for seamless integration with BIM and BEM. The schema captures the building's structure from multiple levels, starting with the root Building element (*ID="aim0013"*, *BuildingType="Office"*) containing precise spatial measurements through the *Area* element (total area: 19,817.78 square feet) with appropriate unit specifications (*AreaUnit="SquareFeet"*). The vertical organization is represented through *BuildingStorey* elements, each containing a *StoreyLevel* element that specifies height level (e.g., *StoreyLevel="12"* with *LengthUnit="Feet"* for the second floor). The schema adopts the *Zone* element from gbXML to maintain thermal zoning consistency with BEM practices, as demonstrated in the example zone (*ID="aim81893"*) containing design parameters such as design heating and cooling temperatures (*DesignHeatT="70°F"*, *DesignCoolT="74°F"*). At the space level (*Space ID="aim67589"*), the schema maintains hierarchical relationships through *BuildingStoreyIdRef* and *ZoneIdRef* attributes, following the gbXML format to preserve spatial and thermal relationships. Each space incorporates detailed characteristics including space conditioning information (*HVACConditionType="HeatedAndCooled"*) and precise area measurements (*Area="241.0" AreaUnit="SquareFeet"*).

The schema's representation of operational schedules demonstrates sophisticated temporal modeling capabilities while maintaining compatibility with BEM through a hierarchical structure aligned with the gbXML format. The *Schedule* element hierarchy, exemplified by the heating operation schedule (*ID="Heatsched-9"*), implements a three-tier temporal organization: *YearSchedule* elements define annual periods, *WeekSchedule* elements specify weekly patterns, and *DaySchedule* elements contain detailed hourly operations. The schedule framework supports *ScheduleType* attribute (*"Temp"*, *"Fraction"*, *"OnOff"*, and *"ResetTemp"*), enabling diverse operational patterns as shown in the *WeekSchedule* element that differentiates between weekday, Saturday, and Sunday operations. These schedule elements are systematically connected to building elements through reference attributes — *Zone* elements maintain HVAC-related schedules through *FanScheduleIdRef*, *HeatScheduleIdRef*, and *CoolScheduleIdRef* attributes, while *Space* elements with *LightScheduleIdRef* and *PlugLoadScheduleIdRef* ensure consistent operational patterns across different modeling domains.

The identification reference system demonstrates streamlined reference handling through gbXML-specific attributes. The *Systems* element under *Space* maintains essential system type information for OB modeling (e.g., *PlugLoadType="OnOff"*, *ThermostatType="Adjustable"*) while preserving connections to detailed system specifications in the gbXML model via reference attributes (e.g., *gbXMLZoneHVACEquipmentIdRef* and *gbXMLLighting-SystemIdRef*). This implementation shows how the enhanced schema efficiently represents building systems for OB simulation while ensuring consistency and integrity of cross-referenced data between BIM and OB documents.

5.2 Case study 2: Standardized OB documentation in the enhanced obXML schema

The enhanced obXML schema's standardized documentation framework is demonstrated by representing a window operation behavior model developed by Achchige et al. (2024), as shown in Appendix C of the ESM. The ModelDescription element implements comprehensive metadata attributes that capture the study's context, including building type (*AppliedBuildingType="MultiFamily"*), location (*DataCollectionCountry="Australia", Data-CollectionCity="Brisbane and Gold Coast"*), and climate characteristics in Köppen–Geiger classification system (*DataCollectionClimateZone="Cfa"*). The framework provides detailed building characterization through *BuildingType-Comment*, while temporal aspects of data collection are precisely documented through *StartDateOfDataCollection*, *EndDateOfDataCollection*, and *DurationOfDataCollection* attributes. The schema's standardized documentation of data collection methods is demonstrated through *TypesDataCollected* attributes, which specify multiple data sources including *"IEQmeasurements"*, *"WindowStateLog"*, *"Occupancy"*, and *"WeatherStation"*. Additionally, the framework captures the study's spatiotemporal resolution through *TimeGranularity="10to19"* and *SpaceGranularity="Room"* attributes, enabling clear documentation of model applicability and limitations.

The *ModelDevelopment* element demonstrates structured documentation of the data processing and modeling methodology, as illustrated in Appendix D of the ESM. Under *DataPreparation*, the schema captures detailed preprocessing steps including data merging ("Merged environmental data to build a 15 min resolution dataset"), cleaning procedures ("Selected only the window datasets with active user engagement"), dimension reduction ("Selected Tout and Tin as predictor variables based on Akaike Information Criterion"), and data partitioning methods ("75% training, 25% testing with randomly chosen dates"). The modeling approach is classified through *ModelingCategory* as *StatisticalStochasticModel* with type "Regression", providing clear categorization of the methodological framework. The *ModelImplementation* element documents the computational environment, specifying the programming language ("MATLAB"), analysis tools ("MATLAB Maximum Likelihood Estimation"), and experimental setup including sensor types (*TemperatureSensor* and *WindowSensor*) and their locations (*Wall* and *WindowFrame*).

The performance documentation feature is implemented through the *ModelEvaluation* element, which provides structured representation of validation metrics. The *PerformanceMetric* element captures evaluation results through structured *AbsoluteMetric* specifications. The implemented framework quantifies the model's prediction accuracy with an Area Under Curve (AUC) value of 0.627, illustrating how standardized metrics can be documented for window operation behavior models.

5.3 Case study 3: gb-obXML schema transformation for an office building

5.3.1 Building description

The gbXML file, named 'OfficeBuilding (Older).xml', was obtained from the GreenBuildingXML repository (gbXML 2024), consisting of a two-story office building located in Boston, MA, with a total floor area of 15,811 square feet (1,469 m^2), as shown in Figure 7. The building contains 33 spaces with various space types, including offices, lobby, lounge, conference rooms, restrooms, mechanical rooms, and stairwells across two floors. This diversity of spaces and multiple stories of the office building make it an ideal test case for validating the converter's ability to handle complex spatial hierarchies and various building systems.

As the BIM model was represented in outdated gbXML schema v0.37 released in February 2010, the data structure of the model is significantly different from the latest version of the schema, v7.04. To ensure valid data structure with

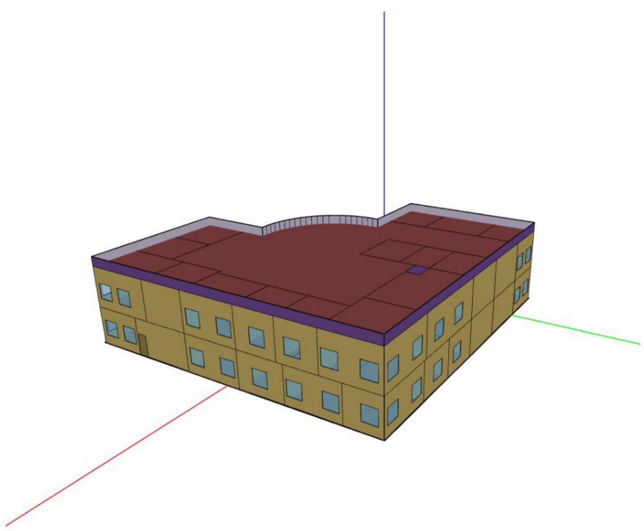

Fig. 7 3D geometries of the two-story office building represented in gbXML format for schema conversion case study

the latest one, we manually updated gbXML model, as the latest version updater of OpenStudio v3.7.0 is compatible with gbXML v7.03. To demonstrate all the functionality of the gb-obXML Schema Converter in this case study, we incorporated detailed space and building system information by adding new elements and attributes, such as *spaceType, unit, Schedule,* and *ZoneHVACEquipmentIdRef*. Additionally, the single zone was divided into seven zones by space type for assessing the translation robustness for different zone types. The updated gbXML file was validated against the latest gbXML schema using XMLSpy software to ensure valid schema for the case study.

5.3.2 Schema conversion results

The gb-obXML converter successfully transformed the building information from the gbXML input file to the enhanced obXML schema, demonstrating preservation of hierarchical relationships and essential building characteristics. Compliance with the obXML v1.4 XSD file was validated using Python 'xmlschema' library. Appendix E of the ESM illustrates example transformations of building, space, and building system elements from the input gbXML file to the output obXML document, highlighting key aspects of the schema conversion.

At the building level, the converter mapped the *Campus/Building* element in gbXML to the *Buildings/ Building* element in obXML while preserving essential attributes such as *ID* and *Name*. The *buildingType* attribute ("Office") was extracted from the gbXML *Building* element and restructured as a dedicated *BuildingType* child element in obXML. Building-level metadata including location information ("Boston, MA") and total floor area (15,811 square feet) were maintained through the conversion process. Moreover, the converter preserved unit specifications by transferring the gbXML attributes to corresponding local attributes in the obXML elements (*LengthUnit="Feet"* and *AreaUnit="SquareFeet"*).

At the space level, the conversion process demonstrated systematic handling of space attributes and references. Space identification was preserved through consistent mapping of *ID*s and *Names* (e.g., "sp-6-Office" and "6 Office"), while space types were mapped with the obXML terminologies according to the schema's specifications (e.g., "OfficeEnclosed" to "Office"). The converter maintained spatial hierarchies by preserving *BuildingStorey* references and zone relationships through *zoneIdRef* attributes. Space-specific characteristics such as HVAC conditioning types were mapped from the *conditionType* attributes to structured *HVACConditionType* child elements. Additionally, the converter preserved precise area measurements for individual spaces (e.g., 201.2 square feet for space "sp-6-Office") while maintaining unit consistency.

For building systems, the converter implemented a structured approach by creating a dedicated *Systems* element under each *Space*. HVAC-related information was transformed from the zone level in the gbXML to space-level system representations in obXML, considering HVAC references (e.g., *zoneHVACEquipmentType= "CAVBox"* and *ZoneHVACEquipmentId*) through attributes like *gbXMLZoneHVACEquipmentIdRef*. The conversion process preserved schedule references for building systems, like lighting (*LightScheduleIdRef*) and equipment (*Plug-LoadScheduleIdRef*), ensuring continuity of operational patterns in the transformed schema. Furthermore, thermal design parameters in zones were accurately maintained, including temperature setpoints (*DesignCoolT*="75°F") and relative humidity levels (*DesignCoolRH*=0.7).

6 Discussion

The case studies demonstrate the implementation of the enhanced obXML schema and the gb-obXML converter, while also highlighting several key implications and directions for future development. This section discusses the major contributions, implications, limitations, and future work.

6.1 Major contributions and implications

The enhanced obXML schema introduces significant advancements in building information representation, enabling more sophisticated OB simulations while maintaining seamless integration with BIM-BEM simulation workflows. Throughout the implementation of 154 enhancements, we carefully balanced expanded representation capabilities against schema complexity, prioritizing elements that address critical interoperability gaps while maintaining a coherent structure aligned with existing workflows. This careful expansion ensures the schema remains accessible for researchers and practitioners while significantly enhancing its representation capabilities for complex occupant-building interactions. This research makes three significant contributions that directly address the critical limitations identified in Section 2.4.

Addressing limited BIM integration through enhanced building representation: The enhanced schema overcomes the limited BIM integration of the existing obXML v1.3.4 by expanding spatial elements to provide essential information for advanced behavioral modeling: the *Area* element with standardized unit specifications enables accurate occupant density calculations and percentage-based occupancy thresholds for different space types; the *BuildingStorey* elements with level specifications (*StoreyLevel*) support modeling vertical movement patterns, elevator usage, and emergency evacuation scenarios in multi-story buildings; the *Zone* elements enable seamless incorporation of thermal zoning information, (e.g., *DesignHeatT* and *DesignCoolT*) and operational schedules through reference attributes (e.g., *FanScheduleIdRef* and *HeatScheduleIdRef*) aligned with BEM practices, allowing OB simulators to consider baseline thermostat setpoints and preset operational schedules for sophisticated simulation scenarios of occupant-system interactions. For instance, when simulating window opening behavior at occupants' arrival, the model can consider whether the HVAC system is operating based on its scheduled status, leading to more realistic predictions of occupant behavior. This comprehensive building representation enables more nuanced modeling of the interplay between automated operations and occupant interventions. Additionally, the schema's reference attributes enable flexible and straightforward data structure for representing element relationships with *Zone* and *Schedule* elements, rather than creating complex hierarchical data structures. gbXML-specific reference attributes (e.g., *gbXMLZoneHVACEquipmentIdRef*, *gbXMLLighting-SystemIdRef,* and *gbXMLInteriorEquipmentIdRef*) enable the data consistency with the original BIM model throughout the building lifecycle, while maintaining essential building system information to focus on behavioral modeling. This allows automatic updates to obXML documents when building information changes that might affect occupant interaction patterns, which is particularly useful during design iterations or retrofits.

Supporting advanced OB models through expanded mathematical framework, environmental parameters, and standardized documentation: The research significantly extends the schema's mathematical capabilities beyond the previous limitation of trivariate models by introducing higher-dimensional mathematical representations (*Logit4D* through *Logit9D*, *Linear4D* through *Linear9D*) and non-linear models (*Quadratic2D*, *Logit2DQuadratic*, *Cubic1D*). Combined with 15 new environmental parameters derived from comprehensive literature review, these enhancements enable implementation of state-of-the-art behavioral models that capture the multifaceted nature of occupant comfort and decision-making processes. Furthermore, to address critical challenges in OB research, we incorporated the standardized documentation framework to enable more effective knowledge sharing and model reproducibility. The comprehensive metadata structure helps researchers to efficiently evaluate model applicability across different contextual factors through detailed documentation of building types, locations (country, city), climate zones, data collection duration, and spatiotemporal granularity. This structured documentation is particularly helpful for researchers to identify the most suitable pre-developed OB models for their specific research conditions by comparing available models based on the similarity of contextual factors to their target application. The schema's *ModelDevelopment* element provides a structured template for documenting complex methodological choices in data-driven approaches, from initial data processing through model training and validation. The *ModelImplementation* element enables detailed documentation of necessary information to reproduce model development processes, such as computational environments, software tools, and experimental setups, while the *ModelEvaluation* element enables systematic comparison of model performance across different implementations using standardized performance metrics for helping researchers to validate their reproductions of published models. The standardized structure ensures that essential model information is consistently captured and readily accessible to the broader research community, enabling researchers and practitioners to compare the model's development process, implementation details, and performance characteristics between different OB models. As the field continues to evolve toward more sophisticated modeling techniques, this structured documentation framework will become increasingly valuable for facilitating the adoption of advanced modeling approaches across different research contexts.

Enabling practical interoperability through automated conversion: The gb-obXML Schema Converter demonstrates practical implementation of improved interoperability by automating the transformation between BIM and OB documents, enabling efficient reuse of building information with minimal manual effort and potential errors in creating obXML files. The converter's systematic mapping rules maintain essential relationships between building elements, while its robust validation process with schema specification files ensures data consistency and structural compliance with both input and output schemas. The modular architecture enables flexible integration of the conversion capabilities into various Python-based applications while keeping the transformation logic separate from implementation codes, facilitating maintenance and updates without modifying the underlying application. The converter and associated resources are publicly available, enabling researchers and practitioners to adopt, extend, and integrate the platform-independent transformation modules into their own applications for BIM-to-BEM simulation workflows. These capabilities collectively establish a foundation for automated synchronization between building information and occupant behavior documents throughout the building lifecycle, from design through operation and retrofit phases.

6.2 Limitations

While the enhanced obXML schema and gb-obXML converter demonstrate significant advancements in OB representation and BIM-BEM interoperability, several limitations should be acknowledged. First, the functional validation of the schema enhancements was limited to three case studies, which may not fully represent the diversity of building types, systems, and operational patterns in real-world applications. Additional functional validation across various building typologies and usage scenarios would strengthen the generalizability of our findings. Moreover, the performance evaluation of enhanced obXML schema as compared to other schemas was not within the scope of this study.

Second, the current implementation of the gb-obXML converter supports one-directional transformation (from gbXML to obXML), limiting bidirectional data exchange capabilities that would be valuable for comprehensive building lifecycle management. The converter also relies on valid data structures in the source gbXML files, which may not always be available in practice due to varying implementation practices across BIM software platforms.

Third, while the enhanced schema supports advanced mathematical representations for OB models, the current enhancements do not include dedicated elements for black-box machine learning models, such as Random Forest, Support Vector Machine, eXtreme Gradient Boosting algorithms that have shown significant effectiveness in predicting occupant behavior in recent research (Zhang et al. 2018; Saha et al. 2019). Incorporating such models would require comprehensive enhancements to both the data schema and OB simulator frameworks, which falls beyond the scope of the current research. This limitation restricts the integration of the black-box machine learning approaches that could potentially capture more nuanced occupant-building interactions.

Finally, adoption barriers related to existing software compatibility present a challenge, as current OB tools like obFMU and OB Model Library use obXML v1.3.3. These technical limitations may restrict the practical application of the enhanced schema's full capabilities in current BIM-to-BEM simulation workflows until supporting tools are updated.

6.3 Future work

Future research can focus on both technical enhancements and practical applications of the improved schema and converter. Technical developments could include extending the schema to handle more complex building systems and control strategies for smart building applications, developing bi-directional conversion capabilities between gbXML and obXML formats, and enhancing the validation framework to ensure data consistency. A critical enhancement would be updating LBNL's obFMU to be compatible with the enhanced obXML schema, as it currently only supports obXML v1.3.3, limiting its ability to leverage the expanded building representation features, advanced mathematical models, and environmental parameters. This enhancement would be particularly important for enabling co-simulation with BPS tools, using the full range of new capabilities including detailed spatial representation and complex building systems. The updated obFMU would allow for more sophisticated simulation scenarios that account for the interplay between automated building operations and occupant interventions, leading to more realistic predictions of building performance. Specifically, an enhanced obFMU would incorporate the new environmental parameters introduced in this research for more comprehensive occupant behavior simulation by facilitating co-simulation between multiple simulation engines. Furthermore, future enhancements to both obXML and obFMU would enable incorporation of black-box OB models and standardized datasets, leading to higher accuracy while maintaining seamless interoperability within the BIM-to-BEM simulation workflow. Additionally, integration with the OB Model Library developed by Deme Belafi et al. (2019) could facilitate automatic assignment of appropriate behavior models based on zone, space types, and building systems from gbXML files, while ensuring semantic consistency and structural schema compliance.

Future work could also focus on developing tools and platforms that leverage standardized documentation and metadata to search for OB models from the library, further enhancing accessibility and reuse across the research community. In this direction, we are developing a publicly available, user-friendly web application for obXML schema conversion to further facilitate adoption within BIM-BEM simulation workflows. This planned application will incorporate more sophisticated mapping rules to handle diverse gbXML structures and OB models from the OB Model Library, offering enhanced integration capabilities and accessibility for researchers and practitioners. The subsequent paper will elaborate technical details and practical applications within BIM-BEM workflows. These advancements would strengthen the bridge between BIM and occupant behavior simulation while promoting more systematic development of occupant-centric design and operation strategies for energy efficient and resilient buildings.

7 Conclusion

This study presents 154 systematic enhancements to the obXML schema that advances the integration of occupant behavior modeling within BIM-to-BEM simulation workflows. The enhanced obXML schema v.1.4 introduces improved integration with BIM models through expanded building system elements, operational schedules, and reference attributes, extends support for advanced occupant behavior models through additional environmental parameters and mathematical frameworks, and implements a standardized documentation framework for reproducibility and knowledge sharing. Key insights from this study include:

- First, we discovered that the fundamental misalignment between gbXML's zone-level and obXML's space-level handling of building systems creates significant interoperability barriers that cannot be overcome through simple element mapping. This realization guided a structured approach with reference attributes to maintain consistency across schemas. The incorporation of *Zone* elements with references to operational schedules ensures alignment of the obXML schema with thermal analysis practices in BIM and BEM.
- Second, our work revealed that standardized documentation frameworks serve as critical enablers for reproducible OB research. Comprehensive documentation captures not only contextual factors but also methodological approaches, implementation details, and performance metrics. By implementing elements for model description, development, evaluation, and implementation, we established a foundation for more effective knowledge sharing across the occupant behavior research community.
- Third, the expanded environmental parameters and mathematical frameworks in the enhanced schema demonstrated the capability of incorporating more sophisticated behavioral models into simulation workflows without requiring manual data transformation or specialized tools, as demonstrated in our documentation of the window operation behavior model. This enhancement enables researchers to implement more nuanced models that capture the multifaceted nature of human-building interactions observed in recent OB studies.
- Lastly, the development of the gb-obXML Schema Converter demonstrated that automated transformation between schemas can maintain essential relationships between building elements while preserving data integrity. Our case study confirmed that this approach enables efficient reuse of building information from BIM models with minimal manual effort, facilitating the incorporation of occupant behavior models into BIM-to-BEM simulation workflows.

These contributions collectively establish a foundation for more seamless integration of occupant behavior modeling into building design and operation practices. By addressing the identified interoperability challenges and implementing standardized documentation frameworks, this work contributes to accelerating the adoption of occupant-centric approaches in building performance simulation, ultimately contributing to more energy-efficient and occupant-responsive buildings.

Electronic Supplementary Material (ESM): The Appendix is available in the online version of this article at https://doi.org/10.1007/s12273-025-1313-z.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article. Tianzhen Hong is an Editorial Board member of *Building Simulation*.

Author contribution statement

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Jihoon Chung, Tianzhen Hong, and Jeetika Malik. The first draft of the manuscript was written by Jihoon Chung and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

Achchige DW, Fiorentini M, Kokogiannakis G, et al. (2024). Comparison of data-driven stochastic window operation models for residential buildings. *Building and Environment*, 261: 111765.

Afsari K, Eastman CM, Castro-Lacouture D (2017). JavaScript Object Notation (JSON) data serialization for IFC schema in web-based BIM data exchange. *Automation in Construction*, 77: 24–51.

Alhammad M, Eames M, Vinai R (2024). Enhancing building energy efficiency through building information modeling (BIM) and building energy modeling (BEM) integration: A systematic review. *Buildings*, 14: 581.

Asif A, Zeeshan M, Khan SR, et al. (2022). Investigating the gender differences in indoor thermal comfort perception for summer and winter seasons and comparison of comfort temperature prediction methods. *Journal of Thermal Biology*, 110: 103357.

Azimi S, O'Brien W (2022). Fit-for-purpose: Measuring occupancy to support commercial building operations: A review. *Building and Environment*, 212: 108767.

Balaji B, Bhattacharya A, Fierro G, et al. (2016). Brick: Towards a unified metadata schema for buildings. In: Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, pp. 41–50.

Balaji B, Bhattacharya A, Fierro G, et al. (2018). Brick: Metadata schema for portable smart building applications. *Applied Energy*, 226: 1273–1292.

Banihashemi F, Weber M, Lang W (2024). Deep learning for predictive window operation modeling in open-plan offices. *Energy and Buildings*, 310: 114109.

Bastos Porsani G, Del Valle de Lersundi K, Sánchez-Ostiz Gutiérrez A, et al. (2021). Interoperability between building information modelling (BIM) and building energy model (BEM). *Applied Sciences*, 11: 2167.

Belazi W, Ouldboukhitine SE, Chateauneuf A, et al. (2019). Experimental and numerical study to evaluate the effect of thermostat settings on building energetic demands during the heating and transition seasons. *Applied Thermal Engineering*, 152: 35–51.

Benner J (2018). CityGML Energy ADE V. 1.0 Specification. Karlsruhe Institute of Technology, Germany.

Boje C, Guerriero A, Kubicki S, et al. (2020). Towards a semantic construction digital twin: Directions for future research. *Automation in Construction*, 114: 103179.

Borile S, Pandharipande A, Caicedo D, et al. (2017). A data-driven daylight estimation approach to lighting control. *IEEE Access*, 5: 21461–21471.

Brick Ontology (2024). Brick: A uniform metadata schema for buildings. Available at https://brickschema.org/. Accessed 28 Nov 2024.

BuildingSMART (2024). buildingSMART data dictionary. Available at https://www.buildingsmart.org/users/services/buildingsmart-data-dictionary/. Accessed 28 Nov 2024.

Calì D, Andersen RK, Müller D, et al. (2016). Analysis of occupants' behavior related to the use of windows in German households. *Building and Environment*, 103: 54–69.

Carlucci S, De Simone M, Firth SK, et al. (2020). Modeling occupant behavior in buildings. *Building and Environment*, 174: 106768.

Chen Y, Hong T, Luo X (2018). An agent-based stochastic occupancy simulator. *Building Simulation*, 11: 37–49.

Ciccozzi A, de Rubeis T, Paoletti D, et al. (2023). BIM to BEM for building energy analysis: A review of interoperability strategies. *Energies*, 16: 7845.

Dave B, Buda A, Nurminen A, et al. (2018). A framework for integrating BIM and IoT through open standards. *Automation in Construction*, 95: 35–45.

DeGraw J, Field-Macumber K, Long N, et al. (2018). BuildingSync® in action: Example implementations. In: Proceedings of the 2018 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA.

Deme Belafi Z, Hong T, Reith A (2019). A library of building occupant behaviour models represented in a standardised schema. *Energy Efficiency*, 12: 637–651.

Di Biccari C, Calcerano F, D'Uffizi F, et al. (2022). Building information modeling and building performance simulation interoperability: State-of-the-art and trends in current literature. *Advanced Engineering Informatics*, 54: 101753.

Dong B, Prakash V, Feng F, et al. (2019). A review of smart building sensing system for better indoor environment control. *Energy and Buildings*, 199: 29–46.

Dong B, Markovic R, Carlucci S, et al. (2022). A guideline to document occupant behavior models for advanced building controls. *Building and Environment*, 219: 109195.

El Asmi E, Robert S, Haas B, et al. (2015). A standardized approach to BIM and energy simulation connection. *International Journal of Design Sciences and Technology*, 21: 59–82.

EnergyADE (2024). Available at http://www.citygmlwiki.org/index.php/CityGML_Energy_ADE. Accessed 28 Nov 2024.

Farjadnia M, Fontan A, Russo A, et al. (2023). What influences occupants' behavior in residential buildings: An experimental study on window operation in the KTH Live-In Lab. In: Proceedings of the 2023 IEEE Conference on Control Technology and Applications (CCTA), Bridgetown, Barbados, pp. 752–758.

Gao H, Koch C, Wu Y (2019). Building information modelling based building energy modelling: A review. *Applied Energy*, 238: 320–343.

gbXML (2024). About gbXML. Green Building XML Schema. Available at https://www.gbxml.org/About_GreenBuildingXML_gbXML. Accessed 28 Nov 2024.

Gilani S, O'Brien W (2018). A preliminary study of occupants' use of manual lighting controls in private offices: A case study. *Energy and Buildings*, 159: 572–586.

Giovannini L, Baracani M, Favoino F, et al. (2021). Occupant-centric control of transparent dynamic façades through an integrated co-simulation framework. In: Proceedings of the 17th International Building Performance Simulation Association Conference (BS2021), Bruges, Belgium, pp. 2679–2686.

González J, Soares C, Najjar M, et al. (2021). BIM and BEM methodologies integration in energy-efficient buildings using experimental design. *Buildings*, 11: 491.

Gonzalez-Caceres A, Hunger F, Forssén J, et al. (2024). Towards digital twinning for multi-domain simulation workflows in urban design: A case study in Gothenburg. *Journal of Building Performance Simulation*, 18: 311–332.

Gu Y, Cui T, Liu K, et al. (2021). Study on influencing factors for occupant window-opening behavior: Case study of an office building in Xi'an during the transition season. *Building and Environment*, 200: 107977.

Gunay HB, O'Brien W, Beausoleil-Morrison I, et al. (2017). Development and implementation of an adaptive lighting and blinds control algorithm. Building and Environment, 113: 185–199.

Haldi F, Calì D, Andersen RK, et al. (2017). Modelling diversity in building occupant behaviour: A novel statistical approach. *Journal of Building Performance Simulation*, 10: 527–544.

Ham Y, Golparvar-Fard M (2015). Mapping actual thermal properties to building elements in gbXML-based BIM for reliable building energy performance modeling. *Automation in Construction*, 49: 214–224.

He Y, Li N, Zhang H, et al. (2020). Air-conditioning use behaviors when elevated air movement is available. *Energy and Buildings*, 225: 110370.

Hong T, D'Oca S, Turner WJN, et al. (2015a). An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework. *Building and Environment*, 92: 764–777.

Hong T, D'Oca S, Taylor-Lange SC, et al. (2015b). An ontology to represent energy-related occupant behavior in buildings. Part II: Implementation of the DNAS framework using an XML schema. *Building and Environment*, 94: 196–205.

Hong T, Sun H, Chen Y, et al. (2016). An occupant behavior modeling tool for co-simulation. *Energy and Buildings*, 117: 272–281.

Hong T, Yan D, D'Oca S, et al. (2017). Ten questions concerning occupant behavior in buildings: The big picture. *Building and Environment*, 114: 518–530.

Hong T, Chen Y, Belafi Z, et al. (2018). Occupant behavior models: A critical review of implementation and representation approaches in building performance simulation programs. *Building Simulation*, 11: 1–14.

Hou H, Pawlak J, Sivakumar A, et al. (2020). An approach for building occupancy modelling considering the urban context. *Building and Environment*, 183: 107126.

Huebner GM, Mahdavi A (2019). A structured open data collection on occupant behaviour in buildings. *Scientific Data*, 6: 292.

IEA (2017). Energy in Buildings and Communities Programme (EBC) Annex 66, Definition and simulation of occupant behavior in buildings. International Energy Agency. Available at www.annex66.org. Accessed 28 Nov 2024.

Imagawa H, Rijal HB, Shukuya M (2020). Development of integrated occupant-behavioural stochastic model including the fan use in Japanese dwellings. *Energy and Buildings*, 226: 110326.

Ivanova IV (2015). Comparison between semi-automated building energy simulation processes using BIM-generated gbXML and IFC formats in EnergyPlus. PhD Thesis, Texas A&M University, USA.

Jalaei F, Jrade A (2014). Integrating building information modeling (BIM) and energy analysis tools with green building certification system to conceptually design sustainable buildings. *Journal of Information Technology in Construction*, 19: 494–519.

Jeong B, Kim J, Chen D, et al. (2022). Developing a window behaviour model incorporating A/C operation states. *Building and Environment*, 214: 108953.

Jeong B, Kim J, de Dear R (2024). A/C usage intensity and adaptive comfort behaviour in residential buildings in South East Queensland, Australia. *Energy and Buildings*, 317: 114407.

Jia M, Srinivasan RS, Raheem AA (2017). From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. *Renewable and Sustainable Energy Reviews*, 68: 525–540.

Jiang Z, Wang X, Li H, et al. (2025). Physics-informed machine learning for building performance simulation—A review of a nascent field. arXiv Preprint. Available at https://doi.org/10.48550/arXiv.2504.00937.

Jones RV, Fuertes A, Gregori E, et al. (2017). Stochastic behavioural models of occupants' main bedroom window operation for UK residential buildings. *Building and Environment*, 118: 144–158.

Justo Alonso M, Dols WS, Mathisen HM (2022). Using co-simulation between EnergyPlus and CONTAM to evaluate recirculation-based, demand-controlled ventilation strategies in an office building. *Building and Environment*, 211: 108737.

Kamel E, Memari AM (2019). Review of BIM's application in energy simulation: Tools, issues, and solutions. *Automation in Construction*, 97: 164–180.

Khare R, Rifkin A (1997). XML: A door to automated Web applications. *IEEE Internet Computing*, 1: 78–87.

Kim JB, Jeong W, Clayton MJ, et al. (2015). Developing a physical BIM library for building thermal energy simulation. *Automation in Construction*, 50: 16–28.

Kim J, Zhou Y, Schiavon S, et al. (2018). Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Building and Environment*, 129: 96–106.

Kim H, Hong T, Kim J (2019). Automatic ventilation control algorithm considering the indoor environmental quality factors and occupant ventilation behavior using a logistic regression model. *Building and Environment*, 153: 46–59.

Kjærgaard MB, Ardakanian O, Carlucci S, et al. (2020). Current practices and infrastructure for open data based research on occupant-centric design and operation of buildings. *Building and Environment*, 177: 106848.

Korsavi SS, Jones RV, Fuertes A (2022). Factors influencing the state of blinds and lights in primary schools: Behavioural models and opportunities to improve children's visual environment. *Journal of Building Engineering*, 61: 105303.

Köppen W, Geiger R (1930). Handbuch der Klimatologie (Vol. 1). Berlin: Gebrüder Borntraeger.

Li H, Zhang J, Xue X, et al. (2022). Issues in bi-directional interoperability between BIM and BEM. In: Construction Research Congress 2022. Arlington, VA, USA: American Society of Civil Engineers, pp. 1355–1364.

Liu Y, Chong W, Yau YH, et al. (2022a). Rethinking the limitations of research on occupants' window-opening behavior: A review. *Energy and Buildings*, 277: 112552.

Liu Y, Chong W, Cao Y, et al. (2022b). Characteristics analysis and modeling of occupants' window operation behavior in hot summer and cold winter region, China. *Building and Environment*, 216: 108998.

Lu K, Jiang X, Yu J, et al. (2021). Integration of life cycle assessment and life cycle cost using building information modeling: A critical review. *Journal of Cleaner Production*, 285: 125438.

Luo N, Pritoni M, Hong T (2021). An overview of data tools for representing and managing building information and performance data. *Renewable and Sustainable Energy Reviews*, 147: 111224.

Luo N, Fierro G, Liu Y, et al. (2022). Extending the Brick schema to represent metadata of occupants. *Automation in Construction*, 139: 104307.

Mahdavi A, Taheri M (2017). An ontology for building monitoring. *Journal of Building Performance Simulation*, 10: 499–508.

Malik J, Azar E, Mahdavi A, et al. (2022). A level-of-details framework for representing occupant behavior in agent-based models. *Automation in Construction*, 139: 104290.

Mousavi ES, Jafarifiroozabadi R, Bayramzadeh S, et al. (2018). An observational study of door motion in operating rooms. *Building and Environment*, 144: 502–507.

Nguyen TH, Ionescu A, Géhin E, et al. (2022). Predicting the opening state of a group of windows in an open-plan office by using machine learning models. *Building and Environment*, 225: 109636.

NIST (2004). Cost analysis of inadequate interoperability in the US capital facilities industry. National Institute of Standards and Technology (NIST), pp. 223–253.

O'Brien W, Wagner A, Schweiker M, et al. (2020). Introducing IEA EBC annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation. *Building and Environment*, 178: 106738.

Osello A, Cangialosi G, Dalmasso D, et al. (2011). Architecture data and energy efficiency simulations: BIM and interoperability standards. In: Proceedings of Building Simulation 2011 IBPSA, pp. 1521–1526.

Pan S, Xiong Y, Han Y, et al. (2018). A study on influential factors of occupant window-opening behavior in an office building in China. *Building and Environment*, 133: 41–50.

Pandey P, Dong AB (2023). Analysis of impacts of window opening behavior on indoor air pollutants in residential dorms through deep neural network. *E3S Web of Conferences*, 396: 01097.

Park JS, An YM, Jeong JW (2023). Determinants of manual window control for natural ventilation in homes. *Building and Environment*, 244: 110741.

Pauwels P, Terkaj W (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction*, 63: 100–133.

Pauwels P, Zhang S, Lee YC (2017). Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, 73: 145–165.

Project Haystack (2024). An open source initiative to streamline working with IoT data. Available at https://project-haystack.org/. Accessed 28 Nov 2024.

Ramaji IJ, Messner JI, Mostavi E (2020). IFC-based BIM-to-BEM model transformation. *Journal of Computing in Civil Engineering*, 34: 04020005.

Roccotelli M, Rinaldi A, Fanti MP, et al. (2021). Building energy management for passive cooling based on stochastic occupants behavior evaluation. *Energies*, 14: 138.

Roth S, Shelden D, Ouellette J (2024a). Integrating occupant behavior data into building information modeling for performance simulation. Report No. 1815-RP, American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE).

Roth S, Shelden D, Ouellette J (2024b). obXML 1.3.4 schema proof-of-concept. Available at https://search.bsdd.buildingsmart.org/uri/ashrae-rp1815/ASHRAE_RP-1815-obXML/1. Accessed 28 Nov 2024.

Saha H, Florita AR, Henze GP, et al. (2019). Occupancy sensing in buildings: A review of data analytics approaches. *Energy and Buildings*, 188: 278–285.

Sansaniwal SK, Mathur J, Mathur S (2021). Quantifying occupant's adaptive actions for controlling indoor environment in naturally ventilated buildings under composite climate of India. *Journal of Building Engineering*, 41: 102399.

Schaumann D, Moon S, Usman M, et al. (2020). JOIN: An integrated platform for joint simulation of occupant-building interactions. *Architectural Science Review*, 63: 339–350.

Schiavon S, Lee KH (2013). Dynamic predictive clothing insulation models based on outdoor air and indoor operative temperatures. *Building and Environment*, 59: 250–260.

Schweiker M, Wagner A (2016). The effect of occupancy on perceived control, neutral temperature, and behavioral patterns. *Energy and Buildings*, 117: 246–259.

Shan X, Luo N, Sun K, et al. (2020). Coupling CFD and building energy modelling to optimize the operation of a large open office space for occupant comfort. *Sustainable Cities and Society*, 60: 102257.

Shi S, Zhao B (2016). Occupants' interactions with windows in 8 residential apartments in Beijing and Nanjing, China. *Building Simulation*, 9: 221–231.

Shi Z, Qian H, Zheng X, et al. (2018). Seasonal variation of window opening behaviors in two naturally ventilated hospital wards. *Building and Environment*, 130: 85–93.

Suermann PC, Issa RR (2010). The US national building information modeling standard. In: Handbook of Research on Building Information Modeling and Construction Informatics: Concepts and Technologies. IGI Global, pp. 138–154.

Sun C, Zhang R, Sharples S, et al. (2018). A longitudinal study of summertime occupant behaviour and thermal comfort in office buildings in northern China. *Building and Environment*, 143: 404–420.

Takasu M, Ooka R, Rijal HB, et al. (2017). Study on adaptive thermal comfort in Japanese offices under various operation modes. *Building and Environment*, 118: 273–288.

Tan Y, Chen P, Shou W, et al. (2022). Digital Twin-driven approach to improving energy efficiency of indoor lighting based on computer vision and dynamic BIM. *Energy and Buildings*, 270: 112271.

Tang S, Shelden DR, Eastman CM, et al. (2019). A review of building information modeling (BIM) and the Internet of Things (IoT) devices integration: Present status and future trends. *Automation in Construction*, 101: 127–139.

Teicholz P (2013). BIM for Facility Managers. John Wiley & Sons.

Torresin S, Aletta F, Oberman T, et al. (2024). Factors influencing window opening behavior and mechanical ventilation usage during summertime: A case study in UK dwellings. *Building and Environment*, 263: 111880.

U.S. GSA (2012). GSA BIM guide for energy performance. U.S. General Services Administration. Available at https://www.gsa.gov/real-estate/design-and-construction/3d4d-building-information-modeling/bim-guides/bim-guide-05-energy-performance. Accessed 28 Nov 2024.

Volk R, Stengel J, Schultmann F (2014). Building information modeling (BIM) for existing buildings—Literature review and future needs. *Automation in Construction*, 38: 109–127.

Wang Y, Petrou G, Symonds P, et al. (2025). Investigating the impacts of home energy retrofit on the indoor environment through co-simulation: A UK case study. *Journal of Building Engineering*, 100: 111794.

Wei Y, Yu H, Pan S, et al. (2019). Comparison of different window behavior modeling approaches during transition season in Beijing, China. *Building and Environment*, 157: 1–15.

Yan D, Hong T, Dong B, et al. (2017). IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. *Energy and Buildings*, 156: 258–270.

Yang J, Santamouris M, Lee SE (2016). Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. *Energy and Buildings*, 121: 344–349.

Yang Y, Pan Y, Zeng F, et al. (2022). A gbXML reconstruction workflow and tool development to improve the geometric interoperability between BIM and BEM. *Buildings*, 12: 221.

Yao M, Zhao B (2017). Window opening behavior of occupants in residential buildings in Beijing. *Building and Environment*, 124: 441–449.

Yoshino H, Hong T, Nord N (2017). IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods. *Energy and Buildings*, 152: 124–136.

Zhang Y, Barrett P (2012). Factors influencing occupants' blind-control behaviour in a naturally ventilated office building. *Building and Environment*, 54: 137–147.

Zhang Y, Bai X, Mills FP, et al. (2018). Rethinking the role of occupant behavior in building energy performance: A review. *Energy and Buildings*, 172: 279–294.

Zhang Z, Lv X, Wei Z, et al. (2022). Effects of cooking and window opening behaviors on indoor ultrafine particle concentrations in urban residences: A field study in Yangtze River Delta region of China. *Building and Environment*, 207: 108488.

Zhang R, Zhou T, Ye H, et al. (2024). Introducing a novel method for simulating stochastic movement and occupancy in residential spaces using time-use survey data. *Energy and Buildings*, 304: 113854.

Zhao W, Chow D, Sharples S (2023). Influences of indoor clothing adaptations on energy consumptions in rural houses. In: Building Simulation Conference Proceedings IBPSA 2023, pp. 509–516.