

Occupant behaviour (OB) modelling to support urban building energy simulation: an overview.

Alessia Banfi¹, Martina Ferrando¹, Francesco Causone¹

¹ Energy Department, Politecnico di Milano, Via Lambruschini 4, Milano, Italy

Abstract

The paper provides an overview of the current state of research on the topic of Occupant Behaviour modelling to support urban building energy simulation, describing the challenges arising when scaling from the building to the urban scale. Available modelling approaches and data sources are discussed and compared to provide readers with a brief modelling guide. Furthermore, research gaps and challenges that need to be addressed in future works are identified, to promote advancement in the field.

Highlights

- An overview of Occupant Behaviour modelling to support UBEM is provided
- Challenges arising when scaling up Occupant Behaviour modelling to the urban scale are highlighted
- Available modelling approaches and data sources are discussed and compared to understand their potential and limitation

Practical implications

An overview of the state of the research on occupant behaviour modelling at large spatial scales can inform the development and refinement of urban building energy models, leading to more accurate predictions of cities' energy consumption and potential savings.

Introduction

Background

In 2021 the building sector alone accounted for up to one-third of final energy demand and almost 40% of energy-related CO₂ emissions (United Nations Environment Programme, 2021). Therefore, there is an ever more pressing need to determine, with sufficient reliability, the current energy performance of groups of buildings at different spatial scales to establish the best solution to decrease the overall energy consumption of cities (Ferrando et al., 2020). To this purpose, multiple Urban Building Energy Modelling (UBEM) tools have been developed to analyse the energy demand and peak load of ten to thousands of buildings together, as well as evaluate different scenarios of intervention (Ferrando et al., 2020). Nevertheless, UBEM tools, like any simulation tool, are characterized by an intrinsic level of approximation, especially related to the uncertainties accompanying the input data (Hong et al., 2020).

In UBEM, to reduce computational burdens, buildings are described with the use of “*archetypes*”, which are fully characterized building models able to represent a cluster (i.e., group) of buildings sharing similar characteristics (Carnieletto et al., 2021). Archetypes comprise information about building use, construction assemblies, systems, fixed values of certain geometric variables, and occupants (Carnieletto et al., 2021).

Occupant-related information, often called Occupant Behaviour (OB), exert a significant impact on the energy use patterns of buildings, affecting demand-side management and design and dimensioning of energy systems (Carlucci et al., 2020). Occupants, in fact, contribute with their presence to increase the sensible and latent heat gains, and with their actions to building systems and appliance usage (Parker et al., 2017).

While in single Building Energy Modelling (BEM) occupants' description emerged to account for up to 30% of the variation of building performance (Mosteiro-Romero et al., 2020), the impact of different OB models on the energy patterns of groups of buildings at different spatial or time scale is still scarcely quantified and need to be further analysed (Ferrando et al., 2022). The assessment of opportunities for energy savings in urban areas requires a comprehensive understanding of the interdependencies between buildings, services, and individuals at a large scale. It is, therefore, essential to investigate the spatiotemporal fluctuations in buildings' energy demand caused by peoples' daily activities and occupancy patterns (Mohammadi & Taylor, 2017).

Despite the limited number of existing studies on the topic of OB modelling in UBEM, some reviews have been already conducted (Dabirian et al., 2022; Doma & Ouf, 2023; Happle et al., 2018; Salim et al., 2020). However, a user-oriented overview is still missing. The objective of this paper is to address this gap by providing readers with a comprehensive understanding of the current state of the research including available modelling approaches, data science techniques used for their development, and data sources. This information will support researchers in the development of reliable and realistic descriptions of occupant behaviour at the urban scale.

Research question and goals

This paper is intended to support the dissemination of the current state of the research of OB in UBEM, to raise awareness on the importance of adopting realistic

occupants' models, to consolidate past findings and identify possible future outlooks to promote advancements in the field. To this purpose, a critical analysis of the main published contributions on the topic is conducted to:

- point out the existing challenges and limitations arising when scaling up building performance simulation to groups of buildings together,
- identify the currently exploited modelling approaches and the data science techniques used for their development,
- discuss the available data and the relative sources that can support urban OB modelling,
- Present future opportunities to help advancing the knowledge of realistic occupants' descriptions at an urban scale.

Occupant Behaviour to support Urban Building Energy Modelling

Definition of "Occupant Behaviour" in the field of building performance simulation

Despite the term "*behaviour*" have been widely adopted in building energy modelling, the lack of an established scientific definition and its cross-disciplinary usage can make it misleading when used to address the ways people interact with buildings (Zhang et al., 2018). Therefore, to avoid misunderstanding, OB in building performance simulation should be interpreted as the description of occupants' presence and energy-related actions.

Occupants' presence (i.e., occupancy) is intended as the headcount of people in a certain space at a certain time while occupants' actions represent the set of activities that could influence buildings' energy use and performance (i.e., windows operation, solar shading adjustment, lighting operation, thermostat adjustment, appliances use, domestic hot water use, and clothing adjustment) (Carlucci et al., 2020; Happle et al., 2018). The further distinction between "adaptive" and "non-adaptive" actions is often proposed, with the former being rooted in people's comfort perception and therefore triggered by certain environmental parameter thresholds (e.g., thermostat adjustment driven by variation in indoor temperature conditions), while the latter being part of individuals' tasks and not oriented to the adaptation to the environment (e.g., appliances usage) (Schweiker et al., 2017). It is important to note, however, that the discrimination between the two is contextual, rather than absolute. Many energy-related actions could be automatic, habitual, constrained by building design and control systems or even influenced by external factors, such as social norms and cultural practices (Schweiker et al., 2017). For example, lighting operation could be adaptive if triggered by visual comfort factors or could be non-adaptive when regulated with event-based schedules or automated control systems; similarly, shading adjustment could be performed to control glare or heat gains but also for privacy or security reasons (e.g., when leaving a room). Following the terminology commonly used in literature,

in the following text, the description of occupants' presence and energy-related actions will be referred to as "Occupant Behaviour". Nonetheless, the abovementioned specification becomes necessary to properly understand the investigated variables in human-building interactions.

From single building to urban level applications: challenges and opportunities

The last decades have seen a growing interest in the field of OB modelling to support building performance simulation but in comparison with the rich body of literature focusing on single-building applications, studies addressing the description of OB at large spatial scales to support UBEM are still limited (Figure 1).

In UBEM the characterization of the building stock is usually dealt with the use of archetypes, which are reference building models inclusive of typical OB attributes (Carnieletto et al., 2021), often in the form of standardized static schedules (Happle et al., 2018). Such schedules can be either customized by modellers, with direct inputs or dedicated editing functions or predefined and extracted from the ones proposed for building-level applications by standards or codes, such as ASHRAE 90.1 (2022). In this way however, buildings described by the same archetype present the same OB patterns, resulting in a model failing of capturing the real stochasticity of urban dynamics, and in a systematic difference between the predicted and the actual energy use of buildings (Ferrando et al., 2022; Happle et al., 2018).

UBEM, in fact, goes beyond the simple linear scaling up of energy modelling from individual to group of buildings but rather tries to capture the inextricably and dynamic interdependencies that exist between buildings, their surrounding environment and the local microclimate (Hong et al., 2020). Therefore, UBEM focuses on people's movement patterns through the urban environment, rather than just within individual buildings. This means that factors such as the location of public transport stations, the design of public spaces, and pedestrian network connectivity, having a significant impact on how people move through the city and access different amenities and services (C. Kang et al., 2012; R. Wang et al., 2022), can, in turn, affect energy use in different parts of the city (Mohammadi & Taylor, 2017).

For these reasons, to obtain realistic energy use patterns of groups of buildings at different spatial and time scales it is important to introduce diversity among buildings described by the same archetype. This diversity should consider the inherent variability in human behaviour and how the surrounding urban environment, socio-economic conditions and subjective values affect human activities and occupancy patterns. For example, identical buildings may have different occupancy levels in different geographical locations, as factors such as transport accessibility or outdoor thermal comfort conditions concur to determine the use of a certain area (Banfi et al., 2022; Happle et al., 2018). Additionally, it's crucial to consider the changes in human behaviour over time when

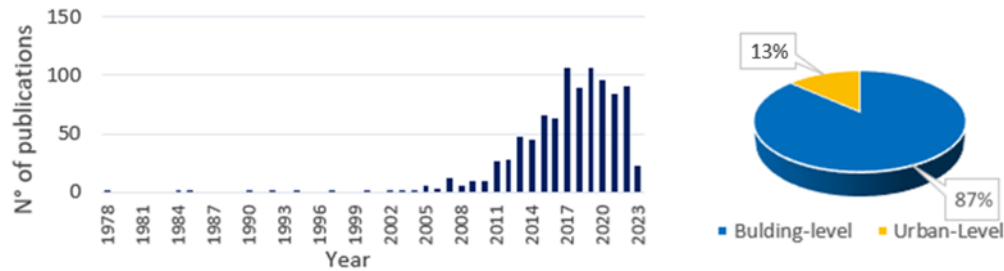


Figure 1: a) publication on the topic of OB to support building performance simulation from 1978 to 2023 b) share of the total publication developed at the building (blue) and urban (yellow) level

conducting long-term urban analyses or scenarios evaluation (Hou et al., 2022).

OB modelling requirements and targets, thus, differ according to the specific research question addressed by the application and the spatial scale considered in the energy modelling. However, the degree of detail required to model urban OB patterns remains unclear (Ferrando et al., 2022). Initial efforts in this sense have been made by Malik et al. (2022), who applied the Level-of-Detail (LoD) technique to understand the detail required to represent OB in building performance simulation. Four LoDs of increasing complexity are proposed, together with suggested use cases. For urban scale analyses, LoD between O-1 and O-2 (i.e., respectively homogeneous rule-based schedules, and heterogeneous static-probabilistic models) are proposed as suitable to capture diversity while avoiding excessive computational burdens (Malik et al., 2022). More details on the nature of different models will be provided in “Modelling approaches” Section.

Despite the importance of this effort to formalize OB LoDs, the quantitative effect of choosing advanced rather than predefined (or vice versa) occupants’ models on the results of the energy simulation at different spatiotemporal resolutions is still unclear, as examples in literature implementing the newly modelled OB schedules in UBEM tools remain limited. Among the preliminary studies, Barbour et al. (2019) found that using mobile-inferred schedules led to a median Energy Use Intensity (i.e., energy consumption per unit floor area) difference of up to -15% and -21% in residential and commercial buildings in Boston, compared to predefined reference schedules, Wu et al. (2020) compared the results of annual energy prediction for a district in San Antonio, Texas, using context-specific and standard occupancy profiles, discovering up to 60% heating energy and 40% cooling energy differences across all building types, while Mosteiro-Romero et al. (2020) suggest that occupant related input can have a significant effect on hourly or peak energy demand predictions but their influence on yearly energy demand simulation is surpassed by the effect of buildings properties. Furthermore, most of the reviewed case studies focus just on occupancy or occupancy-driven electric loads (i.e., appliances and lighting use driven only by occupant’s presence), without detailing occupants’ actions, despite the possibility of setting dedicated schedules, such as windows or shades operations, in some of the available UBEM tools (Doma & Ouf, 2023).

Apart from finding the right trade-off between detail, significance, and computational effort, improving OB modelling at large scales is hindered by the complexity of gathering enough data with sufficient spatiotemporal resolution (Dabirian et al., 2022; Malik et al., 2022). In fact, the dynamic and unpredictable nature of human presence and activities, coupled with privacy concerns that may arise, make it challenging to monitor them (Dabirian et al., 2022). To overcome this limitation, researchers started to investigate the potential of data coming from the proliferation of the Internet of Things (IoT) in everyday life (e.g., location-based service applications, network connectivity, etc.) (Salim et al., 2020). This kind of data has been already employed in other fields (Dashdorj et al., 2018; Su et al., 2020) to effectively model human activity patterns, and proved useful also in pilot studies conducted within the building sector (Barbour et al., 2019; Happle et al., 2020; Hou et al., 2022; X. Kang et al., 2021). Available OB data sources will be discussed in “Data Sources” Section.

Modelling approaches

Building OB modelling makes use of mathematical models and statistical techniques to infer and interpret people’s presence and actions in the built environment (Hou et al., 2022). Starting from the application in BEM, various OB models have been proposed over the years (Carlucci et al., 2020), whose characteristics are somehow reflected in UBEM analyses. Previous studies (Dabirian et al., 2022; Doma & Ouf, 2023; Happle et al., 2018; Hou et al., 2022) have attempted to classify the available urban OB modelling approaches using different terms and categories, often overlapping or combinable, resulting in a complex landscape that can be challenging to navigate. Summarizing the results of the previous analyses, modelling approaches can be differentiated based on:

- The capability of accounting for a certain level of unpredictability or randomness (i.e., stochastic vs deterministic models),
- The adoption of strategies to consider inter-individual diversity (i.e., heterogeneous vs homogeneous models),
- The complexity, intended as the ability to capture the reciprocal influence between building systems and occupants (i.e., static vs dynamic models),
- The level of aggregation, or granularity, of the impact of individuals’ behaviour (i.e., space-based vs agent-based models).

Existing strategies to introduce individual diversity and complexity in OB models are well discussed in the works of Doma & Ouf (2023) and Happle et al. (2018) and therefore excluded from the present study. Conversely, the definition of deterministic and stochastic models is often proposed in the literature, but an overview of the techniques used for their development is somehow missing and it is thus provided together with significant examples. The different granularity of OB models will not be treated in detail, except for Agent-Based Modelling (ABM), which is gaining momentum in the field of OB modelling for its capability of capturing complex and dynamic behaviours (Malik et al., 2022). In fact, while space-based models define OB profiles, of different diversity, complexity and randomness, for specific thermal zones (i.e., the entire building, core/perimeter, etc.), in ABM the focus is moved to the behaviour of individual agents within the urban environment.

Deterministic approaches

Deterministic models represent the simplest way to describe OB in buildings. They consist of fixed schedules, either data-driven or based on statistical trends, and simple sets of rules.

Fixed schedules are static models presented in the form of 24-hour normalized profiles for occupancy and occupancy-related electric loads, typically lighting and appliances use. The choice of the normalization factor is left to the discretion of the modellers and could encompass variables such as nominal building capacity or maximum number of occupants in a certain timeframe (i.e., day, week, month) for occupancy, and peak consumption or installed power density for electric load profiles. Some UBEM tools also allow the integration of deterministic schedules for windows or shade operation, and domestic hot water usage (Doma & Ouf, 2023). Such schedules can be extracted from codes or standards (ASHRAE, 2022; NCM, 2021), or calibrated using context-specific observations.

Static deterministic schedules proposed by standards or codes, or simply “standard schedules”, are based on statistical trends derived from field studies, surveys, and expert judgment but, as noted by other researchers (Hou et al., 2022), may result as outdated or excessively generic. Examples of standard schedules are the ones proposed by ASHRAE 90.1 (2022), which include occupancy and appliance use patterns for average weekdays and weekends for different building typologies.

Data-driven deterministic schedules, instead, are derived from observed data through statistical and data mining techniques. Examples of this procedure can be found in the works of Ferrando et al. (2022), in which schedules representative of the actual energy use and occupancy of 49 residential buildings located in Milan have been obtained through k-means clustering of smart meter registration, Kang et al. (2021), who successfully employed k-means clustering to extract typical weekly occupancy profiles from mobile positioning data, and Buttitta et al. (2019), who managed to derive occupancy

profiles representative of the UK residential building stock analysing survey data through k-mode clustering.

Rule-based models, instead, proved useful to describe adaptive actions linking occupant-building interactions to fixed environmental parameter thresholds. Examples are windows operations driven by trigger values of outdoor temperature or thermostat adjustments based on thermal comfort metrics. In BEM linear or logistic regression are common statistical techniques employed to investigate the relationship between different behaviours and independent predictors or environmental stimuli (Carlucci et al., 2020; Zhang et al., 2018).

Stochastic approaches

Stochastic OB models have been developed to overcome the limitations of deterministic schedules by capturing the randomness of human-related activities. In this way, the probability of a certain event occurring (e.g., building being occupied, windows being open, etc.) can be derived based on historical or monitored data (Dabirian et al., 2022; Happle et al., 2018; Zhang et al., 2018).

In the field of building OB modelling, the Markov Chain (MC) is a widely used stochastic modelling technique, as it allows to describe a sequence of events based on the current state and the probability of the state changing. For example, Richardson et al. (2008) employed MC technique to generate, starting from survey data, active occupancy time-series data representative of UK households, Widén et al. (2009) proposed a stochastic model for domestic lighting demand based on the occupancy of Swedish household members, generated from survey data with a non-homogeneous MC model. Other stochastic modelling methods employed in urban-level applications are Gaussian Mixture Models (GMM) and Survival Analysis. GMM is a type of machine learning algorithm that assumes data to be generated from a combination of several Gaussian distributions. Wang et al. (2020), exploited this method to create a dynamic urban-level occupant density model for commercial buildings in Nanjing, China. Survival Analysis instead, estimates the time duration until an event occurs and was employed coupled with a copula function by Hou et al. (2022) to model occupancy in multiple buildings of a university campus accounting for the characteristic of the surrounding environment.

Stochastic modelling is useful to predict realistic energy patterns at high temporal resolution (Dabirian et al., 2022), but it requires a substantially larger amount of data to be developed compared to deterministic approaches.

Agent-Based Modelling

Agent-Based Modelling is a computational simulation technique used to model systems as comprised of autonomous agents able to interact with each other and their surrounding environment (Malik et al., 2022). ABM is gaining popularity in the context of OB modelling due to its potential to develop models with a high degree of heterogeneity and stochasticity.

In ABM, an agent can represent a single or a group of individuals, whose state in time is defined based on a set of rules, either deterministic or stochastic and static or

dynamic. In general, an agent's state and its evolution encompass a wide set of properties, including position, activity, and comfort preferences, and behavioural capabilities such as sensing, learning, prediction, local interaction, group dynamics and social influence (Malik et al., 2022). However, as recently formalized by Malik et al. (2022), the degree of detail chosen for ABM influences the model complexity, computational and time effort, as well as its outcomes, and should be accurately selected based on the specific application. For urban-level analyses, a lower LoD is suggested (Malik et al., 2022) even though further studies and applications on real case studies are required to test this framework. Nonetheless, regardless of the LoD chosen, ABM tends to be extremely data-intensive, and its implementation may be hindered by the complexity of gathering enough information to characterize the initial set of agents.

ABM can be performed using dedicated toolkits (e.g., AnyLogic, NetLogo, and Repast) or programming languages (Malik et al., 2022). Furthermore, already existing ABMs developed in other fields could be successfully extended and adapted to UBEM applications, as proven by Barbour et al. (2019) and Mosteiro-Romero et al. (2020) with the development of novel occupancy modelling approaches based on transportation ABM frameworks (i.e., TimeGeo and MATSim respectively). Initial efforts made on the side of activity modelling can be found in (Vellei et al., 2021), where a stochastic ABM for thermostat adjustment in Canadian residential buildings is proposed. However, gathering reliable data about people's actions with a sufficient spatiotemporal resolution is far more complex than collecting occupancy information, which can be supported by georeferenced data coming from IoT. Therefore, available urban scale agent-based OB models are mainly focused on occupancy (Happle et al., 2018).

Discussion

The choice of the appropriate OB modelling approach is strongly contextual and depends on the research question, available data, and the degree of detail needed for the specific application. Stochastic modelling is more accurate in capturing the unpredictability of OB than deterministic approaches. Therefore, they have the potential to provide a realistic representation of occupants and, consequently, more reliable predictions of the energy patterns of groups of buildings. On the other hand, stochastic models require a large amount of data to be developed, which may be unavailable, or pose computational burdens due to their complexity.

Deterministic models, even though unable to capture the unpredictability of OB, are simple and data-minimal, being suitable for large spatial (e.g., national) and temporal (e.g., yearly) scale applications, initial design decisions, or all the conditions where data are limited or unavailable. Furthermore, they are easily and directly implementable in all the available UBEM tools, in contrast to stochastic models which could require the development of customized functions for their integration.

Similarly to randomness, model-granularity should be adapted to each specific application. Currently, space-based approaches are the type of model directly integrable into all the available UBEM tools, allowing for an easier and more practical simulation activity. However, even though they can encompass stochasticity, they do not account for occupants' interactions and decision-making processes. ABM overcome this limitation, by providing realistic and flexible OB patterns, inclusive of the effect of group and social dynamics. ABM can also be suitable for long-term scenario evaluation (e.g., energy policy introduction) since they present the potential of accounting for occupants' behavioural adaptation in time (Malik, et al., 2022). However, when deeply detailed, ABM becomes extremely data-intensive and complex, even unnecessarily, and its integration with UBEM tools is an area that requires further research.

Data Sources

Since OB includes the description of the two distinct aspects of people's presence and energy-related activities, the data required for its characterization are typically collected and integrated from multiple heterogeneous sources. The traditional ways to register OB include in situ measurements, either from dedicated sensors or indirect analysis of other measurable parameters, and surveys (Carlucci et al., 2020). In addition, a whole new group of sources coming from the proliferation of IoT and technology into everyday life has uncovered new possibilities for the description of OB.

This Section proposes an overview of the available data sources, specifying whether they provide information on occupant's presence, actions or both. In particular, the following groups of data sources have been identified: in situ measurements, surveys, location-based service (LBS) applications, and network connectivity.

In situ measurements

Information on both occupancy and occupants' activities can be collected through direct and indirect measurements with the use of dedicated sensors.(Carlucci et al., 2020). Multiple technologies have been developed over the years for OB monitoring, but among them, the ones more promising for urban-scale applications are vision-based sensors, smart meters and connected thermostats.

Vision-based sensors can be installed both indoors and outdoors, allowing the direct detection of human presence and mobility patterns in the urban environment (e.g., inference of building occupancy based on video-based parking occupancy detection). The installation of new devices is not always required as also existing cameras, such as the ones installed for surveillance, could be employed. Image-based data can in principle provide information on both people's presence and actions (e.g., detecting human pose), but also due to the complexity of its processing, at the present state of the research, their adoption remains limited to building-level applications (Dabirian et al., 2022).

Smart meters and connected thermostats are indirect OB monitoring technologies. Smart meters are electric devices used to measure electric consumption at high

temporal resolution (e.g., 15-minute time step), while connected thermostats are integrated with remote control and advanced algorithms to learn and adapt to users' preferences and behaviour over time. With the information gathered by such devices, it is possible to gain insight into both occupant's presence and actions, as done by Ferrando et al. (2022) and Vellei et al. (2021).

In situ measurements are commonly employed in building-level studies but their effectiveness on broader scales can be limited by the costs of sensors' installation and maintenance.

Surveys

Surveys represent an important tool in the field of urban OB modelling to collect data on occupants' activities and preferences. They include the collection of demographic information, behavioural patterns, or environmental attitudes and can be performed by government agencies, national statistical institutes, researchers, or private agencies. Time Use Surveys (TUS) are a common type of survey employed to collect information on the duration, timing, and location of different activities throughout the day. TUSs are currently carried out in more than 100 countries around the world, commonly referring to one weekday and one weekend reference days, with a temporal resolution of 10 or 15 minutes (Osman & Ouf, 2021), representing a powerful source of information about the average daily routine of the population. For this reason, they have been used in previous studies to model occupancy, lighting demand, domestic hot water consumption, heating and cooling load, and appliances' electric load (Osman & Ouf, 2021). Examples of large-scale applications of TUS data can be found in the works of Buttitta et al. (2019) and Richardson et al. (2008).

TUS can also be combined with other types of surveys (e.g., travel surveys, working surveys, census data, etc.) to create an even more integrated database of daily activity patterns, inclusive of demographic, economic, and social parameters, as well as information about activities and working time, and travel mode, duration, and purpose (Osman & Ouf, 2021). Furthermore, qualitative sociological surveys, or psychological studies, can provide important insights into the cultural, social, cognitive, emotional, and physical factors that influence human behaviour and energy use in urban environments (Bavaresco et al., 2020). Data collected through surveys can be aggregated at different levels, such as by individual, household, or population. Datasets from surveys conducted by national statistical institutes or government agencies are usually open to the public.

Location-based service applications

Location-based service applications are intended as applications whose functioning requires information about users' geographical location (Salim et al., 2020). Examples include social media, mapping platforms, navigation apps or ride-sharing applications. Data from these sources usually include the ID of the device, latitude and longitude coordinates, and recording time. To avoid privacy issues, this data can be subjected to regulations, and provided in aggregated forms. LBS data, providing

insights into people's location over time, can be predominantly used for occupancy modelling rather than for the description of occupants' energy-related actions.

Currently, in the building sector, only data from social media check-ins and geotagged posts have been exploited to model OB at large spatial scales but GPS traces could open new opportunities in the study of human presence and mobility patterns, as often done in the transportation field (Su et al., 2020). Examples of developed case studies include the use of Google Popular Times to create occupancy schedules (Happle et al., 2020; Parker et al., 2017), Twitter positional data to investigate the relation between urban human mobility and energy use across building types (Mohammadi & Taylor, 2017) and the extraction of weekly occupancy profiles from positional records of social media apps (X. Kang et al., 2021).

Overall, the use of LBS data in urban OB modelling has the potential to provide valuable insights into the complex dynamics of urban areas and their use (especially of open-source datasets such as Google Popular Times or Twitter positional data) will likely continue to grow in popularity.

Network connectivity

Network connectivity data, such as Wi-Fi authentication, or Call-Detail-Records (i.e., information about the network connections made between mobile devices and cell towers), have emerged as a promising data source for modelling occupancy at the urban level. Such data can include information about the number and types of devices connected to the network, duration and timing of connections, and patterns of network traffic. Call-Detail-Records, (CDRs) have been used to identify stay points of users and subsequently estimate building occupancy (Barbour et al., 2019), while Wi-Fi authentications, proved useful to calibrate and validate a novel approach to model occupancy of multiple inter-dependent buildings (Hou et al., 2022). Wi-Fi logs have been used in single-building analyses to study occupant energy-related behaviour (Krishnan et al., 2022), even though studies on the urban scale are missing. The potential and advantages of using network connectivity data in urban occupant behaviour modelling are significant but, similarly to LBS data, they may arise strong privacy concerns.

Discussion

Each of the described OB data sources presents its advantages and disadvantages. Data gathered from in situ measurements, based on dedicated sensors, can provide insights into both occupancy and occupants' activities, with a high level of reliability. On the other hand, the high installation and maintenance costs and the difficulties in calibrating sensors over time may pose strong restrictions to their exploitation on large spatial scales. Furthermore, indoor sensors to be installed in residential buildings are unlikely to be accepted by the population for fear of privacy harm and surveillance. Surveys, instead, can provide high-resolution activity and energy use profiles without spatial limitations. However, they heavily rely on the dimension of the sample and the willingness of participants to respond (Osman & Ouf, 2021). Furthermore, since they are extremely time-consuming to

be conducted, they may be infrequently updated (Osman & Ouf, 2021). Survey data may also be inadequate for applications in which behavioural changes should be considered since they represent average data. Urban sensing provides the advantage of gathering data over a large spatial and temporal scale, allowing for the description of different building typologies and populations. However, further investigation is required to address concerns regarding the quality of the data, potential biases, and privacy issues. Furthermore, data ownership and control by third-party companies can raise questions about market power and transparency.

Conclusion and future outlooks

In this paper, an overview of the current state of the research on the topic of OB modelling to support UBEM is reported, together with a discussion of the available modelling approaches and data sources. Challenges and opportunities arising when scaling up OB modelling from the building to the urban level are highlighted, stressing the need for developing OB patterns able to capture the stochasticity of human behaviour. Different OB modelling approaches are discussed, concluding that selecting the most suitable one depends on the research question addressed, as well as on the quality of the data available. Simpler models, such as the deterministic ones, can be more appropriate for analyses performed at large spatiotemporal resolutions, while stochastic models show the potential to provide a more realistic OB representation. Space-based models are ready to be integrated into the available simulation tools, while agent-based ones require programming skills to be linked with UBEMs. Nonetheless, ABM is gaining momentum for its potential to account for individuals' decision-making processes and behavioural traits. Regarding available data sources to support the modelling, LBS and network connectivity data are becoming increasingly popular for their capability to collect people's location over time on large spatial scales. However, data with sufficient spatiotemporal resolution are rarely openly available, forcing researchers to adopt standard OB schedules in the absence of better data.

Finally, this analysis enabled to identify the following challenges that need to be addressed in future works:

- Different nomenclatures are used by different researchers in the description and classification of modelling approaches. Adopting shared terminology can assist in the interpretation.
- The choice of appropriate LoD for occupant's related input at different spatial and temporal scales is still unclear. Uncertainty analysis of UBEM simulation results using different OB models could prove useful in defining the correct trade-off between model complexity and accuracy.
- Integrability of OB models of different complexity in the available UBEM tool should be improved.
- Reliability of LBS and network connectivity data should be further investigated to dispel any risk of biased results.

- The collection of ground truth OB data should be promoted to enable the validation and calibration of novel approaches

Acknowledgement

The study was developed within the framework of the project URBEM (PRIN, code: 2020ZWKXKE), which has received funding from the Italian Ministry of University and Research (MUR).

References

- ASHRAE. (2022). *ASHRAE Standard 90.1-2022 - Energy Standard for Sites and Buildings Except Low-Rise Residential Buildings*.
- Banfi, A., Tatti, A., Ferrando, M., Fustinoni, D., Zanghirella, F., & Causone, F. (2022). An experimental technique based on globe thermometers for the measurement of mean radiant temperature in urban settings. *Building and Environment*, 222.
- Barbour, E., Davila, C. C., Gupta, S., Reinhart, C., Kaur, J., & González, M. C. (2019). Planning for sustainable cities by estimating building occupancy with mobile phones. *Nature Communications*.
- Bavaresco, M. V., D'Oca, S., Ghisi, E., & Lamberts, R. (2020). Methods used in social sciences that suit energy research: A literature review on qualitative methods to assess the human dimension of energy use in buildings. *Energy and Buildings* (Vol. 209).
- Buttitta, G., Turner, W. J. N., Neu, O., & Finn, D. P. (2019). Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. *Energy and Buildings*, 198, 84–99.
- Carlucci, S., de Simone, M., Firth, S. K., Kjærgaard, M. B., Markovic, R., Rahaman, M. S., Annaqeeb, M. K., Biandrate, S., Das, A., Dziedzic, J. W., Fajilla, G., Favero, M., Ferrando, M., Hahn, J., Han, M., Peng, Y., Salim, F., Schlüter, A., & van Treeck, C. (2020). Modeling occupant behavior in buildings. *Building and Environment*, 174.
- Carnietto, L., Ferrando, M., Teso, L., Sun, K., Zhang, W., Causone, F., Romagnoni, P., Zarrella, A., & Hong, T. (2021). Italian prototype building models for urban scale building performance simulation. *Building and Environment*, 192.
- Dabirian, S., Panchabikesan, K., & Eicker, U. (2022). Occupant-centric urban building energy modeling: Approaches, inputs, and data sources - A review. In *Energy and Buildings* (Vol. 257). Elsevier Ltd.
- Dashdorj, Z., Sobolevsky, S., Lee, S. K., & Ratti, C. (2018). Deriving human activity from geo-located data by ontological and statistical reasoning. *Knowledge-Based Systems*, 143, 225–235.
- Doma, A., & Ouf, M. (2023). Modelling occupant behaviour for urban scale simulation: Review of available approaches and tools. *Building Simulation* (Vol. 16, Issue 2, pp. 169–184).

- Ferrando, M., Causone, F., Hong, T., & Chen, Y. (2020). Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches. *Sustainable Cities and Society* Vol. 62.
- Ferrando, M., Ferroni, S., Pelle, M., Tatti, A., Erba, S., Shi, X., & Causone, F. (2022). UBEM's archetypes improvement via data-driven occupant-related schedules randomly distributed and their impact assessment. *Sustainable Cities and Society*, 87.
- Happle, G., Fonseca, J. A., & Schlueter, A. (2018). A review on occupant behavior in urban building energy models. *Energy and Buildings* (Vol. 174, pp. 276–292). Elsevier Ltd.
- Happle, G., Fonseca, J. A., & Schlueter, A. (2020). Context-specific urban occupancy modeling using location-based services data. *Building and Environment*, 175.
- Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment*, 168.
- Hou, H., Pawlak, J., Sivakumar, A., & Howard, B. (2022). Building occupancy modelling at the district level: A combined copula-nested hazard-based approach. *Building and Environment*, 225.
- Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and Its Applications*, 391(4), 1702–1717.
- Kang, X., Yan, D., An, J., Jin, Y., & Sun, H. (2021). Typical weekly occupancy profiles in non-residential buildings based on mobile positioning data. *Energy and Buildings*, 250.
- Krishnan, D., Kelly, S., & Kim, Y. (2022). A Meta-Analysis Review of Occupant Behaviour Models for Assessing Demand-Side Energy Consumption. In *Energies* (Vol. 15, Issue 3)..
- Malik, J., Azar, E., Mahdavi, A., & Hong, T. (2022). A level-of-details framework for representing occupant behavior in agent-based models. *Automation in Construction*, 139.
- Malik, J., Mahdavi, A., Azar, E., Chandra Putra, H., Berger, C., Andrews, C., & Hong, T. (2022). Ten questions concerning agent-based modeling of occupant behavior for energy and environmental performance of buildings. *Building and Environment*, 217.
- Mohammadi, N., & Taylor, J. E. (2017). Urban infrastructure-mobility energy flux. *Energy*, 140.
- Mosteiro-Romero, M., Hischer, I., Fonseca, J. A., & Schlueter, A. (2020). A novel population-based occupancy modeling approach for district-scale simulations compared to standard-based methods. *Building and Environment*, 181.
- NCM. (2021). *National Calculation Methodology (NCM) modelling guide (for buildings other than dwellings in England)*.
- Osman, M., & Ouf, M. (2021). A comprehensive review of time use surveys in modelling occupant presence and behavior: Data, methods, and applications. In *Building and Environment* (Vol. 196). Elsevier Ltd.
- Parker, J., Hardy, A., Glew, D., & Gorse, C. (2017). A methodology for creating building energy model occupancy schedules using personal location metadata. *Energy and Buildings*, 150, 211–223.
- Richardson, I., Thomson, M., & Infield, D. (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, 40(8), 1560–1566.
- Salim, F. D., Dong, B., Ouf, M., Wang, Q., Pigliautile, I., Kang, X., Hong, T., Wu, W., Liu, Y., Rumi, S. K., Rahaman, M. S., An, J., Deng, H., Shao, W., Dziedzic, J., Sangogboye, F. C., Kjærgaard, M. B., Kong, M., Fabiani, C., ... Yan, D. (2020). Modelling urban-scale occupant behaviour, mobility, and energy in buildings: A survey. *Building and Environment*, 183.
- Schweiker, M., Carlucci, S., Andersen, R. K., Dong, B., & O'Brien, W. (2017). Occupancy and occupants' actions. In *Exploring Occupant Behavior in Buildings: Methods and Challenges* (pp. 7–38).
- Su, R., McBride, E. C., & Goulias, K. G. (2020). Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. *Transportation Research Part C: Emerging Technologies*, 120.
- United Nations Environment Programme. (2021). *2021 GLOBAL STATUS REPORT FOR BUILDINGS AND CONSTRUCTION*
- Vellei, M., Martinez, S., & Le Dréau, J. (2021). Agent-based stochastic model of thermostat adjustments: A demand response application. *Energy and Buildings*, 238.
- Wang, C., Wu, Y., Shi, X., Li, Y., Zhu, S., Jin, X., & Zhou, X. (2020). Dynamic occupant density models of commercial buildings for urban energy simulation. *Building and Environment*, 169.
- Wang, R., Zhang, X., & Li, N. (2022). Zooming into mobility to understand cities: A review of mobility-driven urban studies. *Cities*, 130.
- Widén, J., Nilsson, A. M., & Wäckelgård, E. (2009). A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand. *Energy and Buildings*, 41(10), 1001–1012.
- Wu, W., Dong, B., Wang, Q. (Ryan), Kong, M., Yan, D., An, J., & Liu, Y. (2020). A novel mobility-based approach to derive urban-scale building occupant profiles and analyze impacts on building energy consumption. *Applied Energy*, 278.
- Zhang, Y., Bai, X., Mills, F. P., & Pezzey, J. C. V. (2018). Rethinking the role of occupant behavior in building energy performance: A review. *Energy and Buildings* (Vol. 172, pp. 279–294).